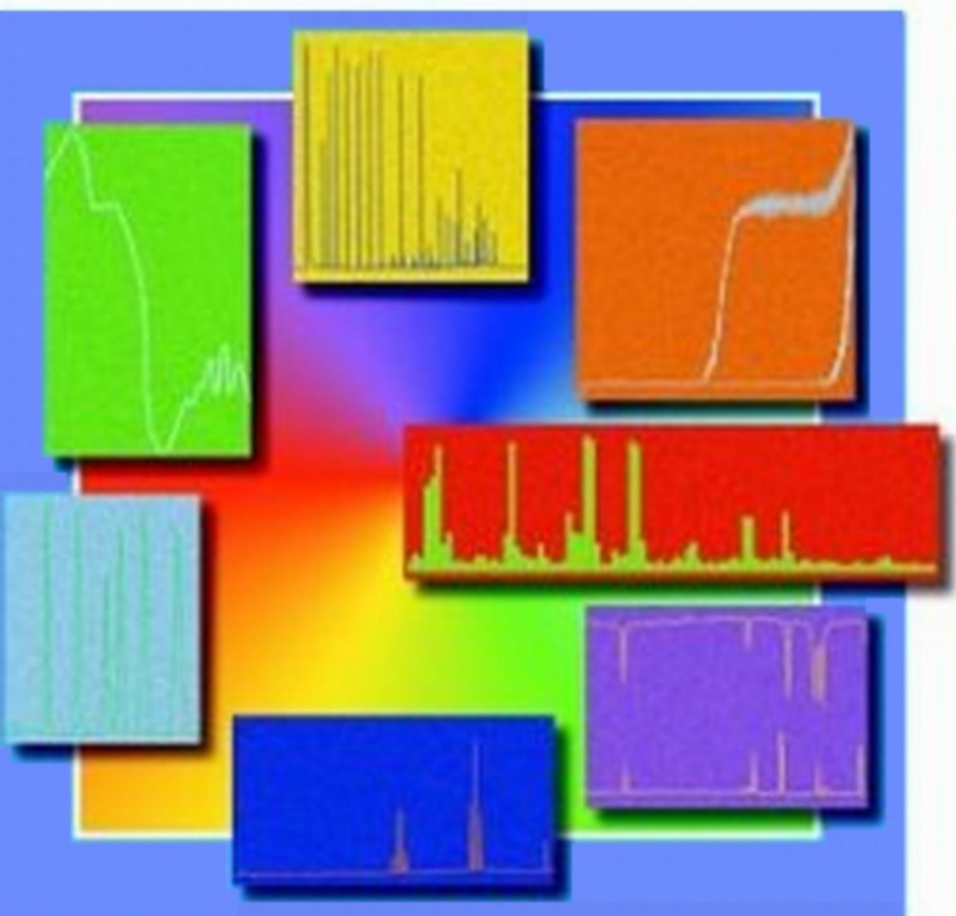# Handbook of Analytical Techniques

Edited by
Helmut Günzler and Alex Williams

Handbook of
Analytical Techniques

edited by Helmut Günzler and
Alex Williams

WILEY-VCH

For more information about analytical techniques please visit our website (www.wiley-vch.de) or register for our free e-mail alerting service (www.wiley-vch.de/home/pas)

Handbook of
Analytical Techniques

edited by Helmut Günzler and
Alex Williams

Prof. Dr. Helmut Günzler
Bismarckstr. 4
D-69469 Weinheim
Germany

Alex Williams
19 Hamesmoor Way, Mytchett
Camberley, Surrey GU16 6JG
United Kingdom

This book was carefully produced. Nevertheless, authors, editors and publisher do not warrant the information contained therein to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Preface

The broad spectrum of analytical techniques available today is covered in this handbook. It starts with general articles on purpose and procedures of analytical chemistry, quality assurance, chemometrics, sampling and sample preparation followed by articles on individual techniques, including not only chromatographic and spectrometric techniques but also e. g. immunoassays, activation analysis, chemical and biochemical sensors, and techniques for DNA-analysis.

Most of the information presented is a thoroughly updated version of that included in the 5th edition of the 36-volume "Ullmann's Encyclopedia of Industrial Chemistry", the last edition that is available in print format. Some chapters were completely rewritten. The wealth of material in that Encyclopedia provides the user with both broad introductory information and in-depth detail of utmost importance in both industrial and academic environments. Due to its sheer size, however, the unabridged Ullmann's is inaccessible to many potential users, particularly individuals, smaller companies, or independent analytical laboratories. In addition there have been significant developments in analytical techniques since the last printed edition of the Encyclopedia was published, which is currently available in its 6th edition in electronic formats only. This is why all the information on analytical techniques has been revised and published in this convenient two-volume set.

Users of the "Handbook of Analytical Techniques" will have the benefit of up-to-date professional information on this topic, written and revised by acknowledged experts. We believe that this new handbook will prove to be very helpful to meet the many challenges that analysts in all fields are facing today.

Weinheim, Germany   Helmut Günzler
Camberley, United Kingdom   Alex Williams
January 2001

Contents

Volume I

11. Gas Chromatography 199

12. Liquid Chromatography 261

13. Thin Layer Chromatography 327

14. Electrophoresis 345

15. Structure Analysis by Diffraction 373

27. Surface Analysis 851

28. Chemical and Biochemical Sensors 951

29. Microscopy 1058

30. Techniques for DNA Analysis 1131

Sucject Index 1151

Symbols and Units

Symbols and units agree with SI standards. The following list gives the most important symbols used in the handbook. Articles with many specific units and symbols have a similar list as front matter.

Symbol	Unit	Physical Quantity
a_B		activity of substance B
A_r		relative atomic mass (atomic weight)
A	m^2	area
c_B	mol/m^3, mol/L (M)	concentration of substance B
C	C/V	electric capacity
c_p, c_v	$J kg^{-1}K^{-1}$	specific heat capacity
d	cm, m	diameter
d		relative density (ϱ/ϱ_{water})
D	m^2/s	diffusion coefficient
D	Gy (= J/kg)	absorbed dose
e	C	elementary charge
E	J	energy
E	V/m	electric field strength
E	V	electromotive force
E_A	J	activation energy
f		activity coefficient
F	C/mol	Faraday constant
F	N	force
g	m/s^2	acceleration due to gravity
G	J	Gibbs free energy
h	m	height
h	$W \cdot s^2$	Planck constant
H	J	enthalpy
I	A	electric current
I	cd	luminous intensity
k	(variable)	rate constant of a chemical reaction
k	J/K	Boltzmann constant
K	(variable)	equilibrium constant
l	m	length
m	g, kg, t	mass
M_r		relative molecular mass (molecular weight)
n_D^{20}		refractive index (sodium D-line, 20 °C)
n	mol	amount of substance
N_A	mol^{-1}	Avogadro constant (6.023×10^{23} mol^{-1})
p	Pa, bar *	pressure
Q	J	quantity of heat
r	m	radius
R	$J K^{-1}mol^{-1}$	gas constant
R	Ω	electric resistance
S	J/K	entropy
t	s, min, h, d, month, a	time
t	°C	temperature
T	K	absolute temperature
u	m/s	velocity
U	V	electric potential
U	J	internal energy
V	m^3, L, mL	volume
w		mass fraction
W	J	work
x_B		mole fraction of substance B
Z		proton number, atomic number
χ		cubic expansion coefficient

Symbol	Unit	Physical Quantity
α	W m^{-2}K^{-1}	heat-transfer coefficient (heat-transfer number)
α		degree of dissociation of electrolyte
$[\alpha]$	10^{-2}deg cm^{2}g^{-1}	specific rotation
η	Pa $\cdot$ s	dynamic viscosity
θ	°C	temperature
κ		c_p/c_v
λ	W m^{-1}K^{-1}	thermal conductivity
λ	nm, m	wavelength
μ		chemical potential
ν	Hz, s^{-1}	frequency
ν	m^2/s	kinematic viscosity (η/ϱ)
π	Pa	osmotic pressure
ϱ	g/cm^3	density
σ	N/m	surface tension
τ	Pa (N/m^2)	shear stress
φ		volume fraction
χ	Pa^{-1} (m^2/N)	compressibility

* The official unit of pressure is the pascal (Pa).

1. Analytical Chemistry: Purpose and Procedures

HANS KELKER formerly Hoechst AG, Frankfurt, Federal Republic of Germany

GÜNTHER TÖLG, formerly Institut für Spektrochemie und Angewandte Spektroskopie, Dortmund, Federal Republic of Germany

HELMUT GÜNZLER, Weinheim, Federal Republic of Germany

ALEX WILLIAMS, Mytchett, Camberley, UK.

1.1. The Evolution of Analytical Chemistry

"Analytical chemistry" (more simply: *analysis*) is understood today as encompassing any examination of chemical material with the goal of eliciting information regarding its constituents: their *character* (form, quality, or pattern of chemical bonding), *quantity* (concentration, content), *distribution* (homogeneity, but also distribution with respect to internal and external boundary surfaces), and *structure* (spatial arrangement of atoms or molecules). This goal is pursued using an appropriate combination of chemical, physical, and biological methods [1] – [6]. From a strategic standpoint the challenge is to solve the analytical problem in question as completely and reliably as possible with the available methods, and then to interpret the results correctly. Sometimes it becomes apparent that none of the methods at hand are in fact suitable, in which case it is the methods themselves that must be improved, perhaps the most important rationale for intensive basic research directed toward the increased effectiveness of problem-oriented analysis in the future.

More comprehensive contemporary definitions of analytical chemistry have been proposed [7], [8], underscoring above all the complexity of the discipline — which the authors of this introduction were also forced to confront.

Consistent with its close historical ties to chemical synthesis, modern analysis is still firmly

embedded within the broader framework of chemistry in general. This is inevitably the case, because systematic analysis depends absolutely upon a solid, factual knowledge of matter. This point is as valid now as it was in 1862 when C. R. FRESENIUS stated in his classic *Introduction to Qualitative Chemical Analysis* [9]: "Chemical analysis is based directly on general chemistry, and it cannot be practiced without a knowledge thereof. At the same time it must be regarded as one of the fundamental pillars upon which the entire scientific edifice rests; for analysis is of almost equal importance with respect to all the branches of chemistry, the theoretical as well as the applied, and its usefulness to doctors, pharmacists, mineralogists, enlightened farmers, technologists, and others requires no discussion."

The *tools* of modern analysis are nevertheless based largely on physical principles. Mathematical techniques related to information theory, systems theory, and chemometrics are also making increasingly important inroads. It would in fact no longer be presumptuous to go so far as to describe "analytical science" as an independent discipline in its own right.

The pathway leading to the present exalted place of analysis within the hierarchy of chemistry specifically and the natural sciences generally has not always been a straight one, however. Indeed, from earliest times until well into the eighteenth century the very concept of "analysis" was purely implicit, representing only one aspect of the work of the alchemists and various practitioners of the healing arts (iatrochemists). Some more tangible objective always served as the driving force in an investigation, and "to analyze" was almost synonymous with the broader aim: a quest for precious metals, a desire to establish the content of something in a particular matrix, or a demonstration of pharmacological activity. Only after the time of LAVOISIER and with the emergence of a separate chemical science—a science largely divorced from external goals—is one able to discern what would today be regarded as typical "analytical" activity. The term "analysis" appears explicitly for the first time around the turn of the nineteenth century in the title of the book, *Handbuch zur chemischen Analyse der Mineralkörper* ("Handbook for the Chemical Analysis of Minerals") by W. A. LAMPADIUS. Further information regarding the history of analysis is available from the monograph by SZABADVARY [10].

Many of the greatest discoveries in chemistry could fairly be described as classic examples of successful analyses, including the discovery of oxygen, the halogens, and several other elements. Well into the nineteenth century, discovering a new chemical element was regarded as the highest and most prestigious achievement possible for an academic chemist, as documented, for example, by desperate attempts to gain further insight into the "rare earths," or to detect the elusive (but accurately predicted) homologues of lanthanum and cerium. MOSANDER in fact devoted his entire life to the latter search.

C. REMIGIUS FRESENIUS once again deserves credit for noting, toward the middle of the nineteenth century, that new analytical techniques invariably lead to fresh sets of discoveries. Whereas the element germanium was found on the basis of "classical" methods (CLEMENS WINKLER, 1886), FRESENIUS' observation clearly applies to the discovery of the alkali metals rubidium and cesium (by ROBERT W. BUNSEN after he and G. R. KIRCHHOFF first developed emission spectroscopy in 1861). Other relevant examples include the discoveries of radium and polonium (by Madame CURIE), hafnium (HEVESY and COSTER, 1922), and rhenium (I. TACKE and W. NODDACK, 1925), all with the aid of newly introduced X-ray spectrometric techniques. This is also an appropriate point to mention the discovery of nuclear fission by OTTO HAHN and FRITZ STRASSMANN (1938), another accomplishment with strongly analytical characteristics [10].

ROBERT BUNSEN is rightfully acknowledged as the harbinger of modern analysis, but much of the discipline's distinctive scientific character was provided by WILHELM OSTWALD [11] building on the activities of J. H. VAN'T HOFF and WALTHER NERNST.

Analytical chemistry in these early decades was often accorded the secondary status of a faithful servant, but even the few examples cited here demonstrate quite convincingly that it also pursued its own unique set of principles—and for its own sake, with a strictly scientific orientation. The principles themselves were shaped by BERZELIUS and WÖHLER; experiment rather than theoretical speculation was the starting point and source of inspiration in this era characterized largely by chemical reactions. Readers of the present essay should in fact take the time to examine the third edition of *Ullmann's* [12] and discover there what the expression "analytical procedure" actually meant even as late as the end of World War II. There can be no mistaking the fact that "purely chemical" methods were still domi-
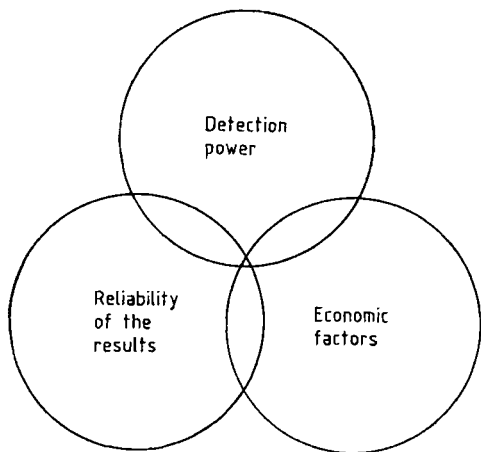
Figure 1. Qualitative criteria for use in evaluating analytical procedures

nant, and that the synthetic process constituted the model, particularly in the field of organic analysis.

Analytical chemistry has been responsible for many important contributions to our basic understanding of matter (e.g., the existence of the various elements, gas theory, stoichiometry, atomic theory, the law of mass action, nuclear fission, etc.), but the growth and development of a separate chemical industry ushered in a phase during which the scientific aspects of analysis suffered serious decline. The demand for analytical services shifted markedly in the direction of routine quality control, particularly with respect to synthetic organic products; indeed, significant resources were invested in the effort to dismember, resolve, and decompose synthetic substances into their simpler constituents (e.g., the chemical elements) — in strict conformity with the original meaning of the word "analysis" (αναλυσισ, *resolution*). For many years organic elementary analysis was virtually the only analytical approach available for characterizing synthetic organic reaction products. The denigration suffered by analysis at that time relative to synthesis (and production) continues to exert a negative influence even today on the university training of analytical chemists.

Elemental analysis in certain other quarters enjoyed a climate much more congenial to further development, especially in the metalworking industry and geochemistry. The indispensable contributions of analysis were recognized here much

earlier, particularly with respect to optimizing product characteristics (e.g., of steels and other alloys), and to providing detailed insight into the composition of the Earth's crust to facilitate the extraction of valuable raw materials. Geochemistry and the steel industry were particularly receptive to BUNSEN's new methods of spectral analysis, for example, which in turn provided a powerful stimulus for the development of other modern instrumental techniques. These techniques encouraged the exploitation of new and innovative technologies, first in the fields of semiconductors and ultrapure metals, then optical fibers and superconductors, and, most recently, in high-temperature and functional ceramics. Extraordinarily stringent demands were imposed upon the various analytical methods with respect to detection limits, extending to the outermost limits what was possible, especially in the attempt to characterize impurities responsible for altering the properties of particular materials. At the same time, the information acquired was expected to reflect the highest possible standards of reliability — and to be available at an affordable price. These three fundamental quality criteria are in fact closely interrelated, as indicated in Figure 1.

The increasing effectiveness of analytical techniques in general led ultimately to progress in the area of organic materials as well, especially with the rapid development of chromatographic and molecular spectroscopic methods. At the same time it also became necessary to acknowledge that technological advances inevitably bring with them new safety and health risks. For this reason analysis today plays an essential role not only in supporting technological progress but also in detection and minimization of the associated risks.

Just as FRESENIUS predicted, analysis has advanced rapidly toward becoming a science in its own right, with interdisciplinary appeal and subject to intense interest extending far beyond the bounds of chemistry itself: to the geological and materials sciences, the biosciences, medicine, environmental research, criminology — even research into the history of civilization, to mention only a few of the most important areas of application. The chemical industry today is the source of only a relatively small fraction of the samples subject to analysis. Rocks, soils, water, air, and biological matrices, not to mention mankind itself and a wide array of consumer goods, together with raw materials and sources of energy constitute the broad spectrum of analytical samples in the modern era (Fig. 2).
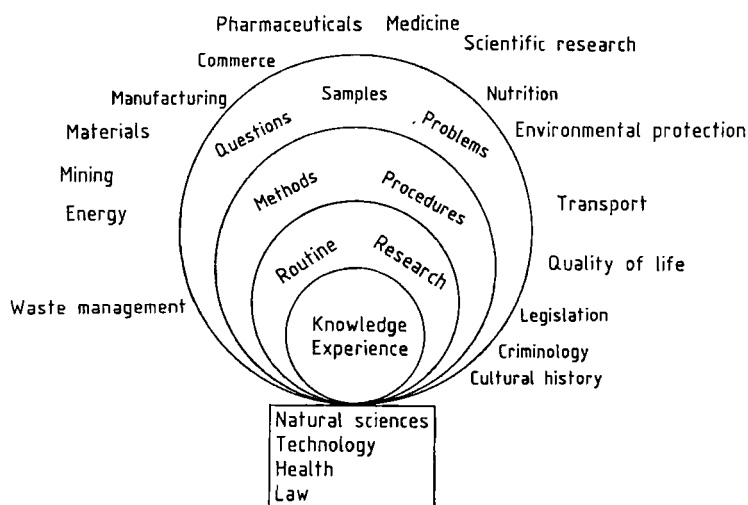
Figure 2. Overall task structure associated with analytical chemistry

Given this diversity of appeal the question has frequently been raised as to whether analysis really is an independent discipline, or if it should not instead be regarded simply as a service activity. The question is of course unrealistic, because analysis by its very nature is clearly both. Equally clear is the crucial importance of analysis to modern society. While the service function is undoubtedly more widely appreciated than other activities characterized by a "strictly scientific" focus, the latter also have an indispensable part to play in future progress.

The diversity characterizing the beneficiaries of analysis has actually remained fairly constant in recent decades, though immediate priorities have undergone a steady shift, particularly during the last 20 years with respect to ecology. Such a "paradigm shift" (THOMAS KUHN), marked by profound changes over time in both motivation and methodology, can occasionally assume revolutionary proportions. It remains an open question whether external change induces analysts to adapt and further develop their methodologies, or if the methodology itself provides the driving force. Here as elsewhere, however, there can be little doubt that "necessity is the mother of invention," capable of mobilizing forces and resources to an extent unimaginable in the absence of pressing problems.

Change also provides an incentive for deeper reflection: should we perhaps reformulate our understanding of the overall significance of analysis,

lift it out of its customary chemico-physical framework and broaden its scope to include, for example, KANT's "analytical judgments," or even psychoanalysis? Some would undoubtedly dismiss the questions as pointless or exaggerated, but from the perspective of the theory of learning they nevertheless provoke a considerable amount of interest and fascination [13], [14].

1.2. The Functional Organization of Analytical Chemistry

Attempting to summarize analytical chemistry in a single comprehensive schematic diagram is a major challenge, one that can only be addressed in an approximate way, and only after considerable simplification (Fig. 2) [5]. The fundamentals supporting the analysis must ultimately be the individual analyst's own store of knowledge, including the basic principles and laws of science and mathematics, together with the scope — and limitations — of existing analytical methods and procedures. Indispensable prerequisites to the successful resolution of an analytical problem include experience, a certain amount of intuition, and thorough acquaintance with a wide variety of modern analytical techniques. Familiarity with the extensive technical literature is also important (including the sources cited at the end of this article), an area in which modern systems of documentation

can be of considerable assistance. For example, an astonishing level of perfection can almost be taken for granted with respect to computer-based systems for locating spectra. Another essential component of the analyst's information base is knowledge regarding the source of each analytical sample — whether it comes from industry, the environment, or from medicine. After all, only the analyst is in a position to provide an overview of the analytical data themselves when the time comes for critical interpretation of the experimental results.

Immediately adjacent to "knowledge" in the functional diagram characterizing analytical chemistry (Fig. 2) is a region occupied by two parallel lines of endeavor: routine analysis on one hand, and research and development on the other, with the latter directed toward new methods and procedures. Both are subject to initiatives and incentives from outside, including other branches of science, medicine, regulatory agencies, commerce, and industry, all of which encourage and foster innovative developments within analysis itself.

Figure 2 also underscores the fact that an analyst's primary activities are of a problem-oriented nature, determined largely by the needs of others. The problems themselves, represented here by the outermost circle, might originate almost anywhere within the material world. Analysis can even play a significant role in the very definition of a scientific investigation. Consider the case of archaeology, for example, a considerable part of which is now "archaeometry," simply a specialized type of analysis.

With respect to the development of new products — such as materials, semiconductors, pharmaceuticals, crop protection agents, or surfactants — analysis plays a companion role at every stage in the progression from research laboratory to market. Studies related to physiological and ecological behavior demand comprehensive analytical efforts as well as intimate knowledge of the materials in question.

1.3. Analysis Today

Figure 3 provides a representative sample of methods to be found in the arsenal of the modern analyst. The figure also highlights the rapid pace of developments in analytical chemistry during the twentieth century [15]. Continued success in meeting present and future analytical challenges involves more than simply the tools, however, most of which have already been perfected to the point of commercialization. Appropriate strategies are required as well, just as a hammer, a chisel, and a block of marble will not suffice to produce a sculpture. Analytical strategies are at least as important as the methods, and the strategies must themselves be devised by qualified analysts, because every complex analytical problem demands its own unique strategic approach.

It is this context that establishes the urgent need for reactivating as quickly as possible the long-neglected training of qualified analysts. New analytical curricula must also be devised in which special emphasis is placed on the close symbiotic relationship in modern analysis between chemistry and physics [6].

Figure 4 depicts in a generalized way the multileveled complex of pathways constituting a typical analytical process and linking a particular problem with its solution. From the diagram it becomes immediately apparent that the "analytical measurement," which is the focal point of most modern physical methods, in fact represents only a very small part of the whole, despite the fact that the treatise to which this essay serves as a preface focuses almost exclusively on the principles of instrumental methods and their limitations.

Physical methods clearly occupy the spotlight at the moment, but *chemical* methods of analysis are just as indispensable today as in the past. Especially when combined with physical methods, chemical techniques frequently represent the only means to achieving a desired end. This is generally the case in extreme trace analysis [16], for example, where attaining maximum sensitivity and reliability usually requires that the element or compound of interest first be isolated from an accompanying matrix and then concentrated within the smallest possible target area or solution volume prior to the physical excitation that leads ultimately to an analytical signal. Combination approaches involving both chemical and physical methods are today commonly referred to as *multistep procedures* (see Section 1.6.11), where some chemical step (e.g., digestion, or enrichment) often precedes an instrumental measurement, or an analysis is facilitated by preliminary chromatographic separation. Chromatographic separation in turn sometimes requires some type of prior chemical transformation [17], as in the gas-chromatographic separation of organic acids, which is usually preceded by esterification.

pre-1850	1850-1900	1900-1925	1925-1950	1950-1975	1975-1992 (selection)

Titrimetry
Gravimetry
Gas-volumetry
	Electrogravimetry
	Coulometry
	Optical emission
	  spectroscopy
	  (OES)
		Raman spectroscopy
		X-Ray emission spec-
		  troscopy
		X-Ray diffraction
		Mass spectroscopy
		IR spectroscopy
		Isotope-dilution analysis
		Radiochemical analysis
		Adsorption chromatography
		Conductometry
			γ-Spectroscopy
			Electron spin resonance
			NMR
			Gas chromatography
			Paper chromatography
			Distribution chromatography
			High-frequency titration
			Electron diffraction
			Activation analysis
			Electrophoresis
			Polarography
			RBS
				Neutron spectroscopy
				Laser spectroscopy
				Electron tunneling spectroscopy
				HPLC
				Candoluminescence
				MECA spectroscopy
				RHEED
				LEED
				Desorptiometry
				SIMS
				Mößbauer spectroscopy
				ESMA
				NQR
				EELS
				Auger electron spectroscopy
				Photoelectron spectroscopy
				Ionography
				FIM
				FEM
				EDX
				Neutron diffraction
				β-Backscattering
				Atomic absorption spectroscopy
				Reflection IRS
				Microwave spectroscopy
				PIXE
					Atomic force microscopy
					Scanning tunnel microscopy
					Atom probe
					Analytical electron microscopy
					FT spectroscopy
					ICR mass spectroscopy
					TXRFA
					IR microscopy
					Scanning auger electron spectroscopy
					Dynamic SIMS
					CARS
					Photoacoustic spectroscopy
					SERS
					FANES
					HONES
					Appearance-potential spectroscopy
					IBLE
					FAB-MS
					GD-AES
					GD-MS
					Ion chromatography
					Ion neutralization spectroscopy
					EXAFS
					SNMS
					UPS, XPS
					Capillary gas chromatography
					Gel chromatography
					Supercritical-fluid chromatography
					HPTLC
					Scanning OC
					Plasma spectroscopy
					ICP-OES, MIP-OES
					ETA-AAS
					2D-NMR
					Multinuclear NMR
					Time-resolution spectroscopy
					GC-MS
					MS-MS-(MS)
					AAS-GC
					ICP-MS
					ICP-GC
					Laser ICP-MS
					Thermospray MS, Electrospray MS
					HPLC-MS
					C-13 NMR
					Charged-particle activation analysis
					Raman microprobe
					RP chromatography
					Solid phase AAS
					Headspace GC
					PIGE
					Static TOFSIMS
					LAMMA

Figure 3. Chronological summary of the arsenal of experimental methods currently available to analytical chemists; based on [15]

The terms "preanalysis" and "postanalysis" have been coined for characterizing steps that precede or follow a "true" analytical operation. Unfortunately, classification in this way tends to denigrate the importance of an operation like sampling or the evaluation of a set of final results, suggesting that these are secondary and relatively peripheral activities — reason enough for exercising considerable caution in use of the terms.

There can be no justification whatsoever for dismissing the importance of chemical reactions

in analysis, as "superprogressive" instrumental analysts occasionally tend to do, treating chemical methods as relics of an outmoded past. Chemical reactions still have a crucial part to play in many operations: sometimes as useful adjuncts, but often enough at the very heart of the determination. It is worth recalling in this context that gravimetry — together with the volumetric methods to which it gave birth — remains virtually the only viable approach to direct and reliable absolute determination (i.e., to *calibration-free analysis*). Such anal-
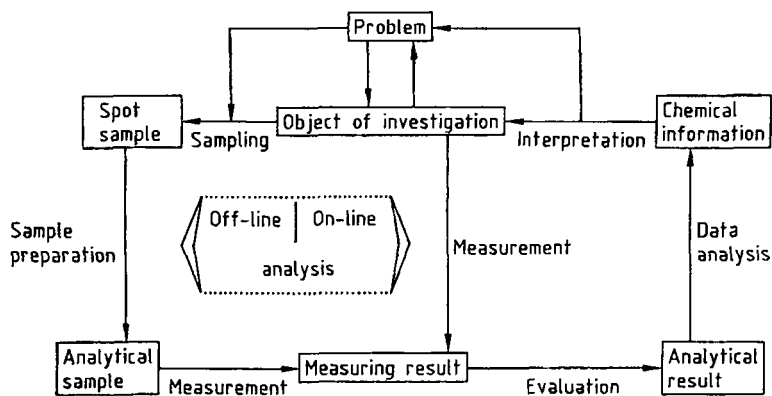
Figure 4. Schematic diagram of the analytical process; based on [15]

yses rely on "stoichiometric factors," which were painstakingly compiled over the course of decades in conjunction with the equally arduous and prolonged quest for an exact set of atomic masses.

Most physical methods, especially those associated with spectroscopy, lead only to *relative* information acquired through a comparison of two signals. This in turn presupposes a procedure involving a *calibration standard*, or reference sample of known composition. The only exceptions to this generalization—at least theoretically—are instrumental activation analysis (which involves the counting of activated atomic nuclei), isotope dilution (especially IDMS—isotope dilution mass spectrometry), and coulometry (assuming the strict validity of Faraday's Law). In view of quality assurance (see → Quality Assurance in Instrumentation), the named methods, jointly with gravimetry, volumetric analysis, and thermoanalysis, were recently designated as *primary methods of measurement* [18]–[21]. They play an important role in achieving traceable results in chemical measurements.

Some may feel that the foregoing observations direct excessive attention to the virtues of classical analytical chemistry. If so, the justification is a continuing need to emphasize the fact that optimal results are achieved when there is a close coupling between chemical and physical methods, and this despite antagonisms that persist between champions devoted to one approach or the other. Even today, classical principles— appropriately adapted—often constitute the most reliable guide.

1.4. Computers

A few remarks are necessary at this point on the subject of *electronic data processing* and the vital supportive role computers now play in analysis.

Developments in this area began with the central mainframe computer, to which a wide variety of isolated analytical devices could be connected. In recent years the trend has shifted strongly toward preliminary data processing via a minicomputer located directly at the site of data collection, followed in some cases by network transfer of the resulting information to a central computing facility. Often, however, the central computer is dispensed with entirely, with all data evaluation occurring on the spot. The powerful impact of electronic data processing on modern analysis dictates that it be addressed elsewhere in the present treatise in greater detail (→ Chemometrics).

The benefit of computers in modern analysis has been clearly established for some time. Computers now provide routine management and control support in a wide variety of analytical operations and procedures, and they are an almost indispensable element in data interpretation, processing, and documentation. Indeed, the lofty goals of "good laboratory practice" (GLP) would probably be beyond reach were it not for the assistance of computers. Computers also have a key role in such wide-ranging activities as automated sample introduction and the control of calibration steps (*robotics*). Process-independent tasks closely related to the ongoing work of a laboratory have long been delegated to computers, including the storage, retrieval, and management of data.

Nevertheless, the claim that we have entered an age of "computer-based analytical chemistry" (COBAC) is inappropriate and overly optimistic; "computer-aided" analysis would be a more satisfactory description, and one more consistent with terminology adopted in other disciplines. "Artificial intelligence," so-called expert systems [22], neural networks, and genetic algorithms will undoubtedly be increasingly important in the analytical chemistry of the future, but in most cases probably in the context of relatively complex routine investigations supported by extensive previous experience. It is unlikely that such methods will prove optimal even in the long term with respect to analytical research in uncharted waters, especially if results are required near the limit of detectability.

1.5. Analytical Tasks and Structures

1.5.1. Formulating the Analytical Problem

Generally speaking, problem-oriented analytical tasks can best be defined with reference to criteria most easily expressed as questions:

1) How has the problem at hand already been stated? Is the problem as stated truly relevant? If so, what is the maximum expenditure that can be justified for its solution, considering both material and economic resources? (Note that not every problem warrants the pursuit of an optimal analytical solution!)
2) What type and size of sample is available? What content range is predicted with respect to the analyte, and what mass of sample would be required to produce an answer?
3) What analytical strategy (including choice of a particular method) is most appropriate within the context set by considerations (1) and (2)?
4) Will critical assessment of the analytical results be possible, with evaluation of an uncertainty budget aiming to determine an *expanded uncertainty* of the analytical result [29], [30]? (see Section 1.6.2)

Ensuring the correctness of a set of results is extremely important, because nothing is more wasteful than acquiring a wrong answer, especially when account is taken of the subsequent interpretation and application of analytical data

with respect to matters of safety, health, and the environment. The ultimate validity of an analytical result can be placed in serious jeopardy as early as the sampling stage, since inappropriate sampling can be a very significant source of error.

Such mathematical tools as statistical tests and uncertainty evaluation are prerequisite to the practical application of an analytical result. In any situation involving verification of compliance with conventions, agreements, regulations, or laws, analysis is expected to provide the meaningful and objective criteria required for assessing the material facts. This means that observed analytical values must be supplemented with quality criteria applicable to the analytical procedure itself, such as the limit of detection, limit of determination, standard deviation, and measurement uncertainty.

1.5.2. Research and Application

Two major branches of analytical chemistry can be distinguished by the types of challenges they address. The first is the problem-oriented service sector, or *routine analysis*. Here one is usually in a position to rely on existing and proven methods and procedures, though some adaptation may be required to accommodate a method to the particular task at hand.

The second area, *basic analytical research*, is the key to resolving an increasingly complex set of problems today and in the future — problems not subject to attack with tools that are currently available, or amenable only to unsatisfactory solutions (with appropriate regard for economic factors). This underscores the high degree of innovative scientific character associated with analysis as a discipline, innovation that often approaches revolutionary proportions. It is perfectly possible for epochal developments to emerge from basic principles that are themselves already well established. A striking example is provided by the path leading from organic elementary analysis as first introduced by JUSTUS LIEBIG, starting with rather large samples, via the work of F. EMICH and F. PREGL, and culminating in today's highly perfected micro techniques, a path that runs parallel to the development of the analytical balance.

It is also interesting to consider in this context the source of some of today's most important innovations, which increasingly result from a close symbiotic relationship between university research centers on one hand, and commercial instrument manufacturers on the other (where the latter often

Figure 5. Strategic organization of the analytical process

at the conclusion of a joint commercialization venture are suppressed or trivialized in the interest of profit, as has unfortunately occurred on more than one occasion.

1.5.3. An Organogram

The two complementary branches of analytical chemistry rely on a common foundation of structure and content, illustrated in the "organogram" of Figure 5.

Starting with an analytical sample (the matrix), and proceeding via the formulation of a specific question regarding the state, nature, mass, concentration, or distribution of that sample, as well as a definition (or at least partial definition) of the true target of the analysis (atoms, ions, molecules, etc.), two different paths might in principle be followed in pursuit of the desired objective. Both commence with the extremely critical steps of sampling (→ Sampling) and sample preparation (→ Sample Preparation for Trace Analysis), which must again be recognized as potential sources of significant error. Under certain conditions it may then be possible to embark immediately on qualitative and/or quantitative analysis of the relevant target(s) through direct application of a physical method in the form of an "instrumental" analysis (e.g., a spectroscopic determination following excitation of the sample with photons, electrons, other charged particles, or neutrons). Such instrumental methods can be subdivided into *simultaneous* and *sequential* methods, according to whether several analytes would be determined at the same time (as in the case of multichannel optical emission spectrometry) or one after another (with the help of a monochromator).

Immediate application of a direct instrumental method (e.g., atomic spectroscopy in one of its many variants) usually represents the most economical approach to elemental analysis provided the procedure in question is essentially unaffected by the sample matrix, or if one has access to appropriate reference materials similar in composition to the substance under investigation [23]–[26]. The alternative is an analytical method consisting of multiple operations separated by either space or time, often referred to as a *multistep procedure*, as indicated on the left in Figure 5. The possibility of combining two or more discrete techniques adds a whole new dimension to chemical analysis, although there is a long tradition of observing a formal distinction between "sep-

have access to extensive in-house research facilities of their own, and may be in a position to introduce important independent initiatives). The reason for the collaborative trend is obvious: continued progress has been accompanied by a disproportionate increase in costs, and the resulting burden can no longer be borne by universities alone. Collaboration between industry and higher education is certainly to be welcomed, but not to the point that technical shortcomings still evident

aration" and true "determination." Separation in this sense has often been understood to include chemical reactions undertaken for the purpose of preparing a new, more readily separable compound—as a solid phase, for example—together with the actual separation step itself (e.g., filtration or extraction), although the term is sometimes interpreted more literally and limited to the latter activity alone. Cases also come to mind in which individual "separation" and "determination" steps cannot be clearly differentiated (e.g., in chromatography).

A separation step might be preceded by some preliminary treatment of the sample, such as a prechromatographic operation [17], and this might also warrant special attention. *Trace enrichment* is typical of the fields in which prechromatographic techniques have much to offer.

Particularly in trace analysis, and in the absence of standard samples for calibration purposes, there still is no satisfactory alternative to relying at least initially on "wet-chemical" multistep procedures. This entails a detour consisting of sample decomposition with subsequent separation and enrichment of the analyte(s) of interest relative to interfering matrix constituents. A suitable form of the analyte(s) is then subjected to the actual determination step, which may ultimately involve one or more of the direct instrumental methods of analysis.

Multistep procedures are even more indispensable in the analysis of organic substances, where a chromatographic separation is often closely coupled with the actual method of determination, such as IR or mass spectrometry. Separations based on chemical reactions designed to generate new phases for subsequent mechanical isolation (e.g., precipitation, liquid–liquid partition) have also not been completely supplanted in elemental and molecular analysis.

Recent progress in analytical chemistry is marked by dramatic developments in two areas: (1) an enormous increase in the number of available analytical methods and opportunities for applying them in combination, and (2) new approaches to mathematical evaluation (chemometrics). As a result, most matrices are now subject to characterization with respect to their components both in terms of the bulk sample and at such internal and external phase interfaces as grain boundaries and surfaces—extending in some cases even into the extreme trace range. As in the past, the safest course of action entails separating the component(s) of interest in weighable form, or taking

an indirect route via gravimetry or titrimetry as a way of establishing a state indicative of complete reaction.

Many modern methods of separation and determination result in the generation of some type of "signal", whereby an appropriate sensor or detector is expected to react in response to concentration or mass flow—perhaps as a function of time, and at least ideally in a linear fashion throughout the range of practical interest. Devices such as photocells, secondary electron multipliers, Golay cells, thermal conductivity cells, thermocouples, and flame ionization detectors convey information related to concentration changes. This information takes the form of an electrical signal (either a voltage or a current), which is fed to some type of measuring system, preferably at a level such that it requires no amplification. Sensor development is an especially timely subject, warranting extensive discussion elsewhere (→ Chemical and Biochemical Sensors).

Further processing of an analytical signal may have any of several objectives, including:

1) Incorporation of a "calibration function" that permits direct output of a concentration value
2) Establishing feedback control as one way of managing the data-acquisition process (e.g., in a *process computer*)
3) Recasting the primary signal to reflect more clearly the true analytical objective (e.g., "on-line" Fourier transformation, a common practice now in both IR and NMR spectroscopy)

1.5.4. Physical Organization of the Analytical Laboratory

Depending on the situation, assignments with which a particular analytical team is confronted might be linked organizationally and physically with the source of the samples in various ways. The following can be regarded as limiting cases:

1) Direct physical integration of the analytical function into the production or organization process, where "on-line" analysis represents the extreme
2) Strict physical separation of the sample source from subsequent analytical activities

It would be pointless to express a general preference for one arrangement or the other, but a few relevant considerations are worth examining.

Analysis "on the spot" eliminates the complications of sample transport, and it offers the potential for saving a considerable amount of time. This in turn facilitates rapid processing, an especially important factor when process control is dependent upon analytical data (e.g., in blast furnace operation). Analysis of this type is always associated with a very specific objective, usually involving a single analytical method and a single specialized type of instrumentation, and its economic viability must be critically evaluated on a case-by-case basis. Costs related to acquisition, amortization, and the repair of expensive equipment must all be considered, as must demands for personnel — who are likely to require special skills and training.

The obvious alternative to integrated analysis is a physically separate, central analytical facility like that traditionally maintained by a large chemical corporation. A laboratory of this sort typically reflects an interest in analysis in the most general sense, with provisions for the utilization of as many as possible — preferably all — of the conventional and fashionable analytical methods in anticipation of a very broad spectrum of assignments. Routine analysis in such a setting can conveniently be combined with the innovative development of new methods and procedures, thereby assuring optimal utilization of equipment that is becoming increasingly sophisticated and expensive. Considering the rapid pace of developments in major instrumentation, and the risks entailed in implementing modern approaches to automation, data processing, and laboratory operations generally, it often becomes apparent that centralization is the only economically justifiable course of action.

Similar considerations underscore the critical importance of continuing education for laboratory personnel, who must of necessity adapt to any changes in hardware. This perspective also sheds additional light on the independent scientific character of analysis, both in the industrial sphere and in academia. The problems encountered are essentially scientific in nature, the questions are fundamental, and the tools engaged in their solution reflect a complex development process that is technically demanding in the extreme.

1.5.5. The Target of Analysis

One of the fields in Figure 5 (the diagram singling out various stages in an analytical procedure) bears the label "Target of Analysis," and its structure deserves closer scrutiny. Until relatively recently the "target of an analysis" was always a list of constituent elements, together with the corresponding overall composition. An arduous trail of analytical research leads from the dualistic theory of matter (BERZELIUS and his contemporaries) to an understanding of the fine structure and conformation of molecules in the solid and liquid (dissolved) states, culminating in direct proof of the existence of atoms. In planning an analysis today it is almost self-evident that the first question to be addressed concerns the particular level in the hierarchically ordered concept "target" at which the investigation is to be conducted.

One important property of this hierarchy is that every higher level of order implies a specific set of properties at each of the lower levels. The reverse is not true, however, since the lower stages are independent and do not presuppose any higher degree of structure. Thus, in order to conduct a molecular structure determination on an organic substrate it is first necessary to ascertain the corresponding elemental composition. Needless to say, analysis at any level in the object hierarchy depends upon the availability of suitable procedures.

Atoms. As shown schematically in Figure 6, the hierarchy of targets begins with *atoms* (and the various *isotopes* of the elements) as the smallest fundamental units with analytical chemical relevance. This is already a rather profound observation in the case of certain geochemical questions, for example, since it is well known that the isotope ratios for such isotopically mixed elements as sulfur or uranium are by no means constant, and an isotope ratio (of chlorine, perhaps) can also be a useful or even indispensable parameter in the practice of mass spectrometry ($\rightarrow$ Mass Spectrometry).

Molecules. *Ions* and *functional groups* have been assigned to a level of their own, located between that of the atoms and that of the *molecules*, which represents the next formal stage in our hierarchical scheme. Chemical reactions long constituted the sole basis for analysis at the molecular level, and together with the methods of atomic and molecular spectroscopy they continue to serve as the foundation of modern analysis.

Macromolecular Species. The transition from high molecular mass substances (*macromolecules*) to the highly ordered macroscopic crystalline

| Mixtures of several phases |
| Rocks, cells |

| Higher order: phases and structural elements | |
| Crystal lattice; unit cell (ionic lattice, molecular crystals, metals) | Secondary, tertiary polypeptide structures; pleated sheets, helix, etc. |

| Macromolecules |
| $[CaSiO_3]_p$, $[-CH_2-CH_2-O-]_p$ |

| Molecules |
| HCl, H_2O, CH_4, CH_3CH_2OH |

| Ions | Functional groups |
| H_3O^+, Ca^{2+}, OH^-, SO_4^{2-} | $-CH_3$, $-OH$, $-NH_2$ |

| Atoms (elements) |
| Ca, Si, O, C, H |

| Isotopes |
| ^{14}C, ^{12}C, ^2H, ^1H |

Figure 6. Hierarchical ordering of the various possible targets of analysis

range is somewhat indistinct, and it is also here that the concept of the molecule loses its validity (ignoring for the moment the notion of molecular crystals). Qualitatively distinct properties emerge, and these can in turn be exploited for purposes of analytical characterization: external form (crystalline habit), symmetry, ionic lattice structure (as well as molecular lattices and the structures of mixed forms), mechanical and electrical characteristics, specific gravity, phenomena associated with phase formation and conversion, melting point, etc. In the course of analysis, a solid material is degraded to a *matrix*, which often turns out to be a troublesome source of interference. At this stage the solid itself is regarded as homogeneous with respect to its composition, although the structure, surface properties, and microdistribution in terms of atoms or other components might also be subjected to analysis.

Heterogeneous multiphase systems (*mixtures*) lay claim to a special status, constituting the highest level in the hierarchical classification system. The prototype for this category is a rock: granite,

for example, which is composed of three homogeneous substances. Precise identification at the mixture level typically requires the use of special techniques drawn from other disciplines, such as mineralogy or petrography.

Analysis of the Various Targets. A formal distinction can be made today between the detection and determination of free atoms, molecules, or ions on the one hand (via mass or optical spectroscopy in the gaseous state), and the analysis of molecules or ions in solution on the other (using such techniques as UV – VIS spectrophotometry, electrochemistry, and NMR or ESR spectroscopy). Mention should also be made of such "colligative" methods as ebullioscopy, cryoscopy, and gas-density determination. The overall importance of colligative methods has declined somewhat, but they continue to provide useful information regarding molecular mass and various dissociation phenomena. Their principal field of application today is the analysis of high polymers (macromolecules), the level immediately above molecules in our hierarchy of targets.

Polymeric molecules and ions have been assigned to a level of their own, primarily because of a unique set of methods that has been developed specifically for dealing with analytical targets in this size range. The most important are ultracentrifugation, certain types of liquid chromatography, light scattering, and — not to be overlooked — chemical approaches to the determination of end groups. Certain colligative methods are also very important here because of their extremely high sensitivity.

With respect to the optical analytical methods it could be claimed (cum grano salis) that *dispersive methods* — which today extend into the short-X-ray region — are increasingly being supplemented by *image-forming methods* (e.g., microscopy, polarizing microscopy, stereoscanning, electron microscopy, ultrasound microscopy). Image-forming methods appear to have reached a plateau (at least for the time being) with the quasimechanical/optical principle as manifested in scanning tunneling microscopy, which has succeeded for the first time in making atomic structures "visible." Methods for the study of surfaces [e.g., Auger spectrometry, X-ray photoelectron spectroscopy (XPS), and secondary ion mass spectroscopy (SIMS)] must also now be incorporated under the heading of general structural analysis (→ Surface Analysis).

Major advances have occurred in recent years in direct instrumental approaches in the *bulk analysis* of condensed matter, providing integrated insight into the various components comprising a complex sample; examples of such methods include X-ray fluorescence, atomic emission, atomic absorption, and atomic fluorescence spectrometry. Especially in optical atomic emission spectrometry and mass spectrometry, traditional sources of excitation such as arcs and sparks are increasingly giving way to alternative techniques: direct-current, high-frequency, and microwave plasmas; glow discharges; and lasers ($\rightarrow$ Laser Analytical Spectroscopy).

Information regarding the qualitative and quantitative distribution of elements within individual phases and at phase interfaces (grain boundaries) is the primary goal of *microdistribution analysis*, in which special probe techniques involving electrons, ions, photons, and neutrons are used to excite the sample under investigation. In the organic realm, biological cells might be singled out as the prototypical analytical substrate. A fundamental distinction must of course be made here between "nondestructive" methods and methods that concentrate on the sample as a whole.

1.6. Definitions and Important Concepts

Terminology plays an important role in analysis, and several technical terms have in fact already appeared in preceding sections of this essay. Here we consider explicitly a select subset of these terms, mainly ones whose widespread usage is relatively recent, as well as a few that are often utilized incorrectly.

1.6.1. Sensitivity, Limit of Detection, and Detection Power

These three expressions tend to be used very loosely, even among analysts — despite the fact that each is subject to very precise definition [1].

The *sensitivity E* of an analytical method expresses the dependency of a measured response on the analytical value of primary interest. It is defined as the first derivative of the measurement function:

$$F(y)E = F'(y) = \mathrm{d}x/\mathrm{d}y$$

In the case of a linear calibration function of the type $x = a_x \cdot y + b_x$, the sensitivity is equal to the slope of the calibration line; i.e.,

$$E = \Delta_x/\Delta_y = a_x$$

The *limit of detection* of an individual analytical procedure is the lowest amount of an analyte in a sample which can be detected but not necessarily quantified as an exact value [27]. Expressed as a concentration c_L or a quantity q_L, the *limit of detection* is derived from the smallest signal x_L which can be detected with reasonable certainty for a given analytical procedure. The value of x_L is given by the equation

$$x_L = x_{bl} + k s_{bl}$$

where x_{bl} is the mean of the blank measurements, s_{bl} is the standard deviation of the blank measurements, and k is a numerical factor chosen according to the level of confidence required. For many purposes, the limit of detection is taken to be $3 s_{bl}$ or 3 x signal to noise ratio, but other values of k may be used if appropriate. For a more detailed discussion see section 7.4.3. The concept "limit of detection" is applicable only to a particular analytical procedure, one that can be precisely defined with respect to all its parameters, whereas the *detection power* is a crude estimate associated with an idealized analysis, in which external interfering factors are largely ignored. This term is therefore reserved for characterizing an analytical *principle*.

The demand for analytical procedures with ever-increasing detection power is especially acute in the context of biologically relevant trace elements because of the ubiquitous concentrations of these materials in all natural matrices. It is the environmental concentrations that effectively establish lowest levels of the corresponding elements that are subject to determination in any biotic matrix. In most cases these levels are in the range > 0.1 ng/g, and thus within a region that could today be regarded as practically accessible — at least in principle. Exceptions include the concentrations of certain elements in Antarctic or Arctic ice samples, for example, or samples from research involving ultrapure sub-

stances, where the goal is to prepare materials (e.g., metals) of the highest possible purity. Relevant impurity concentrations in cases such as these may fall in the pg/g range, often far below the environmental background concentrations of the elements in question.

Establishing the trace-element content for a bulk material typically requires a sample at the upper end of the milligram or lower end of the gram range, but increasing importance is being attached to the acquisition of detailed information regarding the distribution and bonding states of elements within specific microregions (see Sections 1.6.5, 1.6.7, 1.6.14).

Studies related to construction materials and other solids frequently rely on information regarding the distribution of elements at external and internal boundary surfaces, including grain boundaries. With biotic matrices the attention may be focused on tissue compartments, individual cells, or even cell membranes, whereas an environmental analysis might be concerned with individual aerosol particles. Investigations in areas such as these — especially projects involving in situ microdistribution analysis — serve to further promote the ongoing quest for ever lower absolute limits of detection. Assume, for example, one wished to determine the elemental distribution in various protein fractions from blood serum. Blood-serum background levels are in the low ng/g range, so an analytical method would be required with a detection power $10-100$ times greater. Attempting to determine quantitatively an elemental concentration on the order of 1 ng/g in a sample weighing only 1 µg presupposes an absolute detection power in the femtogram range (1 fg $= 10_{-15}$ g).

An interesting question in this context is the minimum mass of an element that would theoretically be required for a successful determination with a given statistical degree of certainty. If the statistical error is not to exceed 1%, for example, then the root-N law of error analysis specifies that at least $10\,000$ atoms must be present for the determination to succeed. For the element zinc (atomic mass 60) this would correspond to a mass of only one attogram (10^{-18} g) [28]. Given the potential of laser spectroscopy ($\rightarrow$ Laser Analytical Spectroscopy), attaining such a goal is no longer considered utopian.

The question "Is it really necessary that analysis continue to strive for greater detection power?" must therefore be answered with an unequivocal "Yes." It is naive to suggest (as some have) that the blame for the many problems with which we are today confronted lies exclusively with the availability of increasingly powerful analytical tools. Such a biased perspective reflects at best uncritical — indeed, irresponsible — misuse of the powers of analysis, although misuse of this type can never be ruled out completely.

1.6.2. Reliability — Measurement Uncertainty

One of the first problems faced by analysts is whether a method will provide a result that is fit for its intended purpose, i.e., whether it will produce a result of the required "accuracy". A quantitative indication of the accuracy is required if the user of the result is to make any judgement on the confidence to be placed in it, or to compare it in a rational way with the results of other analyses. The statement of a result is not complete without information about the "accuracy" or the "uncertainty".

There will always be an uncertainty about the correctness of a stated result, even when all the known or suspected components of error have been evaluated and the appropriate correction factors applied, since there is an uncertainty in the value of these correction factors. In addition, there will the uncertainty arising from random effects.

Recent developments have led to the formulations of consistent and quantitative procedures for evaluating and reporting the uncertainty of a result, which are applicable in all areas of measurement. These procedures have been set out in the ISO Guide to the Expression of Uncertainty in Measurement [29] and their application to analytical chemistry is described in the EURACHEM (Cooperation for Analytical Chemistry in Europe) Guide Quantifying Uncertainty in Analytical Measurement [30].

The approach set out in the ISO Guide treats all sources of uncertainty in a consistent manner and thus avoids the difficulties encountered in some previous approaches to the evaluation of uncertainty, which treated the uncertainty arising from systematic effects in a different manner to that arising from random effects.

In essence, for chemical analysis the ISO definition of uncertainty is:

A parameter associated with the result of an analysis that characterizes the dispersion of the values that could reasonably be attributed to the concentration of the analyte.

Thus the analyst when reporting the result of an analysis is also being asked to provide a parameter that gives a quantitative indication of the range of the values that could reasonably be attributed to the concentration of the analyte. The ISO Guide recommends that this parameter should be reported as either:

A *standard uncertainty* defined as:

uncertainty of the result of a measurement expressed as standard deviation ($\rightarrow$ Chemometrics).

or as *expanded uncertainty* defined as:

a quantity defining an interval about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that could be attributed to the measurand (concentration of the analyte) and which is obtained by multiplying the *standard uncertainty* by a *coverage factor*, which in practice is typically in the range 2 – 3.

It is common practice to report the standard uncertainty using a value of 2 since this gives an interval with confidence level of approximately 95 %.

The evaluation of uncertainty requires a detailed examination of the measurement procedure. The first step is to identify the possible sources of uncertainty. The next step is to evaluate the size of the contribution from each source, or the combined contribution from a number of sources, expressed as a standard deviation. These contributions are then combined to give the standard uncertainty. Detailed examples are given in the EURACHEM Guide [30].

1.6.3. Elemental Analysis

A formal distinction between "elemental analysis" and "elementary analysis" (Section 1.6.4) is seldom carefully observed in English. *Elemental analysis* in the present context is understood to mean a determination of essentially all the elements present in a sample, irrespective of the type of bonding involved or the constitution of the matrix. Means toward that end include not only the classical methods (gravimetric analysis, titrimetry, spectrophotometry, electrochemical and kinetic methods, etc.) but also atomic spectrometric and radioanalytical methods, some of which are essentially nondestructive. From the standpoint of reliability, classical chemical methods are rarely surpassed by instrumental methods, though the latter typically do provide lower limits of detectability, and they are faster and more economical, generally offering the added potential for simultaneous multielement determination.

1.6.4. Elementary Analysis

Elementary analysis in the classical sense refers to quantitative determination of the constituent elements in an organic compound, especially carbon, hydrogen, oxygen, nitrogen, sulfur, the halogens, and phosphorus, although the definition would today be expanded to cover determination of any element present in an organic structure.

Precise characterization of organic substances became possible for the first time as a result of elementary analysis techniques developed by BERZELIUS, LIEBIG, DUMAS, KJELDAHL, and many others. The fundamental principle, which relies essentially on combustion of the material under investigation with subsequent determination of the combustion products, was adapted to microscale analysis by F. PREGL and later developed by other analysts to such a point that "microelementary analysis" now requires only a few milligrams of sample. This has led not only to shorter analysis times, but also the possibility of analyzing extremely small amounts of valuable natural substances.

Microelementary analysis has now been carried to a high degree of perfection as well as almost complete automation. Despite the advent of modern physical methods (MS, NMR, etc.) the classical techniques have lost little of their significance for synthetic chemists and biochemists, thanks mainly to their considerable advantage of providing reliable *absolute* values of actual mass proportions [31], [32].

There has been no lack of attempts to reduce requisite sample sizes even further into the lower microgram range [33], [34], but the associated techniques are often quite time-consuming, and efforts in this direction have been largely superseded by mass spectrometry, especially in its high-resolution mode. Nevertheless, work in the field of "ultramicroelementary analysis" is still of interest, especially since it has produced impressive evidence of the limits of classical microanalysis.

1.6.5. Microanalysis and Micro Procedures

"Microanalysis" is a term originally associated with classical analytical techniques capable of providing very accurate results from as little as ca. 1 mg of substance (relative error $\leq 1\%$; see the discussion of organic microelementary analysis in Section 1.6.4).

The term "micro procedure" is one of several now defined with respect to mass range by a DIN standard (DIN 32 630):

	Sample size
macro procedure	> 100 mg
semimacro procedure (also known as a semimicro procedure)	1 – 100 mg
micro procedure	1 µg – 1 mg
submicro procedure	< 1 µg

Progress in analytical chemistry in recent decades has been so extensive that most procedures could today be included in the "micro" or even "ultramicro" category, especially ones involving chromatography or spectroscopy. Under these circumstances the term "microanalysis" should now be restricted exclusively to the field of elementary analysis. "Chemically" oriented techniques for ultramicro elemental analysis in the microgram range [28] are based on extremely small reaction volumes (e.g., as little as ca. 1 µL) and correspondingly designed facilities for sample manipulation (ultramicro balances, microscopes, capillaries, and extremely tiny tools). Such methods have now been largely superseded by total reflection X-ray fluorescence (TRXRF), which permits rapid, simultaneous, quantitative determination of most elements with atomic numbers > 11 at levels extending into the lower picogram range [35], a striking example of the impressive advance in instrumental microbulk analysis (→ X-Ray Fluorescence Spectrometry).

1.6.6. Stereochemical and Topochemical Analysis

In this case the goal is to describe the target object (an atom, a molecule, or some other component of a solid phase or a solution) with respect to its spatial orientation in its surroundings. Sometimes the frame of reference is an external surface or an internal boundary surface (grain boundary). *Structural analysis* has as its ultimate objective describing all aspects of the overall structure of a particular phase, including the conformations of individual structural elements, perhaps also as a function of time. *Constitutional analysis* produces information regarding relative and absolute arrangements of atoms or atomic groupings (functional groups) within a molecule.

1.6.7. Microdistribution Analysis

Microdistribution analysis is a special type of topological analysis directed toward establishing lateral and depth distributions of the various elements making up a solid—preferably with explicit reference to the ways in which these elements are bonded, all expressed with the highest possible degree of positional resolution. Distribution with respect to phase boundaries may be important here as well. The designated goal is approached with the aid of techniques that permit beams of rays (e.g., laser photons, electrons, ions, or neutrons) to be focused extremely sharply, with a typical cross section of 1 µm or less. Alternative techniques maintain such spatial relationships as may exist within a series of signals, employing a multidimensional detector to transform the crude data into images.

Every effort is made here to achieve the highest possible absolute power of detection. Microdistribution analysis represents the primary field of application for *microprobe* techniques based on beams of laser photons, electrons, or ions, including electron microprobe analysis (EPMA), electron energy-loss spectrometry (EELS), particle-induced X-ray spectrometry (PIXE), secondary ion mass spectrometry (SIMS), and laser vaporization (laser ablation). These are exploited in conjunction with optical atomic emission spectrometry and mass spectrometry, as well as various forms of laser spectrometry that are still under development, such as laser atomic absorption spectrometry (LAAS), resonance ionization spectrometry (RIS), resonance ionization mass spectrometry (RIMS), laser-enhanced ionization (LEI) spectrometry, and laser-induced fluorescence (LIF) spectrometry [36] – [44].

1.6.8. Surface Analysis

Surface analysis is in turn a specialized form of microdistribution analysis, one that provides information on the coverage, distribution, and content of components either at the surface of a solid or in discrete layers located near the surface. Elemental

analysis in this context is conducted in a single plane with no attempt at lateral resolution, utilizing, for example, total reflection X-ray fluorescence (TRXRF), glow-discharge mass spectrometry (GDMS), or secondary neutron mass spectrometry (SNMS). Positional resolution can also be achieved with probe techniques such as Auger electron spectrometry (AES) or secondary ion mass spectrometry (SIMS), and to a limited extent with Rutherford back-scattering (RBS) and X-ray photoelectron spectroscopy (XPS) [45]–[48] (→ Surface Analysis).

1.6.9. Trace Analysis

Here we encounter another term associated with a range whose definition has changed considerably with time. A "trace" was once understood to be a no longer determinable but nonetheless observable concentration of some undesired companion substance (impurity) within a matrix. In the meantime, trace analysis has become an important and very precise field of inquiry — subject to certain restrictions with respect to the achievable reliability, but indispensable in a number of disciplines (→ Trace Analysis). There is little point in attempting to express a "trace" in terms of absolute mass units; data should instead be reported on the basis of content (concentration) in a form such as "μg/kg" (mass proportion) or "μg/L" (mass concentration). These units are to be used in preference to the very popular abbreviations "ppm," "ppb," and "ppt", which need an additional indication to the respective unit (mass, volume, amount of substance).

A warning is in order against the practice of emphasizing wide disparities in content through inappropriate comparisons that make a sensational impression at the expense of reality. Phrases like "a Prussian in Munich," "a needle in a haystack," or "a grain of wheat in a hundredweight of rye" used as metaphors for "ppm" are inconsistent with the fact that what is really at issue is a trace *proportion*; that is, a homogeneous distribution or quasicontinuum within which single individuals can be isolated only hypothetically, and only at the molecular level.

Trace constituents at levels as low as a few micrograms per ton (i.e., in a mass ratio of 1 to 10^{12}) are today subject to meaningful analytical determination thanks to highly developed methods of separation and enrichment (multistep procedures, see Chap. 1.3 and Section 1.6.11)

[49]–[52]. The principal challenge facing the trace analyst is that diminishing concentration leads to a rapid increase in systematic error [16], [52]. Extreme trace analysis with respect to the elements is therefore subject to large systematic deviations from "true" content, even though results obtained with a particular method may be quite reproducible.

Trace determinations based on atomic spectroscopy are usually matrix-dependent, *relative* methods, requiring the availability of standard reference samples for calibration. Unfortunately, no such standards yet exist for extreme ranges, so one is forced to rely instead on multistep procedures whereby trace amounts of elements of interest are excited in isolated form and within the smallest possible volume of analyte. "Limits of detectability" reported in the literature for methods of trace elemental analysis are almost invariably extrapolations based on determinations actually carried out at fairly high concentration. In no sense can these limits be regarded as reflecting real conditions owing to the problem of systematic error. Systematic errors are extremely difficult to detect, so it is advisable that one verify the validity of data acquired at each stage of the work using an alternative analytical approach. Only when data from different procedures agree within the appropriate statistical limits of error should one speak in terms of "reliable" results. Success in identifying and eliminating all the sources of error in an extreme trace analysis therefore presupposes a considerable amount of experience and a well-developed capacity for self-criticism.

Preferred methods in trace determination of the elements include atomic absorption spectrometry (AAS), optical emission spectrometry (OES) with any of a wide variety of excitation sources [e.g., sparks, arcs, high-frequency or microwave plasmas (inductively coupled plasma, ICP; microwave induced plasma, MIP; capacitively coupled microwave plasma, CMP), glow discharges (GD), hollow cathodes, or laser vaporization (laser ablation)], as well as mass spectrometry (again in combination with the various excitation sources listed), together with several types of X-ray fluorescence (XRF) analysis [51].

A special place is reserved for methods of *activation analysis*, involving slow and fast neutrons, charged particles, or photons, applied either directly or in combination with some type of radiochemical separation (Section 1.6.13). These methods quickly became almost indispensable, especially in extreme trace analysis of the ele-

ments, owing to a low risk of contamination and detection levels in at least some cases that are exceptionally favorable ($\rightarrow$ Activation Analysis).

Electrochemical methods continue to be important as well, including inverse voltammetry, coulometry, amperometry, and potentiometry ($\rightarrow$ Analytical Voltammetry and Polarography); indeed, their overall role has actually been expanded with the development of such chemical techniques as ion chromatography and chelate HPLC.

Problems associated with extreme trace analysis of the elements also affect extreme trace analyses of organic compounds, although background levels tend to be less relevant in this case [53]. All the separation methods most commonly applied to organic substances are chromatographic in nature, including thin layer chromatography (TLC; $\rightarrow$ Thin Layer Chromatography), high-performance (or high-pressure) liquid chromatography (HPLC; $\rightarrow$ Liquid Chromatography), gas chromatography (GC; $\rightarrow$ Gas Chromatography), and electrophoresis (more recently: capillary electrophoresis; $\rightarrow$ Electrophoresis), preferably combined with on-line mass or infrared spectrometry.

1.6.10. Trace Elements

The term "trace element" was first introduced in biochemistry after it became apparent during the 1920s that very low levels of certain elements in food can be important to life. Nine such elements had been identified by 1959, whereas today more than twenty different elements are regarded as essential, including several previously recognized only as toxic (e.g., arsenic, lead, and cadmium). Ambivalent physiological characteristics have now been ascribed to many elements, where toxicity may be manifested at high concentrations, but a low concentration is an absolute requirement, since a concentration even lower — or complete absence of such an element — leads directly to symptoms of illness [54]. A more appropriate descriptive term applicable in a nonbiological context might be "elemental traces."

1.6.11. Multistep Procedures

Situations frequently arise in which direct instrumental methods of analysis are inapplicable, perhaps because the corresponding detection power is insufficient, or in the case of a matrix-dependent method because no suitable calibration standards are available to correct for systematic errors. The best recourse is then a *multistep procedure*, in which actual determination is preceded by sample preparation, digestion, separation, or preconcentration steps. Individual operations within such a procedure must be linked as closely as possible, as in a "one-pot method" or one of the flow-injection or continuous-flow techniques that lend themselves so readily to automation. The goal is to concentrate an analyte from a rather large volume of solution (on the order of milliliters) for subsequent analysis on the microliter scale. Systematic errors can be minimized with on-line procedures, permitting highly reliable analysis at the picogram-per-milliliter level.

A search will also continue for elemental analysis techniques based on direct instrumental methods with enhanced powers of detection and more or less complete matrix independence. The motivation for this search goes beyond mere economic factors: direct methods are also less likely to be held hostage to blank readings, because physical sources of excitation (e.g., photons, electrons, charged particles, neutrons), which H. Malissa and M. Grasserbauer [55] characterize as "physical reagents," are essentially free of material contamination.

The trace analysis of organic substances is especially dependent on multistep procedures. In this case losses due to adsorption and vaporization are more worrisome potential sources of systematic error than elevated blank values.

1.6.12. Hyphenated Methods

This unfortunate piece of terminology is intended to emphasize the fact that multiple techniques, usually of an instrumental nature, often lend themselves to direct physical coupling, resulting in combinations whose formal designations contain hyphens (e.g., GC – MS). In contrast to the multistep procedures discussed previously, this type of combination involves a "real time" connection and true physical integration. The greatest challenge is to develop satisfactory interfaces for joining the various separation and detection systems. This particular problem is one that has long plagued the otherwise promising HPLC – MS combination.

1.6.13. Radioanalytical Methods and Activation Analysis

Methods based on nuclear reactions are restricted to laboratories specially equipped for handling radioactive substances (radionuclides) under the close supervision of trained personnel. The first important breakthroughs in trace analysis of the elements (e.g., in semiconductor applications) accompanied the development of activation analysis, which was originally based on excitation with slow or fast neutrons but later broadened to encompass the use of charged particles and photons as well. Activation methods were long held in exceptionally high esteem in the field of trace analysis, although competition eventually surfaced in the form of atomic spectroscopy. The drawbacks of activation methods (long analysis times, high cost, and current exaggerated fears with respect to radioactivity) are now perceived by many to outweigh the advantages (high detection power for many elements and relatively high reliability due to minimal complications from matrix effects or contamination), and the activation technique has recently been demoted to the status of one approach among many. This actually increases the need for stressing that activation analysis must still be regarded as an indispensable technique. Sometimes it in fact represents the only viable solution to a problem in extreme trace elemental analysis (e.g., in high-purity substance studies), whether applied directly (*instrumental* activation analysis) or — more often — in combination with radiochemical methods of separation (*radiochemical* activation analysis). Activation analysis also plays an essential role in the preparation of standard reference samples because of the fact that it is so reliable.

The same considerations apply with even greater force to the use of radioactive tracers in elemental and especially molecular analysis (→ Radionuclides in Analytical Chemistry).

1.6.14. Species Analysis (Speciation)

In biology the term *species* is used to describe a population of organisms with hereditary features that survive even after cross-breeding. The related chemical term *species analysis* was first introduced by biochemists, where a "chemical species" is understood to be a particular molecular form (configuration) of atoms of a single element or a cluster of atoms of different elements. Biologists thus define the term "species" very clearly, whereas chemists apply it in various ways:

1) For the *analysis* of a species, leading to its identification and quantification within some defined region of a sample (*speciation*)
2) For describing the *abundance* or *distribution* of various species of an element within a particular volume
3) In conjunction with the *reactivity* of a given species
4) With respect to the *transformation* of one species into another

Accordingly, various *categories* of chemical species can also be distinguished, including "original" or "conceptional" species, "matrix" species, and "analyte" species.

A species is said to be *original* or *conceptional* if it is resistant to change in contact with other matrices. If a chemical change does occur as a result of such contact, the material is called a *matrix* species. The third category refers to a species that undergoes a change during the course of an analytical procedure, in which case it is some new species (the *analyte* species) that becomes the subject of analytical detection. Consider, for example, the original (conceptional) species CH_3Hg^+, which in soil forms the matrix species CH_3Hg- humic acid, but is subsequently determined as the analyte species CH_3HgCl. Transformations of this type are a major source of the considerable challenge posed by species analysis relative to determining the total content of an element. Transformation may well begin as early as the sampling process, continuing throughout the period of sample storage and at the time of the analysis itself, so that original species present in situ and in vivo are never actually detected [56], [57]. Above all, in both environmental analysis and toxicology, detection of the original species is becoming more and more important (→ Sample Preparation for Trace Analysis).

1.6.15. Chemometrics

Chemometrics is the field encompassing those aspects of chemical analysis associated directly with measurement techniques, especially principles underlying the various types of detection. Opinions differ with respect to the meaning, purpose, and limitations of this discipline, but a relatively clear set of ideas is beginning to prevail. According to K. DOERFFEL et al. [58], chemomet-

rics is concerned with evaluation of observed analytical data with the aid of mathematics, especially statistical methods. Chemometric methods facilitate the extraction of useful information even when the noise level of a signal is high, as well as the establishment of relationships linking multiple observations even when the results themselves seem widely divergent. A good example is provided by the mathematical resolution of partially overlapping signals ("peaks") in a chromatogram.

The tools of chemometrics encompass not only the familiar (univariant) methods of statistics, but especially the various multivariant methods, together with a package of "pattern-recognition" methods for time-series analyses and all the known models for signal detection and signal processing. Chemometric methods of evaluation have now become an essential part of environmental analysis, medicine, process analysis, criminology, and a host of other fields.

Chemometric methods have also been adapted to the development of labor-saving analytical strategies—the establishment, for example, of the ideal sampling frequency in a process analysis, or simplification of a multicomponent analysis so that it reflects only the truly relevant features. In addition, chemometrics plays an important part in quality-assurance programs directed toward analytical investigations. In the future, chemometrics should make a valuable contribution to the design of "legally binding" analyses with statistically assured results [59]–[63] (→ Chemometrics; see also Chap. 1.7).

1.6.16. DNA Analysis

DNA techniques are already being applied in a number of areas of analysis such as human health, identification of sex in certain species, personal identification, environmental and food analysis. The Polymerase Chain Reaction (PCR) is one of the most powerful methods of producing material for analysis from very small samples and can achieve up to a 10^6-fold increase in the target DNA. For example, with this technique it is possible to detect a range of pathogenic micro-organisms with a sensitivity which is orders of magnitude greater than previously achievable, and that is beyond the limits required for public health. It is also a very powerful technique for checking for food adulteration and food speciation. Applications of this technique are growing very rapidly

and it could be one of the most important of the recent developments in analytical science.

1.7. "Legally Binding Analytical Results"

The recently coined phrase "legally binding analytical results" has been enthusiastically adopted by numerous authorities in response to problems raised in the administration of justice by such statistically sophisticated concepts as "confidence coefficient," "confidence interval," and the like. Reproducibility has become the primary criterion applied to analytical test results in a legal setting, not necessarily "correctness." This may appear to be an unscientific development, but it probably must be tolerated, at least within reasonable limits. In any case, this problem has been a subject of intense debate in recent years, offering the promise of welcome changes in the foreseeable future ("the theory of legal substantiation").

1.8. References

[1] K. Danzer, E. Than, D. Molch, L. Küchler: *Analytik-Systematischer Überblick*. 2nd ed., Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig 1987.

[2] R. Bock: *Methoden der Analytischen Chemie*. **vol. 1: "Trennungsmethoden, " vol. 2: parts 1, 2, and 3: "Nachweis- und Bestimmungsmethoden,"** VCH Verlagsgesellschaft, Weinheim 1974–1987.

[3] H. Kienitz et al. (eds.): *Analytiker-Taschenbuch* **vols. 1–21,** Springer-Verlag, Berlin 1980–1999.

[4] G. Svehla (ed.): *Wilson and Wilson's Comprehensive Analytical Chemistry*, **vols. 1–28,** Elsevier, Amsterdam 1959–1991.

[5] *Ullmann's Encyclopedia of Industrial Chemistry*. 5th ed., **Vol. B5, B6,** Wiley-VCH, Weinheim 1994.

[6] R. Kellner, J.-M. Mermet, M. Otto, H. M. Widmer (eds.): *Analytical Chemistry*, Wiley-VCH, Weinheim 1998.

[7] K. Cammann, *Fresenius J. Anal. Chem.* **343** (1992) 812–813.

[8] M. Valcarcel, *Fresenius J. Anal. Chem.* **343** (1992) 814–816.

[9] C. R. Fresenius: *Anleitung zur qualitativen chemischen Analyse*, 12th ed., Vieweg u. Sohn, Braunschweig 1866, p. 4.

[10] F. Szabadvary: *Geschichte der analytischen Chemie*. Vieweg u. Sohn, Braunschweig 1966.

[11] W. Ostwald: *Die wissenschaftlichen Grundlagen der analytischen Chemie*, Leipzig 1894.

[12] Ullmann, 3rd ed., vol. 2/1.

[13] H. Malissa, *Fresenius J. Anal. Chem.* **337** (1991) 159.

[14] H. Malissa, *Fresenius J. Anal. Chem.* **343** (1992) 836.

[15] K. Danzer, *Mitteilungsblatt der Fachgruppe Analytische Chemie der GDCh* **4/1992** M 104 – M 110.
[16] G. Tölg, *Naturwissenschaften* **63** (1976) 99.
[17] W. Dünges: *Prächromatographische Mikromethoden,* Hüthig-Verlag, Heidelberg 1979.
[18] W. Richter in *Report on the Comité Consultatif pour la Quantité de Matière,* 1st meeting, Paris 1995.
[19] X. R. Pan, *Accred. Qual. Assur.* **1** (1996) 181–185.
[20] P. De Bièvre, *Accred. Qual. Assur.* **3** (1998) 481.
[21] W. Wegscheider, *Accred. Qual. Assur.* **4** (1999) 478–479.
[22] J. W. A. Klaessens, G. Kateman, B. G. M. Vanderginste, *TrAC Trends Anal. Chem.* **4** (1985) 114.
[23] B. Griepink, *Fresenius J. Anal. Chem.* **337** (1990) 812.
[24] Ph. Quevauviller, B. Griepink: Reference Materials in Quality Assurance. In H. Günzler (ed.): *Accreditation and Quality Assurance in Analytical Chemistry,* Springer, Berlin, Heidelberg 1996.
[25] B. Griepink, *Fresenius J. Anal. Chem.* **338** (1990) 360–362.
[26] B. Griepink, E. A. Maier, Ph. Quevauviller, H. Muntau, *Fresenius J. Anal. Chem.* **339** (1991) 599–603.
[27] J. Fleming, H. Albus, B. Neidhart, W. Wegscheider, *Accred. Qual. Assur.* **2** (1997) 51–52.
[28] G. Tölg in G. Svehla (ed.): *Wilson and Wilson's Comprehensive Analytical Chemistry,* **vol. III,** Elsevier, Amsterdam 1975, pp. 1 – 184.
[29] *ISO Guide to the Expression of Uncertainty in Measurement,* ISO, Geneva, Switzerland, 1993.
[30] The Quantification of Uncertainty in Chemical Analysis. Available for down loading from the EURACHEM web site (www.vtt.fi/ket/eurachem).
[31] F. Ehrenberger, S. Gorbach: *Quantitative organische Elementaranalyse,* VCH Verlagsgesellschaft, Weinheim 1991.
[32] T. S. Ma, R. C. Rittner: *Modern Organic Elemental Analysis,* Marcel Dekker, New York 1979.
[33] G. Tölg: *Ultramicro Elemental Analysis,* Wiley-Interscience, New York 1970.
[34] W. J. Kirsten: *Organic Elemental Analysis – Ultramicro, Micro, and Trace Methods,* Academic Press, New York 1983.
[35] G. Tölg, R. Klockenkämper, *Spectrochim. Acta Part B* **48 B** (1993) 111 – 127.
[36] K. Kiss: *Problem Solving with Microbeam Analysis,* Elsevier, Amsterdam 1988.
[37] E. Fuchs, H. Oppolzer, H. Rehme: *Particle Beam Microanalysis, Fundamentals, Methods and Applications,* VCH Verlagsgesellschaft, Weinheim 1990.
[38] J. C. Vickerman, A. E. Brown, N. M. Reed: *Secondary Ion Mass Spectrometry: Principles & Applications,* Oxford University Press, Oxford 1990.
[39] A. Benninghoven et al. (eds.): "Secondary Ion Mass Spectrometry SIMS VII," *Proceedings of the 7th International Conference on Secondary Ion Mass Spectrometry,* J. Wiley & Sons, Chichester 1990.
[40] M. Grasserbauer, H. W. Werner (eds.): *Analysis of Microelectronic Materials and Devices,* J. Wiley & Sons, Chichester 1991.
[41] L. Moenke-Blankenburg: "Laser Micro Analysis," *Chemical Analysis,* vol. 105, J. Wiley & Sons, New York 1989.

[42] K. Niemax in: *Analytiker-Taschenbuch,* vol. 10, Springer-Verlag, Heidelberg 1991, pp. 1 – 28.
[43] J. Uebbing, A. Ciocan, K. Niemax, *Spectrochim. Acta Part B* **47 B** (1992) 601.
[44] C. M. Miller, J. E. Parks (eds.): "Resonance Ionization Spectroscopy 1992," *Inst. Phys. Conf. Ser.* **128,** Institute of Physics Publishing, Bristol 1992.
[45] J. M. Watts: *Methods of Surface Analysis,* C.K.P., Cambridge 1989.
[46] M. Grasserbauer, *Philos. Trans. R. Soc. London* **A 333** (1990) 113.
[47] J. C. Riviere: *Surface Analytical Techniques,* Oxford University Press, Oxford 1990.
[48] D. Briggs, M. P. Seah: *Practical Surface Analysis,* 2nd ed., J. Wiley & Sons, Chichester, "Auger and X-Ray Photoelectron Spectroscopy," vol. 1, 1990; "Ion and Neutral Spectroscopy," vol. 2, 1992.
[49] A. Mizuike: *Enrichment Techniques for Inorganic Trace Analysis,* Springer-Verlag, Heidelberg 1983.
[50] J. Minczewski, J. Chwasstowska, R. Dybczinski: "Separation and Preconcentration Methods," in *Inorganic Trace Analysis,* Ellis Horwood Ltd., Chichester 1982.
[51] G. Tölg, *Anal. Chem. Acta* **238** (1993) 3 – 18.
[52] G. Tölg, P. Tschöpel: "Systematic Errors in Trace Analysis," in Z. B. Alfassi (ed.): *Determination of Trace Elements,* VCH, Weinheim, 1994.
[53] K. Beyermann: "Organische Spurenanalyse," in H. Hulpke, H. Hartkamp, G. Tölg (eds.): *Analytische Chemie für die Praxis,* Thieme-Verlag, Stuttgart 1982.
[54] G. Tölg, in H. Malissa, M. Grasserbauer, R. Belcher (eds.): *Nature, Aim and Methods of Microchemistry,* Springer-Verlag, Wien 1981, p. 203.
[55] M. Grasserbauer, *Angew. Chem.* **93** (1981) 1059.
[56] M. Bernhard, F. F. Brinckman, K. J. Irgolic: "The importance of Chemical `Speciation' in Environmental Processes," in M. Bernhard, F. E. Brinckman, P. J. Sadler (eds.): *Dahlemkonferenzen 1984,* Springer-Verlag, Heidelberg 1986.
[57] Group Report, Importance and Determination of Chemical Species in Biological Systems in: "The Importance of Chemical `Speciation' in Environmental Processes," in M. Bernhard, F. E. Brinckman, P. J. Sadler (eds.): *Dahlemkonferenzen 1984,* Springer-Verlag, Heidelberg 1986, pp. 17 – 38.
[58] K. Doerffel, K. Danzer, G. Ehrlich, M. Otto, *Mitteilungsbl. Chem. Ges. DDR* **31** (1984) 3.
[59] D. L. Massart et al.: *Chemometrics: A Textbook,* Elsevier, Amsterdam 1988.
[60] M. A. Sharaf, D. L. Illman, B. R. Kowalski: *Chemometrics,* J. Wiley, New York 1986.
[61] R. G. Brereton: *Chemometrics. Applications of Mathematics and Statistics to Laboratory Systems,* Ellis Horwood, Chichester 1990.
[62] B. G. M. Vandeginste, *Fresenius J. Anal. Chem.* **337** (1990) 786.
[63] St. J. Haswell: *Practical Guide to Chemometrics,* Marcel Dekker, New York 1992.

2. Quality Assurance in Instrumentation

LUDWIG HUBER, Agilent Technologies GmbH, P.O. Box 1280, D-76337 Waldbronn, Germany

2.1. Introduction

Analytical instruments play a major role in the process to achieve high quality and reliable analytical data. Thus everyone in the analytical laboratory should be concerned about the quality assurance of equipment. Quality standards usually applied in analytical laboratories, such as the ISO/IEC 17025 [1] EN 45001 [2], and the NAMAS [3] accreditation standard, stipulate that all instruments used must be adequately designed, well maintained, calibrated, and tested. Furthermore, regulations, principles and directives concerning laboratory work, such as Good Laboratory Practice (GLP) principles and regulations [4], [5] and Good Manufacturing Practice (GMP) [6] directives and regulations, include chapters that specifically deal with equipment.

Unfortunately quality standards and regulations are not specific enough to give clear guidelines on what to do on a day-to-day basis. Owing to this lack of clarity, most laboratories found they had to interpret these standards and regulations themselves. Private and public auditors and inspectors experienced similar problems and there have also been situations whereby the regulations have been interpreted differently by different inspectors.

Nevertheless, there has been some improvement. Inspection guides have been developed by regulatory agencies for use by inspectors, thus gaining a common understanding from the regulatory side. Interpretation guides have been developed by private organizations, providing a common understanding amongst accreditation bodies and users of equipment. Examples of such documents are *The Development and Application of Guidance on Equipment Qualification of Analytical Instruments* [7] and the U.S. *FDA Guide to Inspection of Pharmaceutical Quality Control Laboratories* [8].

This chapter discusses aspects of quality assurance of equipment as used in analytical laboratories. It provides guidelines on how to select a vendor and for installation and operational qualifications, ongoing performance control, maintenance, and error detection and handling that contribute to assuring the quality of analytical laboratory data. It refers mainly to an automated chromatography system as an example, but similar principles can be applied to other instrumentation.

It is not the scope of this chapter to discuss quality measures as applied during development and manufacturing of equipment hardware and software. This cannot be directly influenced by the user. Details on this topic can be found in

We herewith certify that the product/system

Product No: 79855 A/B
Serial No: 1456G3345

successfully passed all our production quality tests.

During final instrument performance verification
the following functional characteristics were
individually tested for conformance with our internal
specifications:

Leak sensor function: [✓]
Solvent leaks: [✓]
Air leaks: [✓]
Sampling precision: [✓]
Operating test: [✓]

Date: Nov. 20. 1992
Test Technician: Schmidt

Declaration of
Conformity

Compliance with EN 45014
and ISO Guide 22

Figure 1. A declaration of conformity according to EN 45014 should be shipped with analytical instruments that show evidence of compliance to documented manufacturing specifications (from reference 9)

published literature [9], [10] and more recently online on the Internet [11].

2.2. Selecting a Vendor

For instruments purchased from a vendor, the quality process starts with the definition of specifications and the selection of the vendor. It is recommended to select vendors recognized as having quality processes in place for instrument design, development, manufacturing, testing, service, and support: for example ISO 9001 registration. Other criteria are the capability of the vendor to provide help in meeting the quality standards' requirements in the users laboratory. As examples, the vendor should provide operating procedures for maintenance and documentation should be available with guidelines on how to test the equipment in the user's environment. For more complex equipment the vendor should provide preventative maintenance and performance verification services in the user's laboratory. Instruments should be selected which have built-in features calibration, for self-diagnosis and for on-site and remote troubleshooting. Software should be available to do the required ongoing performance qualification automatically.

The user should get assurance from the vendor that software has been validated during its development process and that documented quality principles have been applied during manufacturing and testing of the equipment.

A "Declaration of Conformity" should be shipped with all instruments to document that the instrument operated within specification when it was shipped from the factory (Fig. 1). The "Declaration of Conformity" should be an extract from detailed and comprehensive test documentation and include the following information:

– The name and the address of the supplier
– Clear identification of the product (name, type and model number)
– Place and date the declaration was issued
– Name and signature of the supplier's authorized person
– Listing of tested items and check boxes with pass/fail information

Computer systems and software products should be supplied with declarations documenting the evidence of software development validation. The user should also get assurance that development validation procedures and documents can be made available to the user. Critical formulae used in the analytical process should be documented in the user's operating manual.

2.3. Installation and Operation of Equipment

To put equipment in routine operation requires three steps:

1) Preparing the site for installation
2) Installation of hardware and software
3) Operational, acceptance, and performance testing

It is important to do both operational and acceptance tests in the user's environment, even if these tests have been done before installation at the vendors location either as individual modules or as a complete system. Before and during routine use of the system, the performance and suitability of the complete system for the intended use should be verified.

2.3.1. Setting Specifications

Setting the right specifications should ensure that instruments have all the necessary functions and performance criteria that will enable them to be successfully implemented for the intended application and to meet business requirements. This process is also called Design Qualification (DQ). Errors in DQ can have a tremendous technical and business impact, and therefore a sufficient amount of time and resources should be invested in the DQ phase. For example, setting wrong operational specifications can substantially increase the workload for testing.

While IQ (Installation Qualification), OQ (Operational Qualification) and PQ are being performed in most regulated laboratories, DQ is a relatively new concept to many laboratories. It is rarely performed in those cases where the equipment is planned to be used not for a specific but for multiple applications.

This phase should include:

– Description of the analysis problem
– Description of the intended use of the equipment
– Description of the intended environment
– Preliminary selection of the functional and performance specifications (technical, environmental, safety)
– Instrument tests (if the technique is new)
– Final selection of the equipment
– Development and documentation of final functional and operational specifications

To set the functional and performance specifications, the vendor's specification sheets can be used as guidelines. However, it is not recommended to simply write up the vendor's specifications because compliance to the functional and performance specifications must be verified later on in the process during operational qualification and performance qualification. Specifying too many functions and setting the values too stringently, will significantly increase the workload for OQ.

2.3.2. Preparing for Installation

Before the instrument arrives at the user's laboratory, serious thought must be given to its location and space requirements. A full understanding of the new equipment has to be obtained from the vendor well in advance: required bench or floor space and environmental conditions such as humidity and temperature, and in some cases utility needs such as electricity, compressed gases for gas chromatographs, and water. Care should be taken that all these environmental conditions and electrical grounding are within the limits as specified by the vendor and that correct cables are used. Any special safety precautions should be considered, for example for radioactivity measurement devices and the location should also be checked for any devices generating electromagnetic fields nearby. For a summary of all points needing to be considered, see Table 1.

2.3.3. Installation

Once the instrument arrives, the shipment should be checked, by the user for completeness. It should be confirmed that the equipment ordered is what was in fact received. Besides the equipment hardware, other items should be checked for example correct cables, other accessories, and documentation. A visual inspection of the entire hardware should follow to detect any physical damage. For more complex instrumentation, wiring diagrams should be generated, if not obtained from the vendor. An electrical test of all modules and the system should follow. The impact of electrical devices close to the computer system should be considered and evaluated if a need arises. For example, when small voltages are sent between sensors and integrators or computers, electromagnetic energy emitted by poorly shielded nearby

Table 1. Steps towards routine use of instruments (from reference 9)

Before Installation
- Obtain manufacturer's recommendations for installation site requirements
- Check the site for compliance with the manufacturer's recommendations (space, environmental conditions, utilities such as electricity, water, and gases)
- Allow sufficient shelf space for SOPs, operating manuals, and disks

Documents
- ☑ Manufacturer's recommended site preparation document
- ☑ Check-list for site preparation

Installation (installation qualification)
- Compare equipment as received with purchase order (including software, accessories, spare parts, and documentation such as operating manuals and SOPs)
- Check equipment for any damage
- Install hardware (computer, equipment, fittings and tubings for fluid connections, columns in HPLC and GC, cables for power, data flow and instrument control cables)
- Install software on computer following the manufacturer's recommendation
- Make back-up copy of software
- Configure peripherals, e.g., printers and equipment modules
- Evaluate electrical shielding (are their sources for electromagnetic fields nearby?)
- Identify and make a list of all hardware
- Make a list of all software installed on the computer
- List equipment manuals and SOPs
- Prepare installation reports
- Train operator

Documents
- ☑ Copy of original purchase order
- ☑ System schematics and wiring diagrams
- ☑ Equipment identification forms (in-house identification, name and model, manufacturer, serial number, firmware and software revision, location, date of installation)
- ☑ List of software programs with software revisions, disk storage requirements and installation date
- ☑ Accessory and documentation checklist
- ☑ Installation protocol

Pre-operation
Examples:
- Passivate gas chromatograph if necessary
- Flush HPLC fluid path with mobile phase
- Verify wavelength accuracy of HPLC UV/visible detectors and recalibrate, if necessary

Documentation
- ☑ Notebook and/or logbook entries

Operation (acceptance testing, operational qualification)
- Document anticipated functions and operational ranges of modules and systems
- Perform basic functions of the application software, for example, integration, calibration, and reporting using data files supplied on disk (for details see reference 9)
- Perform basic instrument control functions from both the computer and from the instrument's keyboard, for example, for an HPLC system switch on the detector lamp and the pump and set different wavelengths and flow rates
- Test the equipment hardware for proper functioning
- Document all the operation tests
- Sign the installation and operation protocol

Documentation
- ☑ Specifications on intended use and anticipated operating ranges
- ☑ Test procedures of the computer system with limits for acceptance criteria and templates with entry fields for instrument serial number, test results, corrective actions in case the criteria are not met and for printed name and signatures of the test engineer
- ☑ Procedures with templates for operational testing of equipment hardware describing the test details, acceptance criteria, and the definition of operational limits within which the system is expected to operate and against which the testing should be performed
- ☑ Signed installation and operational protocols. If the installation is made by a vendor firm's representative, the protocol should be signed by the vendor and the user firm's representative

After installation
- Affix a sticker on the instrument with the user firm's asset number, manufacturer's model and serial number, and firmware revision
- Develop a procedure and a schedule for an on-going preventative maintenance, calibration, and performance verification
- Prepare and maintain a logbook with entries of instrument problems

Table 1. continued

Documentation	
☑	Instrument sticker
☑	Operating procedures and schedules for preventative maintenance, calibration, and performance verification and procedures for error handling
☑	Notebook and/or logbook entries

On-going performance control (performance qualification, system suitability testing, analytical quality control) for routine analysis

●	· Combine instrumentation with analytical methods, columns, reference material into an analysis system suitable to run the unknown samples
●	Define type and frequency of system suitability tests and/or analytical quality control (AQC) checks
●	Perform the tests as described above and document results
●	Develop procedures for definition of raw data and verification of raw data and processed data (for GLP/GMP)

Documentation	
☑	Test protocols
☑	Data sheets with acceptance criteria for system performance test results
☑	System performance test documents
☑	Quality control charts
☑	SOPs for definition of raw data and verification of processed data (for GLP/GMP)

Table 2. Form for computer system identification

Computer hardware
   Manufacturer
   Model
   Serial number
   Processor
   Coprocessor
   Memory (RAM)
   Graphics adapter
   Video memory
   Mouse
   Hard-disk
   Installed drives
   Space requirement
Printer
   Manufacturer
   Model
   Serial number
   Space requirement
Operating software
   Operating system (version)
   User interface (version)
Application software
   Description
   Product number (version)
   Required disk space

fluorescent lamps or by motors can interfere with the transmitted data.

The installation should end with the generation and sign off of the installation report, in pharmaceutical manufacturing referred to as the Installation Qualification (IQ) document. It is recommended to follow documented procedures with checklists for installation and to use pre-printed forms for the installation report.

When the installation procedure is finished the hardware and software should be well documented with model, serial, and revision numbers.

For larger laboratories with lots of equipment this should be preferably a computer based data base (Table 2). Entries for each instruments should include:

– In-house identification number
– Name of the item of equipment
– The manufacturers name, address, and phone number for service calls, service contract number, if there is any
– Serial number and firmware revision number of equipment
– Software with product and revision number
– Date received
– Date placed in service
– Current location
– Size, weight
– Condition, when received, for example, new, used, reconditioned List with authorized users and responsible person

It is recommended to make copies of all important documentation: one should be placed close to the instrument, the other one should be kept in a safe place. A sticker should be put on the instrument with information on the instrument's serial number and the companies asset number.

2.3.4. Logbook

A bound logbook should be prepared for each instrument in which operators and service technicians record all equipment related activities in a chronological order. Information in the logbook may include:

HPLC instrument verification report

Test method: C\HPCEM\1\VERIF\Check.M
Data file directory: C\HPCEM\1\VERIF\Result.D
Original operator: Dr. Watson

Test item	User limit	Actual	Com
DAD noise	<5 x 10-5 AU	1 x 10-5 AU	pass
Baseline drift	<2 x 10-3 AU/hr	1.5 x 10-4 AU/hr	pass
DAD WL calibration	± 1 nm	± 1 nm	pass
DAD linearity	1.5 AU	2.2 AU	pass
Pump performance	<0.3% RSD RT	0.15% RSD RT	pass
Temp. stability	±0.15°C	±0.15°C	pass
Precision of peak area	<0.5°C RSD	0.09% RSD	pass

Verification test overall results   pass
HP 1100 series system, Friday, November 14, 1997
Test engineer
Name: Signature:

Figure 2. Result of an automated HPLC hardware test. The user selects test items from a menu and defines acceptance limits. The instrument then performs the tasks and prints the actual results together with the limits, as specified by the user

– Logbook identification (number, valid time range)
– Instrument identification (manufacturer, model name/number, serial number, firmware revision, date received, service contact)

Column entry fields for dates, times, and events, for example, initial installation and calibration, updates, column changes, errors, repairs, performance tests, quality control checks, cleaning and maintenance plus fields for the name and signature of the technician making the entry.

2.3.5. Operation

After the installation of hardware and software an operational test should follow, a process which is referred in pharmaceutical manufacturing as Operational Qualification (OQ). The goal is to demonstrate that the system operates "as intended" in the user's environment.

For an automated GC or GC/MS system operational testing can mean, for example, verifying correct communication between the computer and the equipment but also checking the detector response and the precision of the retention times and peak areas. Vendors should provide Operating Procedures for the tests, limits for acceptance criteria, and recommendations in case these criteria cannot be met.

The words "intended ranges" in the definition of operational qualification is important. This means that the instrument does not need to be tested for all functions or the instrument manufacturers specifications, if not all functions of the instrument will be used or if it will not be used within all the instrument's limits. This also means that the user should specify the "intended range" before the tests for the operational qualification begin. For example, an HPLC pump may have been purchased with a 4 channel proportioning valve to switch automatically between different mobile phases. Even though in this case the pump is designed to perform gradient analysis, the gradient performance does not need to be tested if the pump will always be used for isocratic runs. Or if an HPLC UV/visible detector will always be used to measure relatively high concentrations, it is not required to measure the performance close to the detection limit of the instrument.

This also has an impact on the required maintenance efforts. To keep the baseline noise of a UV/visible detector within the manufacturers specification will require frequent cleaning of the flow cell, it will require ultra-pure mobile phases and the lamp may have to be changed frequently. For applications not requiring the greatest sensitivity this effort can be reduced if the acceptance limit for the baseline noise is set higher.

For the systems which consist of several hardware modules such as HPLC systems, operational testing may be done for the entire system or for each module (Fig. 2). Complete system testing is

referred as holistic testing, testing each distinct, important part a computerized analytical system is called modular testing [12].

For this discussion it must be noted that for testing most of the items such as the flow rate precision, injection volume precision or the detector's linearity, more than one HPLC module is required. For example, in practice a pumping system and a detector are always needed for testing the precision of the autosampler. On the other hand, an injection system is required for the testing of the pumps precision, which is usually measured by injecting a series of standards and measuring the standard deviation of the peaks retention time. Some important characteristics are influenced by several parts of the system. An example is the precision of the peak area. This is influenced by the repeatability of the injection volume and by the stability of the solvent delivery system. In some instances especially when using a standard yielding a peak signal to noise ratio of 100 and below, the integration repeatability may determine the system precision. Because of these interdependencies testing using the holistic approach is preferred.

If the system does not perform as expected, individual modules may be interchangeably used to identify the source of the system problem, which means modular testing is recommended for troubleshooting purposes.

The installation and operational tests of equipment purchased from a vendor can be done by either a representative of the vendor's firm or by the user's firm. In any case, the installation should follow written protocols and should be done following the manufacturers recommendations regarding installation and testing.

2.4. Qualification of Software and Computer Systems

Correct function of software loaded on a computer system should be checked in the user's laboratory under typical operating conditions. Usually software and computer systems provide more functionality than those functions of the system which are used in the laboratory. Only those functions of the system which will be used should be tested. Therefore as the first step the intended functions of the system should be documented. Other documentation of such testing should include test conditions, acceptance criteria, sum-

mary of results, and names and signatures of persons who performed the tests.

Preferably such test and documentation of test results should be done automatically using always the same set of test files. In this way users are encouraged to perform the tests more frequently and user specific errors are eliminated. In some cases vendors provide test files and automated test routines for verification of a computer system's performance in the user's laboratory. For example, Agilent provides for the data systems such files and test automated test routines.

The procedure can be used to verify the performance of data systems for formal acceptance testing, operational qualification, or requalification:

- At installation
- After any change to the system (computer hardware, software updates)
- After hardware repair
- After extended use

Successful execution of the procedure ensures that

- Executed program files are loaded correctly on the hard disk
- The actual computer hardware is compatible with the software
- The actual version of the operating system and user interface software is compatible with the data system software

The method tests key functions of the system such as peak integration, quantitation, printout, and file storage and retrieval.

Test chromatograms derived from standards or real samples are stored on disk as the master file. Chromatograms are supplied as part of the software package or can be recorded by the user. This *master data file* goes through normal data evaluation from integration to report generation. Results are stored on the hard disk. The same results should always be obtained when using the same data file and method for testing purposes. The software includes a routine to carry out the performance tests and verification automatically.

To generate the master file, the user selects and defines a *master chromatogram* from a file menu and defines a method for integration and calibration. The instrument will perform the integration of the peaks using the specified method and stores the results in a checksum-protected, binary data file for use as a reference in later verifications. For
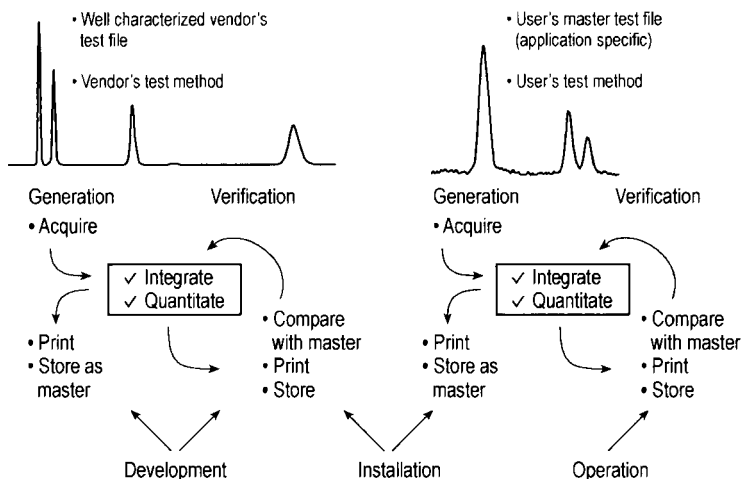
Figure 3. Verification process of chromatographic software. A standard chromatogram is generated by the vendor and well characterized and documented while the software is being developed. The results are verified at installation to make sure that the system works as intended during installation. The user can then generate his/her own application-specific "master" chromatograms to verify the ongoing performance of the system (from reference 9)

performance verification, the user again selects the same master chromatogram and the same method. The program integrates and evaluates peak data, compares it with the master set stored on disk, then prints a report informing the user of the successful verification of the data system (Fig. 3).

The advantages of this method are several fold.

1) A user can select one or more data files which are representative of the laboratory's samples. For example, one file may be a standard with nicely separated peaks which cover a wide range of calibration concentrations, useful to verify the accuracy and the linearity of the integrator. Second or third chromatograms can be used to verify the integrator's capability to integrate real-life samples reproducibly. Chromatograms may be selected with:
   – Peaks close to the detection limits
   – Peaks with shoulders
   – Poorly resolved peaks
   – Peaks on drifting baselines
   – Peaks with tailing
2) The entire process is fully automated, avoiding erroneous results and encouraging users to perform the test more frequently.
3) Results of the verification are documented in such a way that the documentation can be used directly for internal reviews and external audits, assessments and inspections, see for example Figure 4.

2.5. Routine Maintenance and Ongoing Performance Control

When the installation is finished and the equipment and the computer system are proven to operate well, the computerized system is put into routine analysis. Procedures should exist which show that *"... it will continue to do what it purports to do"*.

The characteristics of analytical hardware changes over time due to contamination, and normal usage of parts. Examples are the contamination of a flow cell of a UV detector, the abbreviation of the piston seal of a pump or the loss of light intensity of a UV detector. These changes will have a direct impact on the performance of analytical hardware. Therefore the performance of analytical instruments should be verified during the entire lifetime of the instrument.

Each laboratory should have a quality assurance program which should be well understood and used by individuals as well as by laboratory organizations to prevent, detect, and correct problems. The purpose is to ensure that the results have a high probability of being of acceptable quality. Ongoing activities may include preventative instrument maintenance, performance verification and calibration, system suitability testing, analysis of blanks and quality control samples, and ensuring system security. A plan should be set up to

```
=============================
Data System Verification Report
=============================
```

Tested configuration

Component	Revision	Serial number
Diode-array detector	1.0	3148G00859
HPLC 3D data system	Rev. A.02.00	N/A
Microsoft Windows	3.10 (enhanced mode)	N/A
MS-DOS	5.0	N/A
Processor	i486	N/A
CoProcessor	yes	N/A

Data system *verification test details*

Test name	: C:\HPCHEM\1\VERIFY\CHECK01.VAL
Data file	: C:\HPCHEM\1\VERIFY\CHECK01.VAL\DEMODAD.D
Method	: C:\HPCHEM\1\VERIFY\CHECK01.VAL\DEMODAD.M
Original acquisition Method	: PNASTD
Original operator	: Hans Obbens
Original injection date	: 08/02/1985
Original sample name	: AIRTEST

Signals tested

Signal 1 : DAD B, SIG=305, 190 Ref=550,100 of DEMODAD.D
Signal 1 : DAD B, SIG=270, 4 Ref=550,100 of DEMODAD.D
Signal 1 : DAD B, SIG=310, 4 Ref=550,100 of DEMODAD.D

Data system *verification test results*

Test module	Selected for test	Test result
Digital electronics test	yes	pass
Integration test	yes	pass
Quantification test	yes	pass
Print analytical report	yes	pass

Data system verification test overall results: pass

HP 1050 LC System, Friday, June 18, 1993 12:16:55 PM by _____

Page 1 of 1

Figure 4. A data system verification report

properly calibrate and verify the performance of the complete system after system updates and after repairs and to check the suitability of the system "for it's intended use" at regular intervals. In the following paragraphs different mechanisms to prove that systems perform as expected for their
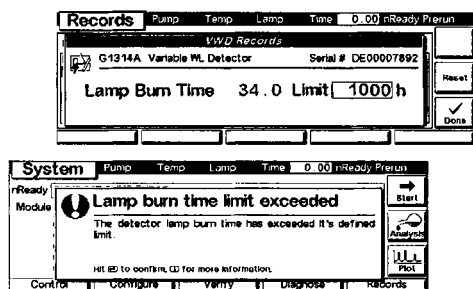
Figure 5. A typical early maintenance feedback system (EMF) informs the user when usage limits are reached, here being the burn time and limits of the detector lamp

intended application will be described. This does not mean that it is recommended to apply all the described procedures. The type and frequency of performance control activities depend on the instrument itself, the application, and on the required performance of the analytical method, as specified as part of the method validation process, or between the laboratory and its clients.

2.5.1. Preventative Maintenance

Operating Procedures for maintenance should be in place for every system component that requires periodic calibration or/and preventative maintenance. Critical parts should be listed and should be available at the user's site. The procedure should describe what should be done, when and what the qualifications of the engineer performing the tasks should be. System components should be labeled with the date of the last and next maintenance. All maintenance activities should be documented in the instrument's log book. Suppliers of equipment should provide a list of recommended maintenance activities and procedures (SOPs) on how to perform the maintenance. Some suppliers also offer maintenance contracts with services for preventative maintenance at scheduled time intervals.

Traditionally maintenance parts are replaced on a time base. For example, an HPLC pump seal every 2 months, a detector's lamp every 3 months or so. This is not economical for the laboratory and not environmentally friendly because frequently a replacement of the parts would not yet be necessary. A better way is to exchange maintenance parts on a usage basis. The user can enter limits for the lamp, the solvent pumped through and the number

of injections. The instruments record the time usage and when the limits are exceeded the user is informed through the user interface. This allows timely exchange of the maintenance parts before the instrument performance goes under the acceptable limit.

All maintenance activities should be recorded in a maintenance logbook. To make this convenient, modern equipment includes electronic maintenance logbooks where the user enters the type of maintenance and the equipment records this activity together with the date and time. An example of such a system is shown in Figure 5.

2.5.2. Calibration

Operating devices may be miscalibrated after a while, for example the temperature accuracy of a GC column oven or the wavelength accuracy the optical unit of a UV/visible detector. This can have an impact on the performance of an instrument. Therefore a calibration program should be in place to recalibrate critical instrument items. All calibrations should follow documented procedures and the results should be recorded in the instrument's logbook. The system components should be labeled with the date of the last and next calibration. The label on the instrument should include the initials of the test engineer, the form should include his/her printed name and the full signature.

2.5.3. Performance Qualification

During performance qualification (PQ) or verification (PV) critical performance parameters of an analysis system are thoroughly tested. Examples are precision of retention times, migration times and peak areas of a chromatographic or electrophoresis system and the baseline noise of an HPLC UV/visible detector.

The performance is tested independently from a specific method using generic test conditions and test samples usually supplied by the instrument vendors. The supplier should also provide recommendations for performance limits (acceptance criteria) and recommended actions in case the criteria cannot be met.

The performance verification should follow documented procedures and the results should be recorded in the instrument's logbook. It is recommended to use templates for the results. An example is shown in Figure 6. The performance verification can be either performed by the user or

Instrument: HP 1050 Series VW detector
Serial number: 1448J3450
Test: Baseline noise
User's specification: 1.0×10^{-4} AU (HP's recommended limit 1.5×10 AU^{-5})
Test frequency: Every 12 months (HP's recom. every 12 months)

Date	Measured value	Corrective action	Final value	Test engineer	
				name	signature
2/3/93	1.4×10-5			Hughes	

Figure 6. Example for a performance verification template

by the supplier or any other third party on behalf of the user. Whoever does the performance verification should have documented evidence that she/he is qualified to do it. This may be certificates of successful participation in training courses.

A frequently asked question is on what performance characteristics should be verified and how often. The frequency of performance checks for a particular instrument depends on the "acceptable limits" as specified by the user. The more stringent the limits are, the sooner the instrument will drift out of the limits, the shorter is the frequency of the performance checks. The time intervals for the checks should be identified and documented for each set up equipment.

Reference 10 lists frequencies and parameters to be checked for chromatographs including liquid and gas chromatographs, electrodes, and for heating/cooling apparatus including freeze dryers, freezers, furnaces, hot air sterilizers, incubators, for spectrometers, autosamplers, balances, volumetric glassware, hydrometers, barometers, and thermometers.

Pre-printed forms should be available for:

– Schedules for preventative maintenance, such as cleaning, and replacement of spare parts
– Maintenance activities
– Calibration activities
– Performance verifications
– LogBooks

2.5.4. Analytical Quality Control (AQC) with Control Samples and Control Charts

The analysis of quality control samples with construction of quality control charts has been suggested as a way to build in quality checks on results as they are being generated. Such tests can then flag those values which may be erroneous for, for example, any of the following reasons:

– Operator is not qualified
– Reagents are contaminated
– GC carrier gas is impure
– HPLC mobile phase is contaminated
– Instrument characteristics have changed over time

For an accurate quality check, Quality Control (QC) samples are interspersed among actual samples at intervals determined by the total number of samples and the precision and reproducibility of the method. The control sample frequency will depend mainly on the known stability of the measurement process, a stable process requiring only occasional monitoring. HUBER [10] recommends that 5 % of sample throughput should consist of quality control samples for routine analysis and 20 – 50 % for more complex procedures.

Control samples should have a high degree of similarity to the actual samples analyzed, otherwise one cannot draw reliable conclusions on the performance of the measurement system. Control samples must be so homogeneous and stable that
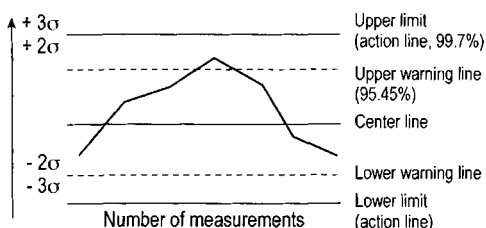
Figure 7. Quality control chart with warning lines and control lines

individual increments measured at various times will have less variability than the measurement process itself. QC samples are prepared by adding known amounts of analytes to blank specimens. They can be purchased as certified reference materials (CRMs) or may be prepared in-house. In the latter case. sufficient quantities should be prepared to allow the same samples to be used over a long period of time. Their stability over time should be proven and their accuracy verified, preferably through interlaboratory tests or by other methods of analysis.

The most widely used procedure for the ongoing equipment through QC samples involves the construction of control charts for quality control (QC) samples (Fig. 7). These are plots of multiple data points versus the number of measurements from the same QC samples using the same processes. Measured concentrations of a single measurement or the average of multiple measurements are plotted on the vertical axis and the sequence number of the measurement on the horizontal axis. Control charts provide a graphical tool to demonstrate statistical control, monitor a measurement process, diagnose measurement problems, and document measurement uncertainty. Many schemes for the construction of such control charts have been presented [10]. The most commonly used control charts are X-charts and R-charts as developed by Shewart. X-charts consist of a central line representing either the known concentration or the mean of 10–20 earlier determinations of the analyte in control material (QC sample). The standard deviation has been determined during method validation and is used to calculate the control lines in the control chart. Control limits define the bounds of virtually all values produced by a system under statistical control.

Control charts often have a center line and two control lines with two pairs of limits: a warning line at $\pm 2\,\sigma$ and an action line at $\pm 3\,\sigma$. Statistics predict that 95.45 and 99.7 % of the data will fall

within the areas enclosed by the $\pm 2\,\sigma$ and $\pm 3\,\sigma$ limits, respectively. The center line is either the mean or the true value. In the ideal case, where unbiased methods are being used, the center line would be the true value. This would apply, for example, to precision control charts for standard solutions.

When the process is under statistical control, the day to day results are normally distributed about the center line. A result outside the warning line indicates that something is wrong. Such a result need not be rejected but documented procedures should be in place for suitable action. Instruments and sampling procedures should be checked for errors. Two successive values of the QC sample falling outside the action line indicate that the process is no longer under statistical control. In this case the results should be rejected and the process investigated for its unusual behavior. Further analyses should be suspended until the problem is resolved.

2.6. Handling of Defective Instruments

Clear instructions should be available to the operator on what to do in case the instrument breaks down or fails to function properly. Recommendations should be given on when the operator should try to fix the problem and when to call for service from the instrument vendor. For each instrument there should be a list of common and un-common failures. In cases of malfunction, it is not sufficient to repair the instrument on-site and to continue the measurements. The failure should be classified into a common or an un-common problem. A common problem like a defective lamp of a UV/visible detector requires short term action. The lamp should be replaced, and after a functional test the instrument can be used for further analyses. The failure and repair, and the result of the functional test should be entered into the instrument's logbook. In cases where there is an un-common failure which cannot be easily classified and repaired by the operator, several steps are required:

- The problem should be reported to the laboratory supervisor or to the person who is responsible for the instrument. The supervisor or the responsible person should decide on further action.
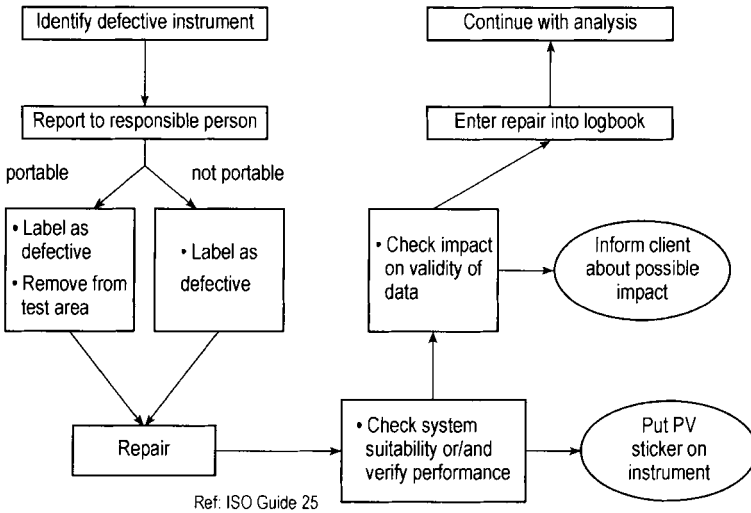
Figure 8. Handling of defective instruments

- The instrument should be taken out of the test area and stored at a specified place, or if this is not practical because of the size, it should be clearly labeled as being defective. For example, small sized portable equipment such as a pH-meter can be taken out of the test area. Larger equipment such as a GC or an ICP-MS system should be labeled as being "out of service".
- After repair, proper functioning should be verified by tests. The type of tests depend on the particular failure and possible impact on the system.
- Effects of the defect on previous test results have to be examined.
- Clients should be informed on any effect the defects could have had on the validity of test data.
- An entry on the defect, repair, and performance verification should be made in the equipment's logbook.

These procedures are summarized in Figure 8.

2.7. References

[1] ISO/IEC Guide 25: *General Requirements for the Competence of Calibration and Testing Laboratories*, International Organization for Standardization, Geneve, Switzerland 2000.

[2] EN 45001: *General Criteria for the Operation of Testing Laboratories, CEN/CENELEC*, The Joint European Standards Institution, Bruessel, Belgium 1989.

[3] NAMAS Accreditation Standard: *General Criteria for Calibration and Testing Laboratories*, M10 of the NAMAS Executive, National Physical Laboratory Teddington, Middlesex 1989.

[4] Food and Drug Administration: *Non-Clinical Laboratory Studies, Good Laboratory Practice Regulations*, U.S. Federal Register. vol. 41, No. 225, November 19, 1976, pp. 51206–51226 (Proposed Regulations) and vol. 43, No. 247, December 22, 1978, pp. 59986–60020, (Final Rule).

[5] OECD Series on Principles of Good Laboratory Practice and Compliance monitoring, Number 1: GLP Consensus Document: *The OECD Principles of Good Laboratory Practice*, Environment Monograph No. 45, Paris 1992.

[6] EC Guide to Good Manufacturing Practice for Medicinal Products in *The Rules Governing Medicinal Products in the European Community*, **vol. IV,** Office for Official Publications for the European Communities, Luxembourg, 1992.

[7] P. Bedson, M. Sargent, *Accred. Qual. Assur.* 1 (1996) 265–274.

[8] *FDA Guide to Inspection of Pharmaceutical Quality Control Laboratories*, The Division of Field Investigations Office of Regional Operations, Office of Regulatory Affairs U.S. Food & Drug Administration, July 1993.

[9] L. Huber: *Validation of Computerized Analytical Instruments*, Interpharm, Buffalo Grove, USA May 1995, 267 pages.

[10] L. Huber: *Validation and Qualification in Analytical Laboratories*, Interpharm, Buffalo Grove, USA Nov. 1998.

[11] Http://www.labcompliance.com, Global Online Resource for Validation and Compliance in Analytical Laboratories, 1999.

[12] ISO/IEC Guide 22: *Information on Manufacturer's Declaration of Conformity with Standards or Other Technical Specifications*, 1st ed., International Organization for Standardization, Geneve, Switzerland 1982.

3. Chemometrics

RENÉ HENRION, Weierstrass Institute of Applied Analysis and Stochastics, Berlin, Federal Republic of Germany

GÜNTHER HENRION, Humboldt University Berlin, Berlin, Federal Republic of Germany

In addition to the symbols listed in the front matter of this volume the following symbols are used:

b	sensitivity ($=dy/dx$)
d	distance
D	distance matrix
n	number of measurements (or objects, Chap. 3.8)
n_i	number of parallel measurements
P	probability
r	correlation coefficient
R	correlation matrix
s	estimate of standard deviation (sdv)
$t(P, f)$	integration limit for Student's distribution
v_i	eigenvectors
x	independent variable (mostly: concentration)
x_d	decision limit
x_D	detection limit
x_i	i-th value (predicting variable, Chap. 3.11)
$\bar{x}$	estimate of mean value
X	matrix
X^T	transpose of the matrix
y	dependent variable (mostly: signal)
λ_j	eigenvalues
μ	mean value
σ	standard deviation

3.1. Introduction

Using a widely accepted definition, *chemometrics* can be understood as "the chemical discipline that uses mathematical, statistical, and other meth-

ods employing formal logic (a) to design or select optimal measurement procedures and experiments, and (b) to provide maximum relevant chemical information by analyzing chemical data" [1]. Defined in this way, chemometrics shares the impact of mathematical modeling with many other disciplines carrying the suffix "metrics", such as biometrics, psychometrics, and econometrics. Nevertheless, the application to problems of chemistry puts emphasis on particular issues not present or less important in these other sciences. For example, keywords like calibration and signal resolution can be mentioned. However, chemometrics is not just the recipient of mathematical progress, it also stimulates the development and the foundation of new mathematical methods. For instance, the use of modern hyphenated analytical methods producing huge amounts of data with increasing complexity of structure, has become a driving force for algorithms in multimodal statistics. Over time, chemometric methods have become indispensable tools, e.g., for quality control, environmental and forensic analysis, medicine, and process control. Chemometrics, however, does not only interpret data or find optimal strategies for analytical work. It also provides a basic for theoretical treatment of analytical chemistry and establishes the field of analytical science as an independent discipline of chemistry. For comprehensive introductions to the field of chemometrics, the reader is referred to the basic monographs [1] and [2].

3.2. Measurements and Statistical Distributions

3.2.1. Measurements

One of the final aims of chemical analysis is the quantitative determination of species concentrations based on appropriate measurements. Usually these measurements exploit a functional relationship between concentrations and a physical (optical, electrochemical, etc.) signal. For a given sample under investigation, the obtained signal will never be constant when repeating the measurement. The reason is that a variety of conditions, such as temperature, conductivity, pH, change slightly over time and their superimposed impact on the measurement is out of control in the sense of a functional model. Therefore, it is useful to understand a measurement as a random variable x

which can have many different outcomes or realizations x_i. In a fixed experiment, the set of realizations $x_1,...,x_N$ is called a sample of x with sample size N. Usually, a sample is characterized by its mean value $\bar{x}$ and its empirical standard deviation s:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i; \quad s = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Example 2.1. The analysis of an element concentration by repeated ($N = 16$) measurements of one laboratory gave the following results:

30.3, 29.9, 32.0, 32.0, 30.0, 31.0, 30.9, 30.3, 30.9, 30.1, 29.4, 29.6, 30.4, 30.9, 28.0, 29.6.

$N = 16$, $\bar{x} = 30.33$, $s = 0.99$.

3.2.2. Statistical Distributions

Typically, the outcomes of random variables are not arranged uniformly but show regions of differing density. The points corresponding to the measurements in Example 2.1 are arranged on a straight line in Fig. 1 a. Their density can be illustrated by means of so-called frequency histograms, where the range of measurements is subdivided into an appropriate number of subintervals of equal length. The number of measurements x_i falling inside a given interval is represented by a rectangle of corresponding height on top of the interval. With increasing sample size N, the histograms take on a more regular shape, which is well approximated by a smooth curve $f(t)$, as in Figure 1 b, where f is the density function of x. Clearly, the realizations of x are most likely to fall inside regions with high density values of f. More precisely, the probability of an outcome of x falling into a set A is given by the integral

$$P(A) = \int_A f(t)dt \tag{2.1}$$

Of particular interest are sets given by left open intervals: $A = (-\infty, \tau)$. In this case, τ is called a $P(A)$ quantile of the distribution of x, which means that the probability for an outcome of x not to exceed τ equals $P(A)$. The density function is always normalized so as to render the integral over the whole real line equal to one. In this way, a density function completely determines the distribution of a random variable. The most prominent representative of probability distributions is the
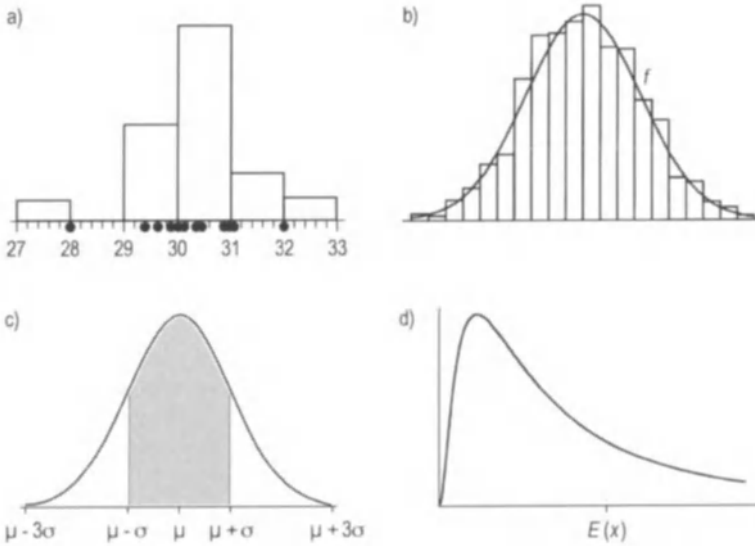
Figure 1. Illustration of densities for probability distributions.
a) Histogram for a small concrete measurement series; b) Histogram for a large series and limiting density function; c) Normal density function with symmetric interval of 68.3 % probability around the mean; d) Density and expected value for the logarithmic normal distribution

normal or Gaussian distribution with density function

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$$

The bell-shaped profile of this function is shown in Figure 1 c. The parameters μ and σ determine the position of the maximum of f and the distance of the inflection points of f from this maximum. Actually, μ corresponds to the expected value $E(x)$, and σ to the standard deviation $\sigma(x)$ of the normal distribution. For general distributions, these quantities are defined as

$$E(x) = \int (t)\, t\, dt$$

$$\sigma(x) = \sqrt{\int (t)[t - E(x)]^2\, dt}$$

The square of the standard deviation $\sigma^2(x)$ is referred to as the variance of x. The expected value, in some sense, represents the most typical outcome of a random variable, whereas the standard deviation measures the average deviation of outcomes from the expected value. For the normal distribution, the probability of a realization falling inside the interval $\mu \pm \sigma$ — that is, closer to the expected value than one standard deviation — is 68.3 %. According to Equation (2.1), this value corresponds to the area portion under the density function for this interval (Fig. 1 c). Taking the larger interval $\mu \pm 3\,\sigma$ gives a probability of 99.7 %. This justifies the $3\,\sigma$ rule, which states that realizations of normally distributed random variables almost surely fall inside this interval.

The importance of the normal distribution relies on the so-called central limit theorem, which roughly states that the distribution of the sum of a large number of independent random variables tends to be a normal one. In particular, measurements which are influenced by a large number of small independent errors are well approximated by a normal distribution with appropriate parameters μ and σ. Nevertheless, other random variables may follow distributions different from the normal one. Counting methods (e.g., X-ray fluorescence) must be described by a Poisson distribution, whereas concentrations in trace analysis are better modeled by a logarithmically normal distribution, that is, the logarithm of the measurements is normally distributed. The latter fact is intuitively clear, since concentrations cannot have negative values, so the density function should be zero on the negative axis (Fig. 1 d). The log-normal distribution also shows that the expected value does not coincide, in general, with the maximum of the density. Other important statistical distributions arise from
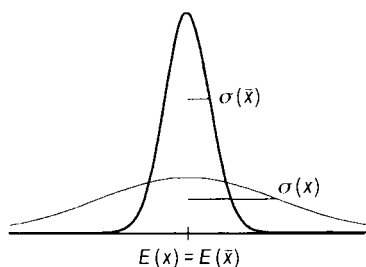
Figure 2. Densities for the normal distribution of a measurements x and of associated mean value $\bar{x}$

transformations of normally distributed random variables, e.g., the t, F and x^2 distributions.

3.2.3. Estimates

It is important to distinguish between parameters of statistical distributions such as density function, expected value, and standard deviation on the one hand, and corresponding sample characteristics such as histograms, mean value and empirical standard deviation on the other. The latter are calculated from actual samples and will converge — in a stochastic sense — with increasing sample size towards the distribution parameters which are the usually unknown characteristics of the abstract probability distribution behind infinitely many samples of one and the same experiment. The sample-based quantities are also called estimates of the corresponding statistical parameters.

Estimates are random variables themselves. For instance, repeating the same experiment as in Example 2.1., one would arrive at a different measurement series with slightly changed values of $\bar{x}$ and s. Repeating this experiment many times, one arrives at a distribution for $\bar{x}$ and s similar to that obtained before for x. Not surprisingly, the expected value of $\bar{x}$ stays the same as that of x: $E(\bar{x}) = E(x)$. The standard deviation, however, decreases by a factor of $\sqrt{N}$, where N is the sample size: $\sigma(\bar{x}) = \sigma(x)/\sqrt{N}$. Hence, the distribution of $\bar{x}$ is much narrower than that of x, (Fig. 2). Increasing the sample size of some measurement series by a factor of nine will therefore decrease the standard deviation of the associated mean value by a factor of three.

3.2.4. Error

Assume that c is the true value of some analytical quantity to be determined. For a given measurement series as in Example 2.1, the error is $|\bar{x}-c|$. We already know that $\bar{x} \to E(x)$ for $N \to \infty$. In the ideal case, where there is no bias, $E(x)=c$; hence, the error will become zero with increasing sample size. Otherwise, the determination has some bias $|E(x) - c| > 0$, and the error can only be reduced to the value of this bias. A statistical test for presence of bias is discussed in Section 3.3.2. The concept of precision refers to the scattering of measurements and is given by the (empirical) standard deviation of a measurement series. As long as precision refers to the results of a single laboratory, it is identified with the concept of *repeatability*. Precision of results provided by different laboratories is identified with the concept of *reproducibility*. For a quantification of both concepts, see Chapter 3.4.

3.3. Statistical Tests

Statistical tests serve the purpose of verifying hypotheses on parameters of distributions. As these parameters are usually unknown, the tests have to rely on estimates which are random variables, hence their outcome has to be interpreted in a probabilistic way.

3.3.1. General Procedure

The general procedure of a test may roughly be sketched as follows:

1) Formulate a "null hypothesis" H_0 usually involving an equality statment about a parameter of interest.
2) Fix a significance level P (usually $P = 95\%$ or $P = 99\%$).
3) Choose an appropriate "test statistic" T which is a random variable depending on the measurement sample and has a well-known distribution under validity of H_0.
4) Select some set A such that the probability of T having values in A equals P.
5) For a given sample, check if the value of T falls outside A. If yes, then reject H_0; otherwise, accept H_0.

Note that, according to this scheme, the probability of rejecting H_0 although H_0 holds true (error of the first kind) is less than $1 - P$ (say 5% or 1%), which in this case gives the probability of the observed fact that T falls outside A. However, nothing is said about the probability of accepting H_0 although H_0 is false (error of the second kind) if we observe that T takes a value that falls in A. The realization of step 4 requires knowledge of quantiles for the distribution of T. Because of limited space, the necessary values for the different tests are not given here; reference is made to standard monographs on applied statistics or handbooks containing the corresponding data tables and also giving a more detailed introduction to statistical theory (e.g., [3], [4]). All tests presented in the following are based on normally distributed data. If this assumption, which should be satisfied in most cases, fails to hold, then one can use nonparametric tests as an alternative (see Section 3.3.3).

Test for Mean with Known Standard Deviation. As an illustration, consider a test for bias in the mean of a set of N laboratory results, when the value of σ is known. As null hypothesis, we formulate coincidence of the expected value behind the measurements with some known true value c:

$$H_0 : E(x) = c$$

A significance level of $P = 95\%$ is fixed and the test statistic $T = \bar{x}$ is chosen. From Section 3.2.3, we know that, $E(T) = E(x)$ and $\sigma(T) = \sigma(x) / \sqrt{N}$. Furthermore, T is normally distributed whenever x is (which one has good reason to assume). It is convenient to select the set A, as required in step 4, in a symmetric way around the expected value (see shaded area in Fig. 1 c). For normal distributions, we already know that the intervals $[\mu - \sigma, \mu + \sigma]$ and $[\mu - 3\sigma, \mu + 3\sigma]$ have probabilities of 68.3 and 99.7%, respectively. To realize the chosen probability of $P = 95\%$, one would have to consider the interval $[\mu - 1.96\sigma, \mu + 1.96\sigma]$, where the value of 1.96 can be read off from appropriate data tables. Since we have to relate the parameters μ and σ here to the expected value and standard deviation of the distribution of $\bar{x}$ rather than that of x, the appropriate choice is (taking into account that $E(T) = E(x) = c$ for validity of H_0)

$$A = \left[c - 1.96\sigma(x) / \sqrt{N}, \quad c + 1.96\sigma(x) / \sqrt{N} \right] \quad (3.1)$$

Now a decision according to step 5 can be made on the basis of an actual sample and a resulting value for T. Note, however, that the described procedure requires knowledge of the distribution parameter $\sigma(x)$, which is rarely available. A related test avoiding this assumption is described in Section 3.3.2.

Example 3.1. For the data of Example 2.1., it shall be checked if there is some bias to an assumed true value of $c = 30$. We also assume that the standard deviation of the (abstract) distribution of the measurement results is known: $\sigma(x) = 0.7$ (e.g., $\sigma(x)$ may have been approximated by empirical standard deviations s on the basis of a large sample size). The test interval calculates as $A = 30 \pm 1.96 \times 0.7/4 = 30 \pm 0.343$. Since the mean value $T = \bar{x} = 30.33$ is contained in this interval, we accept H_0 which means that — up to unavoidable random errors — the laboratory meets the true value (no bias). Recall, that nothing is said about the risk of a false statement here. If, in contrast, the true value were $c = 29.5$, then T is not in A, hence we would reject H_0 and deduce the presence of some bias with a probability of error below $1 - P = 5\%$.

Confidence Intervals. The preceding derivations have shown that the mean value $\bar{x}$ belongs to the interval $E(x) \pm 1.96\sigma(x) / \sqrt{N}$ with probability 95%. But, equivalently, $E(x)$ belongs to the interval $\bar{x} \pm 1.96\sigma(x) / \sqrt{N}$ with the same probability. Hence, given a mean value from a set of measurement results, the unknown distribution parameter $E(x)$ can be included in a symmetric interval around $\bar{x}$ with a given probability. This is called a $P\%$ confidence interval for the statistical parameter. For the data of Example 2.1. one obtains the 95% confidence interval 30.33 ± 0.343. This is the usual way of indicating the result for an analytical concentration etc. For higher probabilities, the interval enlarges accordingly. Choosing, for instance the 99% level, one would have to replace the above factor of 1.96 by 2.58 yielding the interval 30.33 ± 0.452.

One-Sided Tests. Frequently, some additional prior information on the considered statistical parameter is available. There may be, for instance, some danger of bias to the true concentration value as a consequence of sample preparation. From the logical background it might be clear, for instance, that the concentration of the analyte is increased if at all. Then, in the above terminology, we know that $E(x) \geq c$. Endowed with such one-sided information, it is reasonable to test the null hypothesis formulated above with a set A chosen as a one-
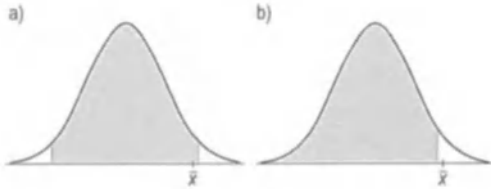
Figure 3. Symmetric (a) and one-sided (b) intervals around the expected value, both having equal probability

sided open interval rather than using the symmetric interval as before. Indeed, here we do not expect our mean value to be much below c, and it is the large values of the mean which are of potential interest now. In this way, we gain information on the crucial side of the inequality to be tested (here: $E(x) > c$) since the test set A is shifted to the left, and this makes it easier to detect excessive deviations on the right. Figure 3 shows that, although both intervals represent 95 % probability, a specific mean value $\bar{x}$ may fall outside the one-sided interval whereas it is inside the two-sided interval. This gives a gain in decision power when exploiting one-sided information. Therefore, we fix A in the form $(-\infty, \tau]$, where τ is chosen such that $P(A) = 95\,\%$. Hence, τ has to be selected as the 95 % quantile (see Section 3.2.2) of the normal distribution with parameters $\mu = E(x)$ and $\sigma = \sigma(x)$. This can be read off from data tables as the value $\tau = 1.65$. Accordingly, for the data of Example 3.1, the mean value now falls outside the interval $A = 30 \pm 1.65 \times 0.7/4 = 30 \pm 0.289$. That is why, in contrast to Example 3.1, the null hypothesis now has to be rejected, and it can be stated that there is a 95 % significant overestimation in the determinations as compared to the true value. The opposite case of knowing that $E(x) \leq c$ is treated analogously by exploiting the symmetry of the normal distribution.

3.3.2. Tests on Parameters of One or Two Measurement Series

Table 1 lists tests on statistical parameters of one or two measurement series. The test intervals are formulated for two-sided tests with probability $P = 95\,\%$. The occuring quantities refer to quantiles of corresponding distributions (F, t, χ^2) which can be found in data tables (e.g., [4]). Indices involving sample sizes refer to so-called degrees of freedom of the corresponding distribution. For instance, $t_{N-1;0.975}$ refers to the 97.5 % quantile of

the t- distribution with $N - 1$ degrees of freedom (note that in the literature quantiles are frequently indicated in the complementary way as critical values, so that 97.5 % correspond to 2.5 %, and vice versa). A two-sided symmetric interval of probability 95 % would then be given by $[t_{N-1;0.025}, t_{N-1;0.975}]$ (since 97.5 % − 2.5 % = 95 %). Owing to the symmetry of the t distribution, $t_{N-1;0.025} = -t_{N-1;0.975}$, whence the interval as indicated in Table 1. There is no problem in changing to other probability levels or one-sided tests (see examples below).

Example 3.2. Using the data of Example 2.1, we want to test the hypotheses $E(x) = 30.8$ and $\sigma(x) = 2$ using the t- and the chi sqd distributions respectively. The 97.5 % quantile of the t-distribution with 15 degrees of freedom equals 2.13, hence $A = [-2.13, 2.13]$. The test statistic calculates for the test on the expected value to $T = (30.33 - 30.8) \times 4/0.99 = -1.89$. Since this value belongs to A, H_0 is accepted and there is no reason to assume a bias in the determinations with respect to the (hypothetical) true result of 30.8. If, however, one knows in advance that the expected value cannot exceed the true one, i.e., $E(x) \leq 30.8$, then A can be chosen as a right-open interval of the same probability $P = 95\,\%$: $A = [t_{N-1;0.05}, \infty) = [-t_{N-1;0.95}, \infty)$ (owing to the symmetry of the t-distribution). For the actual data with $N - 1 = 15$, one gets $A = [-1.75, \infty)$. Now, the value calculated above falls outside the test interval, hence, with the one-sided extra information one can deduce that there is a negative bias to 30.8 in the determinations of the laboratory at the probability level of 95 %. Note that we have just dealt with a one-sided test in the opposite direction to that considered in Section 3.3.1. Another important difference to that simpler test is that now the (theoretical) standard deviation need not be known. Instead, its empirical estimate can be used. To test the hypothesis on the standard deviation, using the chi sqd test gives $A = [6.26, 27.5]$ and $T = 3.68$; hence, the standard deviation behind the given measurement series must be considered different from the 95 % level from the value of 2 which was to be tested. For one-sided tests with additional a priori information that $\sigma(x) \leq 2$ or $\sigma(x) \geq 2$, one would fix the test intervals as $[\chi^2_{N-1;0.05}, \infty)$ and $(-\infty, \chi^2_{N-1;0.95}]$, respectively. Finally, with the given data, confidence intervals for the statistical parameters are easily constructed along the same line as described in Section 3.3.1. For instance, at the 95 % level, one obtains the interval $\bar{x} \pm 2.13$ for covering $E(x)$.

Frequently, it is not the statistical parameters of a single measurement series which have to be tested against specific values, but rather different series are to be tested against each other. For instance, one may ask if the mean values and the standard deviations of the results of two laboratories participating in the same experiment are in agreement. This can be answered by the last two tests recorded in Table 1.

Example 3.3. Consider the measurement data of Example 4.1 below. It shall be tested whether the expected values and standard deviations of the five measurements provided by

Table 1. Sample information, null hypothesis, test statistic, and test interval for common tests on parameters of one or two measurement series

Test	Sample data	H_0	T	A (two-sided, 95 %)
Comparison with fixed standard deviation	s, N	$\sigma(x) = \sigma_0$	$(N - 1)s^2 / \sigma_0^2$	$[\chi^2_{N-1;0.025}, \chi^2_{N-1;0.975}]$
Comparison with fixed expected value	$\bar{x}$, s, N	$E(x) = \mu_0$	$(\bar{x} - \mu_0)\sqrt{N} / s$	$[-t_{N-1;0.975}, t_{N-1;0.975}]$
Comparison of two standard deviations	s_1, N_1, s_2, N_2 $(s_1 \geq s_2)$	$\sigma(x_1) = \sigma(x_2)$	$\dfrac{s_1^2}{s_2^2}$	$[0, F_{N_1-1,N_2-1;0.975}]$
Comparison of two expected values	$\bar{x}_1$, s_1, N_1	$E(x_1) = E(x_2)$	$\dfrac{\bar{x}_1 - \bar{x}_2}{s'} \sqrt{\dfrac{N_1 N_2}{N_1 + N_2}}$	$[-t_{N-1;0.975}, t_{N-1;0.975}]$
$\sigma(x_1) = \sigma(x_2)$	$\bar{x}_2$, s_2, N_2		$s' = \sqrt{\dfrac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}}$	

laboratories A and B are different. Starting with the test on standard deviations, the measurement series with the higher empirical standard deviation is set to be the first one (with respect to index) by definition. So, the symbols s_1, N_1 in Table 1 refer to lab B and s_2, N_2 to lab A. As there is no a-priori information on how the two standard deviations relate to each other, a two-sided test is performed. Accordingly, using the 97.5 % quantile of the F-distribution with $N_1 - 1 = N_2 - 1 = 4$ degrees of freedom which is 9.60, one calculates $T = 0.09^2/0.08^2 = 1.27$ which is clearly inside the test interval. Consequently, there is no reason to assume differences between the underlying theoretical standard deviations of the two labs. For a one-sided test at the same 95 % probability level, the 97.5 % quantile of the F-distribution would have to be replaced by the 95 % quantile.

As for the comparison of expected values, one has to take into account first that the application of this test requires the theoretical standard deviations of the two labs to coincide (more sophisticated tests exist in case this assumption is violated). From the foregoing test, we have no reason to doubt about this coincidence for the data of labs A and B. Following the recipe in Table 1, we calculate $s' = \sqrt{(4 \times 0.09^2 + 4 \times 0.08^2)/8} = 0.085$ and $T = (45.32 - 45.21)\sqrt{5 \times 5 / (5 + 5)} / 0.085 = 2.04$.

With $t_{4;0.975} = 2.78$, we see that T remains inside the test interval. Hence, we cannot deduce 95 % significant differences in the mean values of the results of labs A and B.

3.3.3. Outliers, Trend and Nonparametric Tests

Before extracting statistical characteristics from a measurement series, such as mean value, standard deviation, confidence interval, etc., it has to be checked whether the data are proper with respect to certain criteria. For instance, the sample may contain extremely deviating measurements owing to a gross error (e.g., simply a typing error or improper measurement conditions). Such values are called *outliers* and must be removed form further consideration as they strongly falsify the characteristics of the sample. The same argumentation

holds true for the presence of a *trend* in measurements. This might be caused by directed changes of experimental conditions (e.g., temperature). This leads to a continuous shift of the mean value and increase of the empirical standard deviation with growing sample size, so there is no chance of approaching the theoretical parameters. The presence of outliers and trends can be tested as follows: the test statistic $T = d/s$, where s is the empirical standard deviation and d refers to the maximum absolute deviation of a single value x_i in the measurement series from its mean value, is checked against the values recorded in Table 2 (Grubbs test). If T exceeds the tabulated value, then the measurement x_i is considered as an outlier at the 99 % significance level (smaller levels should not be used, in general, for outlier testing). Similarly, the Trend test according to Neumann and Moore calculates the test statistic

$$T = \sum_{i=2}^{N}(x_i - x_{i-1})^2 / ((N - 1)s^2)$$

If it is smaller than the corresponding value in Table 2, a trend is evident at the 95 % level of probability.

Example 3.4. For the data of Example 2.1., the most outlying measurement with respect to the mean value is $x_{15} = 28.0$, hence $d = 2.33$ and $T = 2.33/0.99 = 2.35$, which is less than the tabulated value 2.75 for $N = 16$. Hence, the series can be considered outlier-free. The trend test statistic for the same data becomes $T = 1.59$ which is smaller than the tabulated value of 1.23. Hence, the series is also free of trends at the 95 % level.

All tests presented so far relied on the assumption of normally distributed measurements. Sometimes, strong deviations from this assumption can

Table 2. Critical values for the Grubbs test and Neumann/Moore test on the presence of outliers and trend, respectively

N	Outlier (99 %)	Trend (95 %)
3	1.16	
4	1.49	0.78
5	1.75	0.82
6	1.94	0.89
7	2.10	0.94
8	2.22	0.98
9	2.32	1.02
10	2.41	1.06
11	2.48	1.10
12	2.55	1.13
13	2.61	1.16
14	2.66	1.18
15	2.71	1.21
16	2.75	1.23

lead to incorrect conclusions from the outcome of these tests. If there is doubt about normality of the data (which itself can be tested as well) then it is recommended to apply so-called nonparametric counterparts of these tests. Nonparametric means that these tests are not based on a distribution. In this way, they are robust with respect to deviations from normality, although of course , less efficient in the presence of normality. Typically, nonparametric tests evaluate some ranking of specifically arranged measurements.

3.4. Comparison of Several Measurement Series

Up to now, we have dealt with statistical characterizations of single measurement series or pairwise comparisons. Frequently, a larger group of measurement series must be analyzed with respect to variance and mean. In the case of commercial products, for example, quality is tested in different laboratories. The different working conditions in individual laboratories provide an additional source of random error. In such situations, it is of interest to analyze the homogeneity of data with respect to standard deviations and mean values. In case that homogeneity can be assumed, much sharper confidence intervals for the precision and accuracy can be obtained from the pooled data rather than from a single lab's results. In the opposite case, the standard deviations relating to repeatability and reproducibility (see Section 3.2.4) can be estimated from *analysis of variance*. Testing for homogeneous data is not only useful, however, in the context of interlaboratory

comparisons but also for the important question of representative sampling of materials to be analyzed.

3.4.1. Homogeneity of Variances

Assume that we are given a group of p measurement series with $N_j, \bar{x}_j, s_j$ denoting sample size, mean value and empirical standard deviation of series j. Further we set $N = \sum_{j=1}^{p} N_j$. We want to check whether all series have equal precision, i.e., H_0:

$$\sigma(x_1) = ... = \sigma(x_p)$$

This may be realized by the Bartlett test, in which, the following averaged standard deviation within the series is first determined:

$$s_g = \sqrt{\frac{\sum_{j=1}^{p}(N_j - 1)s_j^2}{N - p}} \tag{4.1}$$

Now, a test statistic is calculated from the data according to

$$T = \frac{2.306}{c}\left((N - p)\log s_g^2 - \sum_{j=1}^{p}(N_j - 1)\log s_j^2\right)$$

(where "log" = decadic logarithm) and

$$c = 1 + \frac{\left(\sum_{j=1}^{p}\frac{1}{N_j - 1} - \frac{1}{N - p}\right)}{3(p - 1)}$$

H_0 will then be rejected at probability level 95 % if $T > \chi_{p-1;0.95}^2$.

Example 4.1. A sample of FeSi was analyzed in $m = 7$ laboratories with the following results (% Si)

Laboratory	A	B	C	D	E	F	G
	45.09	45.20	45.37	45.23	45.40	45.63	44.93
	45.19	45.27	45.45	45.26	45.40	45.65	44.95
	45.22	45.30	45.48	45.31	45.45	45.73	44.95
	45.25	45.40	45.60	45.39	45.60	45.85	45.14
	45.31	45.43	45.62	45.44	45.60	45.85	45.17
$\bar{x}_j$	45.21	45.32	45.50	45.33	45.49	45.74	45.03
s_j	0.08	0.09	0.11	0.09	0.10	0.11	0.12

For the data from this table, one obtains the results

$p = 7, N_j = 5 (j = 1, ..., p), N = 35, s_g = 0.10, c$
$= 1.10, T = 0.64, \chi^2_{6;0.95} = 12.6$

Consequently, the null hypothesis cannot be rejected at the 95 % level, and one can assume homogeneous variances among the seven different measurement series.

3.4.2. Equality of Expected Values

The hypothesis for homogeneity of expected values is H_0:

$$E(x_1) = ... = E(x_p)$$

This hypothesis is tested by analysis of variance in which — apart from the already determined standard deviation within laboratories (Equation 3) — a so-called standard deviation of laboratory means (standard deviation between laboratories) figures as an important ingredient:

$$s_z = \frac{1}{p-1} \sum_{j=1}^{p} N_j (\bar{x}_j - = x)^2 \tag{4.2}$$

where

$$= x = \frac{\sum_{j=1}^{p} N_j \bar{x}_j}{N} \tag{4.3}$$

is the weighted mean of all single mean values or, equivalently, the overall mean of all measurements in the data table. H_0 will be rejected at probability level 95 % if $T = s_z^2/s_g^2 > F_{p-1,N-p;0.95}$. The reason for using the one-sided test interval here in contrast to the two-sided one in Section 3.3.2 is that the theoretical standard deviations σ_g, σ_z behind their empirical estimates s_g, s_z can be shown always to satisfy the one-sided relation $\sigma_g \leq \sigma_z$, where equality holds exactly in the case that H_0 is true. If H_0 can be rejected on the basis of this test, then the standard deviations in Equations (4.1) and (4.2) represent the repeatability and reproducibility, respectively, of the given laboratory data. Otherwise, the expected values can be considered as being equal, and there is no reason to assume between laboratory bias in the given determinations. As a mean over means, the overall mean of Equation (4.3) has a much smaller variance than the single means (which in turn have smaller variances than the single measurements). This gives rise to very narrow confidence intervals

for the mean and explains why interlaboratory homogeneity comparisons are so valuable in producing reference materials.

Similar to the test on equality of two expected values described in Section 3.3.2, its generalization to the multiple case requires homogeneity of variances, this means a Bartlett test must be performed first.

Example 4.2. For Example 4.1, the Bartlett test gave no reason to doubt the homogeneity of variances, so the test on multiple equality of expected values can be carried out. For the data of the example, one calculates

$$s_z = 0.51, = x = 45.37, T = 26.58, F_{6,28;0.95} = 2.44$$

Consequently, the expected values behind the single measurement series must be considered different at the 95 % level. In other words, there exists a strong laboratory bias which can be characterized by the strongly different values for s_z and s_g.

3.5. Regression and Calibration

3.5.1. Regression Analysis

Regression is a tool for modelling a set of N observed data pairs (x_i, y_i) by means of a functional relationship $y = f(x)$, where f belongs to a specified family of functions (e.g., linear functions, polynomials, bi-exponential functions, etc.). Usually, it is assumed that x is an independent variable which can be fixed without errors, whereas the variable y is dependent on x and is subject to random errors. In analytical chemistry, the role of y is frequently played by measured signals responding to some concentration x or recorded at some wavelength x. The main applications of regression in analytical chemistry are calibration (see Section 3.5.2) and signal resolution. Figure 4a illustrates a calibration problem, while Figure 4c shows the decomposition of a signal profile into a sum of two Gaussian profiles.

The family to which f is supposed to belong to is defined via some set of parameters which are coupled in a particular way with x. Therefore, it is reasonable to extend the above mentioned functional relationship to $y = f(x,p)$, where the parameter p is variable. Some prominent families of regression models are:
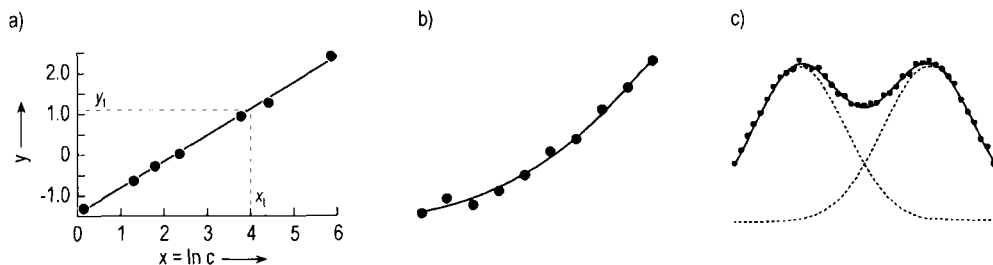
Figure 4. Illustration of different regression models. Linear model a), polynomial model b) and nonlinear model

1) Linear model $f(x,p) = p_1 + p_2 x$

2) Polynomial model $f(x,p) = p_1 + p_2 x + p_3 x^2 + \ldots + p_{n+1} x^n$

3) Multi-exponential model $f(x,p) = p_1 e^{p_2 x} + \ldots + p_{2n-1} e^{p_{2n} x}$

Of course, the linear model is a special case of the polynomial model. Generally, a model is called *quasilinear* when f is a linear function of p. This does not exclude the case that f is nonlinear in x. In particular, the polynomial model is quasilinear although the functional dependence on x may be quadratic, as in Fig. 4 b. Given the data pairs (x_i, y_i), the parameter p yielding the best approximation of f to all these data pairs is found by minimizing the sum of the squares of the deviations between the measured values y_i and their modeled counterparts $f(x_i)$:

$$p \rightarrow \text{minimize} \quad \sum_{i=1}^{N} (y_i - f(x_i, p))^2$$

The advantage of quasilinear models is that the exact solution for p is easily obtained by solving a system of linear equations (see Section 3.11.1). In the simple case of linear models, one can even directly indicate the explicit solution:

$$p_1 = \frac{s_{xx} s_y - s_x s_{xy}}{d}; \quad p_2 = \frac{N s_{xy} - s_x s_y}{d} \qquad (5.1)$$

where

$$s_x = \sum_{i=1}^{N} x_i; \quad s_y = \sum_{i=1}^{N} y_i; \quad s_{xy} = \sum_{i=1}^{N} x_i y_i;$$

$$s_{xx} = \sum_{i=1}^{N} x_i^2; \quad d = N s_{xx} - (s_x)^2$$

The coefficients p_1 and p_2 refer to the intercept and slope of the straight line fitting the data points (see Fig. 4 a). Being based on the random variables x and y, the coefficients p_1 and p_2 are outcomes of random variables themselves. The true coefficients can be covered by the following 95 % confidence intervals:

$$p_1 \pm t_{N-2;0.975} s_R \sqrt{\frac{1}{N} + \frac{\bar{x}^2}{s_Q}}; \quad p_2 \pm \frac{t_{N-2;0.975} s_R}{\sqrt{s_Q}} \qquad (5.2)$$

where

$$s_R = \sqrt{\frac{\sum_{i=1}^{N}(y_i - p_1 - p_2 x_i)^2}{N-2}}$$

and

$$s_Q = s_{xx} - \frac{(s_x)^2}{N}.$$

In contrast to the polynomial model, the multiexponential model is not quasilinear. This makes the determination of the parameters p more complicated. Then iterative methods have to be employed in order to approach a solution. This issue falls into the framework of nonlinear regression analysis. For more details, see [5].

3.5.2. Calibration

The purpose of quantitative chemical analysis is to determine concentrations of certain analytes on the basis of concentration-dependent instrumental responses or signals. The relation between signals (or functions thereof) and concentrations (or functions thereof) is established by calibration. Calibration consists of two steps: in the first step a functional dependence $y = f(x)$ is modeled on the

basis of so-called calibration samples with known concentrations (e.g. reference materials) and corresponding signals, altogether giving a set of data pairs (x_i, y_i). To these data points, an appropriate function is fitted as described in Section 3.5.1. In the second step, the obtained model is applied to a test sample with unknown concentration on the basis of a measured signal. Here, the inverse function is used: $x = f^{-1}(y)$.

Example 5.1. The following calibration results were obtained for the determination of Ni in WO_3 by optical emission spectroscopy (OES):

sample	1	2	3	4	5	6	7
c/ppm	1	3.5	6.0	11.0	51.0	101	501
$x = \ln c$	0.00	1.25	1.79	2.40	3.93	4.62	6.22
y	−1.31	−0.66	−0.33	−0.04	0.84	1.15	2.26

The signal is in good linear relation with the logarithm of concentrations. The data are plotted in Fig. 4 a. From the data table one calculates the values required in Equation (5.1) as

$$s_x = 20.21; \quad s_y = 1.91; \quad s_{xy} = 21.16; \quad s_{xx} = 86.0;$$
$$d = 193.6; \quad p_1 = -1.36; \quad p_2 = 0.57$$

For a test sample, a signal with value $y_t = 1.0$ was measured. Starting from the obtained model $y = -1.36 + 0.57x$, one calculates the inverse function as $x = (y + 1.36)/0.57$; hence, $x_t = 2.36/0.57 = 4.14$. Finally, this last result is transformed back to concentrations via $x = \ln c$. This gives $c_t = e^{x_t} = 62.8$ as the concentration of Ni in the test sample.

In linear calibration, the intercept p_1 corresponds to the signal of the blank, i.e., the signal occuring in absence of the analyte. The slope p_2 is called the *sensitivity* of the signal. For nonlinear calibration curves, this sensitivity changes and has to be calculated as the first derivative dy/dx at a given concentration value x.

The methodology presented so far corresponds to the *classical calibration*. A different approach is *inverse calibration*, where the functional relationship is directly modeled in the form required for prediction of concentrations in test samples: $x = f(y)$. This approach has proven useful in the context of mutlivariate calibration (see Chap 3.11).

3.6. Characterization of Analytical Procedures

An analytical procedure is characterized by its range of concentration (including calibration), pre-

cision (random error), trueness (systematic error), selectivity, and principal limitations.

Trueness [6], [7]. A systematic error can occur as an *additive error* (e.g., an undetected blank) or a *multiplicative error* (e.g., an incorrect titer). Systematic errors are detected by analyzing a short series of m samples with "known" contents x_i and "found" contents y_i as the results. Evaluation by linear regression (Chap. 3.5) yields $y = a + bx$. An intercept $a \neq 0$ is due to an additive systematic error, whereas a slope $b \neq 1$ indicates a multiplicative error. The significance of a and b is tested by verifying that 0 and 1 do not belong to the 95 % confidence intervals around a and b, see Equation (5.2).

If samples with known concentration are not available (e.g., very often in the case of bioproducts), the multiplicative error can be detected in the following way. The sample (solution) is divided into two halves. One half is analyzed directly (result x_1). To the other half an exactly defined amount of the analyte x_+ is added (result x_2). Then the "recovery rate" b is given by

$$b = (x_2 - x_1)/x_+$$

This is performed for m samples, resulting in

$$\bar{b} = \frac{1}{m} \sum b_i; \quad s_b = \sqrt{\sum (b_i - \bar{b})^2 / (m - 1)}$$

A multiplicative systematic error is proved with 95 % probability if

$$t_b = \|1 - \bar{b}\| \sqrt{m}/s_b > t_{m-1,0.975}$$

Example 6.1. Trueness of determination of arsenic levels in yeast by hydride atomic absorption spectroscopy (AAS) should be tested. However, samples with known contents were not available. Therefore, the recovery rate (with $x_+ = 30$ µg As) was determined from $m = 5$ unknown samples.

Sample	1	2	3	4	5
x_1, µg As	5.8	13.8	30.0	43.1	66.8
x_2, µg As	35.2	43.3	59.6	72.4	96.5
b_i	0.9800	0.9833	0.9867	0.9767	0.9900

$\bar{b} = 0.9833$, $s_b = 0.005266$
$t_b = 7.09$; $t_{4;0.975} = 2.78$

Since $t_b > t_{4;0.975}$ the recovery rate differs undoubtedly from $b = 1$. The results gained by hydride-AAS are on average about 2 % too low.

Table 3. Multifactor plan according to PLACKETT and BUR-
MAN for n = 7 components *

Experiment no.	Components							Result
	B	C	D	E	F	G	H	
1	+	+	+	−	+	−	−	y_1
2	+	+	−	+	−	−	+	y_2
3	+	−	+	−	−	+	+	y_3
4	−	+	−	−	+	+	+	y_4
5	+	−	−	+	+	+	−	y_5
6	−	−	+	+	+	−	+	y_6
7	−	+	+	+	−	+	−	y_7
8	−	−	−	−	−	−	−	y_8

* Scheme for a multifactor plan with $n > 7$ see [6].

Selectivity [6]. In a multicomponent system
consisting of an analyte A and other components
(B, C, ..., N), each component contributes to the
signal of the analyte. Selectivity is given if the
signal for A is only randomly influenced by
B, C, ... This can be tested by a simple multifactor
plan according to PLACKETT and BURMAN [8] (see
Table 3) with m measurements for $n = m - 1$ com-
ponents ($m = 8, 12, 16$, etc.). All the m measure-
ments need a solution of the analyte of concen-
tration x_A. Then, B, C, ... are added in the con-
centration $x_B^+, x_C^+, ...$ according to the sign (+) in
the plan. All of these concentrations $x_B^+, x_C^+, ..., x_j^+$
must be similiar to the composition of the samples
that are to be analyzed later. Each of the
$i = 1, ..., m$ measurements gives a result y_i (e.g.,
extinction). The influence W of B, C, ... on the
signal of the analyte A is given by

$$W_j = \left(\sum y_i^+ - \sum y_i^- \right) / m/2 \qquad (6.1)$$

$j = B, C, ...$

Subsequently, the sensitivities b_j for B, C, ...

$$b_j = W_j / x_j^+ \qquad (6.2)$$

and b_A for the analyte A

$$b_A = y_m / x_A \qquad (6.3)$$

are calculated, and the desired polynomial for se-
lectivity is given by

$$y = b_A x_A + b_B x_B + b_C x_C + ... \qquad (6.4)$$

where y is the intensity of the measured analytical
signal. In this way the influence of all accompa-

nying components B, C, ... can be evaluated (for
details, see [9]).

Example 6.3. The selectivity for the determination of
sodium ($x_{Na} = 100$ mg/L) by flame photometry in presence
of Ca^{2+}, K^+, Mg^{2+}, SO_4^{2-}, Cl^-, and PO_4^{3-} had to be tested.
According to the plan (Table) these components were added
for the step (+) of the factors B ... G in the following con-
centration (mg/L):

100 mg Ca^{2+} (B);   100 mg K^+ (C);   100 mg Mg^{2+} (D);
100 mg SO_4^{2-} (E);   1000 mg Cl^- (F);   100 mg PO_4^{3-} (G)

Results y_i of the $m = 8$ measurements (scale reading)

$y_1 = 132$; $y_2 = 137$; $y_3 = 119$; $y_4 = 111$
$y_5 = 125$; $y_6 = 114$; $y_7 = 93$; $y_8 = 129$

Influences W_j (Eq. 6.1)
$W_B = (132+137+119+125)-(111+114+93+129)/4 = 16.5$;
→ $b_B = W_B/100 = 0.165$
$W_C = -3.5$; $W_D = -11.0$; $W_E = -5.5$; $W_F = 1.0$;
$W_G = -16.0$
Then follow all b_j ($j = B, C, ...$) according to Equation (6.2).
The sensitivity for the analyte is finally given by Equa-
tion (6.3) as $b_A = 129/100 = 0.129$. Then the desired polyno-
mial of selectivity (Eq. 6.4) is
$y = 0.129\, x_{Na} + 0.165\, x_{Ca} - 0.035\, x_K - 0.110\, x_{Mg} -$
$- 0.055\, x_{SO_4} + 0.001\, x_{Cl} - 0.160\, x_{PO_4}$. The determination of Na
is strongly influenced by Ca^{2+}, Mg^{2+}, and PO_4^{3-}, less by K^+
and SO_4^{2-}, the influence of Cl^- can be ignored. The analytical
signal is enhanced by Ca^{2+} and depressed by Mg^{2+}, PO_4^{3-}, and
SO_4^{2-}.

Principle Limitations. The analysis of traces is
often influenced by the impurities of the reagents.
Then a blank $\bar{y}_{bl} \neq 0$ (standard deviation σ_{bl}) is
measured at the concentration $x = 0$ of the analyte.
From Equation (3.1) we we know that an analyt-
ical signal $\bar{y}_a$ (n_j parallels) is different from the
blank with 95 % probability if

$$(\bar{y}_a - \bar{y}_{bl}) > 1.96 \frac{\sigma_{bl}}{\sqrt{n_j}}$$

Extending the factor 1.96 to the value 3 would
then even yield a difference with 99.7 % probabil-
ity according to the 3σ-rule (see Section 3.2.2).
Now, taking into account that the expected value
of the signal of some analyte is never smaller than
the blank, one can sharpen the statement above:
the analytical signal is different from the blank
with 99.85 % probability if

$$\bar{y}_a > \bar{y}_{bl} + 3 \frac{\sigma_{bl}}{\sqrt{n_j}} = y_c$$

where y_c is the critical value. The improvment in
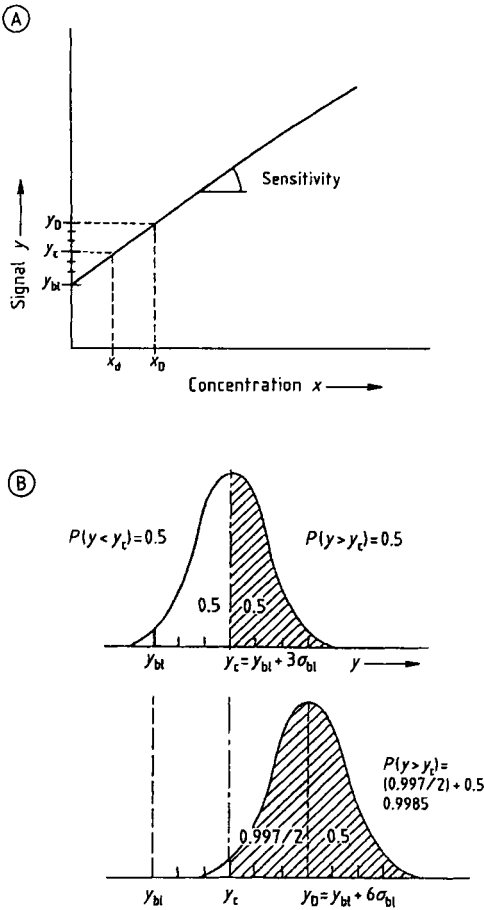probability here is due to using one-sided addi-
tional information.

Figure 5. Limit of decision x_d and of detection x_D
A) Calibration function in presence of a blank; B) Probabilities for the detection of an impurity at limit of decision (upper curve) and at limit of detection (lower curve)

In the case of a blank the calibration function takes the form (see Fig. 5 A)

$$y = a + bx = \bar{y}_{bl} + bx$$

For $y = y_c$ the evaluation function will be

$$x = (y_c - a)/b = x_d$$

x_d — *the limit of decision* [10] — is due to the minimum concentration of the analyte that gives a significant analytical signal.

A signal $y = y_c$ results from a concentration $x = x_d$. The frequently repeated analysis of such a sample yields a signal $y < y_c$ as often as $y > y_c$

caused by the random error of the blank (see Fig. 5 B). Expressed as probabilities gives

$$P(y < y_c) = P(y > y_c) = 0.5$$

Therefore the value x will be interpreted as "analyte present" → (+) as often as "only blank → analyte absent" → (−), or again expressed as probabilities

$$P_x^+ = P_x^- = 0.5$$

Therefore the limit of decision x_d never can be used as guarantee of purity. A sufficiently high reliability for such a guarantee is given for the signal value

$$y_D = \bar{y}_{bl} + \frac{6\sigma_{bl}}{\sqrt{n_j}}$$

or the associated concentration

$$x_D = \frac{6\sigma_{bl}}{b\sqrt{n_j}}$$

x_D is termed the *detection limit* [10]. Then,

$$P_x^+ = P(y > y_c) = 0.9985$$
$$P_x^- = P(y < y_c) = 0.0015 \qquad (6.5)$$

An impurity can be detected now at a very high probability level. Therefore, only this limit of detection allows the characterization of high-purity material.

Example 6.4. The photometric determination of iron with triazin (absorptivity $A = 2.25 \times 10^3$ m^2/mol) yielded $\bar{y}_{bl}$ ($= E_{bl}$) = 0.08 and $\sigma_{bl} = 0.02$. For the analysis with duplicates, the critical value is

$$y_c \ (= E_c) = 0.08 + 3 \times 0.02/\sqrt{2} = 0.122$$

The evaluation function gives $x_d = (0.122 - 0.080)/ 2.25 \times 10^3 = 1.87 \times 10^{-5}$ mol/L and the limit of detection x_D follows from

$$y_D = 0.08 + 2 \times 3 \times 0.02/\sqrt{2} = 0.165$$

$x_D = (0.165 - 0.080)/2.25 \times 10^3 = 3.78 \times 10^{-5}$ mol/L

3.7. Signal Processing

Recording instrumental responses to the presence of some analyte is at the heart of qualitative
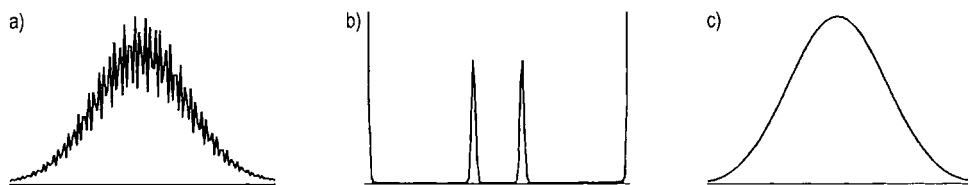
Figure 6. Illustration of Fourier transform
a) Noisy Gaussian signal peak; b) Resulting Fourier transform; c) Filtered signal

and quantitative chemical analysis. The physical nature of these responses or signals may be different according to appropriateness with respect to the given analyte (e.g. optical, electrochemical etc.). In the simplest case, just one signal is recorded, e.g. absorption at some wavelength. Modern instrumentation, however, allows signals to be monitored over a range of physical quantities such as wavelength or time or both of them coupled. Signals are always subjected to perturbations of different kind and origin. Such perturbations can be induced by noise related to the analyzing technique and its apparatus. Furthermore, signals of a certain analyte under investigation are frequently perturbed by interference with signals of different species, which leads to falsifications both in qualitative and quantitative analysis.

3.7.1. Fourier Transform

Contamination of analytical signals by high-frequency noise (e.g., 50 Hz modulated noise caused by the power line) can be detected and corrected by Fourier transformation. Figure 6a shows a Gaussian signal (e.g. spectroscopic peak over the wavelength domain) containing sinoidal noise. For a discretized signal f with n values $f(0),...,f(n-1)$ the (discretized) Fourier transformed signal is defined as

$$F(u) = \frac{1}{n}\sum_{t=0}^{n-1} f(t)e^{-2\pi u t i/n} \quad (u = 0,...,n-1)$$

where i is the imaginary unit and n is the discretization number. As the Fourier transform is complex, one usually plots its absolute value as shown in Figure 6b. The symmetry of the Fourier transform around the center point means that only the first half of the spectrum contains substantial information. Peaks in the Fourier transform indicate the presence of periodic contributions in the orig-

inal signal. The higher the frequency of this contribution, the more this peak shifts to the center. Hence, each of the symmetrically arranged two peaks in the center of Figure 6b corresponds to the presence of a high-frequency contribution in the original signal, whereas the two peaks on the boundary represent the (low-frequency) signal itself. In this way, noise and signal are clearly separated in the Fourier transform. To arrive at such separation in the original representation of the signal, one may exploit so-called filtering: in a first step, the Fourier transformed signal is convoluted with a so-called kernel which typically is some simple function having a peak in the center of the frequency domain; the convoluted signal is transformed back then to the original domain of the signal by Inverse Fourier transformation. The result is shown in Fig. 6c. A clear improvement of the signal-to-noise ratio becomes evident when compared with Fig. 6a.

3.7.2. Data Smoothing

Not all kinds of noise are efficiently removed by Fourier transformation. This is true, in particular, for nonperiodic perturbations such as single spikes. Then, direct data smoothing may be helpful. The simplest method is the moving average, in which each data point of the original signal is replaced by the average of its k nearest neighbors. The appropriate choice of the parameter k is crucial. Enhancing k improves the smoothing effect, which is desirable as far as the reduction of noise is concerned. If k is chosen too large, then information of the signal itself will be smoothed away. Figure 7a shows an originally measured fluorescence spectrum (267 data points) with clear contamination by noise, whereas Figure 7b plots a version smoothed by moving averages with $k = 20$. More advanced techniques are based on polynomial smoothing.
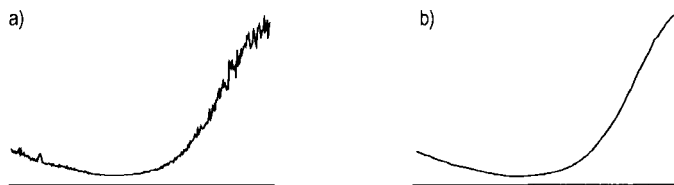
Figure 7. Signal smoothing
a) Noisy fluorescence spectrum; b) Spectrum smoothed by using moving average

Table 4. Small example of a fictive data table (arbitrary units)

Sample no.	[Cd]	[Pb]	[Cu]
1	3	3	1
2	1	2	3
3	4	3	3
4	2	2	0

3.7.3. Signal Resolution

Frequently, the analyte cannot be completely separated by chemical means from other species whose signals interfer with the signal of interest. Then, the resulting superposed signal may be distorted so much that quantitative analysis becomes impossible without additional mathematical tools. In the worst case, even the presence of an interfering component is hidden. Figure 8a shows two idealized Gaussian peaks the superposition of which results in a single peak only. Such junction of peaks occurs when the distance between them is below a critical value. In this case, signal sharpening by differentiation is a possible remedy. Fig. 8b shows the superposed single peak of Fig. 8a in a shrinked domain along with its second derivative. Now, the presence of two underlying components is clearly evident again. In principle one could enhance the order of differentiation, but the gain in signal resolution quickly becomes negligible in comparison with the increase in noise.

3.8. Basic Concepts of Multivariate Methods

3.8.1. Objects, Variables, and Data Sets

The increasing performance of modern instruments allows the characterization of certain objects under investigation by a whole set of properties, which yield a typical fingerprint for each object. These properties are usually statistical variables, whose realization is measured experimentally.

Example 8.1. Several water samples (objects) are analyzed with respect to concentrations of a set of relevant trace elements (variables). Collecting the resulting data gives a constellation as in Table 4. In such data tables or data sets the rows correspond to objects and the columns to variables.

Real life data sets may be generated by very large numbers of objects or variables, depending on the problem. From a more technical point of view a distinction can be made between experimental and instrumental data sets. *Instrumental data sets* are the direct output of instruments, such as digitized spectra, chromatograms, or images in image processing. Here, the number of objects or variables may reach some ten or hundred thousands; therefore, data processing and evaluation generally are automated. *Experimental data* result from single observations (e.g., physico-chemical parameters of natural water at different locations) or from condensation of instrumental data (e.g., one concentration value for a sample instead of a digitized spectrum). As a consequence, acquisition of experimental data is much more expansive, leading to smaller data sets (in the range of several hundreds of objects or variables). As a rule, experimental data tables can be neither processed nor evaluated automatically but require consideration of specific circumstances in the treated problem.

Compared to classical measurement series restricted to a single variable, the multivariate approach yields a considerably greater amount of information for object characterization. This advantage, however, is connected with a loss of direct interpretability. If only two variables were measured, the whole data structure might be still visualized in a diagram, but this approach fails for larger problems. That is why the increase in information must be accompanied by some efficient tool of *data reduction* that enables interpretation
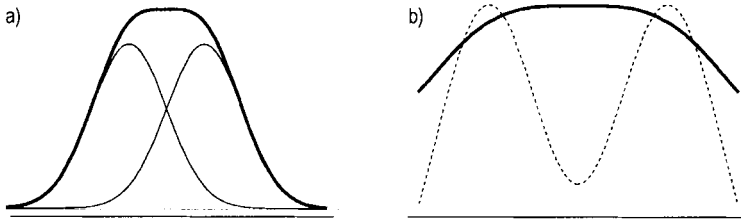
Figure 8. Signal sharpening by differentiation
a) Single peak as a superposition of two close Gaussian peaks; b) Second differential of the superposed signal, yielding two resolved peaks again

or visualization of the essential part of the data structure, thereby separating it from unessential noise. In this sense, data reduction can be understood as the main goal of multivariate methods in chemometrics. Whereas theoretical foundations were developed at the beginning of this century, practical applications of these procedures have revived considerably only during the last few decades (from the 1970s onward) in connection with modern computational equipment (for an introductory survey of multivariate statistics see [11]).

3.8.2. Correlation and Distance Matrices

An important task of multivariate data analysis is classification of objects and variables (i.e., subdivision of the whole data set into homogeneous groups of similar objects or variables, respectively). Similarity of variables is usually measured by their correlation coefficient, whereas similarity of objects is expressed in terms of the geometric distance. The *correlation coefficient r* of two variables x_1, x_2 is computed according to

$$ r(x_1, x_2) = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2} \cdot \sqrt{\sum_{i=1}^{n}(x_{i2} - \bar{x}_2)^2}} \qquad (8.1) $$

where n is the number of realizations (i.e., number of objects or rows of the data table); x_{ij} is the ith realization of the jth variable (i.e., entry of the data table that is located in the ith row and jth column); and $\bar{x}_j$ is the mean of the jth variable.

For the fictive example of Table 4, for instance, the following coefficients r are obtained r(Cd, Pb) = 0.89, r(Cd, Cu) = 0.09, and r(Pb, Cu) = 0.19. All information on correlation can conveniently be collected in the correlation matrix **R**,

which is the basis of many latent variable methods such as principal components, factor, or discriminant analysis:

$$
\begin{array}{cccc}
 & \text{Cd} & \text{Pb} & \text{Cu} \\
\text{Cd} & 1.00 & 0.89 & 0.09 \\
\text{Pb} & 0.89 & 1.00 & 0.19 \\
\text{Cu} & 0.09 & 0.19 & 1.00 \\
\end{array}
$$

Sometimes, instead of correlation the covariance of two variables is used, which is just the numerator of Equation (8.1).

To recognize similarities among objects, they can best be treated as points in p-dimensional Euclidean space, where p is the number of variables (columns) of the data set. If in Table 4 only the concentrations of Cd and Pb had been measured, then all samples could be represented in a diagram as in Figure 9 (i.e., in a two-dimensional space or plane). Although for technical reasons, such representations cannot be realized if the number of variables exceeds 3, the mathematical concept of p-dimensional space is nevertheless useful, for example, when defining the distance of two objects. In a planar display, distances are easily computed by the Pythagorean theorem by taking the square root of the sum of squared coordinate differences between two points (e.g., the distance between samples 4 and 3 equals $\sqrt{5}$). Obviously, the more similar two objects are in all their coordinates the smaller their computed distance is. In this way, homogeneous classes of similar objects may be detected as those that have only small distances between them. Of course, computation of distances is not restricted to the plane but is extended to p-dimensional constellations in a straightforward way:

$$ d(x_a, x_b) = \sqrt{\sum_{j=1}^{p}(x_{aj} - x_{bj})^2} \qquad (8.2) $$
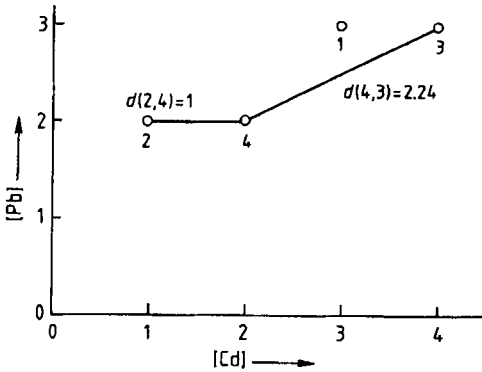
Figure 9. Plot of four samples from Table 4 in an elementary diagram

where x_a, x_b are the objects a and b; and x_{aj} is the entry of the data table in row a and column b. Taking all three coordinates of Table 4 into account gives, for instance:

$$d(4,3) = \sqrt{(2 - 4)^2 + (2 - 3)^2 + (0 - 3)^2} = 3.74$$

All pairwise distances are again collected to give the distance matrix D, which forms the basis of many computational approaches including cluster analysis or multidimensional scaling:

$$
\begin{array}{c}
 \\
1 \\
2 \\
3 \\
4
\end{array}
\begin{array}{cccc}
1 & 2 & 3 & 4 \\
\left(\begin{array}{cccc}
0.00 & 3.00 & 2.24 & 1.73 \\
3.00 & 0.00 & 3.16 & 3.16 \\
2.24 & 3.16 & 0.00 & 3.74 \\
1.73 & 3.16 & 3.74 & 0.00
\end{array} \right)
\end{array}
$$

In this matrix the most similar pair of (different) objects is (1,4), while the most divergent pair is (3,4). Apart from the classical Euclidean distance defined by Equation 8.2, some further relevant measures exist such as Mahalanobis or Manhattan distance. The Mahalanobis distance, for instance, which is important in classification (see Chapter 3.10), is computed according to

$$d(x_a, x_b) = \sqrt{(x_a - x_b)^{\mathrm{T}} S^{-1} (x_a - x_b)} \qquad (8.3)$$

where S is the matrix of all pairwise covariances between variables.

3.8.3. Data Scaling

Data pretreatment is a necessary condition of multivariate analysis. Besides appropriate transformation of data according to their type of distribution (e.g., taking logarithms for concentration values in the range of the detection limit is recommended), data scaling is necessary as a first step of analysis for almost all methods. Scaling means that each variable is linearly transformed to have zero mean and unit standard deviation. In this way the fact of different variables being measured in different units (e.g., concentrations in parts per million or percent is circumvented, to avoid an arbitrary influence in the total variance or the pairwise object distances. After scaling, all variables have equal a priori influence.

3.9. Factorial Methods

3.9.1. Principal Components Analysis

Principal components analysis (PCA) as a statistical method was introduced by HOTELLING in 1933. Principal components are linear combinations of the original variables with optimal features: the first PC defines maximum variance among all possible linear combinations; the second PC defines maximum variance among all linear combinations uncorrelated with the first one, etc. In this way a small set of a few PCs (generally much fewer than the number of original variables) will suffice to represent the greatest part of the total data variation. Furthermore, in contrast to the original variables, different PCs are always uncorrelated, which makes them useful for many statistical applications. Geometrically, PCs may be identified as new coordinate axes that point to directions of large variance. Using the coordinate system defined by the first two PCs is a very popular approach for visualizing data structure in a diagram. From the computational viewpoint, PCs are obtained by solving the eigenvalue problem

$$R \cdot v_j = \lambda_j \cdot v_j \quad (j = 1, \ldots, p) \qquad (9.1)$$

where R is the correlation matrix (see Chap. 3.8); v_j are the eigenvectors; and λ_j are the eigenvalues of R.

The number of linearly independent eigenvectors is at most p. Each v_j consists of p components v_{ij}, which are the coefficients (weights) of the original variables. The ith realization of the jth PC is (with x_{ik} being an entry of the scaled data table)
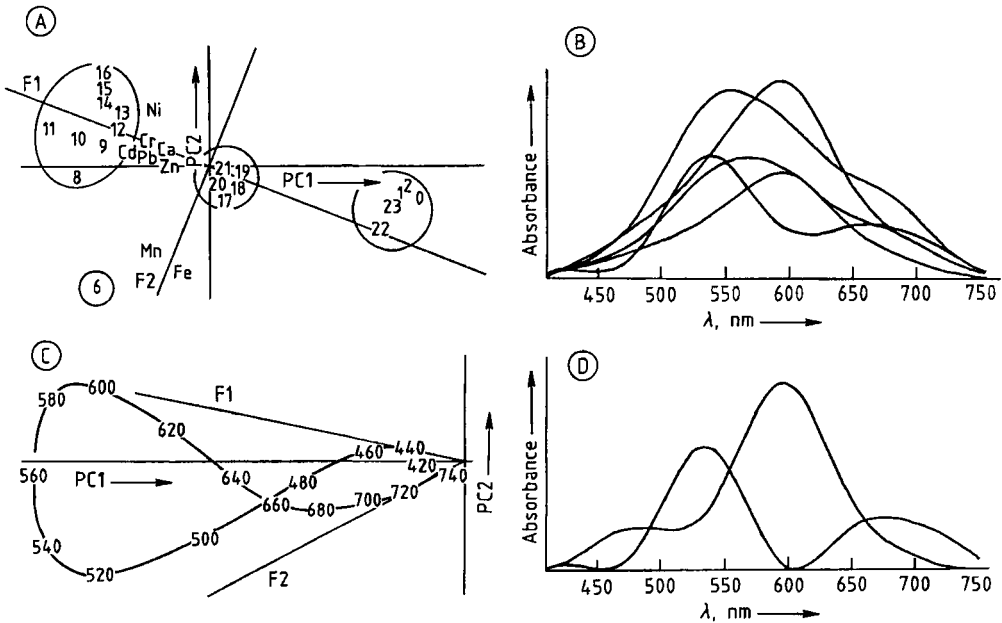
Figure 10. Examples of PCA and FA

A) PC diagram for trace element characterization of wastewater samples; B) Simulated (noise-added) absorbance spectra of five mixtures of two pure components; C) PC wavelength plot for B) including the axes F1 and F2 of pure components; D) Pure component spectra related to B) using the self-modeling method

$$PC_{ij} = \sum_{k=1}^{p} x_{ik} \cdot v_{kj} \qquad (9.2)$$

By using just the first two PCs, a diagram plot for the *objects* is obtained with PC_{i1}, PC_{i2} as coordinates of the ith object. These coordinates are frequently called scores. Similarly, in the same plot a representation of *variables* may be given by using v_{i1}, v_{i2} as coordinates (which are called loadings) for the ith variable. Finally, the eigenvalues contain information on the amount of data structure covered by a PC: λ_j is the variance of the jth PC. Frequently, indication of the variance percentage is preferred by relating one eigenvalue to the sum of all eigenvalues, which in the case of scaled data simply equals p (the number of original variables).

Example 9.1. Figure 10 A provides the PC plot (PC1 versus PC2) for a data set containing analytical results of eight trace elements (variables) in 20 samples (objects) of urban wastewater. The numbers in the plot refer to hours of sampling time in the course of a day. The element patterns of samples do not change continuously, but rather abruptly; four clusters of sampling time are visible: (6 h), (8 – 16 h), (17 – 21 h), and (22 – 2 h). The plot of variables reveals two

major groups of correlating elements: (Fe, Mn), and (Ni, Cr, Cd, Cu, Pb, Zn). Even a joint interpretation of scores and loadings with respect to influence of variable groups in object groups is possible. This aspect is reconsidered below. In this example, the first two of eight eigenvalues were $\lambda_1 = 4.78$, $\lambda_2 = 1.49$; hence, ca. 78 % of total data variation is exhausted by the first two PCs, which is much more than by using any two of the original variables (recall that due to equal weighting by scaling, each pair of original variables covers a percentage of $2/8 = 25\%$ of the total variance).

An important kind of chemometric application of PCA is to find out the number of (unknown) components determining the spectra of a set of mixtures of these components.

Example 9.2. Figure 10 B shows simulated and noise-added spectra of five mixtures as an example. Each of these spectra may be digitized (in the example, 680 equidistant wavelengths in the range between 410 and 750 nm were used) to yield a data table with wavelengths as objects and absorbances of the mixtures at these wavelengths as variables. The total variance within such a data set is composed by a systematic part due to changing proportions of the underlying components within the mixtures and by a generally much smaller part due to noise. Consequently, an eigenvalue analysis should provide a group of large eigenvalues corresponding to the principal components, which reflect chemical composition, and a group of very small eigenvalues relating to PCs, which reflect only noise. Various methods

exist for objective separation of both groups: graphical (e.g., scree-test), statistical (e.g., Bartlett's test on sphericity or cross-validation) or empirical procedures (e.g., Malinowskis indicator function). In the given artificial example the eigenvalue sequence is 4.453, 0.540, 0.004, 0.002, 0.000, making clear, even without any objective evaluation, that the number of underlying chemical components is probably two.

A review on PCA from the chemometric viewpoint may be found in [12].

3.9.2. Factor Analysis

Factor analysis (FA), developed in 1947 by THURSTONE, is a statistical method with different model assumptions compared to PCA. In chemometric applications, however, FA is frequently considered as an additional step after PCA, with the aim of making the results of the latter easier to interpret (a chemometric approach to FA is found in [13]). One important feature of FA is rotation of PCA solutions. In general, PCs are abstract variables without physical meaning. Therefore, an attempt is made to rotate the axes corresponding to the first few significant PCs so as to make the new axes better fit certain groups of variables and facilitate their interpretation as physically meaningful latent factors. The variety of rotations is subdivided into orthogonal (rotations in the true sense) and oblique transformations. *Orthogonal transformations* leave unchanged the variance percentage covered by the rotated axes. *Oblique transformations*, in contrast, do not provide uncorrelated factors but have more degrees of freedom for fitting groups of variables. From the computational viewpoint a successful transformation is achieved by optimizing an appropriate goal function (e.g., Kaisers varimax criterion, compare [13]). In Example (9.1) (Fig. 10 A) the optimal pair of orthogonal factors is denoted by F1, F2, respectively. It obviously fits the variable (and object) structure better than the PCs. Now, the above-mentioned joint interpretation of objects and variables may be given. Obviously, F1 is determined by (Ni, Cr, Cd, Pd, Cu, Zn) and it discriminates three major groups of sampling times with decreasing concentrations in the mentioned six elements from the beginning to the end of a day. This makes it reasonable to identify F1 as a factor in industrial pollution of wastewater. F2 consists of the (Fe, Mn) group, which is related to the 6-h sample. High concentrations of Fe and Mn in wastewater in the early morning might re-

sult from water pipes in private households. With some caution, this could represent a second, independent factor in wastewater exposure.

Another, frequently employed aspect of FA is target testing. This procedure is very useful for deriving new hypotheses from the given data set (for details of the method and applications, particularly in chromatography, see [13]). In Example (9.2) a specific chemical component might be suspected to be present in the mixtures. To verify this, the vector of digitized spectrum of this component could be considered a target to be tested. If the presence of the fixed component is probable, its digitized spectrum vector should be a linear combination of the score vectors of significant PCs since these are, in turn, linear combinations of the spectra of the unknown pure chemical components. The test is performed by multiple linear regresssion (see Chap. 3.11.1), indicating whether the target spectrum may be successfully predicted (within the range of experimental error) by a suitable linear combination of the PC spectra.

A powerful method of FA is *self-modeling of mixture spectra* [14], which allows, under a given assumption, the spectra of underlying pure components to be isolated even if these cannot be modeled by a certain band shape such as Gaussian or Lorentzian. This is a good alternative to conventional least-squares fit. Figure 10 C shows the PC score plot (680 points for different wavelengths in the range 410–750 nm), for Example (9.2). According to Beer's law, absorbances of mixtures are positive combinations of the absorbances of pure components. Therefore, the wavelength points of mixture spectra in a PC plot have positive coordinates with respect to factor axes representing the pure components. Consequently, these pure component axes must define some cone that contains all the wavelength points. This is easily illustrated if, as in the example, the number of pure components equals two. Of course, some degree of freedom exists for choosing the axes F1 and F2. Under the specific assumption, however, that each pure components absorbs radiation uniquely at least at one wavelength (within the considered range), the cone defined by F1 and F2 must be a minimal one (i.e., both axes must pass the wavelength loop tangentially; at 440 and 730 nm in the example). Having found F1 and F2 in this way, the pure component spectra are easily established: The absorbances of the first and second pure components, respectively, are given by projecting corresponding wavelength points of the PC plot onto F1 and F2, respectively. Projection
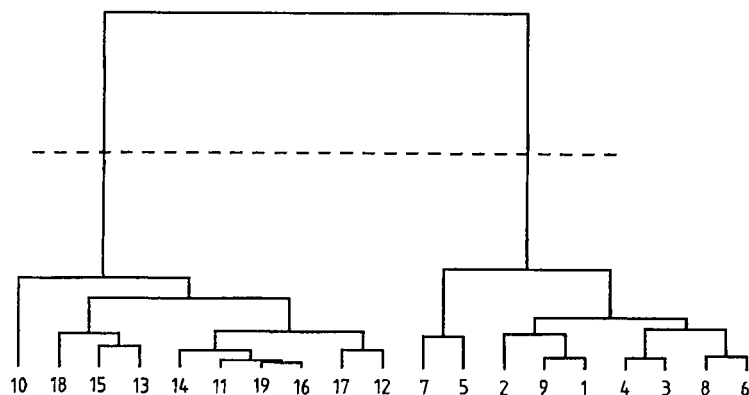
Figure 11. Dendrogram for hierarchic clustering of 19 tungsten powder samples

must be carried out parallel to both axes. The result is illustrated in Figure 10 D.

3.10. Classification Methods

3.10.1. Cluster Analysis

The classification of objects by their patterns in the data set plays a central role in chemometrics. Depending on the prior knowledge concerning the given problem such classification may be performed in a supervised or nonsupervised way. The main tool of nonsupervised or automatic classification is cluster analysis, which produces a partition of the set of objects into several homogeneous subgroups (for an introduction to this field, compare [15]). From a rough viewpoint, cluster analysis may be subdivided into hierarchic and nonhierarchic clustering. *Hierarchic clustering* generates a sequence of partitions that satisfy the definition of a hierarchy; i.e., if any two classes of possibly different partitions are compared, either these classes are disjoint or at least one class is contained in the other (overlapping is excluded). This sequence may be visualized by a so-called dendrogram similar to genealogical trees in taxonomic classification.

Example 10.1. Consider Figure 11, which resulted from hierarchic clustering of 19 tungsten powders being characterized by 11 trace elements. Partitions with different degrees of refinement are obtained by horizontal cuts of the dendrogram. At the lower end, all objects occur as separated classes; at the upper end, all objects are fused in a single class. The level at which certain subclasses are fused (vertical direction) may be interpreted as their dissimilarity. As a

consequence, homogeneous groups are found below long branches of the tree. In the example, a bipartition (1 – 9 and 10 – 19) of the object set is suggested. In fact, the two groups of trace element patterns are in coincidence with two different chemical treatments of tungsten raw materials.

Depending on whether the sequence of partitions is generated by starting at the lower end (successive fusion) or at the uper end (successive splitting), agglomerative and divisive procedures can be distinguished. Most of the well-known methods are *agglomerative* and obey the following type of general algorithm:

1) Given a data set X compute the distance matrix D (see Chap. 3.8) and define each object to form a single class.
2) Find out the most similar pair of classes [e.g., C_1, C_2 ($C_1 \neq C_2$)], realizing the smallest distance $d_{1,2}$ in D. Fuse all elements of C_1 and C_2 to yield a new, larger class $C_{1,2}$. Fix $d_{1,2}$ as the index of fusion (height in the dendrogram). Update D (with a reduced number of classes) by recomputing all distances between the new class $C_{1,2}$ and all old classes that differ from C_1 and C_2. Repeat step 2 until all objects are fused in a single class.

In fact, agglomerative procedures differ only by the way of updating the new distance matrix in each step. As an essential drawback of hierarchic clustering, "mistakes" of fusion at a low level of aggregation are retained, due to hierarchy, in the whole sequence of partitions.

If different partitions are obtained in an independent way by varying some parameter, *nonhierarchic clustering* is used. In so-called partition-making algorithms, such as k-means or Forgy's
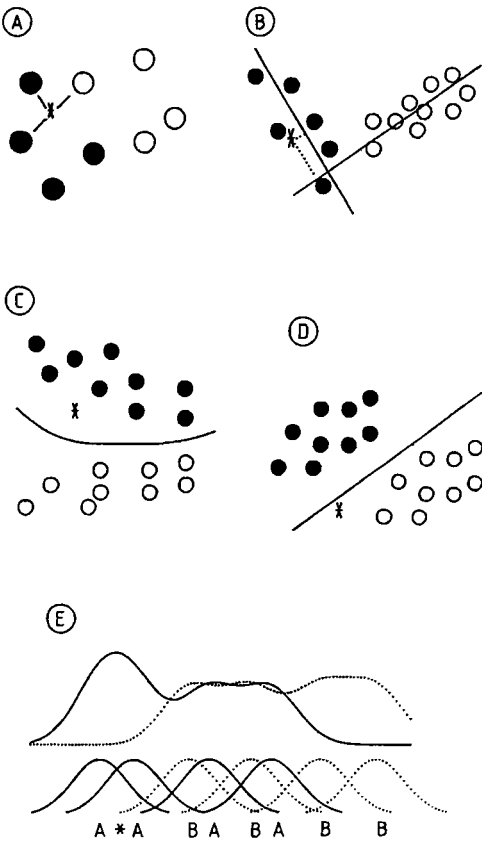
Figure 12. Illustration of supervised classification methods (filled and unfilled circles represent learning objects of different classes; asterisk indicates a test object)
A) The *k*-nearest neighbor (KNN) method; B) SIMCA method; C) Quadratic Bayesian classification; D) Linear Bayesian classification; E) ALLOC method

method, the number of classes expected must be predefined. Different numbers of classes yield different partitions that might have some overlapping; this would indicate an uncertainty in the class structure. Various criteria are used to decide which number of classes or which partition is appropriate. Other important representatives of nonhierarchic clustering are fuzzy clustering and potential clustering (the nonsupervised version of the ALLOC method discussed below).

3.10.2. Supervised Classification

The general principle of supervised classification may be sketched as follows: For specific predefined classes of interest (e.g., different labels

of brandies in food chemistry) a number of samples can be characterized by certain properties (e.g., areas of relevant gas-chromatographic peaks). These samples are called learning objects because their origin is known. Based on the data set of learning objects, rules are derived for subdividing the space of all possible patterns into regions corresponding to the classes. Given any additional objects of unknown origin, they can then—after the same set of properties are determined for test as for learning objects—be assigned to the most probable class (according to the region into which the test object's pattern falls).

An important feature of supervised learning is estimation of the probability of misclassification which has to be kept as small as possible. Since a decision on correctness of classification cannot be made with real test objects having unknown origin, the risk of misclassification must be estimated from the learning objects. The simplest way of doing so would be to plug the data on learning objects into the classification rules that were derived before by the same objects. This method, which is called *resubstitution*, is easily and quickly applicable but it suffers from an underestimation of the real risk of errors. A useful alternative to resubstitution is the *leave-one-out method*, which deletes each learning object one at a time, derives the classification rule by means of the remaining objects, and applies this rule to the variable pattern of the left-out object. The number of misclassifications is much more realistic in this case. Formally applied, the leave-one-out method is very time consuming, but for most classification methods, efficient updating formulas exist to give reasonable computing times. The leave-one-out method, is a special type of cross-validation, where not one but several objects are left out simultaneously.

Some of the most frequently used classification methods are illustrated in Figure 12. For the purpose of visualization the objects are arranged in a plane (e.g., samples that are characterized by just two trace elements or peak areas, etc.). For simplicity, the discussion is restricted to the two-class problem (filled and unfilled circles represent different classes of learning objects). Figure 12 A relates to the *k-nearest neighbor (KNN) method*. Given the pattern (coordinates) of any test object (asterisk) the Euclidean distances can be computed between this test object and all the learning objects by using Equation (8.2). For a certain number *k* (e.g., *k* = 3) the test object is assigned to the class

holding the majority under the k smallest distances. In Figure 12 A the test object would be assigned to the filled-circle class. Given a data set, the appropriate number for k is determined to yield minimum misclassification rate by the leave-one-out procedure. The main disadvantage of the KNN method is that all distances must be recomputed for each new test object, which makes this very time consuming for larger data sets.

Figure 12 B illustrates the soft independent modeling of class analogy (*SIMCA*) *method* [16], where each class of learning objects is modeled separately by a principal components approximation. In the figure, each of the two learning classes is sufficiently described — up to noise — by a straight line or one principal component, respectively. Now, a test object is classified according to its distance (orthogonal projection) to the approximating linear subspaces. In the figure the asterisk is assigned to the filled-circle group. The optimal number of principal components for each class is determined by cross-validation.

A classical method based on the assumption of normal distributions is *Bayesian classification* (Figs. 12 C and D). Roughly speaking (with some simplification), a test object is assigned to the class the centroid of which is nearest in the sense of Mahalanobis distance (see Eq. 8.3). Here, centroid refers to an average object having mean values of a certain class in all variables. The Mahalanobis distance is essentially based on the covariance matrix or, in illustrative terms, the shape of the class distribution. In the case of spherical distribution the Mahalanobis distance coincides with the conventional Euclidean distance (Eq. 8.2). In this special case the test object would be assigned to the class with nearest (in the sense of conventional distance) centroid. This trivial classification rule, however, is not appropriate for nonspherical distributions occurring due to correlations among variables. In contrast to this, the Mahalanobis distance yields an optimal classification rule, provided multivariate normality can be assumed. If the Mahalonobis distance is computed for the covariance matrices of each class separately (by assuming different shapes of distribution), a nonlinear decision curve (surface for more than two variables) is generally obtained (see Fig. 12 C, where the test object is assigned to the filled-circle class). This situation is referred to as quadratic classification. Under the assumption of equal shapes of distribution as in Figure 12 D, decision lines are obtained (hyperplanes for greater dimensions). Therefore this method is called linear

classification. Here the test object would be assigned to the unfilled-circle class. A drawback of quadratic classification is the need to estimate many parameters (different class covariance matrices). If this estimation is not supplied with enough learning objects, the decision surface or classification rule, respectively, will become rather uncertain. This is why linear classification is often used even if the shapes of the distribution are unequal. Then an equalized covariance matrix is generated by averaging.

Another type of classification is the allocation of test objects (*ALLOC*) *method* [17], where the densities of class distributions are estimated by potential functions resulting from so-called density kernels. These kernels, usually Gaussian shaped, are constructed around each of the learning objects. Summing them up pointwise for each class separately yields the desired density estimations. Now a test object is assigned to the class having maximum "influence" in it; this means attaining the highest value of density function at the position of the object. In Figure 12 E, where objects are arranged in one dimension for simplicity, the test object would be assigned to class A. Here, solid curves (Gaussian kernels and summed-up potential function) refer to class A and dashed curves to class B. As for KNN (optimal k) and SIMCA (optimal number of PCs), some parameter must be adjusted in ALLOC, namely, the smoothing parameter, which determines the flatness of the kernels. Obviously, neither too sharp nor too flat kernels are desirable for optimal class separation. The best decision can be evaluated once again by some cross-validation technique.

3.11. Multivariate Regression

3.11.1. Multiple Linear Regression

Multiple linear regression (MLR) is a method that estimates the coefficients $c_1 \ldots c_p$ in a linear dependence

$$y = c_1 \cdot x_1 + \cdot \cdot \cdot + c_p \cdot x_p \qquad (11.1)$$

of some response variable y on several predicting variables x_i. If both predicting variables and responses are observed n times, the resulting data can be collected in a data set X (n rows and p columns) and a vector y (n components). The op-

timal choice of the coefficient vector c in Equation 11.1 (p components) is obtained as the solution of the linear equation $X^TXc = X^Ty$ (where X^T is the transpose of the matrix). By using the resulting coefficients c_j, new responses of y can be predicted from new constellations of the x_i by means of Equation 11.1. Besides statistical evaluations of the computed results (e.g., confidence intervals for coefficients and for predicted values) an essential task of MLR is selecting a significant subset among the possibly numerous predictors [5]. A typical application of MLR arises in multicomponent calibration: Given the spectrum of a sample that is a mixture of p components, the concentrations of these components in the mixture sample need to be determined. Digitizing the spectra of the mixture and of the pure components, respectively, by n distinct wavelengths yields vectors y and x_i, respectively, which are related according to Equation (11.1) by Beer's law, where the coefficients c_i represent the desired concentrations. The pure components need not even be known exactly. It would be sufficient to start computations with a large set of pure components that are possibly included. In an ideal situation, the correct subset could then be detected by some variable selection technique as mentioned above. A serious disadvantage of MLR is the multicollinearity problem. This problem occurs when all or some of the predicting variables are highly correlated. As a consequence, the estimation of regression coefficients in Equation (11.1) becomes more and more unreliable. The fact may be illustrated by the extreme situation in which all predicting variables are identical and, hence, the response variable may be described by arbitrary coefficients. In the example of spectral data, high correlations are generated by components having similar spectra (strong overlapping). Such situations would lead to unreliable computed concentrations.

3.11.2. Latent Variable Regression

The multicollinearity problem can be circumvented by application of *principal components regression* (PCR), which, roughly speaking, is MLR applied not to the original set of predicting variables but to a smaller set of principal components of these variables. Since PCs are always uncorrelated (see Chap. 3.9), multicollinearities may be avoided. The deleted components, however, although describing only a smaller amount of variance in the set of predictors, might account for a

great part of prediction of the response variable that will be lost. A procedure for selecting latent variables as preditors by taking both multicollinearity and predictive ability into consideration is called latent root regression [18].

For multivariate calibration in analytical chemistry, the *partial least squares* (PLS) method [19], is very efficient. Here, the relations between a set of predictors and a set (not just one) of response variables are modeled. In multicomponent calibration the known concentrations of l components in n calibration samples are collected to constitute the response matrix Y (n rows, l columns). Digitization of the spectra of calibration samples using p wavelengths yields the predictor matrix X (n rows, p columns). The relations between X and Y are modeled by latent variables for both data sets. These latent variables (PLS components) are constructed to exhaust maximal variance (information) within both data sets on the one hand and to be maximally correlated for the purpose of good prediction on the other hand. From the computational viewpoint, solutions are obtained by a simple iterative procedure. Having established the model for calibration samples, component concentrations for future mixtures can be predicted from their spectra. A survey of multicomponent regression is contained in [20].

3.12. Multidimensional Arrays

The increasing complexity of experimental designs and hyphenated methods requires a generalization of data tables to higher dimensional arrays. Frequently, a fixed classical objects/variables setting is observed several times under different conditions, yielding a separate data table for each condition. Piecing all these together, one arrives at a cubic or three-dimensional data array, as sketched in Figure 13. In this data example, a set of six trace element concentrations was determined in five different fodder plants for 32 samples (eight parallel detrminations in each of four different sampling regions labeled a to d. While the general entry of a data table may be referred to as x_{ij} (see Chapter 3.8), it must be denoted by x_{ijk} for a three-dimensional array. As before, the indices relate to the corresponding ordered item in each of the three dimensions. In the array of Figure 13, for example, x_{123} represents the concentration of Cd determined in grass for the first sample of region a, where the order of indices
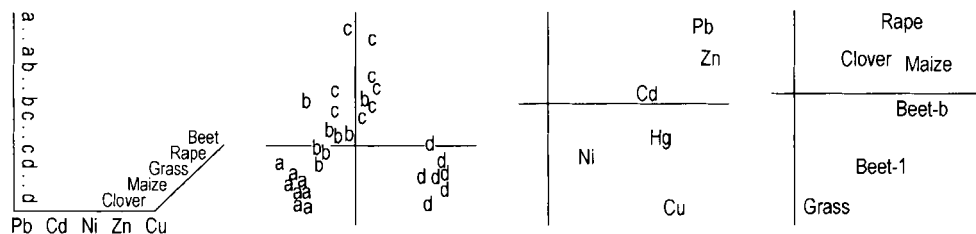
Figure 13. Example of a three-dimensional data array from environmental chemistry (left) and diagram plots for the first two factors of each of the three dimensions (to the right)

is: objects (samples), variables (elements), and "conditions" (fodder plants).

Such arrays raise the question of more generalizations of the table-oriented techniques presented in Chapters 3.9 to 3.11. The most prominent representatives of factorial methods are the so-called Tucker3 [21] and PARAFAC (parallel factor analysis) [22] models. For three-way arrays, the Tucker3 model is expressed as

$$x_{ijk} \approx \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} c_{pqr}\, e_{ip}\, g_{jq}\, h_{kr} \qquad (12.1)$$

where, P, Q, R denote relatively small numbers of components for each of the three dimensions, the e_{ip}, g_{jq}, h_{kr} are entries of so-called component matrices (frequently required to be orthonormal), and the c_{pqr} refer to elements of a so-called core matrix. The approximation sign in Equation (12.1) is to be understood in a least-squares sense. The component matrices contain the scores or loadings of the factors for objects, variables, and conditions, and as such are generalizations of the principal components for data tables indeed, Eq. (9.2) relates to a special case of Eq.(12.1) when reducing the dimension of the array from three to two). In particular, choosing $P = Q = R = 2$, the component matrices contain the coordinates for diagrams of objects, variables, and conditions. In this way, generalizations of plots like Figure 10 a can be obtained for three dimensions. The core elements indicate how single factors of different dimensions are linked with each other. Their squared values c_{pqr}^2 provide the amount of variation in the data explained by jointly interpreting factors p, q, and r, from objects, variables, and conditions, respectively. As an illustration, one may consult the diagram plots for samples, elements and fodder plants collected in Figure 13. For a brief interpretation of a first factor, one recognizes that all elements and fodder plants are given some signed (positive) scores on the first

axis, which at the same time clearly differentiates between sampling regions in the order $d > c$, $b > a$. Hence, this first axis can be interpreted as one of general pollution exposure (not differentiating between elements and plants) with regions affected by pollution according to the just-stated order. A second factor then would differentiate between plants (e.g., grass as opposed to rape) or elements (Pb, Zn as opposed to Cu).

The PARAFAC model is slightly more restrictive, hence simpler, than Tucker 3:

$$x_{ijk} \approx \sum_{p=1}^{P} e_{ip}\, g_{jp}\, h_{kp} \qquad (12.2)$$

This model is particularly important for analysis of data from hyphenated methods. If, for instance, mixed samples of P chemical components are characterized by spectroscopic-chromatographic measurements, then the signal intensity x_{ijk} may be approximated as in Equation (12.2) by means of the spectral and chromatographic profiles g_{jp}, h_{kp} at wavelength j and retention time k of component p at unit concentration and the concentration e_{ip} of component p in mixture i. A peculiarity of the PARAFAC model is that its decomposition according to Equation (12.2) is unique, in contrast to the accordingly reduced decomposition of two-dimensional tables: whereas for data tables the factors can be rotated without changing the result of approximation, this is not possible for three-dimensional arrays. As a consequence, it is not necessary, for the decomposition to be followed by an appropriate rotation to arrive at chemically meaningful factors (e.g., spectra with nonnegative intensities or nonnegative concentration values). The PARAFAC model based on three-dimensional arrays is in principle able to find the correct and unique decomposition of the given data array in a single step. This fact highlights the importance of the model for qualitative (e.g., identification of spectral and chro-

matographic profiles of unknown pure components in the mixtures) and quantitative (e.g., concentrations of pure components in the mixture) analysis at the same time.

The Tucker 3 and the PARAFAC models are easily generalized to arrays of arbitrarily high dimension by using tensorial notation (for a review, see [23])

3.13. References

[1] D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, L. Kaufman: *Chemometrics: a textbook*, Elsevier, Amsterdam 1988.

[2] M. A. Sharaf, D. L. Illmann, B. R. Kowalski: *Chemometrics*, John Wiley & Sons, New York 1986.

[3] D. C. Montgomery, G. C. Runger: *Applied Statistics and Probability for Engineers*, John Wiley & Sons, New York 1999.

[4] L. Sachs: *Angewandte Statistik*, Springer, Berlin 1999.

[5] S. Chatterjee, B. Price: *Regression Analysis by Example*, Wiley-Interscience, New York.

[6] K. Doerffel: *Statistische Methoden in der analytischen Chemie*, 5th ed., Deutscher Verlag für Grundstoffindustrie, Leipzig 1990.

[7] W. J. Youden: "Technique for Testing Accuracy of Analytical Data," *Anal. Chem.* **19** (1947) 946–950.

[8] R. Plackett, J. P. Burman, *Biometrika* **33** (1946) 305–310.

[9] K. Doerffel: "Selektivitätsprüfung von Analysenverfahren mit Hilfe unvollständiger Faktorpläne," *Pharmazie* **49** (1994) 216–218.

[10] J. C. Miller, J. N. Miller: *Statistics for Analytical Chemistry*, 2nd ed., Ellis Horwood Ltd., Chichester 1989.

[11] R. A. Johnson, D. W. Wichern: *Applied Multivariate Statistical Analysis*, Prentice Hall, New Jersey 1982.

[12] S. Wold, K. Esbensen, P. Geladi, *Chemom. Intell. Lab. Syst.* **2** (1987) 37–52.

[13] E. R. Malinowski, D. G. Howery: *Factor Analysis in Chemistry*, J. Wiley and Sons, New York 1980.

[14] W. H. Lawton, E. A. Sylvestre, *Technometrics* **13** (1971) 617.

[15] M. R. Anderberg: *Cluster Analysis for Applications*, Academic Press, New York 1973.

[16] S. Wold, *Pattern Recogn.* **8** (1975) 127–139.

[17] J. Hermans, J. D. F. Habbema: *Manual for the ALLOC Discriminant Analysis Programs*, University of Leiden, Leiden 1976.

[18] R. F. Gunst, R. L. Mason: *Regression Analysis and its Applications*, Marcel Dekker, New York 1980.

[19] H. Wold in K. G. Jöreskog, H. Wold (eds.): *Systems under Indirect Observation*, North Holland Publ., Amsterdam 1981.

[20] H. Martens, T. Næs: *Multivariate Calibration*, J. Wiley and Sons, Chichester 1989.

[21] L. R. Tucker, *Psychometrika* **31** (1966) 279–311.

[22] R. A. Haeshman, *UCLA Working Papers in Phonetics* **16** (1970) 1–84.

[23] R. Henrion, *Chemom. Intell. Lab. Syst.* **25** (1994) 1–23.

4. Weighing

WALTER E. KUPPER, Madison, New Jersey 07940, United States

4.1. Introduction

The field of mass measurement, commonly called weighing, is conveniently structured according to the degree of numerical precision involved. The classification and nomenclature of modern laboratory balances follows a decimal pattern based on the step size d of the associated digital display:

1) Precision balances: $d = 1$, 0.1, 0.01, or 0.001 g
2) (Macro) analytical balances: $d = 0.1$ mg
3) Semimicro balances: $d = 0.01$ mg
4) Microbalances: $d = 0.001$ mg
5) Ultramicrobalances: $d = 0.0001$ mg

Weighing capacities range from 5 g for an ultramicrobalance to >60 kg for the largest precision balances (Figs. 1 – 5). Precision weighing instruments of greater capacity, known as industrial precision scales and not generally used in the laboratory, are available in a weighing range up to 6 t and $d = 0.1$ kg.

The words "balance" and "scale" are often used interchangeably. "Balance" is derived from the Latin *bilanx*, having two pans [1]. "Scale" and "scales" are short forms of "a pair of scales" (old English, meaning dishes or plates). Modern weighing instruments no longer have two weighing pans, but the words "balance" and "scale" survive. The term "balance" is preferred for the more precise weighing instruments found in a laboratory, while "scale" is appropriate for all other weighing equipment.

4.2. The Principle of Magnetic Force Compensation

Analytical and precision balances are based on the principle of electromagnetic force compensation, where the weight of a sample is counterbalanced by a corresponding electromagnetic force. The measuring sensor or transducer in a balance is best described as a linear motor; i.e., it is an electromechanical aggregate that generates a straight-line force and motion on the basis of an electric current.

The operating principle of a loudspeaker (Fig. 6) provides a useful analogy. In a loudspeaker, an oscillatory, voice-modulated current is fed through a coil (c) to create a small magnetic field that interacts with the larger field from a surrounding permanent magnet (d) to generate an oscillatory force that in turn moves a diaphragm or voice cone (a), thus producing sound. In a balance sensor, direct current through a coil (c) similarly produces a static force of precisely the magnitude

Figure 1. Microbalance Mettler MX5, 5100 mg × 1 μg
Automatic self-calibration with built-in reference masses.

Figure 2. Analytical balance Mettler AX201, 220 g × 0.01 mg
Automatic self-calibration with built-in reference masses.

Figure 3. Analytical balance Mettler AB204-S,
210 g × 0.1 mg
Self-calibration with built-in reference mass.

Figure 4. Precision balance Mettler PR8002, DeltaRange,
8100 g × 0.01 g
Self-calibration with built-in reference mass.
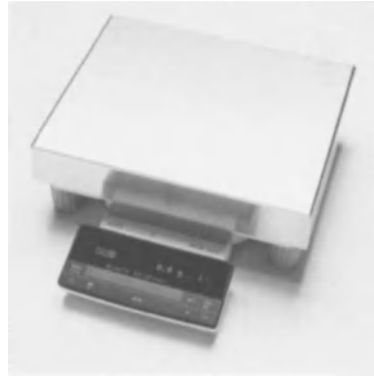
Figure 5. Precision balance Mettler SR32001, 32 100 × 0.1 g
Self-calibration with built-in reference mass.

required to counterbalance the weight of a sample load (b). A position-maintaining feedback circuit (e) controls this current in such a way as to keep the overall magnetic force in equilibrium with the weight. The strength of the required current in this case provides a measure of the weight; i.e., the balance establishes a mass in grams on the basis of a current measured in milliamperes.

A typical electronic balance incorporates a system of links and levers like that shown schematically in Figure 7. The weighing pan (a) is guided by two parallel horizontal links and counterbalanced by the force coil (g) acting through a lever (c). The pivots (b) in this linkage are metallic leaf springs, called *flexures*. In the weighing process, the pan is initially depressed by the sample, after which it is restored to its null level by the servoac-
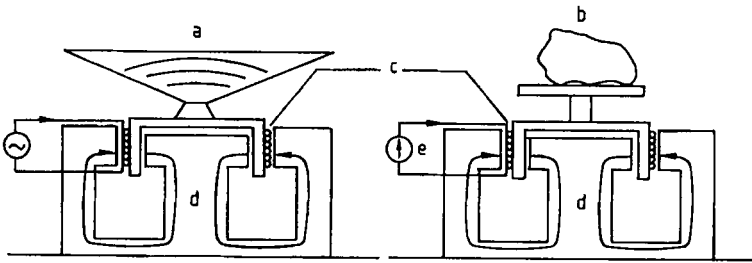
Figure 6. Analogy between a loudspeaker and a laboratory balance, as explained in the text
a) Diaphragm; b) Load; c) Coils; d) Magnets; e) Feedback circuit
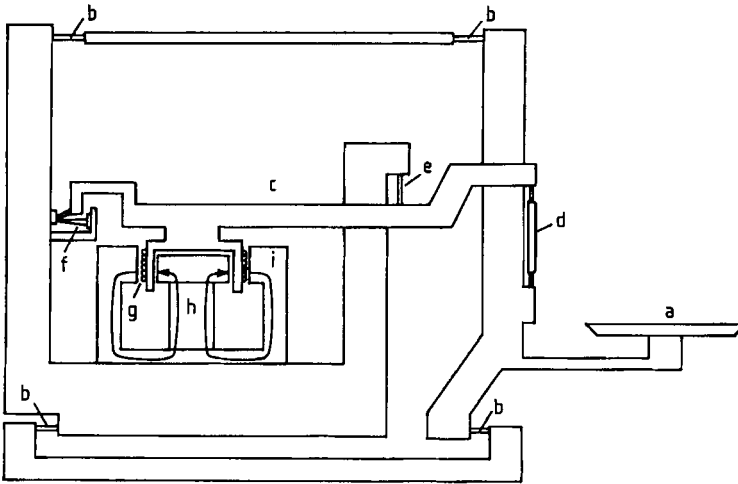
Figure 7. Schematic cross-section through an analytical balance
a) Weighing pan, cantilevered from a parallel-motion linkage; b) Flexure pivots of the parallel-motion linkage; c) Main lever or balance beam; d) Coupling between linkage and beam; e) Main pivot or fulcrum; f) Photoelectric position sensor; g) Force coil; h) Cylindrical permanent magnet; i) Cylindrical soft-iron shell, channeling the magnetic flux

tion of the force motor. This balancing process is controlled with sub-micron sensitivity by a photoelectric position sensor (f).

4.3. Automatic and Semiautomatic Calibration

The electronic signal-processing part of a balance is comparable to a digital voltmeter, except that the display is calibrated in units of mass. The *accuracy* (correspondence between a displayed value and the true mass on the balance) of a modern professional-grade balance is ensured by auto-

matic or semiautomatic self-calibration. In a self-calibration cycle, a reference mass is either deposited on the balance internally or set on the pan manually by the operator, after which the balance's microprocessor updates the value of a calibration factor stored in a nonvolatile (power-independent) memory. This calibration function represents a direct relationship between the weighing object and the displayed value, in that the weighing process is referenced directly to a mass. A more sophisticated form of automatic calibration employs two built-in reference masses to assure both calibration at full capacity and linearity at midrange of the balance (see Figs. 1 and 2).

4.4. Processing and Computing Functions

Thanks to the capabilities of microprocessors, a number of special computation routines are common to most balances:

1) Pushbutton tare for subtracting container weights
2) Automatic zero tracking to keep the display zeroed when no weight is on the pan
3) Vibration filtering to make displayed values more immune to disturbances
4) Stability detection to prevent premature reading or transfer of transient or fluctuating results

In addition to these signal processing features, some balances also provide useful convenience options such as conversion to nonmetric units, parts counting, weight statistics, percentage calculation, and animal weighing.

4.5. Balance Performance

A weighing instrument is commonly tested by loading it with calibrated mass standards and then ascertaining whether the displayed weight values are accurate, meaning that each result corresponds to the value of the test standard within some predetermined tolerance limits. In the most elementary check of balance accuracy, a single test mass is placed once on the balance. This is a test that is often performed for extra assurance in critical applications.

A complete balance test consists of a systematic set of weighings designed to evaluate all aspects of balance performance, including repeatability, eccentric loading errors, linearity, and span calibration. For a balance to be declared in proper working condition, all test results must fall within the tolerance limits specified by the manufacturer. These tolerances normally apply only to new or newly serviced equipment. In other cases a so-called in-service tolerance — equal to twice the original tolerance — is customarily allowed.

Repeatability for successive weighings of the same load is calculated as standard deviation from ten weighings of the same mass, where the balance is reset to zero as necessary before each weighing. The procedure is usually conducted with a mass of approximately one-half the weighing capacity of the balance. This test measures short-term fluctu-

ations and noise of the type present to some degree in every electronic instrument. Excessive standard deviation may be a result of adverse environmental conditions, such as building vibrations or air drafts. A factor equal to three times the standard deviation is customarily called the *uncertainty* of the balance, meaning that virtually all observed values (theoretically, 99.7 %) would fall within this limit on either side of the average.

In the *eccentric load test* a weight equal to one-half the capacity of the balance is placed at the center of the weighing pan and the balance is reset to zero. The weight is then moved, in turn, half-way to the left, right, front, and rear of the pan. Weight readings from the four eccentric positions, if different from zero, are called eccentric loading errors. If not specified separately, the tolerance limit for such errors is of the same order of magnitude as that for nonlinearity. Larger errors indicate that the mechanical geometry of the balance is out of adjustment, usually as a result of abusive treatment. A correctly adjusted balance is practically insensitive to eccentric placement of the load, but any residual errors will be minimized if samples and containers are approximately centered on the balance pan.

Nonlinearity refers to deviations from a mathematically straight line in a graph of display readings versus actual weights, where the line is drawn from the zero point to the full-capacity endpoint. Linearity errors in a balance are caused by practical imperfections in the theoretically linear scheme of electromagnetic force compensation. Nonlinearity manifests itself as a variation in results when the same object is weighed together with differing amounts of tare, and this is, in fact, how the nonlinearity of a balance is tested. A weight equal to one-half the capacity is weighed once without tare, then with a tare of the same magnitude as the weight itself. The mid-range linearity error of the balance is equal to one-half the difference between the two results. This test detects the most common type of balance nonlinearity, an error curve of parabolic shape with maximum positive or negative error at mid range. Higher-order components in the error curve can be detected with four weighings of a test weight equivalent to one-fourth the capacity, using tares equal to 0, 25, 50, and 75 % of the balance capacity. The advantage of such a procedure is that it does not require calibrated weights; the test weight as well as the tares represent nominal values only.

The *calibration error*, also known as sensitivity error or span error, is a small relative factor (a
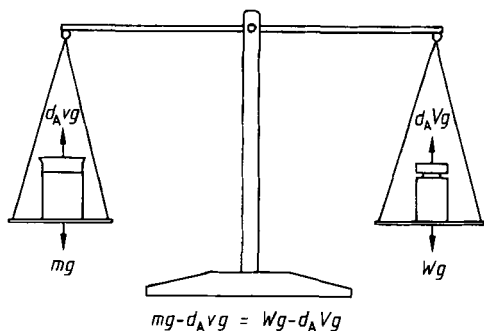
Figure 8. Gravity and buoyancy forces acting on a two-pan balance
v = sample volume. V = volume of the mass standard.
g = acceleration due to gravity.

constant percentage of the applied load) by which all weight readings may be biased. For a balance with a self-calibration device this error is limited to the error tolerance of the reference weight, which is of the order of 0.0001 % (1 ppm) in the case of an analytical balance. If the balance lacks self-calibration, calibration errors may develop:

1) With large changes in room temperature
2) When the balance is moved to a different location
3) As a long-term aging effect in the electromagnetic force sensor

Calibration errors should be ascertained with a certified test weight of substantially lower uncertainty than the calibration tolerance of the balance (insofar as this is possible or necessary in practice). In the finest analytical balances, the achievable calibration-error tolerance is dictated by the accuracy of the available mass standards. However, this limitation is not relevant at the level of 0.1 % to 0.01 % accuracy applicable to most laboratory work.

The temperature dependence of the calibration error (temperature drift) is tested in new balances at the factory. Test weighings are made in an environmental test chamber at various temperatures (e.g., 10, 20, and 30 °C). A typical tolerance for the span drift (sensitivity drift) as a function of temperature for an analytical balance is 1 ppm/ °C, also expressed as $10^{-6}/$°C. With a test weight of 100 g, this means that a correct reading of 100.0000 g at 20 °C could change to 99.9999 g or 100.0001 g (or anywhere in between) if the temperature were to change to either 19 or 21 °C.

4.6. Fitness of a Balance for Its Application

To select an appropriate balance for a given application, one needs to answer two basic questions:

1) How heavy are the samples to be investigated?
2) How accurately must they be weighed?

For example, if the accuracy objective is 0.1 %, then a balance should be selected for which the sample weights would represent at least a thousand times the value of the uncertainty, or three thousand times the standard deviation. Adherence to this rule may be encouraged by placing an appropriate label on the balance, such as "Minimum net sample weight 60 milligrams" for a balance with a standard deviation of 0.02 mg. To weigh smaller samples, one would select a balance with a smaller standard deviation. This requirement of 0.1 % accuracy and its interpretation in terms of the standard deviation are mandated for pharmaceutical weighing according to the *United States Pharmacopeia* [2].

4.7. Gravity and Air Buoyancy

All weighing relies ultimately on the concept of equilibrium with a symmetrical lever (Fig. 8). A *weight*, in the common usage of the term, is represented by the total value of calibrated weighmasses required to put the lever in balance. According to international convention, the reference masses are to be fabricated from stainless steel with a density of 8 g/cm³ [3]. The forces in equilibrium arise primarily from the downward pull of gravity on the two sides of the balance, but small buoyant forces from the surrounding air must also be considered. In Figure 8, the gravity and buoyancy forces are symbolized by arrows. The equation under the diagram expresses the equilibrium in mathematical terms. Solving the equilibrium equation for m leads to the conversion from weight to mass.

$$m = W \cdot (1 - d_A/D)/(1 - d_A/d) = W \cdot k \qquad (1)$$

m sample mass, expressed in the same metric unit as that applicable to W

Table 1. Weights at sea level (W_S) and in Denver, Colorado (W_D), for samples with a mass of one gram and various densities d

Sample density d, g/cm^3	Weight at sea level W_S, g	Weight in Denver W_D, g	Difference $(W_D - W_S)/W_S$, %
0.5	0.997749	0.998124	0.0376
1.0	0.998949	0.999124	0.0175
2.0	0.999549	0.999624	0.0075
4.0	0.999849	0.999874	0.0025
8.0	1.000000	1.000000	0.0000
20.0	1.000090	1.000074	−0.0017

W sample weight, defined in terms of the counterbalancing steel mass
d_A density of ambient air, g/cm^3
D density of the mass standard, normally 8 g/cm^3
d density of the sample, g/cm^3
k weight-to-mass conversion ratio for the sample

According to this expression, there are two cases in which sample mass equals sample weight ($k = 1$): at zero air density (vacuum), or if the sample density d equals the density D of the mass standard.

Typical values for air density are approximately 1.2 mg/cm^3 at sea level and 1.0 mg/cm^3 in Denver, Colorado (altitude 1600 m). The air density may also vary by as much as 6 % with changes in the weather. Given that the ambient air density is variable, it follows from Equation (1) that the same mass will produce slightly different weights at different times and places, as illustrated in Table 1. This variation is too small to be relevant in commercial applications, however, and the same is true for most laboratory weighings as well.

The conclusions derived above for a two-pan balance also apply to modern electronic balances. Weighing still entails comparison of the test object or sample with a calibrated steel mass. However, in the case of a single-pan electronic balance, the standard is placed on the pan at a particular point in time in order to calibrate the balance, and the weighing process occurs on the same pan at a subsequent time. Thus, weighing on an electronic balance involves a sequential rather than simultaneous comparison, which has one important implication: an electronic balance is the functional equivalent of a classical two-pan balance only if the surrounding gravity field and atmosphere are the same during weighing as when the balance was calibrated. Therefore, electronic balances must always be calibrated at the place where they are to be used, particularly since the force of gravity varies by more than 0.1 % over a territory as limited as the continental United States.

4.8. The Distinction Between Mass and Weight

Weight, in the primary definition of the term, means the direct result of a weighing, without any correction for air buoyancy. Mass, in the context of weighing, is understood as an absolute, buoyancy-corrected quantity. A problem arises if a weighing result is reported as a mass, when in fact no buoyancy correction has been made. The discrepancy amounts to about 0.1 % for a material with a density of 1 g/cm^3. In science there is a general desire to avoid the term "weight" in the sense of a weighing result, because scientific nomenclature treats weight as a *force*, properly expressed in dynes or newtons, whereas the gram is a unit of mass. To prevent misunderstandings, if weighing results are reported as "mass" then the context should make it clear whether or not stated values have been subjected to an air-buoyancy correction.

4.9. Qualitative Factors in Weighing

A proper environment and sufficient technical skill on the part of the operator are critical for achieving satisfactory weighing results. Precision balances weighing to 0.1 g perform well in almost any situation, but with analytical and microbalances a number of factors can be detrimental to accurate weighing, including air drafts, direct sunlight and other forms of radiant heat, building vibrations, and electrostatic and magnetic forces. A stable indoor climate is generally sufficient, but balances should be kept away from doors, windows, ovens, and heat and air conditioning outlets, and they should never be placed under ceiling outlets. A mid-range level of relative humidity is best, since this will prevent the buildup of electrostatic charge from excessively dry air as well as

moisture absorption by samples and containers from air that is too humid. The balance table should rest on a solid foundation so the balance will not be affected if the operator moves or leans on the table. There should be no elevators or vibrating machinery nearby. An operator may find it necessary to conduct several experiments in a search for appropriate procedures when weighing hygroscopic, volatile, or ferromagnetic samples. Samples and containers should always be acclimated to room temperature to prevent air convection around the sample that might interfere with weighing.

4.10. Governmental Regulations and Standardization

Weighing is subject to governmental regulation in situations where errors or fraud might result in material or health hazards. Weighing in trade and commerce is commonly supervised by a governmental "Weights and Measures" agency through some system of certification and inspection of scales and weights [4], [5]. In areas other than commerce, the nature and scope of government control over weighing varies depending upon the field, as in the case of pharmaceuticals [2], nuclear materials [6], or military defense contracts [7]. In the United States, critical weighing applications are subject to a system of government-supervised self-control. Any organization conducting weighing processes subject to regulatory control must have adopted formal mass assurance procedures as part of an overall quality assurance system. The government exercises its oversight function through periodic audits. The main elements of a typical mass assurance program are as follows:

1) Balances available to the staff must conform to the accuracy required for the intended use

2) Balances must be maintained and calibrated on a regular cycle, with records kept on file
3) All balance operators must be instructed in correct weighing technique, and training records must be kept on file
4) Formal and documented weighing procedures must be in place
5) Orderly records must be maintained for all weighings that fall within the scope of the regulated activity

The most widely accepted guideline for establishing a system of mass assurance is found in International Standard ISO 9001, Section 4.11., "Inspection, Measuring and Test Equipment" [8]. Under ISO 9001, the quality system of an organization is to be audited and certified by independent quality assessors, analogous to the audits performed by the government on regulated organizations.

4.11. References

[1] B. Kisch: *Scales and Weights: A Historical Outline*. Yale University Press, New Haven 1965.
[2] *The United States Pharmacopeia*, 22nd ed., United States Pharmacopeia Convention, Inc., Rockville, Md., 1990.
[3] OIML Recommendation R-33, Conventional Value of the Result of Weighing in Air, Bureau International de Métrologie Légale, France, 1973.
[4] OIML International Recommendation R 76-1, Non-automatic Weighing Instruments. Part 1: Metrological and Technical Requirements — Tests, Bureau International de Métrologie Légale, Paris, France, 1993.
[5] NIST Handbook 44, U.S. Dept. of Commerce, National Institute of Standards and Technology, Gaithersburg, Md., 1993.
[6] MIL-STD-45662A, Calibration Systems Requirements, Department of Defense, Washington, DC 20360, 1988.
[7] Nuclear Regulatory Commission, Codes 10 CFR 21 and 10 CFR 50, 1989.
[8] ISO 9001.*Quality Systems — Model for Quality Assurance in Design/Development, Production, Installation and Servicing*, 1994.

5. Sampling

JAMES N. MILLER, Department of Chemistry, Loughborough University, Loughborough, United Kingdom

5.1. Introduction and Terminology

The determination of the concentrations of one or more components of a chemical material is a multistage process. Research in analytical chemistry understandably gives emphasis to methods of pretreating materials prior to analysis (e.g., dissolution, concentration, removal of interferences); to the production of standard materials and methods, such as those used to check for bias in measurement; to the development of analytical techniques demonstrating ever-improving selectivity, sensitivity, speed, cost per sample, etc.; and to new methods of optimization and data handling which enhance the information generated by these techniques. Many of these improvements are in vain, however, if the very first stage of the analytical process is not planned and managed properly: this is the *sampling* stage, which should ensure that specimens subjected to analysis are fully representative of the materials of interest. Repeated investigations show that much of the variation encountered when (for example) similar analyses are performed by different laboratories, arises at the sampling stage. This article outlines the general and statistical aspects of the sampling process. Detailed coverage of the chemical aspects of sampling in particular applications is to be found in reviews and monographs [1], [2].

The word *sample* is used in two rather different, but related senses in analytical science. In each case, the general implication is clear — a sample is a small and hopefully representative portion of a larger object. But some differences are also clear. Statisticians use the term to mean a

small number of measurements which are assumed representative of a theoretically infinite *population* of measurements. Chemists and others use the term to mean a small number of specimens, taken in cases where the materials of interest are too large or too numerous to be examined in their entirety. Thus, a water analyst might take ten specimens of water from a particular stretch of river, knowing or assuming that they form a representative sample of the whole of that part of the river at a particular time. Laboratory staff might then make five replicate pH measurements on each specimen: in each case, these five pH values are a sample representing a potentially infinite number of pH values — the population. (It is conventional to use Greek letters for population statistics, and the Roman alphabet for sample statistics.) In this example, the chemical and statistical uses of the word sample are to some degree distinct. In a different application area, an analyst might remove every hundredth capsule from a pharmaceutical production line, and analzye each capsule once to determine the level of the active component. The concentrations obtained would form a sample (in both chemical and statistical senses of the word) representing the behavior of the production line. The latter example also reminds us that sampling theory is closely related to *time series* statistics, and to *quality control* methods.

Once the overall objectives of an analytical process have been set, the sampling steps must be carefully planned: how many specimens (or *sample increments;* see Chap. 5.2) should be taken? How should they be selected? How big should they be (in cases of bulk samples)? One

difficulty is that in some analyses the nature of the sample and the precision of the sampling and measurement steps may not initially be known. In such cases, simple preliminary sampling plans may be used to provide such information, and the resulting data can be used iteratively to improve the sampling procedures for long-term use. Good sampling plans should be available in written form as protocols which describe experimental procedures in detail, set criteria on the circumstances in which a sample can be rejected, and list the additional information that should be gathered when a sample is taken (temperature, weather, etc.). In principle, such protocols allow samples to be taken by untrained staff, but it is desirable (if not always practicable) for a sample to be taken by the analytical scientist who is to perform the subsequent measurements.

5.2. Probability Sampling

Most sampling procedures combine a suitable method of selecting sample items or increments (individual portions) with a knowledge of probability theory. From the analytical data conclusions are drawn on, for example, the risks involved in accepting a particular product. Such methods are known collectively as probability sampling methods. The first step in such a procedure is the selection of the sample items or increments: it is critically important that these are not biased in any conscious or unconscious way, otherwise they will not be representative of the corresponding population. Usually, each part or element of the population must have an equal chance of selection. (This is not absolutely necessary, but if the probabilities of selecting different parts of the population are not equal, those probabilities should be known.) The best method of ensuring such equal probabilities is to use *random sampling*. This method (definitely not the same as haphazard sampling!) involves using a table of random numbers to indicate the items or increments to be selected. The approach can be applied to bulk samples as well as itemized/numbered materials. For example, a truckload of coal could be regarded as being divided into numbered cells by dividing the truck's internal dimensions both horizontally and vertically: the cells to be sampled are then obtained from the random number table. Sampling normally proceeds without replacement, i.e., an item

or cell once sampled is not available for further sampling.

An important component of the sampling process — and an example of how it can be developed and improved with time — is to analyze for any systematic relationships within these random samples. For example, samples taken late in the working day may give results that differ significantly from those taken early in the morning. Such a trend, which if unnoticed might bias the results, can be taken into account once identified. Alternative sampling approaches are less satisfactory than a properly controlled random sample. Simplicity suggests that it might be easier, for example, to sample every hundredth item from a production line: but if the manufacturing process has an undesirable periodicity such a regular sampling plan might conceal the problem.

In some applications it is possible to use a *composite sample*, produced by taking several separate increments from a bulk material, and blending them. This process requires care (especially when solids of different particle size, etc., are involved) to produce a single sample that is analyzed several times. This procedure assumes that it is only the average composition of the bulk material which is of interest. When replicate measurements are made on this composite sample the sampling variance is greater (i.e., worse) than would have been the case without blending, but there is a benefit in that fewer measurements are necessary — considerations of time and cost often arise in the development of sampling plans.

Some materials to be sampled (e.g., ores or minerals) are often heterogeneous, i.e., the components of interest are distributed through the bulk in a nonrandom way. In such cases it is clearly inappropriate to use simple random sampling, and a more complex approach is necessary. This involves dividing the bulk material into a sensible number of segments or strata, and then (using random numbers, as before) taking samples from each stratum. It is usual to use strata of equal size, in which case the sample size for each stratum is also the same, but if the strata are by nature unequal in size, the number of samples taken should be proportional to these sizes. Such *stratified random sampling* clearly involves a compromise between the number of strata chosen and the labor involved in the analyses. If too few strata are chosen, it is possible to take acceptable numbers of samples from each, but some inhomogeneities in the bulk material may remain concealed; if too many strata are chosen, the number of measure-

ments necessary is large, if the sampling variance is not to be too high. This is another instance where prior knowledge and/or iteratively gained experience helps to formulate a sensible sampling plan.

A useful qualitative guide to the principles underlying probability sampling is provided by an ASTM standard [3].

In many cases sampling schemes are regulated by statutes, such as those published by individual governments, the European Community, the Codex Alimentarius Commission, and so on. Such schemes are particularly common in the analysis of foodstuffs, and may specify the mass and number of the sample increments to be taken as well as the statutory limits for the analyte(s) under study.

5.3. Basic Sampling Statistics

The overall random error of an analytical process, expressed as the variance s^2, can be regarded as the sum of two other variances, that due to sampling s_1^2, and that due to the remaining measurement components of the process s_0^2. These variances are *estimates* of the corresponding population variances σ^2 and its components σ_1^2 and σ_0^2. The overall standard deviation s is calculated as usual from single measurements on each of h sample increments, and the confidence limits of the true mean value of the population μ are obtained from:

$$\mu = \bar{x} \pm t\,s/\sqrt{h} \qquad (1)$$

where $\bar{x}$ is the mean of the h measurements, and the t-value is at the appropriate confidence level. This first calculation of the mean and standard deviation allows a sampling plan to be initiated; further measurements then give refined values.

From Equation (1), the variance of $\bar{x}$ is s^2/h, used as an estimate of σ^2/h. To reduce, i.e., improve this variance, it is necessary to take n measurements of each of the h increments. This allows the separation of the sampling and measurement variances by one-way analysis of variance. The replication of the measurements is expected to reduce the measurement variance σ_0^2, and the overall variance of the mean becomes $\sigma_0^2/nh + \sigma_1^2/h$. This relationship provides general guidance on the best practical way of reducing the overall variance. The measurement variance can be mini-

mized by using a high-precision analytical method or by using higher values of n, but in practice the sampling variance is often much larger than the measurement variance, so the choice of h is much more important (see below). If a composite sample is used, i.e., h sample increments blended to form one sample which is then measured n times (see above), the variance of the mean of such replicates is $\sigma_0^2/n + \sigma_1^2/h$. This variance is higher (worse) than when h increments are measured n times each, but the number of measurements is reduced (n as opposed to nh).

When sampling a bulk material, it is clearly important to come to decisions about: (1) how many increments should be taken; and (2) how large they should be. The minimum number of increments h necessary to obtain a given level of confidence is:

$$h = t^2\, s_1^2 / E^2 \qquad (2)$$

where E is the largest permissible difference between the sample estimate $\bar{x}$ and the corresponding population value μ for the determinand (the equation assumes that each increment is measured once only). The t-statistic is used at the desired confidence level, e.g., 95 % or 99.7 %. Since the value of t depends on h, it is necessary to make an initial estimate of h, using the $h = \infty$ value of t (e.g., $t = 1.96$ for 95 % confidence), thus obtaining from Equation (2) a preliminary value of h: a new value of t can then be taken, and a reliable final value of h approached iteratively. The ASTM standard [4] recommends the use of 99.7 % confidence levels. It should be noted that Equation (2) is only applicable if the determinand is distributed according to the Gaussian, or "normal", distribution: different distributions require separate expressions for h.

When a bulk material is examined, the size of each increment is also of importance. Clearly, if each increment is too large, it may conceal the extent of variation within the bulk material; if it is too small, many increments are necessary to reveal the extent of the sampling variance. INGAMELLS [5] utilized the fact that s_1^2 decreased as the increment size increased to develop the equation:

$$W R^2 = K_S \qquad (3)$$

where W is the increment weight, R is the relative standard deviation of the sample composition, and K_S is a constant. The definition of K_S depends on the user (e.g., it might be the weight necessary to
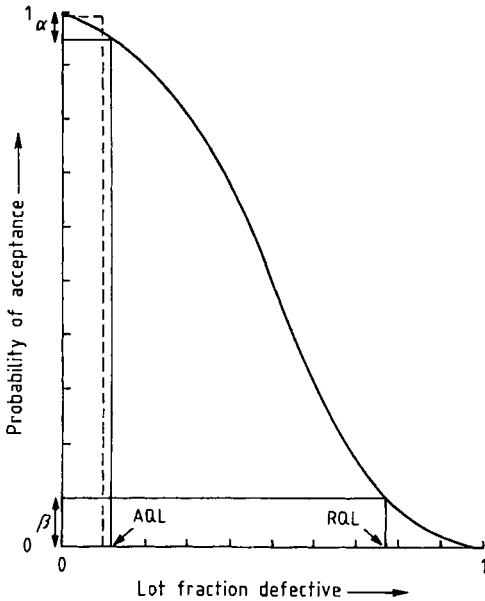
Figure 1. A typical OC curve
The broken line shows the result for an idealized sampling plan

limit the uncertainty in the composition of the material to 2 % with 95 % confidence); it is estimated by carrying out preliminary experiments to determine how s_1 and hence R vary with W. Once K_S is known, the minimum weight W required to produce a particular value of R is readily calculated.

Equations analogous to (2) and (3) can be derived for the more complex situations that arise in stratified random sampling [1], [6]. A computer program has been developed [7] to assist in the solution of sampling problems: it is especially directed at geochemical and other areas where the sizes and shapes of particulate solids may affect R^2.

5.4. Acceptance Sampling

This chapter deals with the problems associated with the acceptance or rejection of a product by a purchaser, on the basis of an examination of a *sample* taken from the whole *lot* of the product. The general aim of the methods summarized here is to protect both the manufacturer (e.g., of a

pharmaceutical product which must contain a specified amount or range of the active ingredient), and also the consumer. The manufacturer is protected if good or valid lots of the product are (wrongly) rejected only infrequently; and the consumer is protected if bad or invalid lots are (wrongly) accepted only infrequently. Since the contents of the components of each lot will show variation (generally assumed to be Gaussian), the term "infrequently" has to be defined numerically (e.g., as 5 % of all occasions), and it is also necessary to give quantitative definition to "good/valid" and "bad/invalid". In some cases, lots are tested by means of a qualitative criterion (sampling by attributes), but more usually a quantitative measure is used (sampling by variables — this approach requires smaller samples than sampling by attributes to achieve a given degree of discrimination), and good and bad are thus defined numerically. Once these definitions are known, it is possible to develop a sampling plan, and describe it by an *operating characteristic* (OC) curve, sometimes known as a power curve, because of its relationship to the concept of the power of a statistical significance test.

An OC curve for sampling by attributes plots the *probability of acceptance* of a lot (y-axis) against the *lot fraction defective* (x-axis), i.e., the proportion of items in a lot which fail to meet the predetermined "good" criterion. A typical curve is shown in Figure 1. The success of a sampling plan in discriminating between good and bad lots depends on the shape of its curve. An ideal plan would give the shape shown in the figure as a broken line: the probability of acceptance would be 1.00 until the lot fraction defective reached a predetermined value, and would then fall abruptly to zero. In practice a curve is always obtained, so two points on it are particularly important. The first is called the *acceptance quality level* (AQL): this is the percentage of defective items that is (just) tolerable with a predetermined frequency, e.g., 5 %. A lot with this proportion of defectives will be rejected with a probability α. This is known as the *producer's risk*. It is often set at 5 %, but this is not universal; in some cases, plans are developed so that $\alpha = 10$ % for small lots, but only 1 % for large lots. A second point of importance is the *rejectable quality level* (RQL), also known as the *lot tolerance percent defective* (LTPD), or the *limiting quality level* (LQL). If β is the probability of acceptance at the RQL, this is clearly the *consumer's risk*. i.e., the risk that the consumer will accept the lot, even though it has a more substan-

tial portion of defective items. Quite commonly, $\beta = 10\%$.

To define a sampling plan when the sampling is by attributes, three numbers must be fixed. These are the sample size h, the *acceptance number*, and the *rejection number*. The acceptance number, sometimes given the symbol c, is the number which must not be exceeded by the number of defective items in the sample if the lot is to be accepted. The rejection number is the number at or above which the number of defectives in the sample leads to rejection of the lot. It might seem that, by definition, the rejection number is $c + 1$, but this is not so, if the sampling plan is a *multiple* one, i.e., a plan in which a second or subsequent sample might be taken, depending on the results obtained for the first sample. In such cases, if the number of defectives in the first sample did not exceed c, the lot would be accepted; if the number of defectives exceeded the rejection number, the lot would be rejected; and if the number of defectives fell between these two extremes, one or more further samples might be taken. It is easy to see that multiple sampling plans might be highly discriminating, but the price paid is the much larger number of measurements needed, and such plans are not very widely used.

In the case of sampling by variables, the criteria needed to define the OC curve are inevitably different, and can best be illustrated by a simple example. Suppose that a manufactured fluorescent dye is of use to a laboratory analyst only if it contains at least 99% of the active fluorophore. Any lot containing $\geq 99\%$ is acceptable, but the analyst is reluctant to buy the dye if the fluorophore content falls below 96%. It is necessary to quantify these requirements: let us say that the manufacturer requires that a dye lot containing 99% fluorophore must have a 95% probability of acceptance, and the analyst requires that the probability of accepting a lot containing only 96% fluorophore is only 1%. It is also postulated that the sample size is h, and that the standard deviation σ for the fluorophore analysis, which reflects both sample inhomogeneities and measurement variation, is known exactly. (This is the case because of long experience of the production and testing of the dye.) It is necessary to estimate h and it is also stipulated that a dye lot will be accepted whenever the average fluorophore content over these h measurements equals or exceeds a value k, which is also to be estimated.

Suppose first that the true fluorophore content of a lot is 99%. The sample average $\bar{x}$ is then normally distributed, with mean 99 and standard deviation σ/h. The probability of acceptance of the lot is the probability that $\bar{x} > k$; this is equal to the probability that the standard normal deviate:

$$z = (\bar{x} - 99)/\left(\sigma/\sqrt{h}\right) > (k - 99)/\left(\sigma/\sqrt{h}\right)$$

This probability has been defined as 0.95, so from the standard statistical tables for z, the result is:

$$(k - 99)/\left(\sigma/\sqrt{h}\right) = -1.64$$

In the alternative situation, suppose that the true fluorophore content of the dye lot is only 96%. By exactly analogous arguments:

$$(k - 96)/\left(\sigma/\sqrt{h}\right) = +2.33$$

the latter number again being obtained from the tables for z at a probability of 0.99. Combining these results:

$$(k - 99)/(k - 96) = -(1.64/2.33) = -0.704$$

from which $k = 97.76$. If it is assumed that $\sigma = 2.0$, h can be obtained from either of the two above relationships. Thus:

$$(97.76 - 99)/\left(2.00/\sqrt{h}\right) = -1.64$$

from which h is 7, almost exactly. The sampling plan thus consists of taking seven measurements, and rejecting a lot whenever the average fluorophore content does not reach 97.76%. A noteworthy feature of these calculations is that σ does not have to be known for k to be determined, so, as in other forms of sampling, a reasonable plan can be obtained, even when σ is initially unknown. It is possible to make a conservative estimate of σ, use it along with k to estimate h, and refine the estimate with increasing experience of the variance.

Sampling plans of this kind generate OC curves of the same form as those described above, although in cases of sampling by variables the x-axis is normally plotted as the mean determinant concentration (the fluorophore level in the example above). Other aspects of OC curves, of multiple sampling plans, and of *sequential* sampling plans, in which samples are taken one at a time, are dealt with in more advanced texts [8] – [10].

5.5. Conclusions

This brief survey of sampling and its application in various circumstances should indicate that the subject is of critical importance to the performance of the overall analytical process, whether the latter is laboratory based or process based. It is also a very good example of the principle that methods based on statistics and probability can, and should, be applied in the planning stage of an analysis — to use such methods only when data have been obtained is almost invariably too late!

5.6. References

[1] B. Kratochvil, J. K. Taylor, *Anal. Chem.* **53** (1981) 924A–938A.
[2] D. T. E. Hunt, A. L. Wilson: *The Chemical Analysis of Water,* 2nd ed., Royal Society of Chemistry, London 1986.
[3] ASTM Standard E 105-58,*Standard Recommended Practice for Probability Sampling of Materials, ASTM, Philadelphia.*
[4] ASTM Standard E 122-72,*Standard Recommended Practice for Choice of Sample Size to Estimate the Average quality of a Lot or Process, ASTM, Philadelphia.*
[5] C. O. Ingamells, *Talanta* **21** (1974) 141–146.
[6] V. Barnett: *Elements of Sampling Theory,* Hodder and Stoughton, London 1974.
[7] P. Minkkinnen, *Anal. Chim. Acta* **196** (1987) 237–245.
[8] D. C. Montgomery: *Introduction to Statistical Quality Control,* Wiley, New York 1985.
[9] E. L. Crow, F. A. Davis, M. W. Maxfield: *Statistics Manual,* Dover, New York 1960.
[10] D. L. Massart, B. G. M. Vandeginste, L. M. Buydens, S. De Jong, P. J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometries & Dualmetries,* **Part 1,** Elsevier, Amsterdam 1997, Chapter 20.

6. Sample Preparation for Trace Analysis

JUTTA BEGEROW, Medizinisches Institut für Umwelthygiene, Düsseldorf, Germany

LOTHAR DUNEMANN, Hygiene-Institut des Ruhrgebiets, Gelsenkirchen. Germany

6.1. Introduction

Modern analysis begins with a definition and outline of the problem and ends only after a detailed critical evaluation of the relevant analytical data is complete, permitting the presentation of a "result" ($\rightarrow$ Analytical Chemistry, Purpose and Procedures). The analyst must therefore retain the ability to monitor a sample conscientiously and knowledgably throughout the analytical process. Only the analyst is in a position to assess the quality of a set of results and the validity of subsequent conclusions, although defining the problem and presenting the conclusions is almost always a cooperative multidisciplinary effort.

Analysis, and particularly trace analysis, thus entails more than the mere qualitative or quantitative detection of a particular element or chemical compound. For example, it presupposes knowledge on the part of the analyst with respect to the origin and structure of the sample matrix. The analyst must also possess a specialist's insight into analogous problems from other disciplines in order to assure the plausibility of the questions raised and critically evaluate and interpret the results, at least in a provisional way.

From these preliminary observations it will already have become clear that trace analysis, the focus of much concern in sample preparation, cannot be regarded as an end in itself. Rather, it is a very relevant and applications-oriented branch of analytical chemistry generally, one that, from a historical perspective, developed from within but became independent of chemical analysis as a whole, which was long regarded simply as a servant to the traditional subspecialities of chemistry and other disciplines ($\rightarrow$ Analytical Chemistry, Purpose and Procedures). Nevertheless, the same tools, equipment, and methodological principles remain common to both general chemical analysis and modern trace analysis.

6.1.1. A Strategy Appropriate to Trace Analysis

Trace analytical efforts directed toward the determination of particular elements or compounds require the analyst steadfastly to pursue a single six-step strategy consisting of

1) Rigorous definition of the problem
2) Problem-oriented selection of a sample
3) Appropriate sample preparation

4) Quantitative determination of the analyte(s)
5) Plausibility tests with respect to the results, together with further evaluation
6) Professional interpretation and presentation of the findings

This strategy must be regarded as an indissoluble whole, with all the constituent parts remaining insofar as possible and practicable in the hands of a responsible trace analyst. Accordingly, it is absolutely essential that the analyst begins with a clear understanding of the problem — if necessary even formulating it. Only after the problem has been defined is it possible to establish the most appropriate approach to sampling and sample preparation. Sampling and sample preparation are followed by the determination itself in the strictest sense of the term, the outcome of which — an analytical result — must be checked for accuracy and plausibility before the findings can be presented. In many cases, the findings will require interpretation developed with the help of specialists from other disciplines, who in turn must familiarize themselves with the analyst's procedures.

Sample preparation plays a central role in the process, but it too often leads a "wall-flower" existence, with primary attention being directed to the determination step. This sense of priorities is reflected all too conspicuously in the equipment and investment planning of many analytical laboratories. However, a welcome trend in recent years points toward fuller recognition of the true importance of sample preparation in the quest for high-quality analytical results and valid conclusions [1], [2].

6.1.2. Avoidance of Systematic Errors

Trace analysis can satisfy the stringent requirements outlined above only by recognizing and minimizing possible sources of error. Unfortunately, the trace analyst almost always suffers from a lack of knowledge of the correct (or "true") analytical value. Nevertheless, it is precisely this fact that is the source of the analyst's peculiar responsibility with respect to the results in question. A purely theoretical examination of the dilemma provides little guidance for routine work, since it leads simply to a recognition that the "true" value represents an unachievable ideal.

At this point only a general consideration is feasible regarding possible sources of systematic

errors that might influence the correctness of an analysis. Every sample preparation technique is associated with its own particular advantages as well as specific sources of error. In practice, the complete exclusion of systematic errors is possible only in rare cases, because optimization at one stage tends to cause problems elsewhere.

Generally speaking, any contact with vessel materials, reagents, or the ambient atmosphere, as well as any change in chemical or physical state, might result in systematic errors. It is therefore crucial that sample preparation be custom-tailored to the problem at hand with particular regard for the nature of the determination step.

Systematic errors arising from contact between a sample and various materials and chemicals are discussed in the context of the analyte (Sections 6.2.1 and 6.3.1), because different analytes require different types of treatment.

6.1.2.1. Trace Losses and Contamination

Trace *losses* are particularly common in the case of highly volatile analytes and as a consequence of adsorption effects, whereas trace *contamination* with otherwise prevalent elements and compounds may arise from laboratory air, vessels, chemicals, and various desorption effects.

The loss of trace analytes can be minimized by conducting all operations in a system that is almost completely sealed hermetically from the ambient atmosphere, and by using vessel materials characterized by a small effective surface area [3]. Surfaces should be free of fissures and preconditioned as necessary to minimize physical adsorption of the analyte (e.g., through ion exchange or hydrophobic interactions). It is also important to consider the possibility of diffusion into or through the vessel walls.

Addition of a radioactive tracer may make it possible to monitor losses that occur during the course of an analysis [4] (→ Radionuclides in Analytical Chemistry). Isotope dilution analysis is also useful for establishing the correctness of a result [5], [6].

Attention must be directed toward contamination in the early stages of laboratory planning and design to ensure that areas and rooms of equivalent sensitivity are situated adjacent to each other and shielded from less sensitive areas by appropriate structural and ventilation provisions. Analyses involving concentrations below the mg/kg or mg/L range, and especially in the ng or pg range, should of course be conducted in a clean-room environment. Precise requirements with respect to ventilation vary from case to case.

6.1.2.2. Uncertainty

The uncertainty on the result arises from both random and systematic effects but in trace analysis systematic effects largely determine the uncertainty of an analytical result. The search for and correction of systematic errors is therefore an important responsibility of every trace analyst. Even after correction for systematic errors the uncertainties on there corrections need to be evaluated and included in the overall uncertainty. Failure to correct for systematic errors leads to the considerable scattering frequently observed with collaborative analyses, and ultimately to inaccurate results. The uncertainty on the result increases disproportionately with decreasing amounts of analyte in the sample.

In trace analysis it is not sufficient simply to report a level of reproducibility for the actual determination of an analyte. Evaluating the quality of an analysis requires a knowledge of the reproducibility and the uncertainty arising from systematic effects (→ Chemometrics). Errors in sampling and/or sample preparation may be orders of magnitude greater than the standard deviation observed in several repetitions of a determination.

Whenever possible, reference materials should be included in the experimental plan as a way of checking for bias [7]. However, it is also important that the chemical state and environment of the analyte be as nearly identical as possible in both reference material and sample. The method of *standard addition* (where the spike should again be in the same chemical state as the native analyte) is particularly useful in a search for systematic errors, as is a comparison involving different analytical methods.

Internal and external quality control measures are a "must" in all trace analytical procedures, and an analytical result can only be accepted if all necessary actions in this respect have been carefully considered and the uncertainty on the result evaluated (Chapter 6.1). Quality control may therefore not only be a part of the determination step itself, but must also include the full analytical procedure including sample preparation. Internal quality control means the intra-laboratory procedure for quality control purposes including the control of uncertainty, whereas external quality control means the inter-laboratory comparison of the results of analyses performed on the same

specimen by several independent laboratories (according to the definition of [8]).

6.2. Sample Preparation and Digestion in Inorganic Analysis

The object of sample preparation in inorganic analysis is to meet the requirements for a substantially trouble-free determination of the analyte. The most important of these with respect to trace analysis in any matrix include:

1) Conversion of the sample into a form consistent with the determination (dissolution)
2) Destruction of the matrix (digestion)
3) Isolation of the analyte from interfering substances that may be present (separation)
4) Enrichment of the sample with respect to trace analytes (concentration)

The dissolution step is designed to compensate for inhomogeneities in the sample. Dissolution and digestion also simplify the subsequent calibration step, ensuring that both the sample and the calibration solutions are in essentially the same chemical and physical state. The extent to which matrix constituents interfere in the determination process is significantly reduced by digestion, leading to a lower limit of detection for the determination. Digestion also facilitates concentration and separation steps.

6.2.1. Sample Treatment after the Sampling Process

6.2.1.1. Stabilization, Drying, and Storage

In trace analysis, sampling must always be followed by an appropriate stabilization step, with due regard for the nature of the matrix and the analyte.

Aqueous samples such as drinking water, surface water, and waste water but also beverages and urine samples, should always be acidified with mineral acids for stabilization purposes immediately after collection. This is especially true for the prevention of desorption processes during sampling and storage of samples in the course of trace metal analysis. Acidification reduces the tendency for ions to be adsorbed onto active sites at the surface of the containment vessel, and it also inhibits bacterial growth [9]). Glacial acetic acid and 65 % nitric acid have been shown to be suitable for this purpose. Approximately 1 mL of acid should be added per 100 mL of sample. Acidified urine in appropriate containers can be safely kept in a deep-frozen state for several months [10], [11]. Stored urine samples often deposit a sediment that may include considerable amounts of analyte (e.g., arsenic, antimony, copper, chromium, mercury, selenium, zinc) [10].

Loss of water from aqueous matrices (e.g., tissue, fruit, vegetables, and soil samples) may occur during storage. For this reason analytical results should always be reported in terms of dry mass to avoid false interpretations. Drying is best conducted immediately after sampling. Water removal can be accomplished by oven drying at elevated temperature, use of desiccating materials, or freeze drying (lyophilization). Freeze drying has been shown to be the most satisfactory procedure since it minimizes the loss of highly volatile elements and compounds. Drying at a temperature as low as 120 °C can result in the loss of up to 10 % of most elements, and losses for mercury, lead, and selenium may be considerably higher (e.g., 20–65 %) with certain matrices [12], [13].

Blood samples must be rendered incoagulable by the addition of an anticoagulant immediately after sampling. Suitable substances for this purpose include salts of ethylenediaminetetraacetic acid (EDTA), citric acid, oxalic acid, and heparin. A particular anticoagulant should be selected with reference to the analyte in question as well as the proposed analytical procedure. Particular attention should be directed to ensuring that the anticoagulant is not itself contaminated with the analyte. For recommendations concerning appropriate levels of anticoagulants see, for example, [14]. Blood sampling kits that already contain the aforementioned anticoagulants are now available commercially. Blood samples in suitable containers can be either refrigerated or deep-frozen prior to analysis. Such samples can normally be kept for several months at − 18 °C without measurable change in the concentration of an analyte. One exception is mercury; in this case reduction to elemental mercury leads to significant volatilization losses within as little as a few days [15]. If storage over several years is envisaged, the sample should either be dried at − 18 °C or quick-frozen in liquid nitrogen at − 196 °C and then stored at or below − 70 °C.

Determination of an analyte in serum or plasma requires that the analyte be isolated subsequent to sampling. This is no longer possible

after deep-freezing because of the hemolytic nature of blood samples. In the case of elements at very low concentration it is preferable that an investigation be conducted on serum rather than plasma. This reduces the risk of contamination, because unlike blood plasma, serum is recovered without addition of an anticoagulant.

6.2.1.2. Homogenization and Aliquoting

In solid samples, elements are normally distributed in an inhomogeneous way. Trace-element determinations are usually restricted to relatively small samples, which requires that a fairly large sample be comminuted and homogenized prior to removal of an aliquot for analysis.

Comminution of the sample presents significant opportunity for contamination. Contamination of the sample from abrasion of the comminution equipment is fundamentally unavoidable, so efforts must be made to select the best possible equipment for each particular analytical task. Equipment is preferred in which the sample comes into contact only with surfaces fabricated from such high-purity plastics as polytetrafluoroethylene (PTFE), since this permits the sample to be used without restriction for the determination of a large number of elements. Friability can be increased by deep-freezing or drying the sample prior to comminution.

6.2.1.3. Requirements with Respect to Materials and Chemicals

Whenever samples, standard solutions, or reagents come into contact with containers, apparatus, or other materials, two opposing physicochemical processes occur at the interface: *adsorption* and *desorption* of ions and molecules.

Ion adsorption at active sites on a contact surface may lead to losses on the order of several nanograms per square centimeter. The importance of this factor therefore increases as the concentration of the analyte decreases. Adsorptive losses are a function of the nature and valence state of the analyte, the nature and concentration of accompanying ions in the solution, pH, temperature, and the duration of contact. Factors related to the containment vessel include the material from which it is constructed, its surface area, and the nature of any pretreatment.

Adsorption can be minimized through the use of quartz, PTFE, polypropylene (PP), perfluoroal-

koxy resin (PFA), or glassy carbon vessels. There is no single ideal material, but quartz is preferred in the case of biological samples provided the use of hydrofluoric acid can be avoided. The surface area of the container and the contact time should be minimized, whereas the analyte concentration should be as high as possible. Cations are adsorbed less strongly from acidic solutions than from solutions that are neutral or basic because such ions are displaced from active sites on the container surface by the much more abundant protons . Adsorptive losses are also lower from solutions with high salt concentrations (e.g., urine) relative to those with low salt concentrations, such as standard solutions. Adsorption can be further reduced by adding a suitable complexing agent so long as this has no adverse affect on the subsequent analysis. For example, mercury can be stabilized as HgI_4^{2-}, and silver as $Ag(CN)_2^-$.

Containers, pipettes, and other laboratory implements contain traces of various elements both in their structure and adsorbed on the surface, and these are subject to desorption on contact with a sample, calibration solutions, or reagents, thereby contributing to a nonreproducible increase in blank results. Table 1 summarizes the inorganic impurities likely to be encountered with various vessel materials.

Careful and thorough cleaning of all vessels is extremely important in trace analysis both for inorganic and organic analytes. Materials must be specially selected with a specific view to the requirements of each task during sample handling and storage, with explicit attention being paid to trace purity. Thorough pretreatment (rinsing, boiling, steaming) with pure nitric, sulfuric, or hydrochloric acid in the case of metal analysis is strongly recommended [18]–[21]. Figure 1 illustrates a steaming apparatus useful for cleaning analytical vessels which are to be prepared for use in inorganic analyses, namely metals but also other inorganic analytes [22]. For organic analyses repeated cleaning of vessel surfaces with suitable solvents is essential.

Reagents for trace analytical applications must also satisfy stringent purity requirements, often exceeding those of commercially available reagents. Acids such as hydrochloric, nitric, sulfuric, hydrofluoric, and perchloric can be prepared relatively easily in the laboratory to a high standard of purity by distillation below the boiling point ("subboiling process"). An apparatus for this purpose is illustrated in Figure 2. Table 2 compares the elemental contents of commercially available

Table 1. Typical inorganic impurities in selected vessel materials [16], [17]

Element	Content, mg/kg				
	Quartz	GC [a]	PTFE [b]	PE [c]	BS [d]
V		4			< 2
Cr		0.5 – 5	10 – 30	15 – 300	< 3
Mn	0.01	0.1	60	10	< 6
Fe	0.2 – 0.8	1 – 15	10 – 35	600 – 2100	200
Co	< 0.001	< 0.01	0.3 – 1.7	0.1 – 5	< 0.1
Ni		0.5			< 2
Cu	< 0.01 – 0.07	0.2	20	4 – 15	1
Zn	< 0.01	< 0.3	8 – 10	25 – 90	2 – 4
As	0.0001 – 0.08	0.06			0.5 – 22
Cd		< 0.01			1
Sn		25 – 50			< 4
Sb	< 0.001	< 0.01	0.4	0.2 – 5	7 – 9
Hg	0.001	0.001			
Pb		0.4			

[a] Glassy carbon.
[b] Polytetrafluoroethylene.
[c] Polyethylene.
[d] Borosilicate glass.

Table 2. Elemental content of selected acids with varying degrees of purity [16]

Acid	Elemental content, µg/L				
	Cd	Pb	Cu	Fe	Zn
10 N HCl, analytically pure	0.1	0.5	1.0	100	8.0
10 N HCl, extra pure	0.03	0.13	0.2	11.0	0.3
2 N HCl, subboiled	0.01	< 0.05	0.07	0.6	0.2
15 N HNO₃, analytically pure	0.1	0.5	2.0	25.0	3.0
15 N, HNO₃, extra pure	0.06	0.7	3.0	14.0	5.0
15 N, HNO₃, subboiled	0.001	< 0.02	0.25	0.2	0.04

acids with those of acids prepared by the subboiling process.

All chemicals and reagents should be checked regularly for purity, and replaced as necessary. It is also important to remember the ever-present danger of contamination during reagent use as a result of interactions with container materials and the entrainment of impurities. For this reason reagents and standard solutions should always be prepared immediately before their use. As a general rule, the addition of reagents should be kept strictly to a minimum.

Possible contamination due to atmospheric dust and particle emission attributable to laboratory staff activities (smoking, use of cosmetics, etc.) can be avoided in trace analytical determinations at the µg/kg, µg/L, and µg/m³ ranges by working under clean-room conditions. A laminar-flow "clean bench" ensures that particles with a diameter > 0.1 µm will be removed with an efficiency of 99.9 % through high-efficiency, sub-micron-particulate air filtration. As indicated in Table 3, this can result in purification factors on the order of 100 with respect to such elements as manganese and vanadium [23].

6.2.2. Sample-Preparation Techniques; General Considerations

Sample-preparation techniques are classified, somewhat arbitrarily, according to external conditions. Categories include wet digestion, dry ashing, pressure digestion, microwave digestion, and cold-plasma ashing. However, these conventional designations are often imprecise or even misleading with respect to the actual mechanism of the process. Several very different names are sometimes applied to a single technique, which presents a considerable obstacle for anyone (particularly a nonspecialist) interested in acquiring a quick overview of systems applicable to a specific task. For
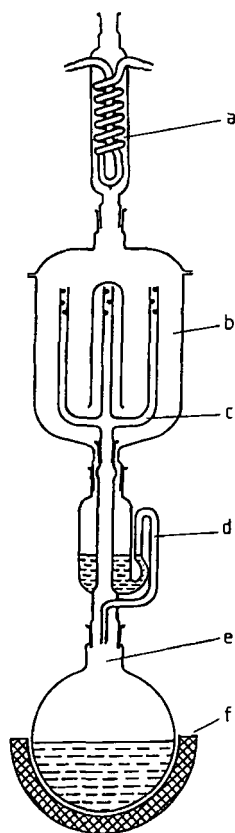
Figure 1. Steaming apparatus for cleaning analytical vessels with nitric acid, hydrochloric acid, or water (with permission from Kürner, Rosenheim/Oberbayern, Germany)
a) Reflux condenser; b) Steam chamber; c) Rack of quartz tubes; d) Overflow; e) Round-bottom flask (1 – 2 L capacity); f) Heating mantle

Figure 2. Quartz apparatus for purifying acids or water by subboiling distillation (with permission from Kürner, Rosenheim/Oberbayern, Germany); this device is also available in PFA for the subboiling distillation of hydrofluoric acid (with permission from Berghof Maassen, Eningen u. A., Germany)
a) Heater; b) Cold-finger condenser

Table 3. Typical levels of contamination by selected elements as a result of dust in the laboratory [23]

Element	Contamination level, ng cm^{-2} d^{-1}	
	Conventional laboratory	Clean workstation
Al	0.82	0.09
V	0.037	0.0003
Mn	0.063	0.002
Fe	1.5	0.2
Co	0.0036	0.0003
Cu	0.032	0.012
As	0.07	0.03
Sb	0.013	0.001

example, the expression "pressure digestion" is inappropriate, because it is a relatively high boiling temperature that ensures more effective digestion, not the associated high pressure. The conventional designation should nevertheless be retained in this case, if only because of its wide acceptance; attempting to rename the procedure now would introduce more confusion than clarity.

Another — very timely — example of incorrect terminology involves uncritical use of the expression "microwave digestion" for both acid digestion with microwave excitation and cold-plasma ashing. Although both techniques make use of microwave radiation, the direct effects of this radiation are of minor importance at most. Expressions like "microwave excitation," "microwave induction," or "microwave assistance" would be preferable.

6.2.2.1. Special Factors Associated with Microwave-Assisted Digestion

Rapid, straightforward digestion processes have long been in demand in a wide variety of disciplines. The disparity between the several hours required for a conventional digestion process and the few seconds or minutes involved in elemental determination has become increasingly obvious in recent years with the increasing automation of analytical procedures. Recently introduced rapid digestion processes, which have profited greatly from modern microwave technology, can contribute to reducing this disparity.

The first published report on microwave ashing was that of KOIRTYOHANN et al., which appeared in 1975 [24]. Initially it attracted little attention because of the lack of appropriate expertise and apparatus.

Modern commercial microwave ovens operate at a frequency of 2.45 GHz. The corresponding energy is so low that such microwaves are incapable of rupturing molecular bonds directly, leading only to rotational excitation of dipoles and molecular motion associated with the migration of ions. No vibrational or electronic excitation occurs [25], [26]. It is for this reason that the designation "microwave-assisted digestion" has been suggested. Microwave energy introduced as radiation is dissipated (converted into heat) by ionic conduction and dipole rotation. Oscillating alignment and subsequent relaxation of dipoles occurs about 5×10^9 times per second at a frequency of 2.45 GHz, causing rapid heating of the sample.

Microwave radiation is generated by means of a magnetron (microwave diode). The radiation is then directed through a waveguide into the interior of the oven. Even distribution of the radiation is facilitated by a "mode stirrer" that prevents the development of standing waves, as well as a turntable and perhaps a rotating antenna located beneath the oven floor.

Starting in the late 1980s, microwave-based techniques became more and more available for analytical purposes [27]–[30]. Today in trace metal and non-metal analyses, microwave-assisted digestion has been successfully applied to the decomposition of broad range of organic matrix components in geological and biological samples as well as analysis and residue monitoring [22].

6.2.2.2. Safety Considerations

Necessary protective measures with respect to the laboratory workforce must be implemented for the handling of acids and other strongly oxidizing or poisonous chemicals. This includes protective clothing and an efficient fume extraction system. Perchloric acid as a digestion reagent offers advantages due to its powerful oxidizing action, but it should be used only when absolutely necessary since it entails extensive safety precautions (e.g., working in a special hood), and a definitive set of regulations must be scrupulously observed.

Special statutory regulations apply to operations involving pressure vessels, and pressure digestion techniques must be carried out with exceptional care and attention. Ashing an organic matrix can cause the pressure to rise to extremely high levels within a short period of time, and the reaction may become uncontrolled. The corresponding risk of a ruptured pressure vessel cannot be overemphasized. When developing a pressure digestion method it is important to begin with a small sample ($\leq$ 100 mg dry mass; even less for pure carbon).

Additional safety considerations apply to the use of microwave technology. First, the equipment should be operated in strict accordance with instructions to ensure that no detectable amounts of microwave radiation will escape. Leak tests should be conducted on a regular basis by a trained technician. Vessels or other objects made of metal should never be placed in the oven chamber. Microwave equipment must also never be operated empty (i.e., in the absence of a sample solution, water, acid, etc.), since the magnetron might be destroyed by reflected radiation. The behavior of unknown samples should be tested on a small scale (see above).

6.2.3. Wet Digestion Techniques

Wet digestion with oxidizing acids is the most common sample-preparation procedure [31], [32]. This category can be extended to include processes involving bases or nonoxidizing acids as ashing reagents. Concentrated acids with the requisite high degree of purity are available commercially, but they can be purified further by subboiling distillation (see Section 6.2.1.3).

The most suitable acid for a digestion is a function of the sample matrix, the analyte, and the proposed determination method. Nitric acid is an almost universal digestion reagent, since it does not interfere with most determinations and is available commercially in sufficient purity. However, nitric acid has a rather low boiling point (122 °C), and its oxidizing power is often insufficient under atmospheric-pressure conditions. Hydrogen peroxide and hydrochloric acid can usefully be employed in conjunction with nitric acid as a way of improving the quality of a digestion. Hydrochloric acid and sulfuric acid may interfere with the determination of certain metals through the formation of stable compounds. As noted previously (Section 6.2.2.2), safety considerations are particularly important when using perchloric acid. Silicate samples require the further addition of hydrofluoric acid.

6.2.3.1. Wet Digestion at Atmospheric Pressure

This category includes all the various ambient-pressure wet digestion techniques involving concentrated acids (or bases). This very inexpensive technique is of inestimable value for routine analysis because it can easily be automated; all the relevant parameters (time, temperature, introduction of digestion reagents) lend themselves to straightforward control [33].

Systems of this type are limited by a low maximum digestion temperature, which cannot exceed the ambient-pressure boiling point of the corresponding acid or acid mixture. As noted previously, the oxidizing power of nitric acid with respect to many matrices is insufficient at such low temperatures. One possible remedy is the addition of sulfuric acid, which significantly increases the temperature of a digestion solution. Whether or not this expedient is practical depends on the matrix and the determination method. High-fat and high-protein samples are generally not subject to complete ashing at atmospheric pressure. Other disadvantages relate to the risk of contamination through laboratory air, the necessarily rather large amounts of required reagents, and the danger of trace losses. Nevertheless, systems operated at atmospheric pressure are preferred from the standpoint of workplace safety.

This category also includes the *solubilization* of simple matrices with low carbon content using special detergents or other solubilization reagents, as well as the enzymatic digestion techniques that are often used in food control and monitoring. Enzymatic digestion in trace analytical chemistry has been described in [34] for the example of the non-specific protease enzyme (pronase), which was used as an alternative sample preparation technique for the determination of trace elements in blood serum by ICP-MS. Using enzymatic digestion, a higher degree of instrument stability was achieved over 3 h period. The reported results are in good agreement with the certified values of the reference material [34]. Also "extraction–digestion" techniques belong to this category, e.g., using aqua regia at relatively low temperatures (e.g., 70 °C) over a long period of time (up to 5 d) as tool for the pseudo-total analysis of metals in soil samples [35]. These techniques are some importance as sample preparation techniques, but one should be aware that they do not lead to complete digestion of the organic matrix. Extraction–digestion techniques have become standard methods for the examination of water, waste water, sludge, and sediments using aqua regia for the subsequent determination of the acid-soluble portion of metals [36].

Thermally Convective Wet Digestion. The conventional approach to wet digestion, which has proven its worth over many years, entails a system equipped with heating units operating either at a fixed temperature or in response to a temperature program. Commonly employed digestion agents include nitric acid, sulfuric acid, hydrogen peroxide, hydrofluoric acid, and perchloric acid, as well as various combinations of these [33]. Most applications of wet digestion involve aqueous or organic matrices, such as surface waters, waste water, biological samples (tissue, body fluids, etc.), food samples, as well as soil and sewage sludge, coal, high-purity materials, and various technical materials [35].

Microwave-Assisted Wet Digestion. Microwave-assisted digestion in open vessels (atmospheric digestion) [37] is generally applicable only with simple matrices or strictly defined objectives, and the results are reproducible only if the specified ashing parameters are strictly observed. Losses may be encountered with mercury and possibly also with organometallic compounds (e.g., those containing arsenic, antimony, or tin). Addition of sulfuric acid is essential in order to achieve a sufficiently high digestion temperature in atmospheric-pressure equipment, where the boiling point of the acid establishes the maximum ashing temperature, although it is important to remember that the presence of sulfate interferes with many procedures for metal determination (e.g., graphite-furnace atomic absorption spectrometry, GF-AAS). Specific experimental conditions for microwave-assisted digestions have been summarized in [9], [22] as a function of the matrix to be ashed.

Ultraviolet Digestion (Photolysis). Ultraviolet (UV) digestion is utilized mainly in conjunction with uncontaminated or slightly contaminated waters, such as sea water and surface water. Liquids or slurries of solids are decomposed by UV radiation in the presence of small amounts of hydrogen peroxide and acids (e.g., beverages, wastewater, soil extracts). The corresponding digestion vessel should be placed in the closest possible proximity to the UV lamp to ensure a high light yield. In photolysis (see Fig. 3) the digestion mechanism can be characterized by the formation
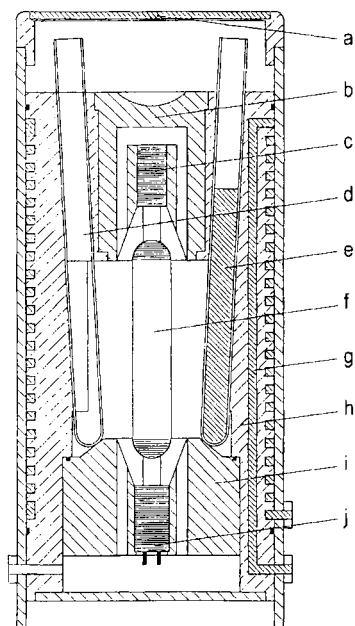
Figure 3. Apparatus for the UV digestion (photolysis) of liquid samples with low carbon content (with permission from Kürner, Rosenheim/Oberbayern, Germany)
a) UV protection glass; b) Lamp cooler; c) Lamp socket;
d) Sample vessel; e) Sample solution; f) UV lamp; g) Coolant (water); h) Sample cooler; i) Lamp cooler; j) Lamp socket

of OH radicals from both water and hydrogen peroxide that is initialized by the aid of the UV radiation [9], [38]. These reactive radicals are able to oxidize the organic matrix of simple matrices containing up to about 100 mg/L of carbon to carbon dioxide and water. Effective cooling of the sample is essential, since losses might otherwise be incurred with highly volatile elements. Contamination can be minimized by the use of a nongas-tight stopper. Hydrogen peroxide addition may need to be repeated several times to produce a clear sample solution [39], [40]. Complete elimination of the matrix is of course possible only with very simple matrices (e.g., slightly contaminated water) or by combining photolysis with other digestion techniques [41], [42].

6.2.3.2. Pressure Digestion

Pressure digestion offers the advantage that the operation is essentially isolated from the laboratory atmosphere, thereby minimizing contamination, and digestion occurs at relatively high tem-

perature due to boiling-point elevation effects. The pressure itself is in fact nothing more than an undesirable — but unavoidable — side effect. The principal argument in favor of this form of digestion is the vast amount of relevant experience acquired in recent decades. The literature is a treasure trove of practical information with respect to virtually every important matrix and a great number of elements (see, for example [43]). Pressure digestion is particularly suitable for trace and ultratrace analysis, especially when the supply of sample is limited.

Since the oxidizing power of a digestion reagent shows a marked dependence on temperature, a distinction must be made between "simple" pressure digestion and high-pressure digestion. Simple pressure digestions (< 20 bar) are limited to a temperature of ca. 180 °C, whereas with high-pressure apparatus (> 70 bar) the digestion temperature may exceed 300 °C.

6.2.3.2.1. Thermally Convective Pressure Digestion

Most sample containers for use in *thermally convective pressure digestion* are constructed from PTFE or PFA, although special quartz containers with PTFE holders are available for trace-analysis purposes. The sample container is mounted in a stainless-steel pressure vessel (autoclave) and then heated in a furnace to the desired temperature. Pressure digestions based on apparatus of the Tölg, Bernas, and Parr type are all feasible below ca. 180 °C, but above this temperature PTFE begins to "flow", rendering it unsuitable for use in high-pressure applications. Investigations with organic matrices have shown that at 180 °C as much as 10 % of the carbon present in high-fat and high-protein samples remains unashed [44]. Such samples must be processed further prior to analysis through evaporation of the digestion solution with perchloric acid.

All thermally initiated digestions have the disadvantage that a considerable amount of time is consumed in preheating the digestion solutions due to the need for heating the autoclave and sample vessels.

High-Pressure Digestion. The introduction of a high-pressure ashing (HPA) technique by Knapp [45] has not only reduced the effective ashing time to ca. two hours but also opened the way to digestion of extremely resistant materials [46]. High-
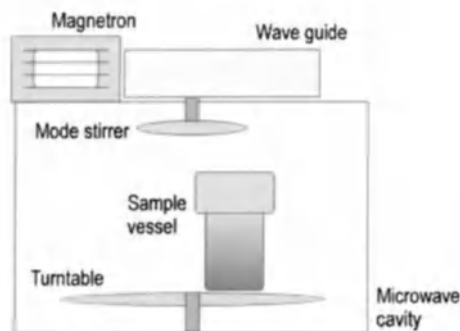
Figure 4. Apparatus for microwave-assisted pressure digestion

pressure ashing is conducted in quartz vessels, with a maximum ashing temperature as high as 320 °C at a pressure of ca. 100 bar. The quartz vessel is stabilized during the digestion process by subjecting it to an external pressure roughly equivalent to that developed within. Even thermally stable organic compounds undergo acid decomposition at such high temperatures. Essentially complete ashing can be accomplished with the vast majority of samples so far investigated. Nitric acid alone is a sufficiently powerful reagent in many cases.

6.2.3.2.2. Microwave-Assisted Pressure Digestion

The digestion of lightly contaminated water, many geological samples, and low-fat plant matrices can be accomplished successfully with a domestic microwave oven and relatively primitive sample holders. The recent introduction of complete commercial microwave systems specifically designed for sample ashing has led to important further developments. Such systems also offer special safety features and improved facilities for controlling and regulating the pressure and/or temperature [47]. Moreover, the distribution of microwave radiation inside the oven cavity tends to be significantly more homogeneous than with domestic devices, in part because the magnetron control is subject to shorter switching cycles.

Nitric acid alone does not always lead to satisfactory microwave-assisted ashing. Addition of hydrochloric acid or hydrogen peroxide is often helpful, and hydrofluoric acid must of course be present in the case of silicate-containing materials [39], [48]. The maximum attainable pressure seldom exceeds 10 bar.

A serious disadvantage of this simple approach is the frequent need for multistep ashing, in which a preliminary digestion is followed by a second treatment, perhaps with added hydrogen peroxide [40].

High-Pressure Microwave Digestion. High-pressure (or high-temperature) microwave-induced processes are designed to support digestion up to a pressure of perhaps 70 bar, opening the way to substantially higher temperatures (e.g., ca. 250 °C) without a corresponding increase in overall processing time. Temperature and digestion time ultimately determine the effectiveness of a digestion, with residual carbon content serving as a useful measure for quantitative assessment. The striking advantage of high-pressure, i.e., high-temperature (!) microwave digestion (see Fig. 4) is that almost complete dissolution can be accomplished with a single-stage process, even in the case of complex matrices. Examples found in the recent literature consist of new successful developments in the technique and apparatus [49] and applications to biological tissues [50] or technical products [51].

Examples of almost completely mineralizable matrices include resistant inorganic and high-fat organic samples. Nevertheless, limitations with respect to subsequent use of electrochemical methods of determination (Section 6.2.5.1) still represents a critical problem [52], and the results are often inferior to those from a conventional high-pressure asher (HPA).

6.2.4. "Dry" Digestion Techniques

The term "dry" digestion is intended to encompass all processes based on gaseous or solid digestion reagents. Such a distinction relative to wet digestion processes is not absolutely essential, but it does offer certain practical advantages [43].

6.2.4.1. Combustion in Air

Strictly speaking, "dry ashing" refers to the oxidation (combustion) of a substance in air at a temperature of several hundred degrees Celsius, often in a muffle furnace or similar apparatus.

Dry Ashing in a Muffle Furnace. Carbon-containing substances generally decompose satisfactorily under these conditions without the use of auxiliary agents. It is advantageous that reasona-
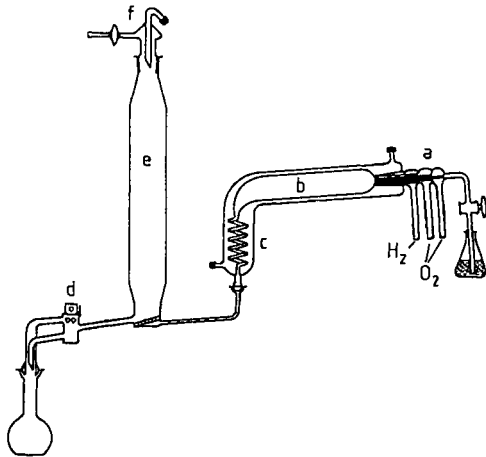
Figure 5. Wickbold combustion apparatus
a) Injection burner; b) Combustion chamber; c) Condenser; d) Multiport stopcock; e) Absorption chamber; f) Flushing port

bly large samples can be digested. In addition to routine biological samples (plants, foods), coal has also been subjected to dry ashing [53]. The importance of this technique has declined in recent years, primarily because it is so time-consuming; it also entails a high risk of contamination since the samples are kept in open vessels [54]. Losses must be anticipated in the case of volatile substances [55]. However, the literature provides numerous examples of decompositions that have been optimized through the addition of "ashing auxiliaries" (salt solutions or acids). Such additives often reduce the ashing time, especially for organic substances, and they may prevent the volatilization of element traces.

Dry Ashing in a Microwave Oven. The time required for dry ashing in air can be reduced by microwave heating of a special block containing the sample [25]. This block is fabricated from a material that is highly absorbent with respect to microwaves (e.g., silicon carbide), and is is surrounded by quartz insulation. At maximum power such a block can be heated to about 1000 °C within ca. 2 min. In addition to time saved during the heat-up phase this system offers the further advantages of a low energy requirement and a procedure that spares the technicians from exposure to intense heat when inserting and removing samples, in sharp contrast to work with a conventional muffle furnace [25].

6.2.4.2. Combustion in Oxygen

Combustion in oxygen is a decomposition procedure that eliminates many of the potential problems of sample contamination associated with digestion reagents [56]. The sole reagent here is gaseous oxygen, which is commercially available in very high purity. Related combustion techniques are based on the use of other oxidizing agents (e.g., sodium peroxide) from which free oxygen is released during the course of the operation. Analogous methods rely on oxidizing agents other than oxygen (e.g., chlorine [57], fluorine), but these have not been widely accepted, in part because they require working with very aggresive materials, often not available with the required level of purity and themselves capable of interfering with the determination process.

Atmospheric-Pressure Combustion. Although combustion in an open vessel is relatively easy to accomplish, it has the disadvantage of possible volatilization of element traces. Significant advantages are associated with the compromise of a quasi-closed quartz glass system consisting of a central combustion chamber, an ignition device (an infrared source), and a reflux condenser charged with liquid nitrogen. The analyte, together with inorganic combustion residues, is redissolved by refluxing with a small amount of acid. Such a system is applicable to almost all the elements, and it leads to nearly complete decomposition [56], [58]. Applications have been described with respect to a large number of organic matrices, including petroleum products, coal, and soils. Combustion systems also exist for analysis of nonmetals. SCHÖNIGER developed a particularly simple combustion process (now bearing his name) for general elemental analysis [59]. A special variant — combustion in an oxyhydrogen flame — is carried out in a Wickbold apparatus (Fig. 5) [60].

Combustion in a Closed Vessel. The risk of volatilization of trace analytes (e.g., mercury) can be reduced further by conducting the combustion in a special combustion bomb (e.g., a Parr bomb) [61]. Oxygen gas is introduced up to some predetermined pressure, after which the sample is subjected to explosive combustion with the aid of a detonation device (ignition wire). The analyte, together with inorganic residues, is again collected in a small volume of acid [62]. This technique is

also applicable to the determination of nonmetals [63].

6.2.4.3. Cold-Plasma Ashing

Decomposition by cold-plasma ashing is accomplished with excited-state oxygen under a pressure of a few millibars. The plasma is formed at reduced pressure (0.01 – 1 mbar) in a high-frequency field established either with a magnetron (2.45 GHz) or with a semiconductor high-frequency generator. In the simplest application samples are ashed in flat dishes (e.g., Petri dishes) arranged in trays [64]. A quasi-closed system with a cold-finger condenser should of course be employed for trace analysis to minimize potential losses [65]. Oxidation of organic matrices occurs by way of reactive, short-lived radicals, leading to more complete matrix decomposition relative to most other approaches. This ashing technique is therefore particularly recommended as a precursor to electrochemical determinations (e.g., polarography). The oxygen plasma is a "cold plasma," which means that the sample temperature never exceeds ca. 150 °C. Risk with respect to the loss of volatile elements is therefore restricted to mercury. A reaction period of ca. 2 – 4 h is usually sufficient to disrupt the organic matrix to such an extent that the residue can be removed from the sample vessel and cold finger by refluxing with a small volume of concentrated acid (2 mL). The most important application of cold-plasma ashing is the mineralization of biological samples [66], although coal, graphite, and plastics (including PTFE) can also be ashed by this technique [65]. Samples as large as one gram or more can be accommodated.

6.2.4.4. Fusion

Fusion is the traditional approach to sample preparation for industrial and geological analyses [67], especially for the analysis of mineral samples (e.g., lithium metaborate melts [68]). The major problems associated with fusion include restricted means for purifying the required (solid) reagents and the presence of high salt concentrations in the resulting analyte solution. Especially for trace analysis, overloading of the matrix in this way is inadvisable.

6.2.5. Illustrative Examples

The sections that follow highlight a few particularly interesting examples of sample preparation, taking into account both the analytical method of choice and effective mineralization of specific matrices.

6.2.5.1. Sample Preparation as a Function of Analytical Method

In the case of *atomic spectrometry* [e.g., atomic absorption (AA), inductively coupled plasma – optical emission spectrometry (ICP – OES), inductively coupled plasma – mass spectrometry (ICP – MS)] matrix interference due to incomplete digestion may manifest itself as differences in the suction rates and aerosol yields for samples relative to calibrating solutions. Such differences may also reflect differences in the bonding states of the elements, which in turn leads to systematic errors and problems with calibration.

Electrochemical determination methods are particularly sensitive to incomplete sample mineralization. The result is matrix interference manifested in the case of inverse voltammetry, for example, as ghost peaks, signal suppression, and an enhanced hydrogen signal. With simple, aqueous samples such as seawater and river water, voltammetric determination of zinc, cadmium, lead, and copper can be effectively coupled with UV decomposition [69], [70]. The same determination method is applicable to these heavy metals in simple plant products such as sugars provided decomposition (ashing) is effected by combustion in a stream of oxygen (typical sample mass: 1.5 g) [71]. Combustion is also appropriate prior to the potentiometric determination of aluminum in organic samples [72]. Biological samples can be prepared for electrochemical analysis by cold-plasma or high-pressure ashing [41], [45], [56], [73].

An improved procedure has been described for atmospheric-pressure combustion of samples from high-purity metals [74]. High-pressure digestion with nitric and hydrofluoric acids [75], [76] and combustion techniques have both been shown to be suitable for the decomposition of samples of silicate-containing materials and fuels. Combustion in this case is carried out either in a stream of oxygen [77] or, to provide a closed system, in an oxygen bomb [78]. Mercury can be satisfactorily determined after prior Wickbold ashing [79].

Table 4. A comparison of the time required for microwave pressure digestion versus thermally convective digestion (for one sample and for six samples)

Operational step	Time, min			
	Microwave digestion (PMD)		Knapp digestion (HPA)	
	One sample	Six samples	One sample	Six samples
Preparation for digestion	10	30	10	30
Ashing, including warm-up	10	30 *	90 – 150	90 – 150
Cool-down	10	30	20	20
Preparation for analysis	5 – 10	30	5 – 10	30
Cleaning the equipment	10 – 20	20	10 – 20	20
Complete procedure	45 – 60	140	135 – 210	190 – 250

* Limited to two parallel samples.

The decomposition of inorganic substances can be accomplished with a mixture of hydrochloric and nitric acids in a closed vessel (pressure digestion). Silicate-containing minerals and glasses require the addition of hydrofluoric acid, and the presence of perchloric acid, phosphoric acid, or sulfuric acid is sometimes useful as well [80]. The resulting metal fluorides can be redissolved by adding a solution of boric acid. The same procedure is effective for ores and slags as well as quartz [81].

Cold-plasma ashing is useful as a way of excluding matrix interferences in the detection of elements with secondary-ion mass spectrometry (SIMS) [82].

6.2.5.2. Combined Use of Multiple Decomposition Techniques

A single ashing procedure is often insufficient for the complete decomposition of a complex matrix, leading some authors to recommend a combination of two or more techniques. One example will suffice to illustrate the principle: pressure ashing followed by UV photolysis. Thus, it has been shown that analysis of olive leaves for heavy metals by voltametric methods leads to distorted results after pressure digestion alone. Reliable data can be obtained only by supple-menting the digestion with UV irradiation to ensure adequate decomposition of the matrix [41].

6.2.5.3. Comparative Merits of the Various Sample-Preparation Techniques

High-fat samples present special problems with respect to decomposition due to the presence of highly resistant components. The result is often incomplete ashing. Moreover, there is always a risk of explosive decomposition caused by the formation of reactive radicals in the digestion solution. Unlike electrochemical methods of determination, which require the complete degradation of organic matrices, atomic absorption (particularly the flame technique) is compatible with a certain amount of residual carbon. Any digestion solution that is colorless and free of particles is a reasonable candidate for AAS analysis.

A systematic comparison of microwave-assisted decomposition techniques with cold-plasma ashing, conventional thermally convective pressure digestion, and high-pressure digestion is presented in [40]. Generally speaking, comparable results can be obtained with high-quality commercial ashing systems of all the common types, with the potential for nearly complete decomposition [e.g., the Paar microwave-assisted pressurized microwave decomposition (PMD) system, the Büchi high-pressure adapter, and the Knapp HPA system]. Microwave ashing of high-fat samples in an open vessel presents a serious risk of significant analyte loss despite the use of a reflux condenser, but losses with a microwave pressure-digestion system have been found to be no greater than with a corresponding convective heating system. Results also suggest that "memory effects" are less noticeable with a microwave system than in the case of a conventional heat source [40].

Table 4 provides a comparison of the time required for various steps in microwave versus convective thermal ashing, taking into account all the associated preparations, the cool-down phase, and cleaning of the apparatus, which has the effect of reducing somewhat the marked time advantage of the microwave-assisted ashing period. The three-fold time advantage of the complete microwave-

Table 5. Overall ashing times associated with various procedures (for one sample)

Sample type	Ashing time, h				
	Pressure microwave *	Open microwave	Tölg **	Knapp **	Cold plasma
Plant matrix	0.75	0.92	4.5	3	10
Sunflower oil	0.75	0.92	4.5	3	20

* Single-stage digestion.
** Thermal convection heating.

assisted procedure is reduced even more if, for example, six samples are digested, because the PMD cannot process more than two samples at once. Table 5 shows that microwave-assisted pressure ashing is the most rapid of the techniques investigated, faster than the frequently used Tölg technique (Section 6.2.3.2.1) by a factor of six. Cold-plasma ashing takes even longer. It should be noted, however, that this comparison ignores the advantages that might be derived from parallel ashing of several samples, and it does not reflect special requirements associated with particular matrices or analytes.

Earlier investigations confirm these results. According to [41], however, complete degradation of many samples is achieved only through cold-plasma ashing or by a combination of wet ashing and UV irradiation. In a determination of aluminum in dialysis liquids it was observed that microwave-assisted ashing produced results comparable to those from conventional pressure ashing [52].

Five different digestion techniques are compared in [83] in an attempt to optimize the determination of mercury in soil via cold-vapor AAS. Only with closed systems was loss-free digestion assured. Open systems resulted either in mercury losses or incomplete recovery due to incomplete ashing. On the other hand, wet ashing in an open system with nitric and sulfuric acids was found to be preferable for the mineralization of biological samples, since it was easier to accomplish and led to lower blank readings [84].

Fusion with lithium metaborate is feasible for geological samples provided the melt is subsequently dissolved in acid in the course of a pressure- or microwave-assisted ashing process [85].

6.2.5.4. Decomposition Procedures for Determining Nonmetals

A pertinent example of an analysis for nonmetallic constituents is the digestion of water or food samples with peroxodisulfate, according to the

Koroleff method [86], which was developed for the determination of all bound nitrogen in seawater, inland waters, rain water, groundwater, wastewater, or effluents. Measurable concentrations are in the range 0.02 – 4 mg of nitrogen per liter. The nitrogen-containing sample is first subjected to a pressure digestion in the course of which organic nitrogen compounds, nitrite ion, and ammonium ion are all oxidized to nitrate. This oxidation proceeds quantitatively only under alkaline conditions, achieved with the aid of a buffer system consisting of boric acid and sodium hydroxide. Reaction commences at pH 9.7 and ends at pH 5 – 6, at which point the nitrate exists in stable form in the acidic digestion solution. Nitrogen is not released in this way from certain five-membered heterocyclic systems [87], but the resulting error is probably negligible with most of the samples subject to investigation. The total nitrogen content in the digestion solutions can be established by photometric or ion-chromatographic methods. A Kjeldahl analysis can be used as a standard for comparison, using either the conventional technique or microwave-assisted heating. The accuracy of the results has been verified on the basis of standard reference material (IAEA A-11, full-cream milkpowder, in the case of foods) and by interlaboratory comparison (river water, wastewater) [88].

The fluoride content of silicates can be established after fusion with lithium metaborate [89], where fluoride is subsequently separated with the aid of superheated steam [90].

6.2.6. Evaluation Criteria

The maxim "one method is no method" is just as applicable to sample preparation as it is to quantitative determinations. A careful comparison of several decomposition techniques is the only way of assuring accurate results, particularly when little experience is available with respect to the decomposition of a specific matrix, or existing reports are contradictory.

6.2.6.1. Completeness

Decomposition in the strictest sense has as its goal a complete elimination of the matrix. Nevertheless, many decomposition procedures fall short of this ideal, permitting traces of the matrix to remain. Residual carbon content can be established with a total organic carbon (TOC) analyzer. This decreases with increasing oxidizing power of the decomposition reagent as well as with an increase in the duration and temperature of the digestion [43].

6.2.6.2. Uncertainty

As with all aspects of trace analysis, strict attention to possible sources of error is important in the successful application of a decomposition technique. The best determination is rendered completely worthless if the sample preparation step leads to deviations from the true value that exceed the reproducibility of the determination itself by several orders of magnitude. The risks are particularly great when one is working with very low concentrations. Even very small amounts of contamination introduced with the reagents, or "memory effects" from previous experiments, can generate utterly incorrect results. The laboratory atmosphere and materials present in the sample containers can also contribute to contamination and loss, and the potential for trace volatilization (of mercury, for example) at elevated temperature must always be taken into account. Sample preparation is an ideal place to apply the principles of "Good Laboratory Practice" (GLP), → Analytical Chemistry, Purpose and Procedures; even better, "Good *Analytical* Practice" (GAP). Construction materials for the equipment used in sample preparation must be selected very carefully, with appropriate consideration of possible adsorption – desorption effects. Steaming with concentrated acid is the most effective method for cleaning the equipment (Section 6.2.1.3).

The nature of any safety devices employed can also play a role with respect to uncertainty in the case of a pressure ashing system. For example, serious losses may be incurred if sample constituents are entrained during a pressure blow-off. With an irreversible safety device like a rupture disc such a problem may be immediately apparent, but the activation of a reversible device like a valve generally cannot be documented even though major losses may be excluded by auto-matic reclosure of the device. The optimal solution is almost always a function of the particular application in question.

6.2.6.3. Time Factors

Sample preparation is increasingly coming to be regarded as the "bottleneck" in analysis, because the actual determination methods are becoming increasingly rapid and ever more stringent demands are imposed with respect to precision and accuracy. The only remedies available are acceleration of the decomposition process and simultaneous preparation of multiple samples. However, the latter expedient inevitably leads to higher investment costs and increased operating expenses. Decomposition can be accelerated by working at higher temperature. In the case of wet and dry decompositions, heating can be accelerated by the use of microwave radiation. The subsequent cool-down phase also consumes less time, because the vessel walls are heated only indirectly via the hot sample, and there is no metal autoclave present, with its correspondingly high heat capacity.

Finally, an important parameter to consider in evaluating a routine sample preparation method is the number of staff involved. Several manufacturers have for some time been making a deliberate attempt to reduce staff requirements by introducing automation (or partial automation) into the regimen of sample preparation. The results have been quite satisfactory in the case of apparatus operating at atmospheric pressure, but the technical challenge is greater with systems operating under pressure, especially high pressure (see Section 6.2.8).

6.2.6.4. The Final Result

It is impossible to separate completely the consideration of a specific decomposition process from the subsequent analytical determination. Every decomposition must be adapted to suit a particular determination, which means there can be no such thing as an ideal decomposition method appropriate to all situations and applications. The preferred method must be established with reference to the following factors:

1) Nature of the analyte(s)
2) Concentration of the analyte
3) Matrix characteristics
4) Sample size

5) Required completeness of matrix degradation
6) Amount of time involved
7) Proposed method of determination

If the greatest possible degree of attention must be directed toward completeness in the trace range, decomposition in a closed system is preferred. Ensuring essentially complete decomposition requires the application of high temperature over a period of at least two hours. If sample preparation must be accomplished more rapidly (as in the case of food analysis, environmental studies, or quality control) tests should be conducted on the feasibility of a microwave-assisted process or direct analysis (see Section 6.2.8.2).

An investigation into the accuracy of an analysis should never be restricted to optimization of the determination itself: equal concern is warranted with respect to sample preparation. The key considerations listed above should be regarded only as rough starting points.

6.2.7. Concentration and Separation of Inorganic Trace Materials

Subsequent to any decomposition, but also in the case of liquid samples such as water and urine, the analytes of interest are generally present in dilute solution together with a large excess of foreign ions (e.g., alkali-metal and alkaline-earth cations). Separation and concentration of the analytes may be necessary to improve the limit of detection and exclude interference. Useful techniques in this regard include liquid–liquid extraction, solid-phase extraction, special precipitation reactions, and electrolytic deposition.

Liquid–liquid extraction is a widely used technique for separating and concentrating traces of various elements. For this purpose the analyte is first complexed with a suitable chelating agent, and the complex is then extracted into a water-immiscible solvent. Generally speaking, any complexing agent can be used provided it leads to stable, extractable complexes of the analyte in question. The complexing agent should be capable of separating the analyte as selectively and quantitatively as possible from accompanying ions. Such an extract can usually be used directly for a determination based on AAS. For example, nickel [91], cobalt [92], and cadmium [93] can all be separated and concentrated directly from urine after chelation with hexamethylene ammonium – hexamethylene dithiocarbamidate (HMA –

HMDC), using the solvent mixture methyl isopropyl ketone – xylene (70:30). The separation and enrichment of nickel and antimony in the form of HMA – HMDC complexes from digestion solutions of livers and kidneys has also been described, where copper and iron are removed in a preliminary step as the corresponding cupferron complexes [94]. Thallium can be extracted from a digestion solution after complexing with HMA – HMDC, diethyl dithiocarbamate, dithizone, and 8-mercaptoquinoline [95], or cupferron [96]. Antimony in urine has been extracted after nitric acid digestion as an ammonium pyrrolidine dithiocarbamate complex [97].

The advantage of this approach to concentration is that it permits very selective separation provided a complexing agent is selected with sufficient regard for accompanying ions. Its principal disadvantage is that the potential concentration factors tend to be rather small.

Metal complexes separable by liquid–liquid extraction can also be concentrated on and subsequently eluted from such adsorber resins as XAD, a reverse phase (e.g., C_{18}) [98], activated charcoal [99], or cellulose [100], leading to the technique of *solid-phase extraction*. Numerous reports describe the concentration of analytes on ion-exchange resins [101], [113] and such chelate-forming ion exchangers as Chelex [101], [114], [115] or Hyphan [116], [117]. Solid-phase extraction may result in concentration factors as high as 1000, particularly with dilute samples like drinking water or rain water.

Separation of element traces is also possible with *precipitation* reactions. This technique permits rapid and extensive concentration in a relatively uniform matrix, but it is not very selective. Aluminum hydroxide, magnesium hydroxide, iron hydroxide, and hydrogen sulfide have all been used for trapping trace amounts of various elements [101], [118]–[120], as have such organic precipitating agents as thionalide, cupferron, and dithiocarbamate [101], [121], [122]. A subsequent determination is carried out either directly on the separated precipitate or after restoring it to solution [118].

The *electrolytic deposition* of ions or complexes has also been described with such electrode materials as carbon, copper, or platinum. For example, this technique is utilized in the method of inverse voltammetric determination, where analyte is concentrated at the cathode [101].

New techniques and strategies have been introduced into sample preparation in the area of
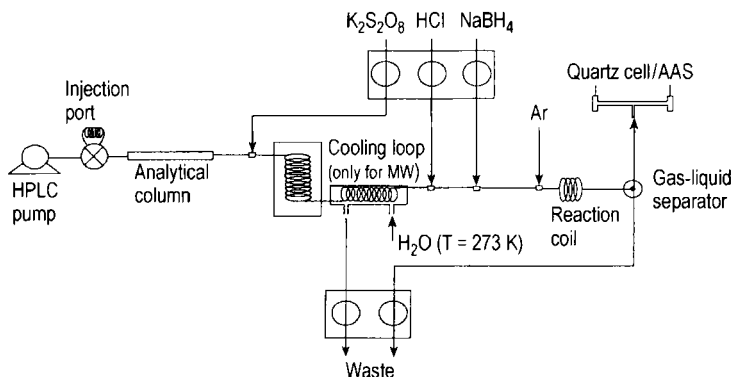
Figure 6. Schematic flow diagram of an on-line coupling of HPLC via UV or microwave-assisted reactors to hydride generation AAS for the determination of As species

preconcentration of inorganic analytes and the separation of such analytes from the matrix [102]–[104]. Very recently MANSFELDT and BIERNATH [105] described a micro-distillation apparatus (Fig. 6) for the determination of total cyanide in soils and sludge.

Another paper has suggested solid-phase extraction (SPE) for the preconcentration and speciation of Cr in waste water samples prior to AAS determination [106]. Interferences of di- and trivalent cations and diverse anions were studied. The detection limits for these species are about 1 µg/L. Other examples for SPE procedures are given in [107] for seawater analysis using on-line ICP-MS and in [108] for the separation and determination of precious metals.

New aspects of liquid–liquid extraction after a complexation step have been discussed for noble metals [109] and for rare earth elements [110].

Also, supercritical-fluid extraction (SFE) has been used for analyte enrichment and matrix separation in metal determinations [111]. The solubilities of metal dithiocarbamates in supercritical carbon dioxide have been characterized [112].

As an additional step in trace analysis, preconcentration techniques are resorted to only when either the sensitivity of the analytical determination is inadequate or severe matrix effects worsen the detection of the analytes [103].

6.2.8. Automation and Direct Analysis

6.2.8.1. Automation

Efforts have long been directed toward automation as a means of minimizing the labor and time involved in an analysis. The introduction of laboratory robots should make it possible to incorporate a significant degree of automation into the time-consuming, labor-intensive area of sample preparation as well, leading to more efficient, reliable, and reproducible sample work-up. An example of an automated system for microwave-assisted pressure ashing is presented in [123].

On-line procedures based on UV-photolysis or microwave-assisted dissolution have been established during recent years for the rapid digestion of relatively simple matrices such as water and waste water samples [130], [131]. In [124] on-line UV digestion with a segmented-flow device was applied for the determination of total cyanide prior to amperometric detection. The detection limit was found to be 0.2 µg/L. The throughput is about 30 analyses per hour.

Another example, that has been published recently by SUR et al. [125], is the on-line microwave and UV digestion of urine samples prior to hydride generation AAS (Fig. 7). The authors managed to determine as many as six arsenic species quantitatively [As(III), As(V), MMA, DMA, arsenobetaine, arsenocholine)]. For more information on species analysis see Section 6.2.9).

Ozone has also been used as a reagent with high quality criterions. For the determination of Hg in a batch cold-vapor system, ozone was used instead of permanganate or peroxidisulfate. The oxidation efficiency of ozone was found to be very high and the digestion was complete in less than 2 min [126].

Microwave decomposition at atmospheric pressure can also be used as an on-line technique, e.g., in a way that the sample passes through a PTFE coil with an inner diameter of about
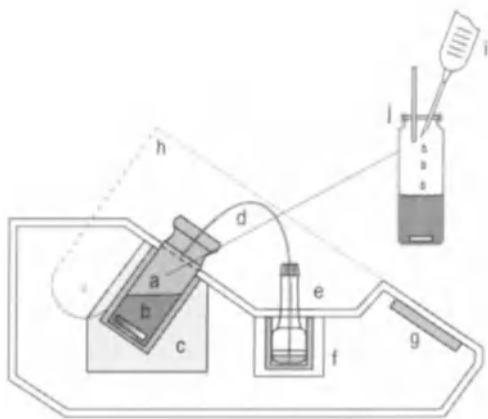
Figure 7. Apparatus for micro-distillation
a) Sample vessel; b) Frit; c) Heater; d) Capillary; e) Absorption vessel; f) Cooler; g) Keyboard and display; h) Protection hood; i) Syringe for reagent feeding; j) Septum (with permission from Dr. Tim Mansfeldt, Ruhr-Universität Bochum, Germany)

1 – 2 mm which is situated in a focused microwave-heated oven. A fully automated on-line sample pretreatment system which combines microwave digestion with sample preconcentration and matrix separation for heavy metals in blood samples is described in [127]. The digested sample solution was transferred on-line to a column packed with iminodiacetate resin for separation of matrix elements (alkaline, alkaline earth, nonmetals) that might interfere with the analyte mass signals in the ICP-MS detection method used. For Ni and Pb the authors reached detection limits in the upper ng/L range, for Fe, Cu, and Zn in the lower µg/L range. The sample throughput is given to be 6 samples/h.

PICHLER et al. [128] have shown that such an on-line microwave procedure can also be adapted to pressurized ashing. In this special case, the sample is injected into a constant stream of nitric acid which passes the microwave heating zone. Expansion of the coil due to the high vapor pressure is prevented by pressurizing its surroundings with nitrogen at up to 35 bar. The temperatures that can be reached in such a system may be as high as 230 °C or even higher.

An on-line SFE/HPLC procedure for the determination of Rh and Pd via chelates has been described in [129].

Flow-injection systems and continuous-flow systems can also make a contribution at the concentration and separation stages as a way of enhancing the potential advantages of automated

atomic spectrometric determinations by the flame and graphite-furnace techniques [132], [133]. Another possible application of flow injection is in the conversion of an analyte into the particular oxidation state called for in some sample preparation or determination procedure, as in hydride generation or cold-vapor AAS [134], [135].

6.2.8.2. Direct Analysis

There are certain situations in which sample preparation can be completely or largely omitted, as in the case of graphite-furnace AAS analysis, which permits the direct introduction of solid samples. Here the "ashing step" has been incorporated into the determination itself, which is then referred to as a "direct process." This integration permits an analysis to be carried out more quickly, and the chance of errors is reduced by limiting the number of individual operations [136]. On the other hand, it is essential that the accuracy of such an analysis be confirmed by comparisons with standard reference materials.

Examples of the AAS analysis of powdered samples, as well as such solid foods as chocolate and flour, are presented in [137], [138]. The direct determination of mercury in soils, coal, and ash is discussed in [139]. ICP – OES (and ICP – MS) can also be transformed into a direct process through solid sample introduction and electrothermal pretreatment [140].

6.2.9. Analysis of Element Species

For the determination of the Hg species methylmercury, phenylmercury and Hg(II), ozone was used successfully in a batch cold-vapor system [126]. The preconcentration and speciation of Cr(III) and Cr(VI) in water sample can be performed using solid-phase extraction (SPE) [105]. An SPME (solid-phase microextraction) technique has been used as the sample preparation system for the innovative simultaneous multielement/multispecies determination of six different mercury, tin, and lead species in waters and urine with GC/MS-MS [141], [142]. The determination of arsenic species [As(III), As(V), MMA, DMA, arsenocholine, arsenobetaine)] has also been shown to work with an on-line digestion step prior to hydride generation AAS [125] (see Fig. 7).

6.3. Sample Preparation in Organic Analysis

Recent decades have witnessed significant advances in the efficiency and productivity of instrumental methods in the field of organic trace analysis. Chromatographic and spectroscopic methods in particular have improved greatly with respect to sensitivity. There has also been constant improvement in selectivity, to the point where some samples can now be subjected to analysis without prior preparation, although this is certainly not true in the majority of cases. The goal of sample preparation in organic trace analysis is to isolate the analyte from the sample matrix and then concentrate it and convert it into a form suitable for analysis by the selected method.

Microwave-based techniques have become versatile tools not only for total digestion of organic and inorganic matrices for use in inorganic analysis, but also in drying processes, solvent extraction, clean-up steps and specific reactions such as sample preparation steps in organic analysis [143].

6.3.1. Sample Treatment after the Sampling Process

6.3.1.1. Stabilization, Drying, and Storage

In the context of organic trace analysis, appropriate stabilization and storage precautions are a function of the nature of the analyte and its concentration. To avoid contamination by ambient air and dust, all operations should be conducted under clean-room conditions (clean-bench environment). It is of paramount importance to assure that contamination from vessels, covers, septa, and stabilizers is rigorously excluded.

Volatilization losses can pose a problem in the case of analytes with high vapor pressure. Evaporative loss can be largely avoided by completely filling each sample container and fitting it with an air-tight seal (water samples, biological fluids). This procedure ensures that there will be no oxidative loss due to oxygen present in the vapor space above the sample. For example, blood samples to be analyzed for volatile organic compounds (VOCs) are first treated with an anticoagulant, after which they can be stored for several months at +4 °C in Sovirel or Pyrex test tubes protected only by screw-cap closures and PTFE seals [144].

Similarly, VOCs adsorbed from air samples onto Tenax or activated charcoal can be sealed and stored for several months prior to analysis. Losses have been reported only with cyclic and aliphatic ketones (catalytic oxidation) and esters (hydrolysis) [145]. For these groups of substances, losses approach a maximum of 20 % after a four-week storage period. No losses were observed with benzene, toluene, and xylenes over a 24-month storage period [146]. Samples containing analytes of lower volatility can be stored in suitable containers for a few days at +4 °C, and for several months deep-frozen at − 18 °C or lower.

If an analyte is to be separated by Soxhlet extraction or supercritical fluid extraction, it is essential that the sample ist first dried. Standard practice is to add a drying agent like anhydrous sodium sulfate. Freeze-drying has also been recommended, but care must be exercised to prevent large nonreproducible losses of the analyte. For example, freeze-drying of milk leads to some loss of most PCB congeners. Losses may be as great as 50 – 74 % for monochlorobiphenyl, dichlorobiphenyl, and trichlorobiphenyl [147], [148].

6.3.1.2. Homogenization and Aliquoting

It is best to avoid homogenization prior to the determination of highly volatile analytes, since this often leads to nonreproducible losses. Liquid samples — preferably solids as well — should be transferred immediately after sampling to gas-tight sample vessels, followed by a direct determination on the basis of headspace gas chromatography (Section 6.3.3). If comminution is absolutely necessary, this should be accomplished in the deep-frozen state [149].

For relatively nonvolatile analytes, in such solid matrices as tissues, foods, or soil, the samples should be homogenized prior to aliquoting. This is often accomplished by trituration with sea sand, although use of a drying agent instead results in simultaneous removal of water, producing a homogenate suitable for immediate liquid extraction in a Soxhlet extractor (Section 6.3.2.3) or for supercritical fluid extraction (SFE; Section 6.3.2.4). Small amounts of sample can be homogenized in an ultrasound bath after the addition of water [150].

6.3.1.3. Requirements with Respect to Materials and Chemicals

Plastic containers are generally inappropriate for storing organic samples, since losses may be incurred through adsorption or migration through the vessel wall, and interfering substances may migrate into the sample from the container itself (plasticizers, antioxidants, monomers, etc.). In the case of analytes with high vapor pressures, volatilization losses must also be anticipated with plastic containers. Samples for organic trace analysis should always be stored in glass vessels with appropriate PTFE-sealed screw-cap closures or glass stoppers.

Materials and chemicals must be carefully checked for contamination before use, and if necessary cleaned and purified. Heating for several hours [151] or several days [152] is absolutely essential and very effective in the trace analysis of volatile analytes. Precleaning with boiling nitric acid may be necessary, as may a rinse with an organic solvent like acetone. Equipment for use in the determination of nonvolatile analytes is most effectively cleaned by rinsing with solvents that will be used subsequently in the course of the preparation process [153].

Glass vessels may also contribute to loss of an analyte, especially from aqueous solution, through adsorption on the vessel surface. This can be largely prevented by silanization of active sites on the glass. Alternatively, the vessel should be rinsed with solvent after removal of the sample, with the rinse solution then added to the sample itself [147].

Liquid–solid extraction procedures are accompanied by a serious risk of contamination. For example, blank tests with commercial extraction columns have revealed traces of plasticizers and antioxidants as well as various alkanes and alkenes [154]. For this reason certain manufacturers now offer solid-phase extraction columns made from glass. The extent of the po-tential contamination depends on the manufac-turer, the adsorbent load, conditioning steps, and the solvent used for the extraction [154], [155]. If no satis-factory commercial column is available, one can easily be prepared in the laboratory from a suitable adsorbent and an appropriate glass tube. For example, XAD can be purified to an extent sufficient for this purpose by Soxhlet extraction [156], [157], which reduces the level of impurities by a factor of 30. Silica gel and aluminum oxide can be purified by washing with solvents, followed by activation in a drying oven [152]. For example, unpurified silica gel has been shown to release as much as 160 ng/g of PCBs [155].

The risk of contamination must also be considered when a diffusive sampler is used to obtain air samples for determining volatile organic compounds. Thus, benzene, n-tetradecane, n-pentadecane, n-hexadecane, n-heptadecane, and dioctyl phthalate [158], along with trichloromethane, 1,1,1-trichloroethane, trichloroethene, and tetrachloroethene [159] have all been detected in extracts from unexposed diffusive samplers. The measured concentrations varied from batch to batch, ranging from 0.005 µg to 0.14 µg per diffusive sampler in the case of the chlorinated hydrocarbons [159], and from 0.1–2.6 µg per diffusion collector for the aliphatic hydrocarbons and dioctyl phthalate [158].

Ultrapure water can be prepared in the laboratory by filtration through an ion exchanger and activated charcoal. According to the manufacturer, one commercial water treatment system of this type reduces the residual hydrocarbon content of water to < 20 µg/L. Further reduction to < 5 µg/L is possible through UV irradiation (as with the "zero-water unit" manufactured by Grätzel, Karlsruhe, Germany). If necessary, additional purification can be achieved with a stream of nitrogen at 80 °C [151], or by extraction with the solvent that will be used for extracting the analyte.

Most solvents are available commercially in sufficient purity, but blank test values should be established on a regular basis to ensure the absence of contaminants.

6.3.2. Separation of the Analyte

6.3.2.1. Hydrolysis

Many of the harmful substances and pollutants in biological samples, including their metabolites, are present wholly or in part in a conjugated (bound) form. This is the case, for example, with pentachlorophenol and phenol, which are excreted in the urine as glucuronide or glucuronide and sulfate, as well as with aromatic amines, present to some extent in the blood as hemoglobin conjugates. In such a situation the first step must accomplish release of the primary analytes. Conjugates can be cleaved by acid hydrolysis [160]–[163], basic hydrolysis [162], or enzymatic hydrolysis [162], [164]. Hydrolysis with the aid of

enzymes, such as sulfatase and glucuronidase, has the significant advantage that bonds are cleaved extremely selectively, and information can also be obtained regarding the binding partners of the analyte. A disadvantage is the considerable amount of time required, typically several hours.

In many cases a portion of the matrix can be removed or converted into some more easily removable form before actual separation of the analytes is attempted. Relatively involatile, chemically stable analytes [e.g., organochlorine pesticides, PCBs, or polychlorinated dibenzo-*p*-dioxines (PCDDs)] can frequently be released from a matrix by acidic or alkaline hydrolysis with concentrated acid or base. Proteins can often be precipitated from biological samples by addition of acid, salt, or an organic solvent, or they can be broken down by enzymatic hydrolysis [165]–[167].

6.3.2.2. Liquid – Liquid Extraction

Extraction with organic solvents is the most common extraction technique for isolating an analyte from a liquid sample. The technique is based on distribution of the analyte between two immiscible liquid phases.

The decisive parameter with respect to extraction yield is the *distribution coefficient* for the analyte between the particular phases involved. A distribution coefficient can often be influenced advantageously by establishing a specific pH, thereby dividing the sample into strongly or weakly acidic, neutral, or basic fractions [168]. An example is provided by the extraction of aromatic amines from blood and urine [169]. Further possibilities include formation of ion pairs [170], complexation with metal salts, or salting-out of an aqueous phase.

If the distribution coefficient is sufficiently large, the simplest approach to liquid – liquid extraction is shaking the sample with an appropriate amount of an organic solvent. With smaller distribution coefficients or large sample volumes, continuous extraction or countercurrent extraction is required to achieve a complete separation. The apparatus for continuous extraction causes a liquid immiscible with the sample solution to circulate continuously and in finely divided form through the sample [171]–[173]. Extracted analytes are concentrated by distillation at appropriate times between individual extraction cycles.

Disadvantages of liquid – liquid extraction include the high dilution of the extract, which must

subsequently be concentrated, and a high rate of solvent consumption. Both problems can be minimized by employing microextraction methods. Devices for extracting aqueous samples as large as 1 L with 200 µL of organic solvent have been described [174], [175].

Another variant of liquid – liquid extraction takes advantage of a liquid phase immobilized on a solid sorbent such as kieselguhr, Celite, Chromosorb W, or Chromosorb P [176]. The immobilized phase may be either aqueous or nonaqueous.

The formation of emulsions frequently presents problems in liquid – liquid extraction. Often such emulsions can be broken by centrifugation, freezing, or the addition of salt.

6.3.2.3. Soxhlet Extraction

Soxhlet extraction is chiefly applied to the separation of relatively nonvolatile analytes from solid samples. For example, this technique is useful for extracting pesticides, PCBs, and PCDD/PCDFs from various matrices, including fatty tissue, soil, and paper [177], [178]. The advantage of the method is high yield, achieved by continuous extraction of the sample with fresh extractant. Disadvantages include the considerable amount of time required (typically several hours to several days), the thermal stress to which analytes are subjected, and the fact that the analytes are obtained in very dilute form in a solvent, and must therefore be concentrated in a subsequent step. Any fat present in the sample is co-extracted, and this must also be removed separately.

6.3.2.4. Supercritical Fluid Extraction (SFE)

Recently, an extraction method long practiced in industry — supercritical fluid extraction (SFE) — has also been introduced into the analytical sector. As the name suggests, SFE employs supercritical fluids for extraction purposes in place of the organic solvents of conventional extraction.

A compound is said to be in a supercritical state when the critical pressure and critical temperature for that particular material are exceeded. Critical parameters for several relevant compounds are listed in Table 6. Supercritical fluids exhibit simultaneously properties associated with both gases and liquids (see Table 7). Thus, like gases, they are compressible, but they also display solvencies similar to those of liquids. Any increase in temperature at constant pressure reduces the

Table 6. Critical parameters for selected substances [179]

Substance	Critical temperature, °C	Critical pressure, MPa
Xe	16.6	5.83
CHF$_3$	25.9	4.83
CClF$_3$	28.8	3.92
CO$_2$	31.1	7.37
N$_2$O	36.4	7.24
NH$_3$	132.2	11.27
CH$_3$OH	239.4	8.09
H$_2$O	374.1	22.04

Table 7. Typical characteristics of gases, liquids, and supercritical fluids

Property	Gas	Liquid	Supercritical fluid
Density, g/mL	$10^{-4} - 10^{-3}$	0.6 – 1.4	0.1 – 1
Diffusion coefficient, cm^2/s	10^{-1}	10^{-5}	$10^{-3} - 10^{-4}$
Solvency	no	yes	yes
Compressibility	yes	no	yes

solvent power of a supercritical fluid, but it also leads to an increase in diffusion rate, which tends to lower the minimum required extraction time. Compared to conventional extractants, supercritical fluids have low viscosities and diffusion rates that are higher by a factor of 10–100, both of which contribute to reduced extraction times. With supercritical CO$_2$ and N$_2$O, which are gases under normal conditions, the extractant is separated by reducing the pressure to atmospheric levels, leading to simultaneous concentration of the extract.

Supercritical CO$_2$ is the most frequently used extractant for SFE. It has the advantage of being chemically rather inert, and its critical temperature is low, so it is valuable for the extraction of such thermolabile analytes as steroids and fragrances. The low critical pressure for CO$_2$ opens the way to a relatively broad range over which the solvency can be varied through adjustment of the pressure. Other advantages of CO$_2$ as an extractant include low toxicity, high purity, and low cost. The principal disadvantage of CO$_2$ is a relatively low polarity. Its solvent power with respect to polar analytes can be improved, however, by adding such polar *modifiers* as methanol, acetone, hexane, or dichloromethane.

Extraction with supercritical CO$_2$ has been recommended for separating a wide variety of analytes, including pesticides, PCBs, vitamins, and fragrances from meat, fish, baby food, and animal feed [180]–[185]. PAHs, PCBs, PCDO/PCDFs, and other substances have been extracted from soil, fly ash, sediment, air particles, polymers, and plants using supercritical CO$_2$ together with a modifier [186]–[196].

6.3.2.5. Solid-Phase Extraction (SPE)

Solid-phase extraction (SPE) is used for the selective separation and concentration of analytes from liquid samples. Extraction of the analytes is based in this case on the distribution of dissolved substances between a solid-phase surface and the sample liquid. Separation of various sample constituents may be a result of differing polarities, differences in molecular size, or differences with respect to ion-exchange capacity.

Many of the comments above regarding liquid–liquid extraction also apply to solid-phase extraction. However, the latter has the advantages that it can be accomplished more rapidly, requires less solvent, provides more highly concentrated extracts, and is relatively easy to automate. Laboratory robots and suitable column arrangement have rendered large parts of the sample preparation process automatic, thereby facilitating reproducibility and efficient operation [197], [198].

Many different adsorbents are applicable in this context (see Table 8). The most common adsorbents for solid-phase extraction are based on silica gel, the surface of which has been modified in some way.

Octadecyl surface phases (C$_{18}$) are used for the reversed-phase extraction of nonpolar substances from aqueous solutions. Typical applications include the extraction of organochlorine pesticides [199]–[201], organophosphorus pesticides [202], chlorinated and unchlorinated hydrocarbons, triazines [200], [201], PAHs and nitro-PAHs [200], carbamates [200], chlorophenols [200], [203], aflatoxins [204], plasticizers [205], vitamins [206], and medicaments such as barbiturates and antibiotics [207], [208].

Shorter octyl phases (C$_8$) are used for extracting substances of medium polarity. Substances that bind irreversibly to C$_{18}$ phases can often be concentrated and re-eluted successfully with C$_8$ phases.

Normal phase materials, including unmodified silica gel, aluminum oxide, and Florisil, separate sample constituents into fractions of comparable polarity. They are often utilized to separate and concentrate pesticides [209]–[211], PCBs [147], [209], and PCDD/PCDFs [212]–[216] from such

Table 8. Important adsorption agents classified according to type

Nonpolar	Polar	Anion exchangers	Cation exchangers
Octadecyl	Cyano	Primary amine	Carboxylic acid
Octyl	Kieselguhr	Secondary amine	Sulfonic acid
Butyl	Silica gel	Quaternary ammonium salt	
Cyclohexyl	Florisil		
Phenyl	Aluminum oxide		
Amino			
Diol			

biological samples as blood, breast milk, fatty tissue, and foods. Normal phases are also used to extract polar sample constituents, such as amines, alcohols, phenols, dyes, medicaments, or vitamins [217].

Anion exchangers are mainly employed in the extraction of carbohydrates, peptides, nucleosides, and amino acids. Cation exchangers are useful for extracting amino acids and nucleosides [217].

Apart from modified silica gel, the most frequently used solid-phase adsorbents are activated charcoal, macroreticular resins (XAD), ordinary silica gel, aluminum oxide, and Florisil. Activated charcoal is a universal adsorbent for concentrating trace organic materials in aqueous solutions and air. XAD resins are also commonly employed for extracting organic trace constituents, such as alkoxyacetic acids [157], organochlorine pesticides [200], [218], carbamates [198], [202], and triazines [218] from aqueous matrices.

The activity of these solids must be accurately adjusted to ensure reproducible results and well-defined fractions, and sample solutions must always be carefully dried prior to extraction.

Exceptionally low limits of detectability can be achieved with automated on-line systems. Here the complete eluate is analyzed rather than simply an aliquot, which lowers the risk of evaporative losses and sample contamination. Automated sample preparation through solid-phase extraction is becoming increasingly important in extreme trace analysis, as in the determination of PCDDs in tissue or in the blood of uncontaminated persons [219].

A recently developed variant of solid-phase extraction is extraction with the aid of *extraction disks* [220], [221], membrane filter plates on which an appropriate solid-phase material has been immobilized. Advantages of extraction disks include higher flow rates and more rapid material exchange. Another variant involves the use of SPE for simultaneous extraction and derivatization. For this purpose the analyte is adsorbed onto a solid

phase that has been impregnated with a derivatizing agent, permitting the separation and concentration of analytes that would otherwise not be adsorbed at all. Solid-supported reagents are useful in such applications as the determination of organic acids or cannabinoids in plasma [222], [223].

6.3.2.6. Solid-Phase Microextraction (SPME)

A new technique, which is applicable for sampling in air and liquids or in the headspace above a liquid or a solid sample, is solid-phase microextraction (SPME). The mechanism of SPME, which has been developed by PAWLISZYN et al. [225], [226], is based on the partition equilibrium of the analytes between the sample or the headspace above the sample, respectively, and a fused silica fiber coated with a suitable stationary phase. The amount of analyte extracted by the fiber is proportional to the initial analyte concentration in the sample and depends on the type of fiber. After sampling, the fiber can be thermally desorbed directly into the injector of a gas chromatograph. SPME combines sampling, analyte enrichment, matrix separation, and sample introduction within one step [226]. Since its development, this innovative technique has found widespread use in environmental analysis. It has, for example, been applied in the determination of volatile organic compounds [227], [228], phenols [229], pesticides [230], polyaromatic hydrocarbons, and polychlorinated biphenyls [231] in water.

SPME fibers have been used as air sampling devices for volatile organic compounds in ambient and workplace air and give results that are in good agreement with traditional sampling methods [232]. Furthermore, grab sampling is used with stainless steel canisters or glass bulbs in combination with SPME [226], [233]. The dependence of the sampling rate on humidity and air temperature can be eliminated by correction factors [233]. The main drawback is the low storage stability of the

samples, due to uncontrolled losses of analytes by adsorption on the walls of the canisters or by evaporation from the loaded fiber.

6.3.2.7. Miscellaneous Techniques

Microwave-assisted extraction (MAE) is an upcoming trend in rapid extraction techniques. Is has been applied recently, for example, to the extraction and determination of polycyclic hydrocarbons in marine sediments [234] and in wood samples [235] prior to an HPLC/UV determination.

An alternative approach to separating volatile compounds from liquid and solid samples is distillation. Simple distillation can accomplish the isolation of a volatile analyte from a nonvolatile residue, or separate multiple sample constituents with widely differing boiling points. A further application is separation of a sample into fractions with different boiling ranges.

Steam distillation is an effective way to separate from a matrix such steam-volatile analytes as phenol [224], alkylphenols [160], or formic acid [236]. Steam distillation offers the advantage that the analyte is recovered in a nearly matrix-free condition, although it is diluted with a large amount of water from which it must subsequently be separated and concentrated. Solid-phase extraction is particularly useful for this purpose (Section 6.3.2.5).

Another technique involving distillative concentration of sample constituents is *sweep co-distillation*. Here the sample is treated with a highly volatile solvent introduced with the aid of a stream of carrier gas. The solvent in turn transports soluble components of the sample to a cooled distillation receiver. The method is useful for such applications as the isolation of volatile pesticides from animal and plant fats [237], [238].

Membrane techniques, including dialysis, ultrafiltration, and reverse osmosis, are also applicable to sample preparation problems [239] – [241]. *Dialysis* separates analytes on the basis of their ability to diffuse through a membrane as a result of a concentration gradient. It is most frequently used for concentrating analytes in biological liquids. Automation and on-line coupling with liquid chromatography have both been reported [242], [243]. *Reverse osmosis* is very similar to dialysis, but in this case a pressure difference is used to cause the solvent to migrate through a membrane from a region of high analyte concentration to one of low analyte concentration. Reverse osmosis is frequently called upon for concentrating large volumes of very dilute solutions. *Ultrafiltration* is a method for concentrating large sample molecules present in dilute solution. Separation in this case depends directly on molecular size, and occurs when a sample solution is filtered through a membrane with an appropriate pore size. One possible application is in the separation of free molecules present in a biological fluid from similar molecules that are bound to receptors [244]; another is the desalination of a liquid sample.

6.3.3. Headspace Techniques

6.3.3.1. Static Headspace Technique

Highly volatile analytes can be separated for subsequent quantitative determination by one of several *headspace techniques*. All headspace techniques are based on the Henry – Dalton law, which states that, in a closed system, the vapor-space concentration of an analyte depends exclusively on the temperature and the corresponding analyte concentration in solution or on a solid surface. Consequently, raising the temperature causes volatile analytes to separate from their matrix and become concentrated in the surrounding vapor space. In the *static* headspace technique, a gas chromatographic determination is made of the analyte once equilibrium has been established. The separation efficiency characteristic of gas chromatography and the virtual absence of matrix-derived detector noise means that chlorinated, aromatic, and aliphatic hydrocarbons, along with alcohols, ketones, and esters, are all subject to determination essentially free of interference regardless of the matrix. Limits of detectability in the µg/L range permit volatile organic compounds to be determined in body fluids from persons occupationally exposed to contaminants [245], [246], as well as in foods [247] and water [248]. Nevertheless, the distribution coefficient does depend on the matrix, so a relatively complicated and time-consuming calibration is essential. Depending on the matrix, calibration can be based on matrix standards, standard addition, or internal standardization.

6.3.3.2. Dynamic Headspace Technique (Purge and Trap)

In the dynamic headspace technique, a carrier gas is passed continuously through the sample under investigation. Constant contact with fresh

carrier gas causes volatile analytes to be removed gradually and almost completely from the sample. These are then concentrated on an adsorbent such as Tenax or Chromosorb. Quantitative transfer of the analytes to a gas chromatographic column is accomplished by thermodesorption, using either on-line or off-line techniques.

The concentration that results (typically by a factor of ca. 100) leads to limits of detectability in the ng/L range, permitting the simultaneous analysis of environmentally conditioned concentrations of as many as 48 volatile organic compounds in water and such body fluids as blood and urine [144], [249]–[251].

6.3.4. Determination of Trace Organic Materials in Air Samples

Air sampling for analysis of gases and vapors is accomplished either with the aid of a pump and subsequent concentration on a solid adsorbent (*active sampling*), or with a collector, in which case analytes reach the collecting phase via diffusion or permeation (*passive sampling*). Gaseous trace organic materials in air are often collected by adsorption on activated charcoal, Tenax, or XAD resin, although other adsorbents such as aluminum oxide or Florisil can also be used.

Analytes are usually desorbed from the adsorbent by washing with appropriate solvents, after which the eluates can be analyzed by gas chromatography. Carbon disulfide is the most commonly employed desorption agent [252]. For the desorption of polar analytes, small amounts of such polar solvents as methanol, 2-propanol, or 2-butanol can be added to the carbon disulfide as a way of increasing the desorption yield [253].

Analytes can also be desorbed thermally and analyzed on-line with a gas chromatograph. This expedient results in a significant increase in sensitivity, since the entire sample is utilized for the analysis. One disadvantage is that the binding power of activated charcoal (the most suitable adsorption material for most analytes) is too great to be overcome with conventional thermodesorbers, eliminating the possibility of using this combination for quantitative analysis. Quantitative desorption from charcoal can be accomplished with a microwave thermodesorption device, however, and this microwave thermodesorption is a mild process, suitable even for such labile analytes as fragrances [254].

Particle-bound substances like polycyclic aromatic hydrocarbons (PAH) are determined by collecting particulate air constituents on glass-fiber or PTFE filters, or in impactors, followed by extraction.

6.3.5. Analyte Concentration

Subsequent to one of the aforementioned extraction procedures (e.g., liquid–liquid extraction, steam distillation, Soxhlet extraction) the analytes are obtained in extremely dilute form in a large volume of solvent. Various approaches to concentration are possible depending on the nature of the analyte, the solvent, the initial volume, and the target volume.

Aqueous analytes, such as those from a steam distillation, can be concentrated by either solid-phase or liquid–liquid extraction (Sections 6.3.2.2 and 6.3.2.5). An analyte dissolved in a highly volatile solvent (as with liquid–liquid or Soxhlet extraction) is concentrated most effectively by evaporation of the solvent in a gas stream (e.g., nitrogen, helium), or with a rotary evaporator or a Kuderna–Danish concentrator. Losses must be anticipated with a rotary evaporator, however, especially of volatile analytes, though even less volatile constituents may also be lost through codistillation. Recoveries of volatile analytes are higher with a Kuderna–Danish concentrator. Concentrating an extract by solvent evaporation in a gas stream is appropriate only with relatively small volumes on account of the low vaporization rate. Analyte loss is possible through aerosol formation or evaporation.

6.3.6. Derivatization

Separation and concentration of an analyte must often be followed by some type of *derivatization*. Derivatization is conducted with one or more of the following objectives:

1) To facilitate chromatographic separation
2) To increase selectivity
3) To improve the limit of detection

The first objective applies mainly to gas chromatography, and involves the preparation of a more volatile or less polar form of the analyte. Improved selectivity is generally less important in this case due to the large number of theoretical plates available with capillary gas chromatography columns,

in contrast to HPLC. However, the limit of detection plays an important role in both gas and liquid chromatography.

Most derivatizations for gas chromatography are esterifications or etherifications. For example, analytes containing carboxyl groups are often converted into methyl [156], [161], pentafluorobenzyl [255] – [258], or triethyl esters [259] – [261]. Analytes containing acidic hydroxyl groups (phenols, chlorophenols, glycol ethers) or amino groups (e.g., aniline) can be derivatized with such perfluorinated compounds as heptafluorobutyric anhydride, pentafluoropropionic anhydride [169], or pentafluorobenzoyl chloride [262].

Derivatization with halogenated compounds offers the advantage that the derivatives are subject to extremely sensitive detection with an electron capture detector (ECD). The disadvantage is that excess derivatization agent remains in the sample solution after the reaction, necessitating its removal prior to gas chromatographic determination because of the potential for interference during ECD detection.

Derivatizations for HPLC are designed mainly to improve the limit of detection, permitting the use of highly sensitive or selective detectors inapplicable to the analytes themselves. Enhanced absorption of UV/visible light is achieved by the introduction of chromophoric groups. Analytes can also be rendered fluorescent by the introduction of fluorophoric groups.

Carboxylic acids, including formic, acetic, lactic, propionic, malic, tartaric, and citric acids, can be transformed with benzyl, naphthacyl, phenacyl, or bromophenacyl bromides [263], [264]; p-nitrobenzyl bromide [265]; p-nitrophenacyl bromide; methoxyaniline; or p-nitrobenzyl-N,N'-diisopropylurea [266], [267] into esters that absorb UV or visible light. α-Keto acids (e.g., glycolic, glyoxylic acids) are detectable with UV light after derivatization with phenylhydrazine [268], [269].

Fluorescent compounds are obtained by reacting carboxylic acids with 4-bromomethyl-7-methoxycoumarin [270] – [273] or 4-bromomethyl-7-acetoxycoumarin [274], [275].

Analytes containing hydroxyl groups, such as phenols, glycols, and alcohols, can be converted with 3,5-dinitrobenzoyl chloride [276] or dabsyl chloride [277] – [279] into compounds that absorb UV or visible light. Fluorescent derivatives can be obtained with 7-(chlorocarbonyl) methoxy-4-methylcoumarin [280].

Derivatizations for HPLC purposes are accomplished either off-line or on-line. An on-line process may involve either pre-column or post-column reaction depending on the objective.

6.3.7. Coupled Techniques

Various coupled sample preparation and determination processes are increasingly utilized in trace organic analysis, whereby separated analytes pass directly in an on-line way into a chromatographic analysis system. The main advantages of such on-line procedures reflect the quantitative nature of analyte transfer from the sample preparation stage to the analytical determination, leading to an optimal limit of detection. This approach also minimizes the risk of sample loss and contamination. However, coupled techniques are relatively difficult to imple-ment, since parameters must be optimized not only with respect to sample preparation but also for the subsequent chromatographic separation. One of the most important coupled techniques at the present time is LC – GC. Automatic sample preparation is in this case accomplished via HPLC, permitting the separation of a complex matrix, or even a concentration of trace material. A subsequent capillary gas chromatographic analysis results in a high degree of resolution (with isomeric materials, for example) and the high signal-to-noise ratio essential for trace analysis thanks to the availability of such extremely sensitive and selective devices as electron capture detectors (ECD), flame photometry detectors (FPD), phosphorus – nitrogen-selective detectors (PND), and mass spectrometric detectors (MSD) [281] – [283]. Coupling is achieved with a sample loop (loop technique), which is used to transfer a small portion of the LC eluate to an uncoated GC precolumn (retention gap) [284]. Determinations of pesticide metabolites from maize [285] and organochlorine pesticides from fat [286] constitute two examples of the widespread application of this technique.

LC – MS coupling is also becoming an increasingly common analytical method now that effective interfaces have been developed (particle beam, thermospray, electrospray). Determinations of pesticides in sea water [220], carbamates and phenylurea derivatives in water samples [287], and organophosphorus pesticides and chlorophenols [288] are just a few examples of the use of this method.

LC – TLC coupling constitutes one possibility for carrying out multidimensional LC. Thus HPLC (usually reversed-phase chromatography) is coupled with thin layer (adsorption) chromatography in such a way that the eluate from the HPLC column is transferred via a capillary column to a mechanically transported TLC plate [289], [290].

The coupling of SFE with gas chromatography has also been described, as has chromatography with supercritical fluids (SFC). Examples include the separation and determination of PCBs, PAHs, and pesticides in such environmental samples as soil and sediments [291] – [294]. A frequently employed technique for determining traces of pesticide in aqueous samples is the on-line coupling of solid-phase extraction or dialysis with HPLC [242], [243], [295].

6.4. References

[1] P. Tschöpel: "Sample Treatment" in M. Stoeppler (ed.): *Hazardous Metals in the Environment*, vol. 12, Elsevier, Amsterdam 1992.

[2] O. Behne, *J. Clin. Chem. Clin. Biochem.* **19** (1981) 115.

[3] G. Knapp, *Int. J. Environ. Anal. Chem.* **22** (1985) 71.

[4] D. v. Renterghem, R. Cornelis, R. Vanholder, *Anal. Chim. Acta* **257** (1992) 1.

[5] J. Trettenbach, K. G. Heumann, *Fresenius Z. Anal. Chem.* **322** (1985) 306.

[6] L. B. Fischer, *Anal. Chem.* **58** (1986) 261.

[7] B. Grieping, H. Marchandise in: *Analytiker Taschenbuch*, vol. 6, Springer Verlag, Berlin 1986, p. 3.

[8] DFG Deutsche Forschungsgemeinschaft in J. Angerer, K. H. Schaller (eds.): *Analyses of Hazardous Substances in Biological Materials*. **vol. 5**, VCH Verlagsgesellschaft, Weinheim 1997, pp. XXI – XXII.

[9] J. Begerow, L. Dunemann, in L. Matter (ed.): *Elementspurenanalytik in biologischen Matrices*, Spektrum Akademischer Verlag. Heidelberg – Berlin – Oxford 1997, p. 27.

[10] D. Behne, G. V. Iyengar: "Spurenelementanalyse in biologischen Proben", *Analytiker-Taschenbuch*, vol. 6, Springer Verlag, Berlin 1986.

[11] M. Kiilunen, J. Järvisalo, O. Mäkitie, A. Aitio, *Int. Arch. Occup. Environ. Health* **59** (1987) 43.

[12] G. V. Iyengar, B. Sansoni: "Elemental Analysis of Biological Materials: Current Problems and Techniques with Special Reference to Trace Elements," *IAEA Technical Report 197*. Wien 1980, p. 73.

[13] G. V. Iyengar, K. Kasperek, L. E. Feindegen, *Sci. Total Environ.* **10** (1978) 1.

[14] J. Angerer: "Spezielle Vorbemerkungen," in D. Henschler (ed.): *Analysen in biologischem Material*, 10th suppl., VCH Verlagsgesellschaft, Weinheim 1991.

[15] J. C. Meranger, B. L. Hollebone, G. A. Blanchette, *J. Anal. Toxicol.* **33** (1981) 5.

[16] P. Tschöpel et al., *Fresenius Z. Anal. Chem.* **302** (1980) 1.

[17] K. Heydorn: *Neutron Activation Analysis for Clinical Trace Element Research*. **vol. 1**, CRC Press, Boca Raton, Fla. 1984, p. 42.

[18] J. R. Moody, P. Lindström, *Anal. Chem.* **47** (1977) 2264.

[19] S. P. Ericson et al., *Clin. Chem. (Winston-Salem N.C.)* **32** (1986) 1350.

[20] S. B. Adeloju, A. M. Bond, *Anal. Chem.* **57** (1985) 1728.

[21] R. Cornelis et al. in: *Nuclear Activation Techniques in the Life Sciences*. IAEA, Wien 1979, p. 165.

[22] J. Begerow, L. Dunemann in M. Stoeppler (ed.): *Sampling and Sample Preparation*, Springer-Verlag, Berlin – Heidelberg – New York 1997, p. 155.

[23] P. Lievens, J. Versieck, R. Cornelis, J. Hoste, *J. Radioanal. Chem.* **37** (1977) 483.

[24] A. Abu-Samra, J. S. Morris, S. R. Koirtyohann, *Anal. Chem.* **47** (1975) 1475.

[25] E. D. Neas, M. J. Collins in H. M. Kingston, L. B. Jassie (eds.): *Introduction to Microwave Sample Preparation*. Am. Chem. Soc., Washington, D.C., 1988, p. 7.

[26] R. A. Nadkarni, *Anal. Chem.* **56** (1984) 2233.

[27] Z. Sulcek, P. Povondra: *Methods of Decomposition in Inorganic Analysis*, CRC Press, Boca Raton, Fla. 1989.

[28] P. Aysola, P. Anderson, C. H. Langford, *Anal. Chem.* **59** (1987) 1582.

[29] R. T. White, Jr., G. E. Douthit, *J. Assoc. Off. Anal. Chem.* **68** (1985) 766.

[30] L.-Q. Xu, W.-X. Shen, *Fresenius Z. Anal. Chem.* **332** (1988) 45.

[31] G. Knapp, *Int. J. Environ. Anal. Chem.* **22** (1984) 71.

[32] G. Knapp, *Fresenius Z. Anal. Chem.* **317** (1984) 213.

[33] K. W. Budna, G. Knapp, *Fresenius Z. Anal. Chem.* **294** (1979) 122.

[34] F. R. Abou-Shakra, M. P. Rayman, N. I. Ward, V. Hotton, G. Bastian, *J. Anal. At. Spectrom.* **12** (1997) 429.

[35] H. Hodrejarv, A. Vaarmann, *Anal. Chim. Acta* **396** (1999) 293.

[36] M. Bettinelli, C. Baffi, G. M. Beone, S. Spezia, *At. Spectrosc.* **21** (2000) 50.

[37] A. C. Grillo, *Spectrosc. Int.* **1** (1989) 16.

[38] M. Kolb, P. Rach, J. Schäfer, A. Wild, *Fresenius J. Anal. Chem.* **342** (1992) 341.

[39] L. Dunemann in B. Welz (ed.): *5th Colloquium Atomspektrometrische Spurenanalytik*, Bodenseewerk Perkin-Elmer, Überlingen 1989, p. 593.

[40] L. Dunemann, M. Meinerling, *Fresenius J. Anal. Chem.* **342** (1992) 714.

[41] J. Hertz, R. Pani, *Fresenius Z. Anal. Chem.* **328** (1987) 487.

[42] F. Wahdat, R. Neeb, *Fresenius Z. Anal. Chem.* **335** (1989) 748.

[43] L. Dunemann, *Nachr. Chem. Tech. Lab.* **39** (1991) no. 10, M1.

[44] M. Würfels, E. Jackwerth, M. Stoeppler, *Fresenius Z. Anal. Chem.* **329** (1987) 459.

[45] P. Schramel, S. Haase, G. Knapp, *Fresenius Z. Anal. Chem.* **326** (1987) 142.

[46] P. Schramel, G. Lill, R. Seif, *Fresenius Z. Anal. Chem.* **326** (1987) 135.

[47] B. Zunk, *Anal. Chim. Acta* **236** (1990) 337.

[48] L. B. Fischer, *Anal. Chem.* **58** (1986) 261.

[49] H. Matusiewicz, *Anal. Chem.* **71** (1999) 3145.

[50] M. Deaker, W. Maher, *J. Anal. At. Spectrom.* **14** (1999) 1193.

[51] D. Merton, J. A. C. Broekaert, R. Brandt, N. Jakubowski, *J. Anal. At. Spectrom.* **14** (1999) 1093.

[52] R. A. Romero, J. E. Tahán, A. J. Moronta, *Anal. Chim. Acta* **257** (1992) 147.

[53] M. S. Chaudhary, S. Ahmad, A. Mannan, I. H. Qureshi, *J. Radioanal. Nucl. Chem.* **83** (1984) 387.

[54] M. Feinberg, C. Ducauze, *Anal. Chem.* **52** (1980) 207.

[55] D. Huljev, B. Huljev, Z. Rajkovic-Huljev, *Radiol. Iugosl.* **22** (1988) 403.

[56] G. Kaiser, G. Tölg, *Fresenius Z. Anal. Chem.* **325** (1986) 32.

[57] P. Barth, R. Caletka, V. Krivan, *Fresenius Z. Anal. Chem.* **319** (1984) 560.

[58] S. E. Raptis, G. Kaiser, G. Tölg, *Anal. Chim. Acta* **138** (1982) 93.

[59] J. Binkowski, P. Rutkowski, *Mikrochim. Acta* **1** (1987) 245.

[60] M. Kulke, F. Umland, *Fresenius Z. Anal. Chem.* **288** (1977) 273.

[61] M. Fujita et al., *Anal. Chem.* **40** (1968) 2042.

[62] S. S. Q. Hee, J. R. Boyle, *Anal. Chem.* **60** (1988) 1033.

[63] D. A. Levaggi, W. Qyung, M. Feldstein, *J. Air. Pollut. Control Assoc.* **21** (1971) 277.

[64] G. Schwedt, L. Dunemann, *LaborPraxis*, June 1990, 476.

[65] S. E. Raptis, G. Knapp, A. P. Schalk, *Fresenius Z. Anal. Chem.* **316** (1983) 482.

[66] D. Behne, P. A. Matamba, *Fresenius Z. Anal. Chem.* **274** (1975) 195.

[67] M. Gallorini, E. Orvini, A. Rolla, M. Burdisso, *Analyst (London)* **106** (1981) 328.

[68] C. Feldmann, *Anal. Chem.* **55** (1983) 2451.

[69] H. Brügmann, T. X. Gian, H. Berge, *Acta Hydrochim. Hydrobiol.* **16** (1988) 457.

[70] M. Weidenauer, K. H. Lieser, *Fresenius Z. Anal. Chem.* **320** (1985) 550.

[71] T. M. Karadahki, F. M. Najib, F. A. Mohammed, *Talanta* **34** (1987) 995.

[72] A. Campiglio, *Mikrochim. Acta* **3** (1987) 425.

[73] P. Ostapczuk, M. Froning, M. Stoeppler, *Fresenius Z. Anal. Chem.* **334** (1989) 61.

[74] K. Gretzinger, E. Grallath, G. Tölg, *Anal. Chim. Acta* **193** (1987) 1.

[75] H. Heinrichs, H. J. Brumsack, N. Loftfield, N. Kö-nig, *Z. Pflanzenernähr. Bodenkd.* **149** (1986) 350.

[76] S. Sprung, G. Kirchner, W. Rechenberg, *Zem.-Kalk-Gips, Ed. B* **37** (1984) 513.

[77] B. Morsches, G. Tölg, *Fresenius Z. Anal. Chem.* **219** (1966) 61.

[78] P. C. Lindahl, A. M. Bishop, *Fuel* **61** (1982) 658.

[79] R. Wickbold, *Angew. Chem.* **64** (1952) 134.

[80] J. Dolezal, J. Lenz, Z. Sulcek, *Anal. Chim. Acta* **47** (1969) 517.

[81] G. Tölg, *Pure Appl. Chem.* **44** (1975) 645.

[82] J. T. Brenna, G. H. Morrison, *Microbeam Anal.* **19th** (1984) 265.

[83] W. van Delft, G. Vos, *Anal. Chim. Acta* **209** (1988) 147.

[84] L. Vos, R. van Grieken, *Anal. Chim. Acta* **164** (1984) 83.

[85] M. Bettinelli, U. Baroni, N. Pastorelli, *J. Anal. At. Spectrom.* **2** (1987) 485.

[86] F. Koroleff, Baltic Intercalibration Workshop, Kiel March 7 – 14, 1977, p. 30.

[87] K. Grasshoff, M. Ehrhardt, K. Kremling (eds.): *Methods of Seawater Analysis*, Verlag Chemie, Weinheim 1983, p. 164.

[88] L. Dunemann, M. Meinerling, unpublished results (1992).

[89] W. Rechenberg, *Zem. Kalk Gips* **25** (1972) 410.

[90] F. Seel, E. Steigner, J. Burger, *Angew. Chem.* **76** (1964) 532.

[91] J. Angerer, R. Heinrich-Ramm, G. Lehnert, *Int. J. Environ. Anal. Chem.* **35** (1989) 81.

[92] E. Schumacher-Wittkopf, J. Angerer, *Int. Arch. Occup. Environ. Health* **49** (1981) 77.

[93] R. Heinrich, J. Angerer in P. Brätter, P. Schramel (eds.): *Trace Element-Analytical Chemistry in Medicine and Biology*, vol. 2, Walter De Gruyter, Berlin 1983.

[94] A. Dornemann, H. Kleist, *Fresenius Z. Anal. Chem.* **300** (1980) 197.

[95] B. Griepink, M. Sager, G. Tölg, *Pure Appl. Chem.* **60** (1988) 1425.

[96] M. Buratti et al., *Clin. Chim. Acta* **150** (1985) 53.

[97] R. Kobayashi, K. Imaizumi, *Anal. Sci.* **5** (1985) 61.

[98] H. Watanabe, *Anal. Chem.* **53** (1981) 738.

[99] H. Berndt, U. Harms, M. Sonneborn, *Fresenius Z. Anal. Chem.* **322** (1985) 329.

[100] P. Burba, P. G. Willmer, *Fresenius Z. Anal. Chem.* **329** (1987) 539.

[101] G. Schwedt: *Methoden der Spurenanreicherung anorganischer und organischer Stoffe aus Wässern*, Vogel-Verlag, Würzburg 1988.

[102] A. Taylor, S. Branch, H. M. Crews, D. F. Halls, L. M. Owen, M. White, *J. Anal. At. Spectrom.* **12** (1997) 119R.

[103] M. Hoenig, A.-M. de Kersabiec, *Spectrochim. Acta B* **51** (1996) 1297.

[104] J. Begerow, L. Dunemann in H. Günzler (ed.): *Analytiker-Taschenbuch*, **vol. 18**, Springer-Verlag, Berlin – Heidelberg – New York 1998, p. 67.

[105] T. Mansfeld, H. Biernath, *Anal. Chim. Acta* **406** (2000) 283.

[106] D. M. Adriá-Cerezo, M. Llobat-Estellés, A. M. Mauri-Aucejo, *Talanta* **51** (2000) 531.

[107] A. Seubert, G. Petzold, J. W. McLaren, *J. Anal. At. Spectrom.* **10** (1995) 371.

[108] I. Jarvis, M. M. Totland, K. E. Jarvis, *Analyst* **122** (1997) 19.

[109] J. Begerow, M. Turfeld, L. Dunemann, *Anal. Chim. Acta* **340** (1997) 277.

[110] V. K. Panday, K. Hoppstock, J. S. Becker, H.-J. Dietze, *At. Spectrosc.* **17** (1996) 98.

[111] K. E. Laintz, C. M. Wai, C. R. Yonker, R. D. Smith, *Anal. Chem.* **64** (1992) 2875.

[112] C. M. Wai, S. Wang, J.-J. Yu, *Anal. Chem.* **68** (1996) 3516.

[113] G. Schulze, O. Elsholz, *Fresenius Z. Anal. Chem.* **335** (1989) 724.

[114] M. Agarwal, R. B. Bennett, I. G. Stump, J. M.D'Auria, *Anal. Chem.* **47** (1975) 924.

[115] M. M. Kingston, I. L. Barnes, T. J. Brady, T. C. Rains, *Anal. Chem.* **50** (1978) 2064.

[116] P. Burba, K. H. Lieser, *Fresenius Z. Anal. Chem.* **297** (1979) 374.

[117] H. J. Fischer, K. H. Lieser, *Fresenius Z. Anal. Chem.* **335** (1989) 738.

[118] A. Disam, P. Tschöpel, G. Tölg, *Fresenius Z. Anal. Chem.* **295** (1979) 97.

[119] R. Chakrovorty, R. van Grieken, *Int. J. Environ. Anal. Chem.* **11** (1982) 67.

[120] B. Andresen, B. Salbu, *Radiochem. Radioanal. Lett.* **52** (1982) 19.

[121] C. L. Smith, J. M. Motooka, W. R. Willson, *Anal. Lett.* **17** (1984) 1715.

[122] J. Bandovskis, M. Vircavs, O. Veveris, A. Pelne, *Talanta* **34** (1987) 179.

[123] J. M. Labrecque in H. M. Kingston, L. B. Jassie (eds.): *Introduction to Microwave Sample Preparation*, Am. Chem. Soc., Washington D.C. 1988, p. 203.

[124] L. Solujic, E. B. Milosavljevic, M. R. Straka, *Analyst* **124** (1999) 1255.

[125] R. Sur, J. Begerow, L. Dunemann, *Fresenius' J. Anal. Chem.* **363** (1999) 526.

[126] K. Sasaki, G. E. Pacey, *Talanta* **50** (1999) 175.

[127] C.-C.Huang, M.-H. Yang, T.-S. Shih, *Anal. Chem.* **69** (1997) 3930.

[128] U. Pichler, A. Haase, G. Knapp, *Anal. Chem.* **71** (1999) 4050.

[129] B. W. Wenclawiak, T. Hees, C. E. Zöller, H.-P. Kabus, *Fresenius' J. Anal. Chem.* **358** (1997) 471.

[130] M. Burguera, J. L. Burguera, O. M. Alarcon, *Anal. Chim. Acta* **179** (1986) 351.

[131] V. Karanassios, F. H. Li, B. Liu, E. D. Salin, *J. Anal. At. Spectrom.* **6** (1991) 457.

[132] B. Welz, M. Schubert-Jacobs in B. Welz (ed.): *5th Colloquium Atomspektrometrische Spurenanalytik*, Bodenseewerk Perkin-Elmer, Überlingen 1989, p. 327.

[133] B. Welz, M. Sperling, X. Yin in B. Welz (ed.): *6th Colloquium Atomspektrometrische Spurenanalytik*, Bodenseewerk Perkin-Elmer, Überlingen 1992, p. 203.

[134] G. Schulze, H. Tessmer, O. Elsholz in B. Welz (ed.): *5th Colloquium Atomspektrometrische Spurenanalytik*, Bodenseewerk Perkin-Elmer, Überlingen 1989, p. 347.

[135] A. Meyer, L. Dunemann in B. Welz (ed.): *6th Colloquium Atomspektrometrische Spurenanalytik*, Bodenseewerk Perkin-Elmer, Überlingen 1991, p. 115.

[136] J. A. Broekaert, *Anal. Proc. (London)* **27** (1990) 336.

[137] I. Atsuya, K. Itoh, K. Akatsuka, *Fresenius Z. Anal. Chem.* **328** (1987) 338.

[138] P. Fecher, C. Malcherek in B. Welz (ed.): *5th Colloquium Atomspektrometrische Spurenanalytik*, Bodenseewerk Perkin-Elmer, Überlingen 1989, p. 501.

[139] K. H. Tobies, W. Großmann in B. Welz (ed.): *5th Colloquium Atomspektrometrische Spurenanalytik*, Bodenseewerk Perkin-Elmer, Überlingen 1989, p. 513.

[140] A. Sugimae, R. M. Barnes, *Anal. Chem.* **58** (1986) 785.

[141] L. Dunemann, H. Hajimiragha, J. Begerow, *Fresenius' J. Anal. Chem.* **363** (1999) 466.

[142] L. Moens, T. de Smaele, R. Dams, P. Van Den Broeck, P. Sandra, *Anal. Chem.* **69** (1997) 1604.

[143] Y. Jin, F. Liang, H. Zhang, L. Zhao, Y. Huan, D. Song, *Trends Anal. Chem.* **18** (1999) 479.

[144] F. Brugnone et al., *Int. Arch. Occup. Environ. Health* **64** (1992) 179.

[145] J. Rudling, E. Bjoerkholm, B.-O. Lundmark, *Ann. Occup. Hyg.* **30** (1986) 319.

[146] P. J. H. D. Verkoelen, M. W. F. Nielen, *J. High Resolut. Chromatogr. Chromatogr. Comm.* **11** (1988) 291.

[147] M. D. Erickson: *Analytical Chemistry of PCBs*, Ann Arbor Science Publishers, Stoneham, Mass. 1986.

[148] B. Bush, J. T. Snow, S. Connor, *J. Assoc. Off. Anal. Chem.* **66** (1983) 258.

[149] R. C. Entz, H. C. Hollifield, *J. Agric. Food Chem.* **30** (1982) 84.

[150] W. Wittfoht, W. J. Scott, H. Nau, *J. Chromatogr.* **448** (1988) 433.

[151] H. Hajimiragha, U. Ewers, R. Jansen-Rosseck, A. Brockhaus, *Int. Arch. Occup. Environ. Health* **58** (1986) 141.

[152] J. Angerer, G. Scherer, K. H. Schaller, J. Müller, *Fresenius J. Anal. Chem.* **339** (1991) 740.

[153] D. G. Patterson et al., *Anal. Chem.* **58** (1986) 705.

[154] G. A. Junk, M. J. Avery, J. J. Richard, *Anal. Chem.* **60** (1988) 1347.

[155] A. Bergmann, L. Reutegardh, M. Ahlman, *J. Chromatogr.* **291** (1984) 392.

[156] B. Wigilius et al., *J. Chromatogr.* **391** (1987) 169.

[157] J. Begerow, R. Heinrich-Ramm, J. Angerer, *Fresenius Z. Anal. Chem.* **331** (1988) 818.

[158] H. C. Shields, C. J. Weschler, *J. Air Pollut. Control Assoc.* **37** (1987) 1039.

[159] J. Begerow, E. Jermann, T. Keles, L. Dunemann, *Fresenius J. Anal. Chem.* (1994), in press.

[160] R. Heinrich, J. Angerer, *Fresenius Z. Anal. Chem.* **322** (1985) 766.

[161] G. Birner, H. G. Neumann, *Arch. Toxicol.* **62** (1988) 110.

[162] K. M. Engström, *Scand. J. Work Environ. Health* **10** (1984) 75.

[163] M. Balikova, J. Kohlicek, *J. Chromatogr.* **497** (1989) 159.

[164] E. R. Adlard, C. B. Milne, P. E. Tindle, *Chromatographia* **14** (1981) 507.

[165] J. A. F. de Silva, *J. Chromatogr.* **273** (1983) 345.

[166] K. G. Wahlund, T. Arvidson, *J. Chromatogr.* **282** (1983) 527.

[167] A. M. Rustum, *J. Chromatogr. Sci.* **27** (1989) 18.

[168] S. G. Colgrove, J. H. Svec, *Anal. Chem.* **53** (1981) 1737.

[169] J. Lewalter, U. Korallus, *Int. Arch. Occup. Environ. Health* **56** (1985) 179.

[170] M. Akerblom, G. Alex, *J. Assoc. Off. Anal. Chem.* **67** (1984) 653.

[171] J. Czuczwa et al., *J. Chromatogr.* **403** (1987) 233.

[172] T. L. Peters, *Anal. Chem.* **54** (1982) 1913.

[173] W. G. Jennings, A. Rapp: *Sample Preparation for Gas Chromatographic Analysis*, Hüthig Verlag, Heidelberg 1983.

[174] D. A. J. Murray, *J. Chromatogr.* **177** (1979) 135.

[175] J. F. J. van Rensberg, A. J. Hasset, *J. High Resolut. Chromatogr. Chromatogr. Comm.* **5** (1982) 574.

[176] N. F. Wood, *Analyst (London)* **94** (1969) 399.

[177] M. Teufel et al, *Arch. Environ. Contam. Toxicol.* **19** (1990) 646.

[178] S. Hashimoto, H. Ito, M. Morita, *Chemosphere* **25** (1992) 297.

[179] R. C. Reid, J. M. Prausnitz, B. E. Poling: *The Properties of Gases and Liquids*, 4th ed., McGraw-Hill, New York 1987.

[180] J. W. King, *J. Chromatogr. Sci.* **27** (1989) 355.

[181] K. S. Nam et al.: *Proceedings of the International Symposium on Supercritical Fluids*, French Chemical Society: Paris Index, France 1988, p. 743.

[182] J. W. King, J. H. Johnson, J. P. Friedrich, *J. Agric. Food Chem.* **37** (1989) 951.

[183] M. A. Schneiderman, A. K. Sharma, K. R. R. Mahanama, D. C. Locke, *J. Assoc. Off. Anal. Chem.* **71** (1988) 815.

[184] M. A. Schneiderman, A. K. Sharma, D. C. Locke, *J. Chromatogr. Sci.* **26** (1988) 458.

[185] K. Sugiyama, M. Saito, *J. Chromatogr.* **442** (1988) 121.

[186] P. Capriel, A. Haisch, S. U. Khan, *J. Agric. Food Chem.* **34** (1986) 70.
[187] V. Janda, G. Steenbeke, P. Sandra, *J. Chromatogr.* **479** (1989) 200.
[188] S. B. Hawthorne, D. J. Miller, *J. Chromatogr. Sci.* **24** (1986) 258.
[189] B. W. Wright, C. W. Wright, J. S. Fruchter, *Energy Fuels* **3** (1989) 474.
[190] S. B. Hawthorne, D. J. Miller, *Anal. Chem.* **59** (1987) 1705.
[191] M. M. Schantz, S. N. Chesler, *J. Chromatogr.* **363** (1986) 397.
[192] B. O. Brady, C. P. C. Kao, K. M. Dooley, F. C. Knopf, *Ind. Eng. Chem. Prod. Res. Dev.* **26** (1987) 261.
[193] N. Alexandrou, J. Pawliszyn, *Anal. Chem.* **61** (1989) 2770.
[194] F. I. Onuska, K. A. Terry, *J. High Resolut. Chromatogr. Chromatogr. Comm.* **12** (1989) 357.
[195] B. W. Wright, C. W. Wright, R. W. Gale, R. D. Smith, *Anal. Chem.* **59** (1987) 38.
[196] S. B. Hawthorne, M. S. Krieger, D. J. Miller, *Anal. Chem.* **61** (1989) 736.
[197] S. Forbes, *Anal. Chim. Acta* **196** (1987) 75.
[198] U. Juergens, *J. Chromatogr.* **371** (1986) 307.
[199] J. J. Richard, G. A. Junk, *Mikrochim. Acta* 1986 no. 1, 387.
[200] G. A. Junk, J. J. Richard, *Anal. Chem.* **60** (1988) 451.
[201] R. Bagnati, E. Benfenati, E. Davoli, R. Fanelli, *Chemosphere* **17** (1988) 59.
[202] J. Manes Vinuesa, J. C. Molto Cortes, C. Igualada Canas, G. Font Perez, *J. Chromatogr.* **472** (1989) 365.
[203] J. Angerer et al., *Fresenius J. Anal. Chem.* **342** (1992) 433.
[204] K. I. Tomlins, K. Jewers, R. D. Coker, *Chromatographia* **27** (1989) 67.
[205] M. R. Khan, C. P. Ong, S. F. Y. Li, H. D. Lee, *J. Chromatogr.* **513** (1990) 360.
[206] K. E. Savolainen, *J. Pharm. Sci.* **77** (1988) 802.
[207] R. Schmidt, P. Kupferschmidt, *Clin. Chem. (Winston-Salem N.C.)* **35** (1989) 1352.
[208] V. Marko, L. Soltes, K. Radova, *J. Chromatogr. Sci.* **28** (1990) 403.
[209] V. W. Burse et al., *J. Anal. Toxicol.* **14** (1990) 143.
[210] W. Butte, C. Fooken, *Fresenius J. Anal. Chem.* **336** (1990) 511.
[211] H. Bouwman, R. M. Coopan, A. J. Reinecke, *Chemosphere* **19** (1989) 1563.
[212] P. H. Cramer et al., *Chemosphere* **20** (1990) 821.
[213] A. K. D. Liem et al., *Chemosphere* **20** (1990) 843.
[214] P. Fuerst, H. A. Meemken, W. Groebel, *Chemosphere* **15** (1986) 1977.
[215] H. Beck et al., *Chemosphere* **16** (1987) 1977.
[216] R. G. Heath et al., *Anal. Chem.* **58** (1986) 463.
[217] H. F. Walton, R. D. Rocklin: *Ion Exchange in Analytical Chemistry*, CRC Press, Boca Raton, Fla. 1990.
[218] T. G. Kreindl, H. Malissa, K. Winsauer, *Mikro-chim. Acta* 1986 no. 1, 1.
[219] W. E. Turner, S. G. Isaacs, D. G. Patterson, Jr., *Chemosphere* **25** (1992) 805.
[220] D. Barceló, G. Durand, V. Bouvot, M. Nielen, *Environ. Sci. Technol.* **27** (1993) 271.
[221] E. R. Brouwer, H. Lingemann, U. A. T. Brinkmann, *Chromatographia* **29** (1990) 415.
[222] J. Rosenfeld, M. Mureika-Russell, S. Yeroushalmi, *J. Chromatogr.* **358** (1986) 137.
[223] J. M. Rosenfeld, M. Mureika-Russell, A. Phatak, *J. Chromatogr.* **283** (1984) 127.

[224] Z. Bardodej, *Arbeitsmed. Sozialmed. Arbeitshyg.* **3** (1968) 141.
[225] J. Pawliszyn: *Solid Phase Microextraction: Theory and Practice*, Wiley-VCH, New York 1997.
[226] C. Grote, J. Pawliszyn, *Anal. Chem.* **69** (1997) 587.
[227] I. Valor, C. Cortada, J. C. Moltó, *J. High Resolut. Chromatogr.* **19** (1996) 472.
[228] T. Nilsson, F. Pelusion, L. Montanarella, B. Larsen, S. Facchetti, J. Madsen, *J. High Resolut. Chromatogr.* **18** (1995) 617.
[229] M. Möder, S. Schrader, U. Franck, P. Popp, *Fresenius' J. Anal. Chem.* **357** (1997) 326.
[230] M. T. Sng, F. K. Lee, H. A. Lakso, *J. Chromatogr. A* **759** (1997) 225.
[231] D. W. Potter, J. Pawliszyn, *Environ. Sci. Technol.* **28** (1997) 298.
[232] P. A. Martos, J. Pawliszyn, *Anal. Chem.* **69** (1997) 206.
[233] F. Mangani, R. Cenciarini, *Chromatographia* **41** (1995) 678.
[234] V. Pino, J. H. Ayala, A. M. Afonso, V. Gonzáles, *J. Chromatogr. A* **869** (2000) 515.
[235] V. Pensado, C. Casais, C. Mejuto, R. Cela, *J. Chromatogr. A* **869** (2000) 505.
[236] J. Angerer: "Ameisensäure" in D. Henschler (ed.): *Analysen in biologischem Material*, 4. ed., Verlag Chemie, Weinheim 1980.
[237] A. B. Heath, R. R. Black, *J. Assoc. Off. Anal. Chem.* **70** (1987) 862.
[238] R. L. Brown, C. L. Farmer, R. G. Millar, *J. Assoc. Off. Anal. Chem.* **70** (1987) 442.
[239] R. A. Minear, L. H. Keith (eds.): *Water Analysis*, **vol. 3**, Academic Press, Orlando 1984, p. 84.
[240] T. N. Eisenberg, E. J. Middlebrooks: *Reverse Osmosis Treatment of Drinking Water*, Butterworths, London 1986.
[241] T. D. Brook: *Membrane Filtration: A User's Guide and Reference Manual*, Science Tech., Madison 1983.
[242] M. M. L. Aerts, W. M. J. Beek, U. A. T. Brinkmann, *J. Chromatogr.* **435** (1988) 613.
[243] D. C. Turnell, J. D. H. Cooper, *J. Chromatogr.* **395** (1987) 613.
[244] A. C. Metha, *Tr. AC Trends Anal. Chem. (Pers. Ed.)* **8** (1989) 107.
[245] K. Pekari, M. L. Riekkola, A. Aitio, *J. Chromatogr.* **491** (1989) 309.
[246] G. Machata, J. Angerer: "Gaschromatographie, Headspace-Technik" in D. Henschler (ed.): *Analysen in biologischem Material*, 7th suppl., Verlag Chemie, Weinheim 1983.
[247] V. C. Stein, R. S. Narang, *Arch. Environ. Contam. Toxicol.* **19** (1990) 593.
[248] S. L. Friant, I. H. Suffet, *Anal. Chem.* **51** (1979) 2167.
[249] L. Dunemann, H. Hajimiragha, *Anal. Chim. Acta* **283** (1993) 199.
[250] D. L. Ashley et al., *Anal. Chem.* **64** (1992) 1021.
[251] H. Hajimiragha, U. Ewers, A. Brockhaus, A. Boettger, *Int. Arch. Occup. Environ. Health* **61** (1989) 513.
[252] K. H. Pannwitz, *Drägerheft* **332** (1985) 10.
[253] K. H. Pannwitz, *Drägerheft* **327** (1983) 6.
[254] G. A. Reineccius, R. Liardon in R. G. Berger, S. Nitz, P. Schreier (eds.): "Topics in Flavor Research," Proceedings of *International Conference Freising-Weihenstephan*, 1985, p. 125.
[255] K. H. Schaller: "Mandelsäure" in: D. Henschler (ed.): *Analysen in biologischem Material*, 2nd suppl., Verlag Chemie, Weinheim 1978.

[256] J. Begerow, J. Angerer, *Fresenius J. Anal. Chem.* **366** (1990) 42.

[257] M. I. Daneshvar, J. B. Brooks, *J. Chromatogr.* **433** (1988) 248.

[258] H. B. Lee, T. E. Peart, J. M. Carron, *J. Chromatogr.* **498** (1990) 367.

[259] S. Jacobbson, A. Larsson, A. Arbin, A. Hagman, *J. Chromatogr.* **358** (1986) 137.

[260] R. v. Smith, S. Tsai, *J. Chromatogr.* **61** (1971) 29.

[261] P. Pfäffli, H. Savolainen, H. Keskinen, *Chromatographia* **27** (1989) 483.

[262] G. Johanson, H. Kronborg, P. Näslund, M. Byfalt Nordquist, *Scand. J. Work Environ. Health* **12** (1986) 594.

[263] Y. L. Emeillat, J. Menez, F. Berthou, L. Bardou, *J. Chromatogr.* **206** (1981) 89.

[264] R. Patience, J. Thomas, *J. Chromatogr.* **234** (1982) 225.

[265] E. Grushka, H. Durst, E. Kikta, *J. Chromatogr.* **112** (1975) 673.

[266] T. Jupile, *J. Chromatogr. Sci.* **17** (1979) 160.

[267] W. Steiner, E. Müller, D. Fröhlich, R. Battaglia, *Mitt. Geb. Lebensmitteluntersuch. Hyg.* **75** (1984) 37.

[268] M. Petrarulo, S. Pellegrino, M. Marangella, F. Linari, *J. Chromatogr.* **432** (1988) 37.

[269] M. Petrarulo et al., *J. Chromatogr.* **465** (1989) 87.

[270] W. Dünges, *Anal. Chem.* **49** (1977) 442.

[271] W. Dünges, *Chromatographia* **9** (1976) 624.

[272] E. Grushka, S. Lam, J. Chassin, *Anal. Chem.* **50** (1978) 1398.

[273] A. Crozier, J. Zaerr, R. Morris, *J. Chromatogr.* **238** (1982) 157.

[274] H. Tsuchiya, T. Hayastsi, H. Naruse, N. Takagi, *J. Chromatogr.* **234** (1982).

[275] H. D. Winkeler, K. Levsen, *Fresenius Z. Anal. Chem.* **333** (1989) 716.

[276] M. A. Carey, H. E. Perisinger, *J. Chromatogr. Sci.* **10** (1972) 537.

[277] J. K. Lin, J. Y. Yang, *Anal. Chem.* **47** (1975) 1634.

[278] J. Y. Chang, H. L. Creaser, *J. Chromatogr.* **116** (1976) 215.

[279] J. K. Lin, C. C. Lai, *Anal. Chem.* **52** (1980) 630.

[280] K. Karlson, M. Alsandro, M. Novotny, *Anal. Chem.* **57** (1985) 229.

[281] K. Grob, Jr., D. Fröhlich, B. Schilling, *J. Chromatogr.* **295** (1984) 55.

[282] K. Grob, Jr., C. Walder, B. Schilling, *J. High Resolut. Chromatogr. Chromatogr. Comm.* **9** (1986) 95.

[283] K. Grob, Jr., B. Schilling, *J. High Resolut. Chromatogr. Chromatogr. Comm.* **8** (1985) 726.

[284] K. Grob, Jr., G. Karrer, M. L. Riekkola, *J. Chromatogr.* **334** (1985) 129.

[285] H. J. Cortes, E. L. Olberding, J. H. Wetters, *Anal. Chim. Acta* **236** (1990) 173.

[286] R. Barcaralo, *J. High Resolut. Chromatogr. Chromatogr. Comm.* **518** (1990) 465.

[287] B. A. Anderson, *Am. Environ. Lab.* **1** (1989) 41.

[288] D. Barcelo et al., *Anal. Chem.* **62** (1990) 1696.

[289] S. A. Soper, K. L. Ratzlaff, T. Kuwana, *Anal. Chem.* **62** (1990) 1438.

[290] J. Strojek, S. A. Soper, K. L. Ratzlaff, T. Kuwana, *Anal. Sci.* **6** (1990) 121.

[291] S. B. Hawthorne, D. J. Miller, J. J. Langefeld, *J. Chromatogr. Sci.* **28** (1990) 2.

[292] J. R. Wheeler, M. E. McNally, *J. Chromatogr. Sci.* **27** (1989) 534.

[293] F. I. Onuska, K. A. Tery, *J. High Resolut. Chromatogr. Chromatogr. Commun.* **12** (1989) 527.

[294] M. R. Andersen, J. T. Swanson, N. C. Porter, B. E. Richter, *J. Chromatogr. Sci.* **27** (1989) 371.

[295] C. H. Marvin et al., *J. Chromatogr.* **518** (1990) 242.

7. Trace Analysis

HELMUT MÜLLER, Universität Halle-Wittenberg, Fachbereich Chemie, Merseburg, Federal Republic of Germany

HEINZ W. ZWANZIGER, Fachhochschule Merseburg, Fachbereich Chemie- und Umweltingenieurwesen, Merseburg, Federal Republic of Germany

JOHANNES FLACHOWSKY, Umweltforschungszentrum, Leipzig-Halle GmbH, Leipzig, Federal Republic of Germany

Symbols

1) Symbols of features (quantities, variables):
X — analyte feature (e.g., concentration)
Y — signal, response

2) Symbols of measured or calculated values of features:
x — concentration value
x_A — analytical value; analyte concentration
x_C — calibration value
$x_{decision}$ — limit of decision
$x_{detection}$ — limit of detection
$x_{quantitation}$ — limit of quantitation
y — signal or response value
y_{blank} — blank value
y_{crit} — critical value

3) Symbols for running indices:
b — with respect to blank values
c — with respect to calibration
n — general running index

4) Symbols for terminal values of indices:
N — number of ...

N_a — ... analytical values
N_b — ... blank values
N_c — ... calibration values

5) Symbols of statistical features and parameters:
symbols with a bar: mean values
symbols with a hat: calculated values according to a model
a_0 — intercept in regression models
a_1 — slope; sensitivity; regression coefficient
k — statistical factor (quantile of a distribution to be specified)
q — order of quantile
res — residual value (difference of measured and calculated values)
s — calculated ("sample") standard deviation
s^2 — calculated ("sample") variance
sgn (cor_{XY}) — sign of the correlation coefficient of X and Y
SS — read as "sum of squares" ...
SS_{XX} ... of centered values of feature X
SS_{YY} ... of centered values of feature Y
SS_{XY} — read as "sum of products of centered values of features X and Y
t, $t_q(v)$ — values of Student's distribution (quantiles)
α — error probability (risk of first kind)

β	error probability (risk of second kind)
v	degrees of freedom
σ	"population" standard deviation

7.1. Subject and Scope

The determination of analytes in very low concentration (traces) plays a fundamental role in many areas of science and technology.

A "trace" means a very small level of a substance that is present alongside a large excess of other components, which causes certain effects and must be detected (with qualitative information) and determined (with quantitative information).

A distinction is made between organic and inorganic "traces." Examples of *organic traces* are dioxins or furans in waste gases from refuse incineration plants, fluorochlorohydrocarbons (FCHCs) in the atmosphere, chlorinated hydrocarbons (CHCs) in water, and many others. *Inorganic traces* are referred to as trace elements, a concept that includes all elements (i.e., metals, semimetals, and nonmetals). Despite the different objects that must be monitored analytically, the methods and instruments in this field have so much in common that a special branch of analysis — trace analysis — has developed [1], [2].

For trace analysis, analytical methods with extreme powers of detection are required, which must often be coupled with enrichment steps. Another requirement follows from the fact that the traces occur alongside a large excess of other substances (matrix). The main components are present in $10^2 - 10^{10}$-fold excess. Therefore, methods must be developed that permit trace determinations without interference by the main component (i.e., specific methods); otherwise, separation of the trace and main components must be carried out before the determination.

The possibility of determining traces of elements arose in the 1920s with the development of spectrophotometry, spectroscopic methods, and polarography. Trace analysis was carried out very intensively and with great sophistication in connection with the development and production of the first atomic bomb. Later, it received an important stimulus from the development of novel materials, especially semiconductors.

Today, more areas than ever depend on the results of trace analysis: Nuclear energy, production of semiconductors and ultrapure substances, metallurgy, materials research and production, geology, mineralogy, oceanography, medicine, animal and plant physiology, criminology, ecology, and environmental research.

Currently, ecological research and routine analysis relevant to the environment provide trace analysis with imporant stimuli. At the same time, the centers of attention are shifting more and more from element trace analysis to organic trace analysis.

7.2. Fields of Work

Modern technology has many possibilities at its disposal for solving problems in trace analysis. A uniform characterization and assessment of trace-analytical methods requires the clearest possible definitions: For reporting a content G, mass relations — of which the simplest is percentage by mass — are unequivocal and independent of any additional information. The contents of the components to be determined can vary within wide limits. Very diverse proposals for the boundaries of these ranges are found. For practical purposes, the following classification has been established:

Main fraction	$100\% - 10\%$
Minor fraction	$10\% - 0.1\%$
Traces	less than 0.1%
Micro fraction range	$1000 - 1$ µg/g
Nano fraction range	$1000 - 1$ ng/g

For nanofractions (contents below ng/g) the term "extreme trace analysis" has become generally accepted.

In addition to the specification of content ranges in parts by mass w, parts by volume σ, or parts by amount of substance x, in trace analysis the term ppm (parts per million; $1 \text{ ppm} = 1:10^6 = 10^{-4}\%$) is generally used for parts by mass w. Still lower concentrations are expressed in ppb (parts per billion); errors can arise because the intended reference quantity 10^9 ($1 \text{ ppb} = 10^{-7}\%$) is called "billion" mainly in the United States, but "milliard" in Europe. This quantity is sometimes represented as ppM (parts per milliard).

In the literature, ppm by mass is still encountered occasionally, and the ppm term is then used without further identification. Unfortunately, the terms ppm and ppb, which preferably should be used only for problems in solid analysis, are also used, instead of mass concentration β (or C) in milligrams or micrograms per liter, for specifying

Table 1. Specification of content units in trace analysis

Unit	Designation	Percentage
Parts by mass w		
mg/kg	1 ppm	10^{-4}
µg/kg	1 ppb	10^{-7}
ng/kg	1 ppt *	10^{-10}
pg/kg	1 ppq **	10^{-13}
Part by volume σ		
mL/m^3	ppm	10^{-4}
µL/m^3	ppb	10^{-7}
nL/m^3	ppt *	10^{-10}
pL/m^3	ppq **	10^{-13}

* ppt = Parts per trillion.
** ppq = Parts per quadrillion.

the concentration of trace contents in dissolved samples.

The content of trace components in gases is indicated as volume concentration σ or mass concentration β. Useful units are milliliters per cubic meter (ppm) and milligrams per cubic meter. The conversion is carried out according to

$$\sigma\left[mL/m^3\right] = \frac{\text{molar volume [L/mol]}}{\text{molar mass [g/mol]}} \cdot \beta\left[mg/m^3\right]$$

The molar volume is usually related to 25 °C and 101.3 kPa and is 22.47 L/mol. For dusts, the mass concentration β (mg/m^3) or the particle concentration C (particles/m^3 or fibers/m^3) is used.

Table 1 surveys the specification of content data in trace analysis. Limiting factors for the optimum choice of an analytical method for a trace-analytical problem are the portion of sample available and the expected mass concentration β of the analyte.

The sample portions or sample volumes most frequently used are in the range of a minimum of grams or milliliters to micrograms or microliters. Modern methods (e.g., capillary zone electrophoresis, CZE) require sample volumes only in the nanoliter range. This results in a high *mass sensitivity* that leads into the range of detection of individual molecules or atoms. By coupling capillary zone electrophoresis with laser-induced fluorescence detection, an absolute total of 600 molecules has even been detected.

Clearly distinguished from this is the "*concentration sensitivity*" of the individual methods. The current state of the art is ppm analysis of a wide range of analytes in a variety of matrices. The trend to ppb and in a few cases ppt (10^{-10}%) (parts per trillion, 10^{-12}) without enrichment is expected.

7.3. Methods of Modern Trace Analysis

Modern trace analysis is characterized by two different strategies, whose orientation is more physical or chemical, respectively:

1) Application of instrumental direct methods
2) Development of chemical-analytical compound methods

Instrumental Direct Methods. In instrumental direct methods, after sampling and possibly sample preparation, the sample is analyzed directly. Instrumental direct determination methods are usually matrix-dependent relative methods. A mathematical correction of the matrix effect is possible only in particular cases (e.g., X-ray fluorescence analysis, XRF). To compensate for systematic errors, therefore, standard reference materials are required, which must be very similar in composition to the sample to be analyzed (see Section 7.4.5). However, an optimum power of detection at high accuracy (freedom from systematic error) can be achieved only when the trace to be determined is present in isolated form in the highest possible mass concentration. In many cases of trace analysis, therefore, more complex multistage (compound) methods are required.

Compound Methods. In compound methods, after dissolution or digestion of the sample, the analyte is separated from the matrix, usually enriched, and then determined (Fig. 1). The calibration is relatively simple because the analyte components have been isolated. The only standards required are analyte solutions of known concentration, since the matrix can no longer interfere with the determination. The advantage of easy calibration is lessened, however, by the fairly large number of operations and the correspondingly greater risk of systematic errors (Fig. 1). This requires very careful and complicated optimization of compound methods to reduce such errors, which today are a central problem of extreme trace analysis. Compound methods are therefore often carried out in closed systems (digestion, enrichment, and/or separation in one apparatus) to prevent errors in the determination of trace contents resulting from the sample environment and sample preparation steps. The necessary trace-analytical hygiene is as important for samples as for standards, solvents, reagents, vessels, instruments, and the laboratory atmosphere. If the highest purity is required, particle emission by laboratory person-
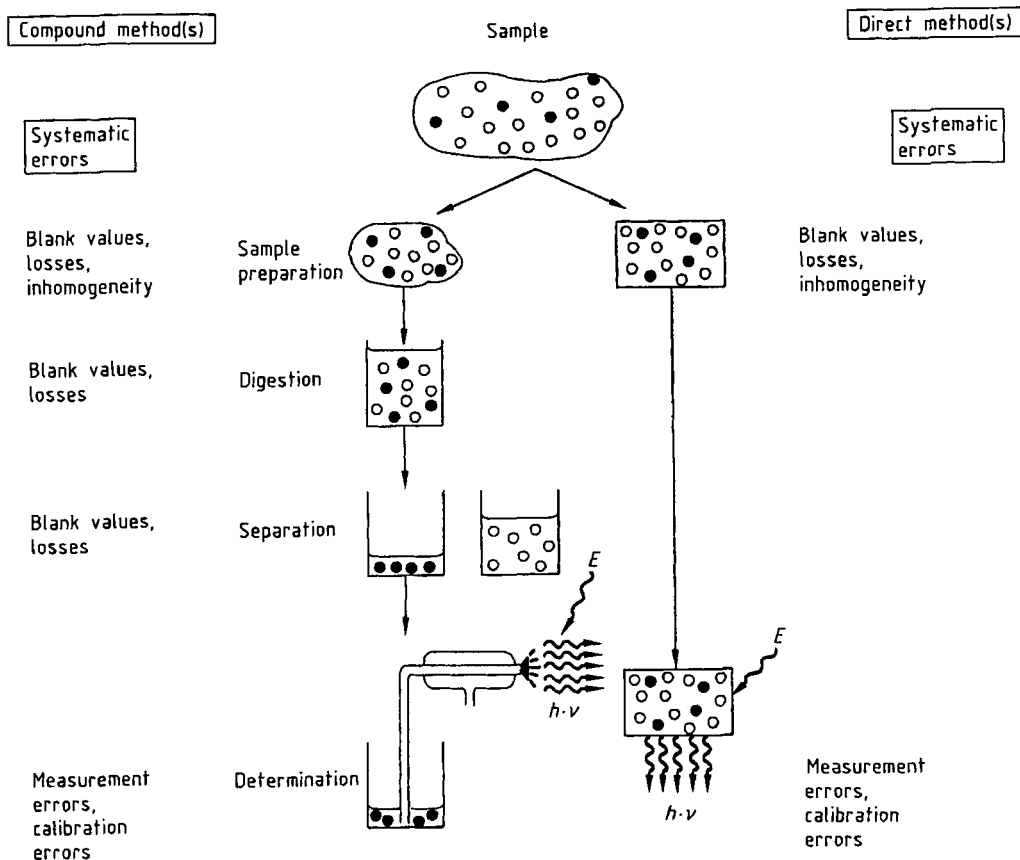
Figure 1. Systematic errors in direct and compound methods

nel must also be considered. Work is therefore carried out in clean rooms or clean boxes, through which laminar flow of filtered air occurs.

New operating techniques (e.g., flow methods), in combination with modern instrumental methods, enable the separation and enrichment processes to be automated in quasi-closed systems, and have contributed substantially to the increasing establishment of such composite methods in trace-analytical practice [3].

Despite the recognizable advantages of compound methods, work is increasingly being carried out on the (further) development of multielement methods with high powers of detection, especially for micro samples. Table 2 shows a survey of modern instrumental multianalyte methods.

Particularly for determining organic trace contents, analytical methods are used that combine systems with different measurement principles and therefore different selectivity (coupled or hy-

phenated techniques). Chromatographic methods, particularly gas chromatography, supercritical fluid chromatography (SFC), high-performance liquid chromatography, and capillary electrophoresis (CE) are often coupled with spectroscopic methods [7], [8]. The most important coupled systems in trace and environmental analysis are the following:

GC	– MS, FTIR, AES/AAS, FTIR – MS, ICP – MS,
SFC	– MS, FTIR, AES/AAS, FTIR – MS, ICP – MS,
HPLC	– MS, FTIR, AES/AAS,
CE	– MS, UV

Coupling of Chromatographic Methods
HPLC – GC, HPLC – GC – MS, SFC – GC – MS, HPLC – TLC – FTIR

The advantages of such coupled techniques lie in a frequently effective separation of the matrix (improvement of the accuracy), an increased

Table 2. Multianalyte methods [4]–[6]

Method	Concentration range, ppb	Matrix effects [a]	Simultaneous	Sequential
Mass spectrometry (MS)	1 – 10	++	+	
Spark ion source MS (SSMS)				
Secondary ion MS (SIMS)				
Inductively coupled plasma MS (ICP – MS)				
Glow discharge source MS (GD – MS)				
Laser-induced MS (LMS)				
Neutron activation analysis (NAA)	1 – 10	(+)	+	
X-ray fluorescence analysis (XRF)	10^b – 1000	++	+	
Total reflection X-ray fluorescence (TRXRF)	10 – 100	++	+	
Atomic emission spectrometry (AES)	10 – 100	(+)		
Inductively coupled plasma AES (AES – ICP)	1 – 10	+++	+	+
Atomic fluorescence spectrometry (AFS)	100 – 1000	++		+
Gas chromatography (GC)	10 – 100 [c]		+	
Capillary GC	1 – 10 [c]		+	
High-performance liquid chromatography (HPLC)	10 – 100 [c]		+	
Capillary zone electrophoresis (CZE)	100 – 1000		+	

[a] + = low … +++ = high.
[b] Wavelength-dispersive XRF (WDXRF).
[c] Dependent on detector principle.

power of detection, and a contribution to speciation analysis [9].

7.4. Calibration and Validation (see also → Chemometrics)

7.4.1. Conceptual Problems

A fully defined trace-analytical procedure is characterized by the concentration range, as well as the calibration function, the precision and accuracy, and the limits of decision, detection, and possibly quantitation. The use of the three last-mentioned terms in the literature unfortunately is not uniform and requires international standardization.

The *limit of decision* is an important criterion in the optimization of a trace-analytical procedure and the comparison of methods. It depends greatly on possible errors in individual steps. It can be determined in many different ways, which may vary in their complexity [10], [11] and whose general applicability on national level is still being elaborated [11], [12]. The precise specification of the means of calculating the limit of decision should therefore be a part of the analytical instructions. Determination of the limit of decision and determination of an analyte are based, except in the case of absolute methods (gravimetry, coulometry), not only on the actual course of the

analysis but also essentially on the results of the calibration process.

7.4.2. Errors

Like any analytical procedure, trace analysis is subject to sources of error that can lead to systematic and random falsification of the observed values or test results. The reliability of results is therefore determined by the accuracy and precision. Measures of the precision are repeatability and reproducibility.

Observed values or results of measurement means a set $\{y_1, y_2, \ldots, y_N\}$ of N recorded values y_n (with $n = 1, 2, \ldots, N$) of a signal quantity Y (e.g., absorbance) that result from N given values x_n ($n = 1, 2, \ldots, N$) of a concentration quantity X or from which an analytical result x_A can be derived (calculated).

Systematic errors lead to one-sided displacement of results. Two cases exist. Case A: If the trace component to be determined is absent from a sample ($x_A = 0$), its presence can be simulated (e.g., if the value y_0 of the signal quantity Y specific to the analyte exceeds a critical level y_{crit}). For case B, in the presence of the analyte ($x_A > 0$), two extreme situations are conceivable: The "complete" loss of analyte (which leads to $y_A < y_{crit}$), or an analyte content that is no longer in the trace range ($y_A \gg y_{crit}$).

The risk of a wrong decision (in case A to accept the presence of the analyte, or in case B
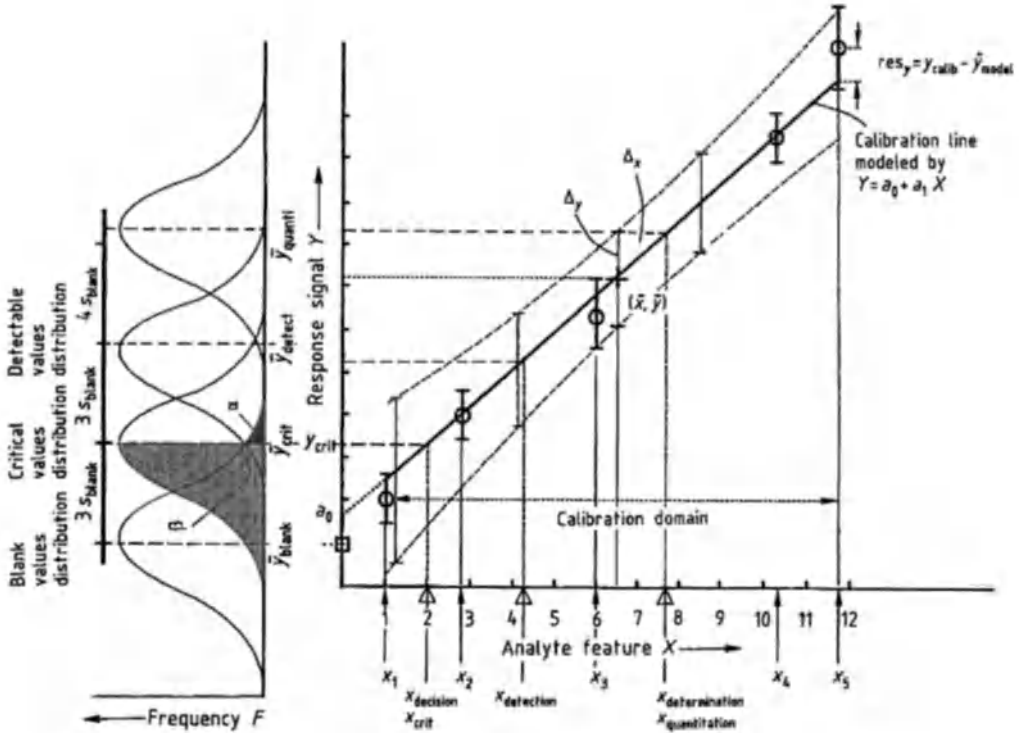
Figure 2. Illustration of problems in simultaneous consideration of the signal and calibration domains

to exclude its presence) can have quite different importance in the two cases. This is not discussed further here, since systematic sources of error can in principle be recognized and eliminated (see Section 7.4.4).

Unavoidable *random errors* cause a two-sided, random deviation of individual observed values from a particular value (target value, mean, median) and also result in different risks in the assessment of the signal or the reporting of analytical results.

7.4.3. The Critical Signal Value and Limits of the Procedure

The critical signal value y_{crit} plays a central part in determining the characteristics of the procedure. Via an experimentally supported or calculated calibration relation (e.g., $Y = a_0 + a_1 X$) it leads to the limit of decision $x_{crit} = x_{decision}$ (see Fig. 2).

The critical signal value is determined most simply from repeated analysis of several blank

analytical portions or from repeated measurement of one blank analytical portion, the latter, however, provides less real characteristic quantities. Also, in the case of an absent analyte ($x = 0$), a total of N_b randomly determined signal values y_b ($b = 1$, ..., N_b), are measured, which must not be zero and whose important characteristics can be calculated easily.

Mean
$$\bar{y}_{blank} = \frac{1}{N_b} \sum_{b=1}^{N_b} y_b$$

Variance
$$s_{blank}^2 = \frac{1}{N_b - 1} \sum_{b=1}^{N_b} (y_b - \bar{y}_{blank})^2$$

Standard deviation
$$s_{blank} = \sqrt{s_{blank}^2}$$

The upper limit of the confidence interval of the mean of the blank value is then regarded as the critical value y_{crit} of the response value.

$$y_{crit} = \bar{y}_{blank} + s_{blank} t_q(v) \sqrt{\frac{1}{N_b}}$$

It can be calculated by means of a tabulated value for Student's t distribution with $v = N_b - 1$ degrees of freedom and order $q = 1 - \alpha$ (one-sided values) ($\rightarrow$ Chemometrics).

The notation for the t values is according to [13]. This notation assumes a symmetrical distribution of the observed value signals. However, for trace determinations, this idea, and particularly that of normally distributed values, are frequently unjustified. A logarithmic transformation may be necessary [14], [15].

If a signal value of any test portion falls above this critical value, one can assume, with a certain risk of error α, usually set at 5 % ($\alpha = 0.05$), that the signal value is not a blank value but results from the presence of analyte. At this point the qualitative decision about the presence of an analyte is made with respect to the signal.

The corresponding critical concentration $x_{crit} = x_{decision}$, the *limit of decision*, is obtained by combining the calibration relation with the expression for y_{crit}, by assuming that $\bar{y}_{blank} = a_0$.

$$x_{decision} = \frac{s_{blank}}{a_1} t_q(v) \sqrt{\frac{1}{N_b}}$$

Optimization of the limit of decision is thus possible by reducing the standard deviation of the blank value, increasing the sensitivity a_1, and increasing the number N_b of blank value measurements. The numerical value of the limit of decision can also be improved by taking future analytical steps into account (i.e., the number N_a of analyses of the analytical portion). If the standard deviation is assumed to be the same for the actual analysis as for the blank value measurements, then

$$x_{decision} = \frac{s_{blank}}{a_1} t_q(v) \sqrt{\frac{1}{N_b} + \frac{1}{N_a}}$$

A more sophisticated way to obtain the critical signal value is to calculate the upper confidence limit of the intercept a_0 of the calibration relation that was derived from N_c measurements of analytical calibration portions [11]:

$$y_{crit} = a_0 + s_{res} t_q(v = N_c - 2) \sqrt{\frac{1}{N_c} + \frac{1}{N_a} + \frac{\bar{x}_c^2}{SS_{XX}}}$$

(The residual standard deviation s_{res} of the calibration model and the sum of the squares of deviations SS_{XX} are defined in Section 7.4.4; $\bar{x}_c^2$ is the mean of the concentration in the calibration step.)

By inserting this expression in the calibration model, the limit of decision is again obtained. This limit could be lowered if the mean of the calibration range lay as close as possible to the limit of decision. In fact, the highest concentration value should be lower than $10 \times x_{decision}$. Figure 2 shows that the statistical distribution of observed values near the limit of decision provides values that fall with 50 % probability within the spread of the mean blank value. In other words, at this point the risk of wrongly accepting a blank signal as an analyte signal (an error of the first kind with a probability of α) is very low, but the risk of still rejecting an analyte signal as a blank value signal (an error of the second kind with a probability of β) is very high. An analyte signal is thus reliably detected when the last-mentioned risk is as small as possible. For $\alpha = \beta$, therefore, the *limit of detection* is twice the limit of decision ($\rightarrow$ Chemometrics).

As can be seen in Figure 2, for a particular observed value $y_{quantitation}$ an acceptable analyte interval $\pm \Delta x$ (for $x_{quantitation}$) can be indicated only if $\Delta x \approx s_{res}/a_1$ is not too large. This quotient is a critical quantity for the practical application of an analytical method. In addition, the relative standard deviation of an analyte result must sometimes not exceed a specified maximum. These two requirements lead to the definition or pragmatic setting of so-called limits of *quantitation*. Since such limits depend on the particular value observed, they cannot be regarded as meaningful characteristics of the method.

Historically, the simple $k\sigma$ criteria used in Figure 2 are based on numerous simplifications. They are usually based on the standard deviation σ of blank value measurements: The limit of decision is defined with $k = 3$, the limit of detection with $k = 6$, and as one possibility, the limit of quantitation with $k = 10$ [16]. (The k values take into consideration the probability α of erroneous statistical and, therefore, erroneous analytical decisions. Thus, by fixing k or α, the purpose of the particular trace-analytical procedure can be taken into account.) The basic considerations in such definitions of method limits extend back to H. KAISER [14] and G. EHRLICH [17]. KAISER also tried to define characteristic quantities of methods for multicomponent analysis. For reasons of space, simultaneous multicomponent analysis cannot be discussed here [18], [19]. The previous discussion reveals how delicate results of trace analyses are in general. To achieve a responsible discussion in public it should be at least reported together with

confidence limits ($\rightarrow$ Chemometrics) and single values (means) should never be compared to critical target values.

7.4.4. Adequate Calibration Models

Calibration means the recording of observed values of response feature Y (signal quantity) typical of the method as a function of the analyte feature X (concentration or content) from N_c solid, liquid, or gaseous test portions. If possible, the analytical calibration portions should originate from different standard reference materials (SRMs), whose residual composition (matrix) corresponds largely to that of the expected analytical portions. If only one standard material or one standard portion is available, the analytical calibration portions should be prepared by spiking a blank test sample (standard addition) rather than by dilution of the standard (see Section 7.4.5). Here, analytical calibration portions for which the sampling error is negligible are assumed. (For nomenclature of sampling steps and of the aliquots arising from sample preparation steps, see [20].)

The objective of the calibration is to set up and statistically ensure the mathematical model of the functional dependence of the signal quantity Y on analyte content X. Linear models of the type

$$Y = a_0 + a_1 X$$

are most widely used.

Calibration in the narrower sense means determination of the sensitivity a_1. If N_c pairs of values (x_n, y_n) are available ($n = 1, 2, \ldots, N_c$), from which the arithmetic means ($\bar{x}, \bar{y}$) have been calculated, the sensitivity a_1 (regression coefficient, slope) can be calculated easily with the sum of squares of deviations from $n = 1$ to N_c, $SS_{XX} = (x_n - \bar{x})^2$ and $SS_{YY} = (y_n - \bar{y})^2$ and the sum of products of deviations $SS_{XY} = (x_n - \bar{x})(y_n - \bar{y})$ where $a_1 = SS_{XY}/SS_{XX}$.

Since all straight lines of calibration pass through the center of gravity of the data ($\bar{x}, \bar{y}$), the constant a_0 is obtained from the model equation.

Strictly speaking, a_0 and a_1 are estimates of the true but unknown characteristic quantities of the calibration model, which represents an unknown relationship between Y and X. For that reason, $\hat{a}_0$ and $\hat{a}_1$ are used. The $\hat{y}_n$ calculated by this model for every given x_n generally differs from the measured y_n by the residual $res_n = y_n - \hat{y}_n$.

This model is based on several assumptions:

1) The quantity X can be set at points x_n without error
2) The variance of Y (at points x_n) is constant in the calibration range ("homoscedastic", which can be tested with the Cochran test [13], [21]; if the assumption is not confirmed, weighted regression models should be used)
3) The quantity Y (at the points x_n) shows a normal distribution (see Fig. 2, left) and is outlier-free (the latter situation is tested with the Dixon test [13], [21]; if the assumption is not confirmed, a robust regression method can be used, for example)
4) The residuals res_n (see Figure 2) are normally distributed and do not correlate with the values of x_n (which can be tested qualitatively by simply plotting res_n against x_n; if the assumption is not met, a nonlinear analyte–signal relation rather than linear may be valid)

In general, a linear model is naturally assumed to be justified. If tests [for coefficients of determination, goodness of fit (GOF) and lack of fit (LOF); see below] show this to be improbable, a nonlinear calibration must be selected, in the simplest case by using quasi-linear models (consideration of higher powers of X; polynomials), or piecewise linear calibration can be performed [22]. (A survey of various basic regression models can be found in [23]).

In trace-analytical methods the nonconfirmation of item 1 above must be reckoned with. The characteristic quantities a_0 and a_1 can then still be calculated within the framework of a linear model:

1) $a_1 = \text{sgn}\,(\text{cor}_{XY})\,\sqrt{SS_{YY}/SS_{XX}}$, if the error ratio (corresponding to the long-term standard deviations) of the two response features Y and X is equal to the slope.
2) $a_1 = \frac{d}{2} + \text{sgn}\,(\text{cor}_{XY})\,\sqrt{\left(\frac{d}{2}\right)^2 + 1}$ where $d = \frac{SS_{YY} - SS_{XX}}{S_{XY}}$ if the errors of Y and X are equal (orthogonal regression)
3) $a_1 = SS_{XY}/(SS_{XX} - N_c \cdot \sigma^2_X)$, or
4) $a_1 = (SS_{YY} - N_c \cdot \sigma^2_Y)/SS_{XY}$, if the (long-term) variances σ^2 of the response features are known from previous experiments. For more details and alternatives see [24].

The characteristic quantities obtained for the straight lines must be tested for their statistical significance and should be quoted with a confidence interval. Their standard deviation is a function of the residual standard deviation s_{res} of the regression model. This is defined as:

$$s_{\text{res}} = \sqrt{\frac{1}{N_c - 2} \sum_{n=1}^{N_c} \text{res}_n^2}$$

and also determines the width of the confidence band around the straight line of calibration.

If at the points x_n of the calibration range, only single measurements y_n are available, the quality of the calibration model can be evaluated only as a whole (sum of systematic model errors and random experimental errors) usually by testing the goodness of fit and by testing the coefficient of determination.

Only if replicate measurements of y_n have been made for at least one point x_n model errors and experimental errors can be weighed against each other by LOF and complete analysis of variances, ANOVA ($\rightarrow$ Chemometrics). Replicate measurements also reduce the width of the confidence band.

Modeling, tests, calculation of the characteristic quantities of the method, and graphical visualization are supported by computer programs (e.g., [25]), which however usually consider only certain aspects of the calibration. An alternative method of calibration for imperfect response features and concentration quantities is to apply fuzzy theory [26].

7.4.5. Quality Assurance and Standard Reference Materials

Perhaps the greatest problem in trace analysis is assurance of the accuracy of the results (i.e., the avoidance of systematic errors). Systematic sources of error are possible in every step of an analytical process. The most reliable method for detecting systematic errors is continuous and comprehensive *quality assurance*, particularly by occasional analysis of (certified) standard reference materials. Strictly speaking, an analytical method cannot be calibrated if suitable (i.e., representative) standard reference materials adequately representing the matrix of the expected test samples are not available. However, *internal laboratory reference materials* can then usually be prepared, whose matrix largely resembles the matrix of the test portions expected. If problems occur in the preparation of such reference samples, the standard addition method (SAM) can be applied, in which internal laboratory standards are added stepwise to the test sample (analyte and matrix)

and the sum of the two analyte contributions (from sample and standard) is determined.

In these cases, however, the danger exists that the speciation of the analyte in the added standard and that in the real sample are not identical.

Internal quality assurance can also be carried out by comparison of methods. In this case, two methods or procedures that are completely different in analytical principle should be applied to the same analytical problem. Comparison of one's own method with a certified method, is most useful. If the results of the two methods are plotted against each other in a scattergram, in the ideal case all values fall on a straight line with slope $a_1 = 1$ and intercept $a_0 = 0$. Here again, depending on the ratio of the random errors of the methods being compared, the characteristic quantities of the straight line corresponding to the actual data can be calculated as shown above, and interpreted and tested (statistically assured) as systematic additive (a_0) and multiplicative ($a_1 \cdot X$) error contributions.

External quality assurance, usually by participation in interlaboratory (round robin) tests, also contributes to the detection and avoidance of systematic errors.

However, the mere use of standard reference materials offers no guarantee of accurate, error-free calibration. Thus, many solid SRMs are useless for the calibration of solid microanalyses if the danger of microinhomogeneities is not recognized and if the minimum amount of test portion is not chosen accordingly [27].

This illustrates not only the difficulty of carrying out "perfect" quality assurance but once again the problems of reliable trace analysis in general.

7.5. Environmental Analysis

7.5.1. The Problem

The task of environmental analysis is the identification and quantification (screening and monitoring) of contaminants [28]. The analytical characterization and evaluation of dangerous wastes from the past are typical examples of applied environmental analysis. Traditionally, for the risk assessments of old waste deposits, analytical methods (in the form of costly laboratory analysis) are used remote from the site of investigation [29]. At abandoned waste deposits and industrial sites, contaminant distributions are extraordinarily het-

Technology	Number of Sites
Immunoassay	43
X-ray fluorescence	39
Cone penetrometer sensor	34
Gas chromatography	24
Fourier-transformed infrared spectrometry	3
Colorimetric test strip	3
Fiber-optic chemical sensor	3
Mercury vapor analyzer	2
Biosensor	1

* Some typical reports:
Cone penetrometer coupled with laser induced fluorescence probe
EPA/600/R97/019 (SCAPS-LIF), EPA/600/R97/020 (ROST)
Field portable gas chromatograph/ mass spectrometer
Bruker-Franzen EM 640 (EPA/600/R-97/149)
Viking SpectraTrak 672 (EPA/600/R-97/148)
Field portable X-ray fluorescence spectrometer
(EPA/600/R-97-144,-145,-146)
Draft EPA Method 6200, March 1996: "Field Portable X-Ray Fluorescence Spectrometry for the Determination of Elemental Concentrations in Soil and Sediment".

erogeneous in relation to surface, depth, and composition. As a result, by inexpert and not directly verifiable sampling, errors of many hundred percent are caused that are strikingly disproportionate to the complexity of the analysis of the sample material [30]. This inconsistency between precise but cost-intensive single-sample analysis and an incorrect sample can be avoided by applying on-site measurement techniques with hand-held devices, portable devices, or mobile devices with full laboratory capability installed in a vehicle [31]. In the case of on-site work, the analytical results are mostly of limited precision, but since sampling, sample handling, sample preparation, and analytical work are combined directly at the site the information density is higher.

The cost of surveying and subsequent redevelopment has induced the EPA, within the framework of its SITE program, to start a monitoring and measurement technology program (MMTP). Part of this is the Environmental Technology Verification Program (ETV). The objective of this program is to organize the application of modern analytical techniques for field capability, to create new analytical methods for the field, and to apply these methods in addition to standard EPA specifications for contaminated sites [32].

The EPA has founded an online Field Analytic Technologies Encyclopedia (FATE), available on the Internet at http://fate.clu-in.org/. Summaries of applications of field analytical and site characterization technologies are available at http://clu-in.com. Some activities of the EPA are listed in Table 3. The newest results will be presented yearly at ON-SITE ANALYSIS conferences (http://www.infoscience.com). In Europe, too, there is some activity to establish such field screening methods [33].

7.5.2. Possibilities of Mobile Analysis

7.5.2.1. Definition and Requirements

Mobile analysis is a special form of on-site analysis in the environmental sphere. For this purpose the devices (sampling, sample preparation, measurement devices) are transported in a vehicle. The equipment can be integrated permanently into the car or operated hand-held.

In the screening mode, it needs methods which can identify unknown components and also quantify them. In the monitoring mode, the contaminant is known but not its distribution. In this case the analytical methods must be selective for the quantification of the contaminant. A typical example is the monitoring of organics (e.g., TNT, PAH) with immunoassay kits accepted by the EPA.

Mostly the field methods have higher detection limits than laboratory methods. Sample preparation and analytical procedure must be fast without extensive clean-up and separation steps.

The large amount of analytical information collected in the field is typical for mobile on-site analysis. The analytical techniques enable immediate analysis at the contaminated site, rapid change of site, and the necessary flexibility (surveying, site description, drill-core analysis, incident and damage analysis) in different measurement positions and at different sites.

Requirements for the use of field measurements are summarized below.

Analytical Parameter
  Determination of field and sum parameters, toxic metals, anions and cations,
  Organics: volatile organic compounds (VOC), total petroleum hydrocarbons (TPH), polynuclear aromatic hydrocarbons (PAH), polychlorinated biphenyls (PCB), pesticides
  in soil, water, and soil vapor

Process Conditions

Sample preparation without or with only little chemical work

Linearity of the analytical signal over several orders of concentration

Large range of variation with small response of the analytical signal (robustness of the analytical method)

Multiparameter analysis with high selectivity

Simple procedures that require little knowledge of analytical chemistry

Low cost of the analytical procedure

Equipment Technology

Rugged design toward mechanical and climatic stress

Low-energy configuration

Minimum space requirement of the equipment with high operating convenience

Autonomous electrical energy generation

Availability of working surfaces and techniques for handling samples

Computer technique, modem connection, global position system device (GPS)

Geo-Environmental software (visualization and grid optimization)

Possible analytical methods for use in the field (ppm range or higher) are listed in what follows:

Field and global parameters

Hand-held devices for the water field parameters [T, pH, electroconductivity (EC), dissolved oxygen (DO), oxidation – reduction potential (ORP)]

Hand-held monitoring techniques for VOC and gases: flame ionization detection (FID), photoionization detector (PID), thermal conductivity detector (TCD), infrared sensor (IR), see Figure 3.

Transportable equipment for measuring adsorbable organic halogen compounds (AOX), chemical oxygen demand (COD), total organic carbon (TOC)

Anions/Cations

Ion-selective sensors, reaction kits, photometer, voltammeter for the determination of ions in water or leachates

Toxic Metals

EDXRF analyzer (hand-held or transportable, X-ray tube or radioactive isotope source) for elements in soils

Organics

GC, preferably with MS coupling and sample preparation [extraction, purge & trap, spray & trap, solid phase extraction (SPE), solid phase microextraction, headspace, thermodesorption, soil air trap], nondispersive infrared spectroscopy (NDIR), FT-IR, TLC, immunoassays and biosensors (for determination of defined substances)

7.5.2.2. Example of Equipment

Only in some cases are devices designed for work under field conditions. Generally, they are applicable for the determination of inorganic traces in water or leachate and also for the determination of gases and VOC in the soil air. Some special instruments (NDIR) exist for the determination of TPH in soils under field conditions. However, modern laboratory analytical devices are generally so safely transportable that they can be carried along, secured for transport, in vehicles and installed at the site of application. Part of the relevant technology has already been designed for mobile application. Examples include metal analysis by using EDXRFA (e.g., Spectrace QuanX, SpecTrace 6000, SpecTrace 9000) and the GC/MS EM 640 (membrane inlet and direct inlet) from Bruker Daltonik Bremen and Viking SpectraTrak 573 from Bruker Daltonics. The equipment is also fully suitable for work under laboratory conditions.

The equipment plan of a vehicle is determined largely by the planned task. For a widespread use of mobile on-site analysis, GC/MS and EDXRF are the base instruments. The devices work also with autosamplers. In the case of EDXRFA, the measuring results in the ppm range are obtained in 15 min, including sample preparation (drying, milling). In the case of GC/MS studies it is a function of the analytical problem and the quality, the amount, and the composition of the organics. The upper time limit for obtaining the analytical information is not less than one minute.

7.5.3. Selected Applications of Mobile Analysis

7.5.3.1. GC-MS Screening and IMS monitoring

A typical example of the application of GC-MS is the field screening of soil gas composition because of its simple sample preparation tech-
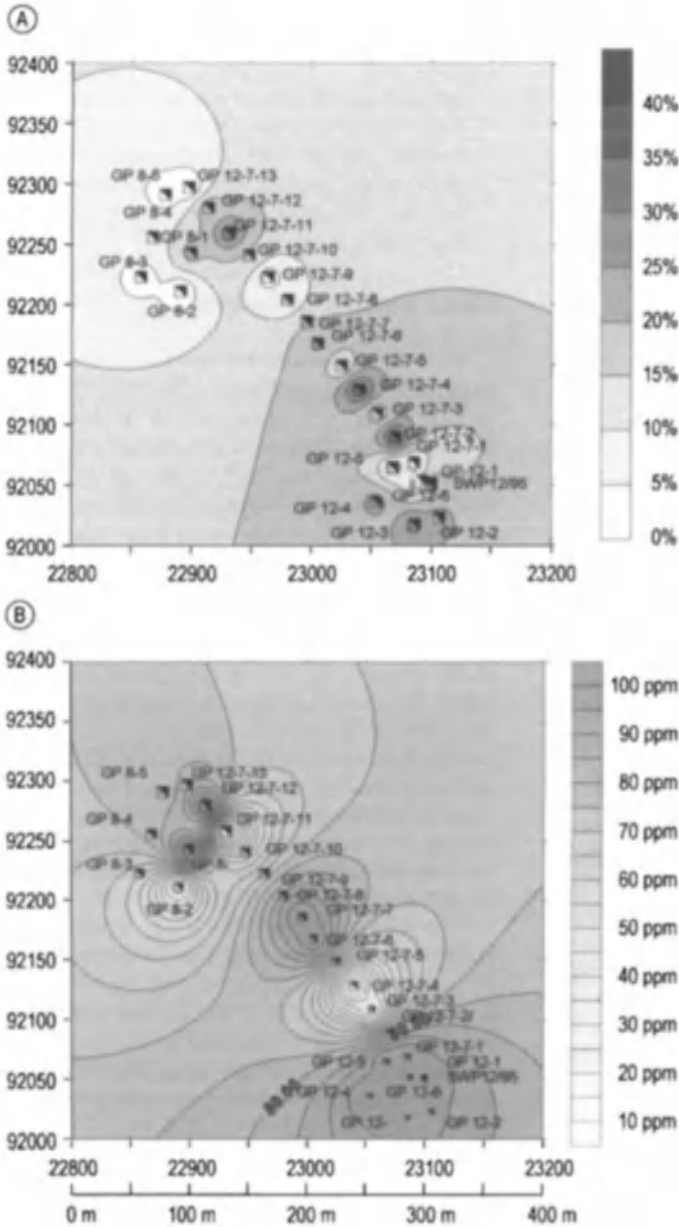
Figure 3. Typical example of the application of hand-held monitoring technique for the soil gas emission of a landfill site (depth of exploration of the drilling points: 1.50–1.80 m; visualisation by using SURFER software)
A) Determination of CH_4 distribution with an IR sensor; B) Determination of H_2S distribution with an electrochemical sensor

nique. A case study was to determine the VOC distribution, especially of tetrachloromethane (TCM), in a housing estate in the neighborhood of an old factory site. The first step of such studies is the determination of the soil gas composition and of the major components. The next step con-

sists of choosing a fast, simple, and inexpensive measuring method for monitoring and controlling the major risk component, in this case TCM. Possible methods of TCM detection are gas reaction tubes, a hand-held photoionization detector (PID), a mobile gas chromatograph, a hand-held ion mo-
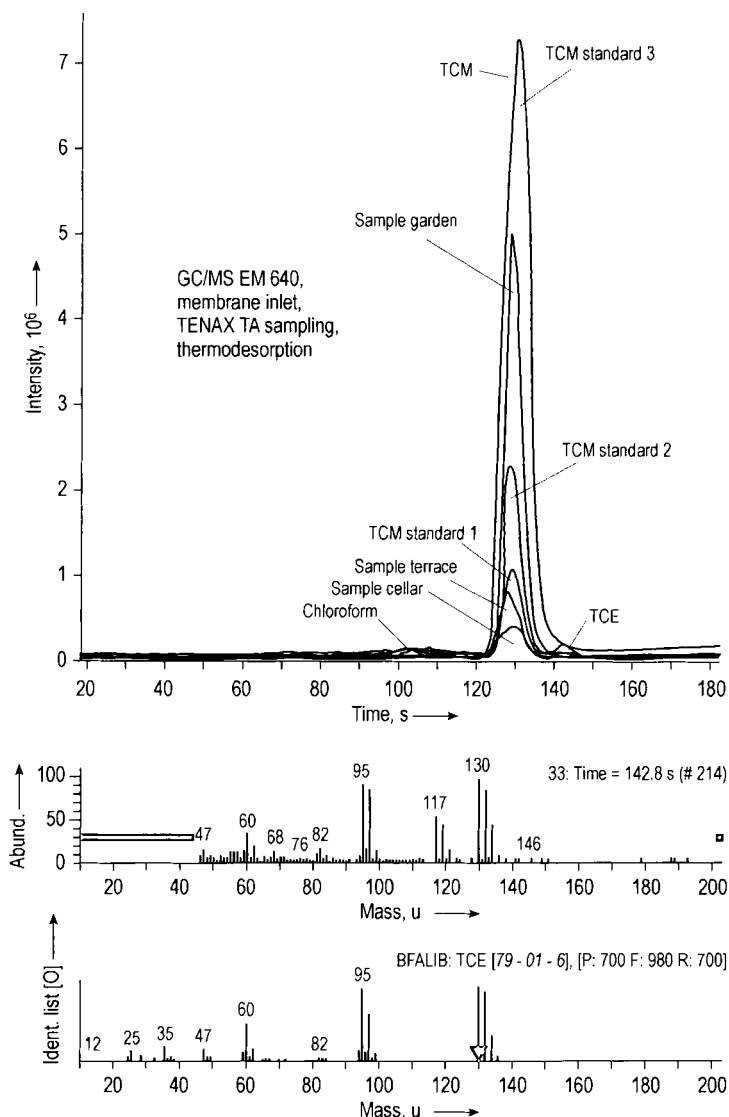
Figure 4. Presentation of some TIC runs of soil gas samples (garden, terrace; bore hole depth 2 m), gas standard and indoor air samples (cellar)
TCM = tetrachloromethane; TCE = trichloroethylene

bility spectrometer (IMS), and a mobile GC-MS. The application of gas reaction tubes is simple, but time consuming and expensive with poor reproducibility of results. The linear range of determination lies in an order of magnitude, and there is a pronounced nonconformity between the different tube types. The PID with the typically used 10.6 eV UV lamp is not suitable for TCM determination. The IMS is the fastest instrument (over-

all analysis time < 2 min). It gives only a sum signal for halogenated hydrocarbons in the negative operation mode (selective detection of a Cl peak) with high sensitivity (< 0.2 vol ppm), but with a small range of linearity of determination and the possibility of extension by using a dynamic gas dilution. The mobile GC-MS system is the best technique in all the cases of field screening of nonpolar compounds with a high dy-
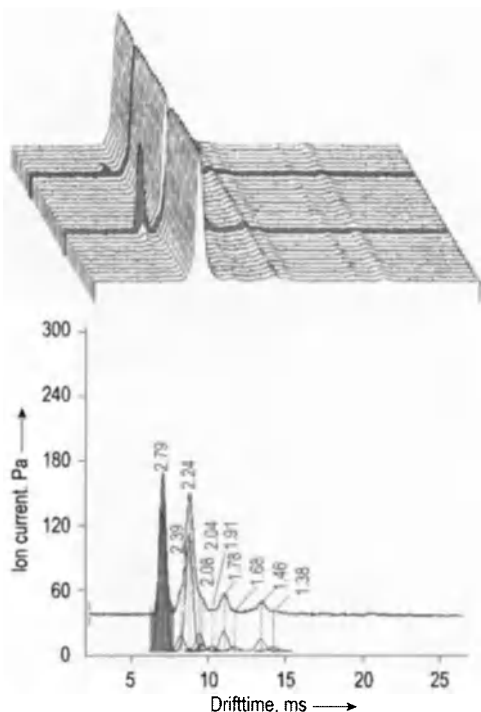
Figure 5. CCl$_4$ detection in soil gas (sampling depth 1.80 m) by using ion mobility spectrometry (8 bit IMS type RAID 1 from Bruker-Saxonia in the negative mode, push injection of 1 mL soil gas, measurement of the Cl-peak
red peak: sample taken near a walnut tree; Green peak: sample taken at the garden fence
Spectral range with two different sample injections (top) and a typical single spectrum (bottom) with a Cl peak (green)

namic range of more than three orders of magnitude under real field conditions and also allows the detection and identification of other contaminants. The system is expensive and its operation is not simple. A good compromise is to use GC-MS as a master instrument and IMS as a fast monitoring instrument. The accordance of results between GC-MS and IMS is better then 20 %. Figure 4 summarizes the application of GC-MS in this field. After collection (indoor air and soil gas of boreholes) of the volatile organic compounds on TENAX TA the organics are thermodesorbed and separated by gas chromatography. The total ion chromatogram (TIC) is analyzed with the aid of a MS spectra data bank and analytical software specially set up for the system (in this case Bruker DataAnalysis software), and the substances are identified. The system is also calibrated for quantification and the concentrations of substances are calculated with the aid of the software.

Figure 5 shows the detection of TCM by ion mobility. The term ion mobility spectrometry refers to the principles, practice, and instrumentation for characterizing chemical substances by means of gas-phase ion mobilities [34]. Ion mobilities are determined from ion velocities that are measured in a drift tube with supporting electronics. Ion mobilities are characteristic of substances and can provide a means for detecting and identifying vapors. In practice, a vapor sample is introduced into the reaction region of a drift tube in which neutral molecules of the vapor undergo ionization, and the resultant ions, i.e., product ions, are injected into the drift region for mobility analysis. Mobility K is determined from the drift velocity attained by ions in a weak electric field of the drift region at atmospheric pressure. The normalized value K_0 is a characteristic constant of the product ion (in this case $K_0 = 2.79$ cm^2 V^{-1} s^{-1} for the Cl peak).

By using a mass spectrometer with direct inlet (e.g., Bruker EM 640S) it is possible to inject directly the soil air samples and to report only the typical mass fragments of TCM in the SIM mode (Fig. 6). In this way the analytical procedure is faster and more sensitive. The TCM distribution using all results of IMS and GC-MS measuring can be visualized with special software (e.g., SURFER software from Golden Software Inc.). An example of TCM distribution is incorporated in Figure 6.

7.5.3.2. Elemental Analysis by Mobile EDXRF

Energy-dispersive X-ray fluorescence analysis (EDXRF) has proven to be ideal for on-site analysis [35]. With the availability of Peltier-cooled semiconductor detectors in EDXRF, its mobile application to field screening of toxic metals is unproblematic. In combination with geo-statistical methods or geographic information systems this technique provides important and meaningful information about the distribution of contaminants on hazardous waste sites. In Figure 7 this is illustrated by the determination of the metal distribution in a flood plain area in the neighborhood of a landfill site. A radioisotope is used to excite the characteristic fluorescent X-rays in a sample in hand-held devices. The instruments are rugged, fast, and easy to use. The sensitivity is lower than that of instruments equipped with adjustable X-ray tube sources, which have a detection limit in the ppm range. EDXRF can simultaneously determine 20 or more elements from
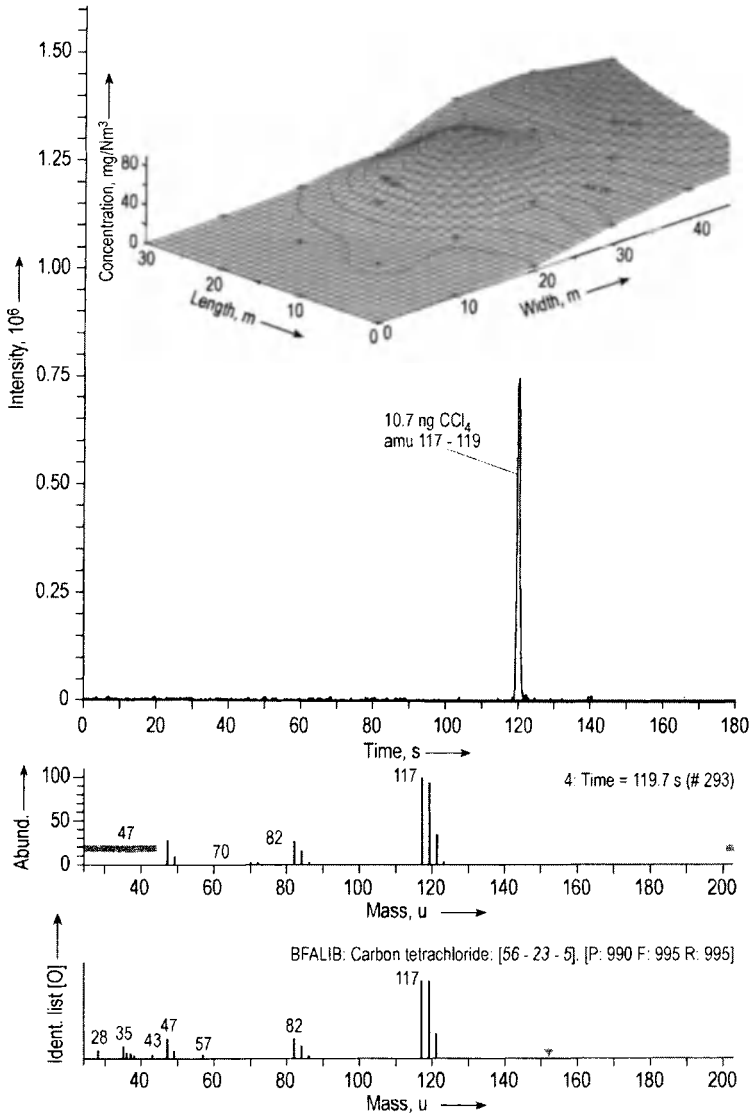
Figure 6. Selective detection of tetrachloromethane (TCM) contaminations by using SIM mode (measuring conditions: Bruker Daltonik GC/MS EM 640S with direct inlet; split 1 : 10; direct injection of 0.2 mL gas samples, 30 m HP 5 ms, $d_i = 0.25$ mm, $f_{th} = 0.25$ mm). The upper part of the figure incorporates the 3D visualisation of TCM distribution of soil gas (sampling depth 1.80 m; combination of IMS and GC-MS data) of part of the contaminated area in the neighborhood of the housing estate

sodium to uranium within only a few minutes. The instruments are also applicable in the laboratory. Some of these (e.g., the QuanX from Spectrace) also work with a 20-position sampler for automatic operation. The sample preparation is simple and fast and consists of drying and milling (particle size < 100 μm). The accuracy of the results depends on the quality of the used standards and the difference between the matrix of the standards

and the sample (Fig. 8). Therefore, the results were calculated with the so-called Fundamental Parameter algorithm supplied with the spectrometer software. This algorithm corrects automatically for interelement and matrix effects. For this reason, no site-specific calibration is required, which is an enormous advantage over empirical calibration models.
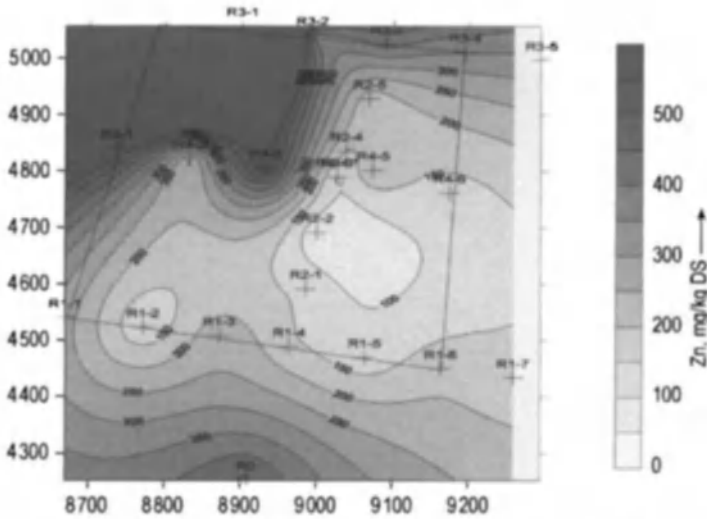
Figure 7. Determination of element distribution (e.g., Zn) in the subsoil of a landfill site by using mobile XRF (SPECTRACE 6000) and geostatistical methods
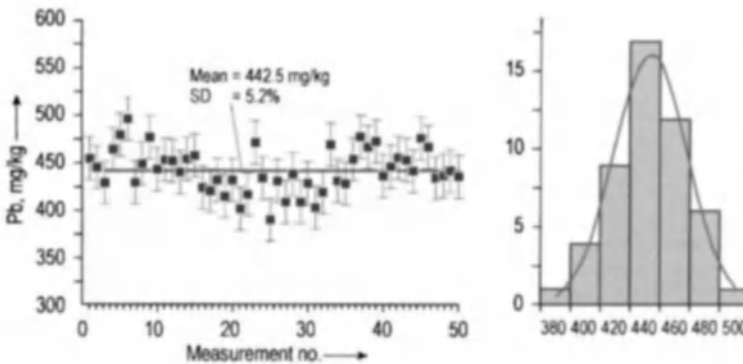
Figure 8. Reproducibility of the field screening EDXRF method [mean value, standard deviation (SD) and gaussian distribution of Pb content in excavated material from a waste disposal site]: determination of the repetition accuracy over a time scale of two months

Typical for the application of mobile EDRFA is the fast determination of element distributions in drilling cores or in finding hot spots in soil profiles (Fig. 9).

7.5.4. Conclusions

Field screening methods are used to determine the presence or absence of typical components at a given site. Needs for screening methods exist not only for site characterization, but also for determining the progress of remediation efforts. A driving force in mobile screening technology devel-opment and commercialization is the ever-increas-ing costs associated with environmental compli-ance. For the different contaminants, there many different field screening instruments. Many com-panies offer field test kits and a variety of detector tubes. Battery-operated spectrophotometers are available to support field analyses. Gas chroma-tography is one of the most mature technologies being applied in this field. The most recent ad-vances involve the use of kits based on immuno-chemistry, GC-MS coupling, and XRFA. Attrac-tive features include a high degree of analyte se-lectivity and ppb to ppm sensitivity in most ma-trices.
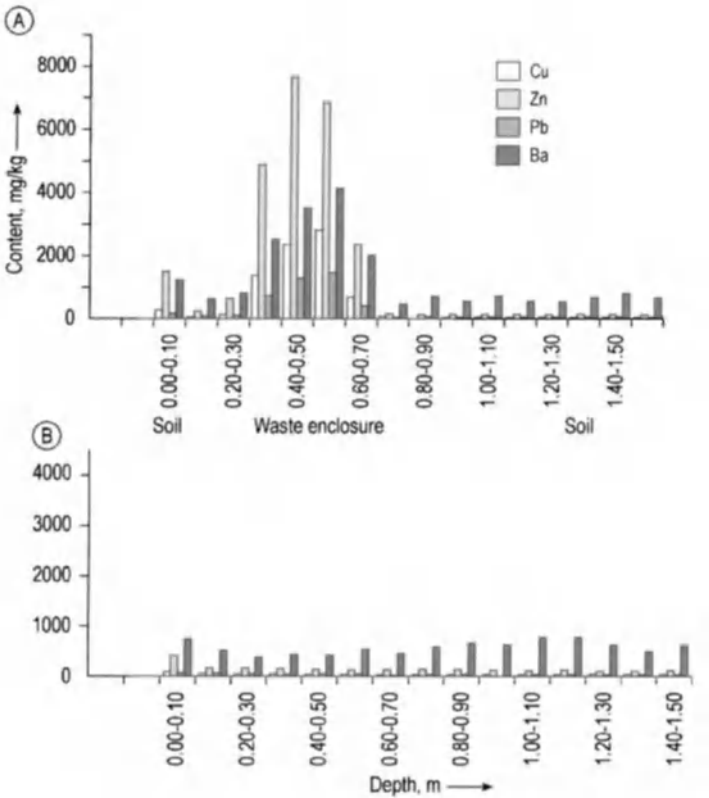
Figure 9. On-site determination of waste enclosure in a flood plain area in neighbourhood of a landfill site by using transportable EDXRF spectrometer (sampling by digging in the depth, waste contamination in profile 1)
A) Digging profile 1; B) Digging profile 2

7.6. References

[1] O. G. Koch, G. A. Koch-Dedic: *Handbuch der Spurenanalyse*, 2nd ed., **vol. 1,** 1, Springer Verlag, Berlin 1974.

[2] K. Beyermann: *Organische Spurenanalyse,* G. Thieme, Stuttgart 1982.

[3] G. Knapp, M. Michaelis: "Chemie in Labor und Biotechnik," *CLB Chem. Labor. Betr.* **44** (1993).

[4] G. Tölg, *Analyst (London)* **112** (1987) 365–376.

[5] B. Sansoni, *Pure Appl. Chem.* **59** (1987) 579–610.

[6] R. Kellner et al. (ed.): *Analytical Chemistry,* Wiley-VCH, Weinheim 1998.

[7] G. Schwedt, A. Meyer, *LaborPraxis* **16** (1992) 712–721.

[8] U. Thiele, *Lab. Trend.* 1999 10–14.

[9] L. Dunemann, J. Begerow: *Kopplungstechniken zur Elementspeziesanalytik,* VCH-Verlagsgesellschaft, Weinheim 1995.

[10] M. Luthardt, E. Than, H. Heckendorff, *Fresenius' Z. Anal. Chem.* **326** (1987) 331–339.

[11] DIN 32645 (May 1994): *Chemische Analytik; Nachweis-, Erfassungs- und Bestimmungsgrenze; Ermittlung unter Wiederholbedingungen; Begriffe, Verfahren, Auswertung.*

[12] DIN 55 350, part 34 (Feb. 1991): *Begriffe der Qualitätssicherung und Statistik; Erkennungsgrenze, Erfassungsgrenze und Erfassungsvermögen.*

[13] *ISO Standards Handbook 3: Statistical Methods,* 3rd ed., ISO, Genf 1989.

[14] H. Kaiser, *Spectrochim. Acta* **3** (1947) 40–67.

[15] K. Doerffel, G. Michaelis, *Fresenius' Z. Anal. Chem.* **328** (1987) 226–227.

[16] G. L. Long, J. D. Winefordner, *Anal. Chem.* **55** (1983) 712A–724A.

[17] G. Ehrlich, *Fresenius' Z. Anal. Chem.* **232** (1967) 1–17.

[18] K. R. Beebe, B. R. Kowalski, *Anal. Chem.* **59** (1987) 1007A–1017A.

[19] C. W. Brown, R. J. Obremski, *Appl. Spectr. Rev.* **20** (1984) 373–418.

[20] W. Horwitz, *Pure Appl. Chem.* **62** (1990) 1193–1208.

[21] DIN ISO 5725 (April 1988): *Präzision von Meßverfahren; Ermittlung der Wiederhol- und Vergleichspräzision von festgelegten Meßverfahren durch Ringversuche.*

[22] L. M. Schwartz, *Anal. Chem.* **49** (1977) 2062–2068.

[23] J. D. Hwang, J. D. Winefordner, *Prog. Anal. Spectrosc.* **11** (1988) 209–249.

[24] C. Hartmann, J. Smeyers-Verbeke, D. L. Massart, *Analysis* **21** (1993) 125–132.

[25] J. Kramer: *STATCAL*, Umschau Verlag, Frankfurt/Main 1992.

[26] Y. Hu, J. Smeyers-Verbeke, D. L. Massart, *Chemom. Intell. Lab. Systems* **8** (1990) 143–155.

[27] J. Pauwels, C. Vandecasteele, *Fresenius J. Anal. Chem.* **345** (1993) 121–123.

[28] I. L. Marr, M. C. Cresser, L. J. O. Ottendorfer: *Umweltanalytik*, Thieme, Stuttgart 1988.

[29] V. Franzius, R. Stegmann, K. Wolf, E. Brandt: *Handbuch der Altlastensanierung*, **vol. 2,** R. v. Decker's Verlag, Heidelberg 1992 chap. 4.

[30] L. H. Keith (ed.): "Principles of Environmental Sampling," *ACS Professional Reference Book*, 2nd ed., American Chemical Society, Washington D.C., 1991.

[31] J. Flachowsky: "Mobile Umweltanalytik," in H. Günzler et al. (eds.): *Analytiker-Taschenbuch*, **vol. 18,** Springer-Verlag, Berlin – Heidelberg – New York 1998, pp. 143–180.

[32] Abstracts of the Eight International Conference ON-SITE ANALYSIS, Lake Las Vegas, NV, January 2000, www.ifpac.com.

[33] H. J. Gottlieb et al. (ed.): "Field Screening Europe," Proc. of the first international conference on strategies and techniques for the investigation and monitoring of contaminated sites, Kluwer Academic, Dordrecht 1997.

[34] G. A. Eiceman, Z. Karpas: *Ion Mobility Spectrometry*, CRC Press, Boca Raton – Ann Arbor – London – Tokyo 1994.

[35] G. Matz, W. Schröder, J. Flachowsky: "On-Site Investigation of Contaminated Soil by GC-MS and EDXRF-Techniques," in F. Arendt, G. J. Annokkée, R. Bosman, W. J. van den Brink (eds.): *Contaminated Soil '93, 4th Int. KfK/TNO Conf. on Contaminated Soil, Berlin May 1993*, vol. I, pp. 657–664, Kluwer Academic Publishers, Dordrecht 1993.

8. Radionuclides in Analytical Chemistry

RAINER P. H. GARTEN, Max-Planck-Institut für Metallforschung, Dortmund, Federal Republic of Germany

JURAJ TÖLGYESSY, Department of Chemistry, Faculty of Natural Sciences, Matej Bel University, Banská Bystrica, Slovak Republic

Abbreviations

A	analyte
A_1	activity of component 1
AA	activation analysis
AAS	atomic absorption spectrometry
B	reagent
IDA	isotope dilution analysis
R*	labeled radioreagent
SRM	standard reference material

8.1. Introduction

8.1.1. Definition and Purpose

Radioanalytical chemistry covers the use of radioactive nuclides and nuclear radiation for analytical purposes:

1) In the analytical determination of any chemical species or element

2) In the development, improvement, and quality assurance of analytical procedures for the benefit of analytical principles

8.1.2. History

The use of radionuclide techniques in analytical chemistry was first reported in 1913 by G. HEVESY and F. PANETH in a study of the solubility of lead sulfide in water, using the natural lead isotope ^{210}Pb as indicator [67]. Isotope dilution analysis was introduced by O. HAHN in 1923 [68], using ^{231}Pa to determine the yield of ^{234}Pa. The development of radioreagent methods followed, and further development of radioanalytical chemistry has established a range of analytical methods and techniques [1]–[4], [61], [65], [87], [93], [95], [97]. These include the use of artificial radionuclides and labeled compounds, the principles of nuclear activation [4]–[10], [66] (→ Activation Analysis), and absorption and scattering of radiation [11], [12]. The most important procedures are shown in Table 1.

8.1.3. General Features

Radiochemical analysis is based on two outstanding features of radioactivity:

1) The high sensitivity and ease of measurement of radioactive radiation [13], [14]
2) The possibility of labeling chemical compounds with radioactive tracers [15]–[17], [69], [70]

Radionuclide techniques often have higher sensitivity than other analytical methods. The amounts of nuclides, correlated to an activity of 1000 Bq (see Table 2), can be derived from the law of radioactive decay. These amounts vary considerably, corresponding to the wide range of radioactive half-lifes. For 90 % of the commonly used nuclides [17], half-lifes range from several minutes to several years, so the corresponding masses are extremely low.

Radionuclides are often diluted with inactive isotopes, but specific activities (i.e., activity per total mass of element) are still very high. Since the background in nuclear spectroscopy is very low, and sensitivity is high, activities as low as 0.2 Bq can be readily detected (^{3}H, ^{125}I, ^{132}I), and even 0.01 Bq (several γ emitters, e.g. ^{24}Na, ^{38}Cl, ^{42}K, ^{46}Sc, ^{59}Fe, ^{60}Co, ^{65}Zn, ^{110m}Ag, ^{182}Ta, ^{187}W,

^{198}Au). Total activities of the order of 0.1 – 20 kBq (3 – 500 nCi) are often sufficient in analytical applications. The precision of the final result can be better than 2 %, if the counting error is kept below 1.5 %. This can be obtained from counting times between 100 s and 5 h.

Analytical applications of radionuclide techniques rely on the assumption that different isotopes of the same element exhibit the same properties in any macroscopic physical or chemical process, and that radioactive labeling does not influence the other properties of a chemical species. This is generally the case, with deviations below 1 % (with the exception of hydrogen isotopes) owing to isotopic fractionation or radiation effects. For analytical purposes, the radiotracer and the analyte must be present in the same chemical form. This is usually to achieve, but specialized preparative techniques may be necessary for radioactive labeling of more complex organic compounds.

8.1.4. Importance and Current Trends

Radionuclides are used in many subdivisions of analytical chemistry (see Table 1). Of major importance are radiotracers in methodological and pathway studies, isotope dilution analysis (IDA), radioimmunoassay, and nuclear activation analysis (AA) (→ Activation Analysis) [66]. They are all especially suited to analyze the extremely small amounts of substances encountered in ultratrace analysis or in trace analysis of microsamples.

Over the past two decades, the emphasis has shifted from high detection power in routine analysis toward an independent approach, applying this high detection power to the development of analytical procedures and reference materials.

Radiochemical methods for routine analysis have lost ground to other, primarily spectroscopic, methods (see Chap. 8.3) [18]–[21]. Nonradiochemical methods often yield highly reproducible results, but may involve systematic errors.

The need for increased reliability and analytical quality control has emphasized the usefulness of radiochemical methods for the certification of standard reference materials (SRM) [9], [22], [23]. Radioanalytical methods are often suitable for homogeneity testing and distribution analysis of traces in SRMs. Activation analysis, radiotracer techniques, and isotope dilution analysis are becoming increasingly important for assessment of analytical quality.

Table 1. The scope of radioanalytical chemistry

Division	Section	Examples
Tracer labeling in organic analysis	labeling with tritium, ^{14}C, and heteroatoms (^{32}P, ^{36}Cl, ^{35}S, ^{125}I, ^{131}I, etc.)	random molecular sites selective molecular sites specific molecular sites
Tracer techniques in inorganic and organic analysis (addition of radioactive compounds to the sample)	chemical methodological studies and quality control of fundamental analytical processes (multistage procedures)	combustion and decomposition
		equilibrium constants
		phase separation efficiencies
		extraction
		volatilization
		precipitation
		redissolving precipitates
		coprecipitation
		electrochemical separation
		stability of solutions
		evaluation of sources of errors in trace and
		ultra-trace processes
		adsorption
		desorption
		volatilization
		contamination
		yields of complex analytical procedures
	kinetic studies	diffusion and forced diffusion (including electromigration) in solutions and gases
		diffusion, forced diffusion, and self-diffusion[a]
		chemical reactions: from long reaction times down to the ms regime (μs in special cases); macroscopic reaction order and rate
		reactive intermediate studies
		reaction mechanisms
		metabolism in organims and organelles
		rates of emission, immission, and migration of chemical species in environmental systems: natural, anthropogenic, and intentionally introduced radionuclides
	isotope dilution analysis	stoichiometric
		substoichiometric
		sub- and superequivalence method
	radiotracer- and radioreagent based analytical methods and pathway studies	radiochromatography
		radioreagent methods
		radiorelease techniques
		radioimmunoassay
		radiometric titration
		isotope exchange
		autoradiography
		studies of metabolism
		pathways of tracers in the environment (generally short-lived, for safety reasons)
		wear and wear protection in engineering materials corrosion and corrosion protection in engineering, construction, and biomaterials
Activation analysis (AA), (production of radionuclides in the sample by nuclear reactions, their separation and measurement)	discriminating types of AA	Induced by
		thermal neutrons (reactors)
		epithermal neutrons (reactors)
		fast neutrons (generators, cyclotrons, isotope sources)
		protons (accelerators)
		α particles (sources, accelerators)

Table 1. (continued)

Division	Section	Examples
		fast ions of low to mid-range Z (^{6}Li, ^{9}Be, ^{12}C, ^{14}N, and others from accelerators)
		γ-radiation (reactors, sources, cyclotrons, synchrotrons)
		delayed (ordinary) AA vs. prompt[b] (prompt nuclear reaction analysis NRA) neutron- or charged particle-induced reactions
		fast[c] (spectroscopy of short-lived nuclides) vs. intermediate[d] and long lived[e] AA
		purely instrumental (mostly γ-spectrometric) vs. radiochemical i.e., group separations
		or specific element separations (including substoichiometric calibration techniques)
	applications	certification analysis of standard reference materials
		basic research in analytical science and technology
		routine control of high-purity materials
		trace elements in biomaterials (including medicine)
		routine production control of commercial high-performance materials
		on-line and in-line process control (e.g., bomb detection)
		trace elements from environmental materials (water, dust, soil, sludge)
		trace elements in plants (pollutants, uptake, storage, mobilization paths)
		trace elements in geochemical and cosmochemical matter
Determination of natural (H, C, K, Th, U, Ra, etc.) and anthropogenic (Cs, Sr, Tc, Np, Pu, etc.) radioactivity		levels and distribution of activity in environmental materials (air, water, soil, etc.)
		dating in archeology, fine art, historical, and environmental research (^{14}C, ^{3}H, ^{7}Be, ^{10}Be, ^{40}K, ^{232}Th, ^{235}U, ^{238}U)
Nuclide radiation sources in analysis (on-line and remote sensing devices)		radionuclide-induced X-ray fluorescence analysis (on-line product flow control, outdoor and remote sensing)
		resonance scattering of γ rays
		resonance absorption of γ rays
		absorption and backscattering of X, α, β and γ radiation (on-line control):
		mass density monitoring
		product mass flow control
		monitoring of coatings and films
		radiographic testing of materials and devices
		detection of radiation induced radicals
		radiation dosimetry
		particle size distribution
		absorption and moderation of neutrons for light element detection
		remote B- and Cd-sensing
		humidity measurement of product flows
		plasma desorption mass spectrometry (soft ionization technique induced by fast heavy fission nuclei from ^{252}Cf-source)
		prompt (n, γ) reactions (for prospecting, oil and gas well logging)
		α- and β-ionization detector devices (pressure gauges, gas flow metering, smoke sensors, electron capture detectors).

[a] Diffusion of a compound or element in a matrix containing an excess of the same species; (active and inactive) isotopic tracer methods are the only methods available.   [b] $\Delta t < 10^6$ s.   [c] half-lifes $\tau < 1$ h.   [d] half-lifes $\tau < 3$ d.   [e] Half-lives $\tau > 20$ d.

Table 2. Mass corresponding to 1 kBq for carrier-free radionuclides

Nuclide	Mass ^{a}, ng	Half-live
^{36}Cl	820	3×10^5 a
^{14}C	6.1	5.8×10^3 a
^{63}Ni	0.57	120 a
^{85}Kr	6.8×10^{-2}	10.6 a
^{125}I	1.6×10^{-3}	60 d
^{3}H	2.8×10^{-3}	12.4 a
^{131}I	2.2×10^{-4}	8.04 d
^{35}S	6.3×10^{-5}	87.2 d
^{32}P	9.5×10^{-5} *	14.3 d
^{99m}Te	5.1×10^{-6}	6.02 h
^{18}F	2.8×10^{-7}	1.83 h

* Commercially available with maximum specific activity of 2×10^{14} Bq/g.

Radioanalytical methods are well suited to the determination of basic analytical data, such as equilibrium constants, or kinetic data (see Section 8.3.1). These data are important for development and optimization of new analytical procedures [61].

The most frequent analytical use of tracers is in the biomedical sciences. Radioimmunoassays, in the form of rapid diagnostic test kits have led to the development of nonradioactive assays based on similar principles [2], [24].

Many important tracer applications can be substituted by other methods. Even IDA and tracer applications in self-diffusion studies can be replaced by inactive isotope tracer methods using mass spectrometry and other methods for isotope ratio determination. However, because of the extremely high sensitivity of IDA, radioactive tracers are of unique usefulness in radioimmunoassays, radiorelease reagents, radiochromatography, AA, and for systematic studies in trace and ultratrace analysis, physiological chemistry, IDA, diffusion, isotope exchange, and physical chemistry of solids.

Radioanalytical chemistry will continue to occupy an important position, despite increasing criticism of nuclear technology.

Recently, there has been little innovative work in the field of methodology, but the applications field has continued to expand [7], [25], [66]. Highlights include application of isotope enrichment techniques in environmental studies [26], [71], [72], improved accuracy in the characterization of SRM [23], [27], [28], preirradiation chemistry with high-purity reagents under clean room conditions to permit speciation from AA [7], [26] (sometimes termed molecular AA [25]), derivative AA [7], accelerator-based dating methods combined with IDA, short-time and pulsed reactor activation, extension of the sub- and superequivalence method [73]–[75], in-vivo analysis by prompt-γ neutron AA [7] and by short-lived AA [8], [29], and extended use of the single comparator (k_0) standardization method in neutron AA [30].

8.2. Requirements for Analytical Use of Radionuclides

8.2.1. Safety and Operational Aspects

Safety and operational aspects involve the hazards of ingestion of radionuclides and exposure to radiation [76].

8.2.2. The Labeled Substance

Aspects of the labeled substance involve the following requirements [14], [20], [26], [31]–[33], [66]:

1) Radionuclide and radiochemical *purity* are essential. Radiochemical purity can be checked by reverse IDA (see Section 8.4.2). The labeled substance must react identically to the analyte.

2) Complete *homogenization* and *isotope exchange* must be achieved, requiring successive transformation of the active tracer as well as of the inactive analyte into all chemical species that are considered to occur in the complete system [20], [26], [34].

3) *Isotope effects* must be negligible.

4) For most elements, suitable radionuclides for analytical use are commercially available. There are only seven elements of interest (B, He, Li, N, Ne, O, Mg), for which this is not the case [17].

5) The *preparation* of a specific isotope, of an isotopically modified (i.e., mixture of labeled and unlabeled), isotopically substituted, or even site-specifically labeled compound is often mandatory in analytical applications.

6) *Synthesis* (including biosynthesis) of these labeled compounds often requires much more effort than the actual radioanalytical experiment [34]. A variety of labeled compounds and pharmaceuticals are available commerically from stock, and a further range

of compounds will be prepared on request by a number of suppliers [35]. In special cases, doubly and triply labeled molecules have been used to follow the reactions of different functional groups.

7) Costs of radionuclides for analytical applications are typically in the range $ 150 – 400 per 40 MBq (= 1 mCi), and higher where preparation is difficult or time consuming (e.g., ^{54}Mn, ^{85}Sr, ^{36}Cl: $ 700 – 4000 per 40 MBq). Labeled substances from stock usually cost between $ 2000 (^{14}C) and $ 200 (^{3}H) per 40 MBq; the price depends on the cost of preparation.

8.2.3. Activity Measurements

Commercially available instrumentation is generally used, with the exception of some research work and highly specialized routine analytical applications.

Quantitative measurement of radioactive substances is generally straightforward with the usual counting devices and relatively simple with γ ray, X ray, and hard β emitters. The *detection limit DL* of the concentration of the analyte is given [36], [77] by:

$$DL = \frac{f}{s} \sqrt{B} \qquad (1)$$

where $f = 3$ is appropriate for detection at the 99.9 % confidence level. B is the blank level, including analytical blank and spectral background of the detection device, and s is the analytical sensitivity given by the slope of the analytical calibration function, i.e., under idealized conditions:

$$s = \varepsilon \eta \, t \, a \, m_{sample} \qquad (2)$$

where ε is the detection efficiency, η is the emission probability, t is the counting time, a is the specific activity of the analyte, and m_{sample} is the mass.

Specimen preparation for radioactivity measurements aims, in general, to obtain the analyte:

1) As completely as possible and with the highest possible degree of purity from extraneous radioactivity and from excess matrix matter

2) Concentrated on the smallest possible area or volume of a solid, liquid, or gaseous (for H, C, S, Ar, Kr, Xe, Rn) target

3) On a specimen support that is compatible with the chemical nature of the species analyzed, with the specific radiation to be measured, and with the detector system, e.g., liquid scintillation cocktails [37] for weak β emitters, or aluminum trays for dried layers of hard β emitters [13].

These preparation requirements are compatible with the more general rules for specimen preparation in trace and ultra-trace analysis by any technique [18], [40], [45]. Requirements are different for measurements of: γ radiation of energy >50 keV; hard β radiation >200 keV; α radiation; weak β radiation below 200 keV; and X radiation below 70 keV. In this respect, specimen preparation follows the general outline of preparation procedures for small counting samples [13], [37].

8.2.4. Choice of Radionuclide

The choice of radionuclide involves:

1) Measurement, possibly including discrimination between different indicator nuclides or indicator and interfering nuclides (e.g., from byproducts or blank level impurities)

2) Suitable half-life

3) Chemical species

γ *Emitters* are preferred for multitracer experiments (most of all in nuclear AA) for applications where the tracer is dispersed in, or shielded by a material (e.g., in-vivo diagnostics in biology and medicine, measurement from bulky solid material), but also from tracer solutions.

Pure β *emitters* are especially useful where the shorter range of their radiation is utilized, e.g., for autoradiography [39], and for surface and thin film studies, including adsorption, corrosion, and catalysis. Measurements of β-emitting tracer solutions are performed with high efficiency by liquid scintillation.

α *Emitters* are only used as tracers in special cases, either where their very low range is important (e.g., surface and thin film techniques), or where they represent the only available nuclides of the element (e.g., actinides).

Imperfect tracers are often used when labeling of a specific chemical site is inappropriate with isotopic nuclides. Imperfect tracers are widely used to label gas streams with inert gases (^{85}Kr, ^{41}Ar, ^{133}Xe), water with dissolved anionic tracers (^{82}Br, ^{131}I, ^{24}Na), hydrocarbon fuels with dis-

solved ^{60}Co-naphthenate or 1,2-^{82}Br-dibromoethane. Similarly, loss of metal ions at ultra-trace levels due to adsorption on the vessel walls and by volatilization have been modeled by ^{195}Au, ^{203}Hg, ^{60}Co [18], [20], [40], [78], which are easily measured, and which are available at high specific activity.

8.2.5. Appraisal of Radionuclide Use in Analysis

8.2.5.1. Advantages

Utilization of radionuclides in analysis has the following advantages:

1) Cost effectiveness [4], [9], [40]–[44], [66] (Section 8.2.2).
2) Element specificity and chemical site specificity (Section 8.2.2).
3) High dynamic range in a single experimental run. Radionuclide blanks are often extraordinarily low, far below detectability. Multielement analysis by simultaneous use of different tracer nuclides, often without separation.
4) Quality and quantity of the signal-producing decay are independent of influences such as temperature, pressure, external fields, and the chemical environment of the nucleus.
5) Locally resolved identification. Distribution detection, e.g., from within an analytical decomposition apparatus, a growing plant, a thin-layer chromatogram, or sections from biological tissue or engineering materials. Lateral resolution as low as 10 μm by autoradiography [39].
6) Predetermination of counting precision [13], [49].
7) Following reaction schemes or metabolites: in vitro and in vivo with fast element detection, often from outside the analytical system at comparable levels of sensitivity and accuracy.
8) No interference with analysis by other methods (nondestructive, conservative).

8.2.5.2. Disadvantages

Disadvantages of radionuclides in analytical chemistry are as follows:

1) Synthesis of labeled compounds may be difficult
2) Safety considerations (see Section 8.2.1) are required to avoid incorporation of radiotoxic

substances and radiation exposure, specialized laboratory and equipment, qualified and skilled staff, disposal of radioactive waste
3) Boron, aluminum, nitrogen, and oxygen are of analytical interest, but no nuclides are available with half-lifes longer than 10 min
4) Specific sources of error due to radiotracers (see Section 8.2.5.3) have to be controlled
5) No distinction between different chemical species without additional separation
6) In general, sample material cannot be returned to normal use after analysis. Safe handling may restrict use of some analytical processes

8.2.5.3. Sources of Error

In analytical applications of radionuclides, specific sources of error can generally be ascribed to either analytical and trace-analytical considerations, or particular features of radionuclides and their measurement.

Difficulties in handling sub-nanogram amounts of analyte are well known [40], [45]. Because of the very low concentrations considered, chemical techniques such as solvent extraction, electrodeposition, volatilization, coprecipitation, chromatography, and ion exchange may be necessary. Additionally, surface adsorption on vessel walls, colloidal matter, and dust particles in the system may be a significant source of error. Errors may be prevented by blank experiments, high acid concentration above 1 mol/L (diminishing adsorption), selection and preconditioning of vessel materials [20], [40], [44]–[46]. Since these problems are not specific to radiotracers, they can be most effectively reduced by addition of an inactive isotopic carrier of the same chemical species.

Difficulties in Obtaining Representative Labeling. For a tracer experiment, it is necessary to process the tracer so that it is chemically identical to, and in the same phase as the unlabeled substance. Merely mixing the radiotracer with the analyte sample is generally not sufficient (see Section 8.2.2) [20], [26], [33], [34], [47].

Radiation-induced effects in labeled molecules occur with compounds of high specific activity, labeled with weak β emitters or even α emitters. Since autoradiolysis takes place mainly via secondary radical reactions of fragments formed from solvents or major components, radiolytic damage

can be reduced by cooling and by adding a radical scavenger.

Impurities from decay products (i.e., from radioactive decay and from radiation-induced decay of chemical compounds) can interfere with the activity measurements, or with the chemical reactions employed to process the sample. Radionuclide purity is a function of time.

Limited radiochemical purity may result in detection of radioactivity from chemical species different from the intentionally labeled one. This is of prime importance in specifically labeled enantiomers. Labeled compounds should be purified immediately before use. Chromatography and reverse IDA (see Section 8.4.2) are important for checking radiochemical purity.

Isotope effects are important in reactions where tritium is involved in the rate-determining step [48], because it reacts more slowly than hydrogen. Similar effects, but to a much lesser degree, are reported for ^{14}C, ^{32}S, and other nuclides [4], [17], [48].

Errors Due to Measurement Setup. Some sources of error that are specific to radionuclides result from the measuring technique:

1) Errors due to radioactive decay are generally smaller than 0.2% since half-lifes are accurately known [3], [13], [30], [49].
2) *Statistical uncertainty* results rom counting a limited number N of radioactive decay events. Mostly, there is no problem in obtaining $N \approx 10^4$ with precision at the 1% level, as long as background is low.
3) *Geometric errors* result from uncertainties in the dimensions of the sample, detector, and the distance between them. They are often difficult to calculate on an absolute scale, and are therefore calibrated empirically. Depending on the calibration precision, on the type of nuclide, sample, and detector, the relative uncertainty can range from 50 to 0.2%. Calibration on an absolute scale depends on the use of accurate standards, which are rarely available with a certified accuracy better than 2–3%.
4) *Errors from absorption and backscattering* of radiation are generally corrected empirically as well.
5) *Errors from instrumental sources, quenching, and luminescence* can be kept very low (ca.

0.1%) by control of temperature and humidity, uniform count rates from samples and standards to suppress differences in dead time and pulse pile-up, and uniform matrix of samples and standards in liquid scintillation counting [37]. Otherwise, errors as high as 40% [37], [79] can be reduced to a few percent, either by utilizing fast detectors and electronics, or by empirically correcting for pulse pile-up losses, dead time, and spectrum instability [5], [13], [79].

8.3. Radiotracers in Methodological Studies

Radioanalytical chemistry offers a variety of accurate and highly sensitive methods. Even with the increased performance of other trace analysis methods [18]–[21], including electrothermal atomic absorption spectrometry, atomic emission, and fluorescence spectrometry (→ Atomic Spectroscopy), coupling methods, and high-performance sample introduction in organic and inorganic mass spectrometry (→ Mass Spectrometry), chromatographic separations (→ Gas Chromatography; → Liquid Chromatography), and neutron AA (→ Activation Analysis), there are still a number of areas where radionuclide applications are unsurpassed (see Section 8.1.4).

Picogram amounts of chemical species can be traced through a complete chemical process, and problems in process performance are often revealed and overcome rapidly and simply [18], [20], [22], [23], [28], [33], [40], [41], [45], [47], [50], [61], [78], [93], [97].

Important information can be obtained rapidly and cheaply by radioindicators. Specific procedures for quantitatie determinations based on radiotracers (see Chaps. 8.4, 8.5) involve additional principles.

8.3.1. Principles and Importance

Radioactive tracer analysis is based on the simple proportionality between the mass m_x of analyte x and the radioactivity A_1 initially added to the sample:

$$m_x = k \cdot A_1 \tag{3}$$

where k is a proportionality constant, measured for an individual experiment. This is based on the condition that the radioactive substance m_1 added to the sample does not increase the mass of the analyte, i.e., $m_x \gg m_1$. Otherwise, for determination of m_x from A_1 the formalism of IDA (Chap. 8.4) is applied, but very often the simple proportionality between mass $(m_x + m_1)$ and A_1 is sufficient for a methodological investigation.

By adding a radioactive tracer before the individual process stage, and following it by radioactivity measurements, the recovery, distribution, and loss are measured independently of any additional risk of contamination. By means of nuclear activation, radioactive tracers can be introduced into solid samples and can be used to determine the recovery and loss during decomposition of solids, and separation from solid samples by volatilization, leaching, and solvent extraction [9], [10], [40], [51] – [64], [67] – [80], [93], [97].

A variety of examples from the general analytical field [2], [7], [17], [25] and from trace element analysis [20], [33], [44], [46] covers separate stages of combined analytical procedures such as decomposition [20], [44], [45], evaporation [81], coprecipitation, adsorption, and washout of precipitates; contamination and adsorption [18], [33], separation by chromatography, ion-exchange, liquid – liquid distribution, and electrodeposition [4], [33], [34], [38], [40], [41], [78].

Radioanalytical indicator methods (including the radioreagent method, see Chap. 8.5) experienced a revival in the 1960s and early 1970s. Modern direct instrumental methods [19] underemphasized the importance of radioanalytical methods, which are necessary to establish the accuracy of spectroscopic methods, to avoid systematic errors, and to assess trace content in SRMs. An important advantage is the ability to reveal accidental losses, even at the ultra-trace level.

The most common areas of analytical applications are:

1) Determination of equilibrium constants, including solubility and complex dissociation constants, phase distribution in solvent extraction, adsorption, ion exchange, coprecipitation; and kinetic data, such as rate constants
2) Tracing the distribution of a chemical species in an analytical system for decomposition, chromatography, evaporation, and separation
3) Determination of recovery, yield, and loss of a chemical species in a process

8.3.2. Control of Sampling

Inaccuracies in sampling procedures occur, owing to contamination, loss, and nonrepresentative sampling. Sampling, pretreatment, storage, etc., all affect the accuracy.

Extended radioindicator studies on trace element sampling (e.g., of Al, Cr, Mn, Fe, Co, Zn, Sr, Cs, As, Hg, Pb) have been reviewed [4], [5], [20], [32]. In environmental analysis (e.g., solids, water, aerosols, biological tissue), adequate sampling and sample treatment is of primary importance to account for variation in biological activity due to the distribution of chemical species between different phases and solid fractions [5], [16], [20], [22], [26], [32], [46]. Sample homogeneity may be a significant limitation to the quality of analytical results. It is of increasing importance with decreasing concentrations from the µg/g to the ng/g level, and it limits the amounts of subsamples and the number of replicates. Studies on sample homogeneity, and on sampling and storage of very dilute solutions, are conveniently performed with radiotracers for numerous trace elements [16], [18], [20], [33], [44], [45].

8.3.3. Control of Contamination and Loss

Contamination generally poses the most important problem in ultra-trace analysis. It causes fluctuations of the blank values, thus defining the lower limits of detection, but also introduces systematic inaccuracies. Tracer techniques can be used to study the sources of contamination: reagent blanks, vessel walls, airborne pollutants, etc.

Contamination cannot be overcome by radiochemical means. Purification procedures for reagents and vessels and a clean working environment are required and SRMs should be used. Tracer techiques are unique tools in its investigation and control (for reviews see [18], [34], [40], [44]), owing to the advantages mentioned in Chapter 8.2 and Section 8.3.1 [18], [44]. Radioindicator studies have been conducted on reagent purification by electrolysis, sublimation, sub-boiling distillation, liquid – liquid distribution, ion exchange, and on vessel cleaning by rinsing, leaching, and steaming.

In general, the same principles are applied to radiochemical investigations of losses by adsorption (i.e., during storage, pretreatment, precipitation, filtration), by volatilization (during decomposition, ashing, storage, digestion, drying), and

by chemical reaction (complexation, ion exchange, photochemical and redox reaction) using radioindicators [20], [40], [45].

8.3.4. Separation Procedures

Separation stages and chemical reactions, e.g., for decomposition or phase transformations, are essential components of combined analytical procedures [52], [53]. They should be reproducible, quantitative, selective, and unequivocal.

Separation procedures are based on the principles of volatilization, liquid–liquid distribution, adsorption, diffusion, chromatography, ion exchange, electrophoresis, precipitation, coprecipitation, and electrodeposition. In all of these, radiotracers provide the best tool for methodological investigations, determination of equilibrium constants, kinetic data, and optimization of applied analytical data (yield, interference levels, etc.) [54]. Use of radiotracers in complex multielement separation schemes is reviewed in [4], [17], [20], [41], [54], radiochromatography is reviewed in [55], [61], [93], [97].

8.3.5. Control of the Determination Stage

Radionuclides in methodological studies of determination methods are sometimes useful. These are the measurement processes associated with the determination stage of a combined analytical procedure. Sources of systematic error in atomic absorption spectroscopy, optical emission spectrometry, and electrochemical methods, as well as optimization of the determination procedures have been examined.

Examples are the behavior of trace elements in graphite furnace atomizers [81] and investigations of the double layer structure on analytical electrodes and its exchange reactions with the solution [56].

8.4. Isotope Dilution Analysis

The principle of isotope dilution analysis (IDA) [31], [47], [90], [97], [98] involves measurement of the change in isotopic ratio when portions of a radiolabeled and nonlabeled form of the same chemical species are mixed. To perform a radioisotope IDA, an aliquot of a radioactive spike substance of known specific activity $a_i = A_i/m_i$ is

added to the test sample containing an unknown analyte mass m_x. This analyte mass is then calculated from a determination of the specific activity a_x of the resulting mixture after complete homogenization. The important advantage of IDA is that the analyte need not be isolated quantitatively. This is often significant at the trace concentration level, where quantitative separation is not feasible or inconvenient, or if interferences occur, e.g., in the analysis of mixtures of chemically similar compounds. Only a portion of the analyte is separated, so that the separation reagent is not used in excess, and this generally improves selectivity [31], [54], [57]–[59]. Selective determination of *chemical species* in various states (e.g., tri- and pentavalent As, Sb, tri- and tetravalent Ru, Ir) by IDA is favored by the supplementary selectivity gained from the use of substoichiometric amounts of reagent. Losses during purification and other steps are taken into account. A suitable separation procedure is needed to isolate part of the homogenized sample [2], [31].

In more advanced variants of IDA, measurements of masses are substituted by volumetric and complementary activity measurements.

The same principle is also applied to inactive stable isotope IDA. Mass spectrometric determination of isotope ratios then replaces the activity measurements [82] (→ Mass Spectrometry).

8.4.1. Direct Isotope Dilution Analysis

In this most frequently used version of IDA (single IDA), the mass m_x of inactive analyte substance can be determined by using the labeled substance of mass m_1, radioactivity A_1, and the specific activity a_1:

$$a_1 = A_1/m_1 \tag{4}$$

After homogenization with the analyte mass m_x, the specific activity has decreased to:

$$a_2 = \frac{A_1}{m_1 + m_x} = \frac{A_2}{m_2} \tag{5}$$

The total radioactivity of the system is unchanged:

$$(m_1 + m_x)a_2 = m_1 a_1 \tag{6}$$

The unknown mass m_x is calculated from:

$$m_x = m_1 \left(\frac{a_1}{a_2} - 1 \right) \tag{7}$$

The specific activities a_1 and a_2 are obtained from activities A_1 and A_2 and masses m_1 and m_2 of the fractions isolated from the initial labeled substance (Eq. 4) and from the homogenized solution (Eq. 5).

Equation (7) shows that IDA depends on the relative change in specific activity due to isotope dilution. Low-yield separations may be sufficient. For high precision, a_1 should be high, with mass $m_1 \ll m_x$. Then Equation (7) reduces to

$$m_x \approx m_2 \frac{A_1}{A_2} = \frac{A_1}{a_2} \qquad (8)$$

Equation (8) shows that the accuracy of m_x is limited by the physicochemical determination of m_2.

Example. Determination of glycine in hydrolyzed protein [98]: 152.6 mg hydrolyzed protein homogenized with 5.07 mg ^{14}C-labeled glycine of specific activity $a_1 = (96.2 \pm 1.2)$ counts min^{-1} mg^{-1} (relative activities are sufficient). After separation of a portion of glycine from the mixture, specific activity of $a_2 = (51.3 \pm 0.9)$ counts min^{-1} mg^{-1}. Equation (7) gives:

$$m_x = 5.07 \,\text{mg} \left(\frac{96.2}{51.3} - 1 \right) = (4.44 \pm 0.17)\,\text{mg}$$

The percentage of glycine in the protein is:

$$\frac{4.44}{152.6} \cdot 100 = (2.91 \pm 0.11)\%$$

8.4.2. Reverse Isotope Dilution Analysis

Another application of IDA is the determination of the activity A_y of a radioactive substance y from a complex mixture of radioactive substances by adding a known mass m_c of an inactive carrier substance. This technique is also based on Equation (7), reversely solved for activity A_1. Homogenization and separation of a portion m_2 (activity A_2, specific activity a_2) gives:

$$A_y = a_2 \left(m_c + m_y \right) \qquad (9)$$

from the specific activity a_y of substance y. If the inactive carrier is used in significant excess (i.e., $m_c \gg m_y$) the separation yield is:

$$A_y = A_2 \frac{m_c}{m_2} \qquad (10)$$

To obtain m_y, the specific activity a_y is required: $m_y = A_y / a_y$.

Reverse IDA (indirect IDA, or dilution with inactive isotopes) is particularly important in organic analysis and biochemistry [24], [32] to test for radiochemical purity and stability of labeled compounds. It is often used in radiochemical nuclear AA for the separation of activated element traces from a variety of interfering radionuclides in the analytical sample, for determination of the isolated activity, in comparison with that of a reference sample (i.e., to determine the radiochemical yield according to Eq. 10).

In contrast to direct IDA, reverse IDA is not limited by the sensitivity of the analytical determination of the isolated mass m_2, but merely by the specific activity a_y of the radioactive substance. By adjusting the carrier mass m_c, the sensitivity and accuracy can be increased, so reverse IDA is suitable for trace and microanalysis. Losses by absorption from carrier-free solutions must be taken into account. In separations from mixtures of radioactive substances, high-quality purification is important.

Double (Multiple) IDA. Since the specific activity a_y is often unknown, double IDA can be used to determine m_y. Two equal aliquots containing the same, but unknown mass m_y are diluted with different known carrier masses, m_c and m'_c. After separation of portions m_2, m'_2 and measurement of specific activities a_2, a'_2 one obtains from Equation (9):

$$a_2 = \frac{A_y}{m_c + m_y} = \frac{a_y m_y}{m_c + m_y} \qquad (11)$$

$$a'_2 = \frac{a_y m_y}{m'_c + m_y} \qquad (12)$$

Thus:

$$m_y = \frac{a'_2 m'_c - a_2 m_c}{a_2 - a'_2} \qquad (13)$$

$$a_y = \frac{\left(m_c - m'_c \right) a_2 a'_2}{a'_2 m_c - a_2 m'_c} \qquad (14)$$

Multiple (direct and reverse) IDA produces a calibration graph $m_x = f(1/a_2)$.

8.4.3. Derivative Isotope Dilution Analysis

If the preparation of labeled substances is too difficult, derivative IDA may be applied. This technique combines a radioreagent and IDA, pref-

erably of low selectivity, to react with one or more analytes. It is mainly used with mixtures of complex organic compounds. The basic stages are:

1) Reaction of analyte A with radioreagent B*, of known specific activity, to form a radioactive product AB*. Excess B* may be removed in an optional purification step:

$$A + B^* \longrightarrow AB^* + B^*_{ex} \qquad (15)$$
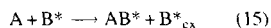
2) The mass of AB* is determined by mixture with a known mass of inactive AB by reverse IDA:

$$AB^* + B^*_{ex} + AB \longrightarrow (AB^* + AB) + B^*_{ex} \qquad (16)$$
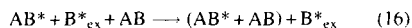
3) Separation of diluted product (AB* + AB) and determination of its specific activity.

The principal requirements for this technique are:

1) Reaction (Eq. 15) should be quantitative or of known yield.
2) Isotope exchange between substances AB and B* must be prevented. Since inorganic substances are quite amenable to such exchange, application of derivative IDA to inorganic analytes is impossible.
3) Purification from excess reagent must be quantitative to avoid significant bias, even from minor impurities.

As a variant, dilution with A* is applied prior to the first reaction (Eq. 15) with radioreagent B*.

Determination of the reaction yield and analysis of inorganic substances are possible with this variant technique [2]. Its main importance is in the determination of trace amounts in complex mixtures. It is routinely used in biochemistry and physiological chemistry [2], [31], [32], [75]. The simpler approaches of radioimmunoassays are generally preferred.

Further variants are based on combinations with double derivative IDA, or addition of an inactive derivative prior to direct IDA.

8.4.4. Substoichiometric Isotope Dilution Analysis

Utilization of conventional IDA for trace determinations is limited by the necessity to isolate macroscopic amounts of the analyte substance for determination of m_1 and m_2 by physicochemical methods. The principle of substoichiometry developed by Ružička and Starý [57] avoids this limitation.

Exactly equal (low) masses m_1 and m_2 are isolated from the spike solution of initial specific activity $a_1 = A_1/m_1$ and from the solution after homogenization with the analyte ($a_2 = A_2/m_2$). Equation (7) reduces to:

$$m_x = m_1 \left(\frac{A_1}{A_2} - 1 \right) \qquad (17)$$

The masses m_1 and m_2 are isolated using the same amount of the separation reagent, which must be stoichiometrically insufficient and consumed quantitatively (or to exactly the same amount) in this reaction. It must form a product which can easily be separated from the excess of unreacted substance. Separation reagents should have high chemical stability, no tendency to adsorption at very low concentrations, and sufficient selectivity for the analyte. Separations with enrichment factors of $10^4 - 10^5$ are typical. Since the mass m_1 of the added radioactive substance must be accurately known, reverse IDA can be applied.

Error propagation is comprehensively discussed by Tölgyessy and Kyrš [2]. At trace levels below 1 µg, standard deviations of ca. 3 % and accuracies of ca. 6 % are typical; precision and accuracy can be 1.5 % or better at higher concentrations.

8.4.4.1. Substoichiometric Separation by Liquid – Liquid Distribution

The extraction of metal chelates is suitable for separating equal masses of metals from solutions of different concentrations. Extraction of a metal chelate MA from a metal M^{n+} by an organic reagent HA is generally described by [47], [57], [58]:

$$M^{n+} + n(HA)_{org} \rightleftharpoons (MA_n)_{org} + nH^+ \qquad (18)$$
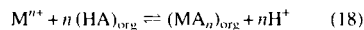
with the extraction constant K:

$$K = \frac{[MA_n]_{org} [H]^n}{[M^{n+}][HA]^n_{org}} \qquad (19)$$

where subscript org denotes the organic phase. In trace element analysis by substoichiometric IDA, concentrations of $10^{-6} - 10^{-9}$ mol/L are typically extracted from aqueous samples of volumes V in the range 10 mL to 100 µL into organic phases of smaller volumes (typically one-tenth of the sample

volume) of $10^{-5} - 10^{-8}$ mol/L solutions of chelating reagent.

The dependence of Equation (18) on the pH value of the solution makes it necessary to keep within an optimum pH range. The substoichiometric principle requires a higher than 99 % consumption of the initial concentration c_{iHA} in Equation (18), so that $[HA]_{org} < 0.01 c_{iHA}$. Thus the solution must satisfy the condition:

$$pH > -\log(0.01 c_{iHA}) - \frac{1}{n} \log K \qquad (20)$$

With increasing pH, dissociation of acid HA (dissociation constant K_{HA}) in the aqueous phase increases. This dissociation is negligible at:

$$pH < pK_{HA} + \log q_{HA} + \log \frac{V_{org}}{V} \qquad (21)$$

where q_{HA} is the distribution coefficient of the reagent, V is the sample volume, and V_{org} is the volume of the organic phase. Hydrolysis of the metal at increasing pH must not interfere with the separation.

The reagent (e.g., diphenylthiocarbazone, diethyldithiocarbamic acid (2-thenoxy)-3,3,3-trifluoroacetone, 8-hydroxyquinoline) must be stable to decomposition, even at very low concentrations in acid and neutral media, and must form an extractable chelate with a sufficiently high value of K (for further useful ligands see [38], [40], [58], [83]). Optimum conditions for substoichiometric extraction are calculated from K, K_{HA}, and q_{HA}. Selectivity of the separation can be improved by use of additional masking reagents [54], [57].

Substoichiometric extraction of ion-association complexes is generally restricted to higher concentration levels (e.g., determination of Na, P). Water-soluble complexes of more than 20 metals with ethylenediaminotetraacetic acid can be used for substoichiometric determination at concentrations below 1 ng/g, provided that the excess of unreacted metal is separated by liquid–liquid distribution, ion-exchange, carrier coprecipitation, or other suitable means. Selectivity is high and can be further increased by masking reagents.

8.4.4.2. Redox Substoichiometry

Oxidation or reduction by a substoichiometric mass of reagent is followed by separation, extraction, coprecipitation, etc. [73]. $KMnO_4$, $K_2Cr_2O_7$, and $FeSO_4$ have been used as substoichiometric redox reagents.

8.4.4.3. Displacement Substoichiometry

Complete extraction of an analyte metal M_1 with excess complexing reagent is followed by removal of the excess and subsequent displacement of M_1 by a substoichiometric amount of another metal M_2, provided that M_2 has a sufficiently higher extraction constant K_2. Further variants are discussed elsewhere [2].

8.4.4.4. Applications

Applications include trace element determination in rocks, soils, biogenic samples, and pure metals by direct IDA, i.e., added radiotracer [2], [54], [60]. Another field of increasing importance is in radiochemical AA for assessment and improvement of accuracy at trace and ultra-trace levels by determination of the radiochemical yield. This applies primarily to biomedical samples, high-purity materials, and environmental materials 7]–[9], [32], [84].

8.4.5. Sub- and Superequivalence Method

The sub- and superequivalence method is based on the isoconcentration principle. Three versions of this method exist: the direct method for the determination of nonradioactive substances [99], the reverse method for the determination of radioactive substances [100], and the universal isotope dilution method for determination of both [101].

Historically, the *reverse method* has been developed first [100]. In the reverse method, two series of aliquots are formed from a solution of the radioactive (labeled) analyte. The first series is formed by equal volumes of the analyte solution [containing $x = (x,...,x,...,x)$ amounts of labeled analyte] while the (identical) aliquots of the second series contain ξ times higher volumes of the same solution ($x = \xi x,..., \xi x,..., \xi x$). The members of the first series are isotopically diluted by successively adding fixed amounts of the same substance in inactive form $y = (y_1,...,y_\xi,...,y_a)$, and so the members of the first series contain $(x + y) = (x + y_1, ..., x + y_\xi,..., x + y_a)$ amounts of the substance. After establishing the same separation conditions for all aliquots of both series, e.g., by adding masking and buffer solutions, and after dilution to the same volume, a fixed amount of reagent is added to all aliquots. Then the separation is performed at the same temperature and pressure.

The added amount of reagent is fixed well in the aliquot containing the highest total concentration of analyte (substoichiometric regime), but it is in excess in the aliquot containing the lowest total amount of substance. The ratio of the activities of the isolated fractions from the second series (which are theoretically identical) and the activities of the isolated fractions of particular members of the first series are plotted as a function of the dilution increment and the mass of analyte is determined from this plot [2], [73], [89].

In the *direct method* two series of aliquots are formed from a solution of the labeled counterpart of the substance to be analyzed (which is not radioactive). The first series is formed by increasing volumes of this solution containing ξy_1, ξy_2, ...ξy_a amounts of the labeled substance. The second series is formed by increasing, but ξ times lower amounts of the same solution. To the aliquots of the second series constant amounts of the substance to be analyzed (in nonradioactive form) are added. After establishing the same separation conditions for all aliquots of both series, the separation is performed. The evaluation of the analytical results is identical to that in the reverse method.

The *universal method* (two partially different procedures) is (are) the combination of both methods.

Besides a large variety of metals, several chemical species of specific interest, e.g., cyclic 3',5-adenosine monophosphate, thyroxine and 3,5,3'-trioiodothyronine have been determined by sub- and superequivalence IDA and the results show higher accuracy of the method over that of radioimmunoassay [102], [103]. The principle of the method also allows the determination of certain physicochemical characteristics, e.g., estimation of the stability constants of complexes and the solubility of precipitates.

The method is a calibration curve approach. It does not require that all the separated amounts of the aliquots mentioned are exactly the same, as in substoichiometric IDA (Section 8.4.4). This is the reason why separation reactions can be used that are not quantitative, a sufficient prerequisite is reproducibility [2], [24], [74].

An alternative calibration curve approach for nonquantitative reactions is to determine the distribution coefficient by a concentration-dependent distribution method [2], [75], which is a variant of the radioreagent method (Section 8.5.1).

8.5. Radioreagent Methods

Radioreagent methods (RRM) are based on the use of a radioactive species in a quantitative reaction, and measurement of the change in activity of that species in the course of the reaction. The radioactive species may be a labeled reagent, the analyte, or a substance able to undergo an exchange reaction with some compound of the analyte.

After separation from excess reagent (by liquid-liquid distribution, chromatography, precipitation, etc.), the mass or concentration of this product is determined from activity measurement. The determination is based on a radioactive substance chemically different from the analyte substance (in contrast to IDA), therefore the chemical reaction is of prime importance. By variation of this key reaction, the principle can be adapted to various procedures. The superiority of radioreagent methods over classical separation techniques arises from the use of an inactive carrier and the high sensitivity of the activity measurements, which are not subject to interference by the carrier or other substances [1], [16], [24], [34], [60], [87], [92], [94], [95], [97].

Depending on the type of chemical interaction and relationship of the analyte to the radioactive substance measured, RRM can be divided into three basic group:

1) The radioactive substance is a typical reagent which is able to react with the compound to be determined.
2) The method is based on "nonequivalent" competition; i.e., the radioactive substance is able to exchange with the analyte.
3) The method is based on "equivalent" competition; i.e., the labeled substance is chemically identical with the substance to be determined.

From the practical point of view, it is convenient to divide RRM into the following groups:

1) Simple RRM in which the reactions take place quantitatively, i.e., either the analyte or the reagent is completely consumed and compounds of definite composition are formed.
2) Methods of concentration-dependent distribution utilizing reactions where products of unstable composition are formed, but in which the extent of reaction is determined by the corresponding equilibrium constant or reaction time.

Table 3. Useful concentration regime of the radioreagent method

Analyte	Radionuclide	Labeled reagent	Separation	Concentration regime
H^+	^{131}I	KI	extr.I_2 into pyridine/chloroform	> 0.1 µg/mL
H_3BO_3	^{18}F	HF	extr.HBF_4 into 1,2-dichloroethane[b]	≈ 0.1 µg B
NaLS[c]	^{59}Fe	$[Fe(II)(phen)_3]^{c}$	extr.$Fe(phen)_3(LS)_2$ into $ChCl_3$	5 – 300 ng/mL Na
Cl^-, Br^-, I^-	^{203}Hg	$C_6H_5Hg^+$	extr.C_6H_5HgCl into benzene	0.5 – 15 mg/mL
Cl^-, I^-	^{110m}Ag	$AgNO_3$	pptn.$AgCl$; AgI	0.4 – 250 µg/mL
Cl^-	^{203}Hg	$HgNO_3$	pptn.Hg_2Cl_2	0.8 – 13 µg/mL
Bi	^{131}I	KI^{d}	extr.$HBiI_4$ into n-butylacetate	0.04 – 4 µg
Cationic detergents	^{131}I	Rose Bengal	extr. ion-associate into $CHCl_3$	

[a] Extr. = solvent extraction; pptn. = precipitation. [b] In presence of methylene blue. [c] LS = lauryl sulfate; phen = 1,10 phenanthroline. [d] In presence of ascorbic acid + Na_2SO_3 in 0.6 mol/L H_2SO_4.

3) Isotope exchange methods based on the exchange of isotopes between two different compounds of one element, with a radioactive isotope in one of the compounds and a nonradioactive isotope in the other, usually the analyte.

4) Radiorelease methods comprising procedures in which the analyte reacts with a radioactive reagent, thus releasing an aliquot of the reagent activity, in most cases into the gas phase.

5) Radiometric titration usually involves use of radioactive reagents to determine the equivalence point, but there are variants that do not use a radioreagent; instead they are based on a change in the intensity of radiation caused by its absorption or scattering in a medium containing the analyte [2], [32], [87], [92], [95], [96], [97].

8.5.1. Simple Radioreagent Methods

Simple radioreagent methods (SRRM) utilize a reaction with a suitable reagent, and sometimes with a third substance; one of the reagents is radioactive and the reaction is quantitative [2], [32], [87], [95]. SRRM can be classified into the following three groups:

1) Determination with labeled reagents,
2) Determination with labeled analyte,
3) Determination with labeled competing substance

8.5.1.1. Determination with Labeled Reagents

An excess of radioactive reagent solution of known analytical concentration is usually used. The active product is separated from the excess of radioactive reagents by precipitation, formation of extractable chelates, or sorption.

The substance to be determined (n_x mol) reacts with a radioactive reagent (n_R mol), forming a precipitate, extractable compound, or other separable substance. After the separation of the product, the radioactivity of the product (A_P), that of the excess of unreacted reagent (A_E), or both are measured. The following relationship is valid:

$$\frac{A_R}{n_R} = \frac{A_P}{z\,n_x} = \frac{A_E}{n_R - z\,n_x}$$

Thus

$$n_x = \frac{n_R A_P}{z A_R} = \frac{n_R}{z}\left(1 - \frac{A_E}{A_R}\right)$$

where A_R is the radioactivity of the reagent and z denotes the stoichiometric ratio in the compound formed (the number of moles of the reagent interacting with 1 mol of the test substance). If n is not exactly known, a calibration curve approach can be used. Quantitative isolation of trace amounts of the reaction products poses another important problem (Table 3, [2], [41], [47], [60]).

In this way, the concentration of chloride in water was determined by precipitation with silver labeled with ^{110m}Ag; ^{14}CO was used for the determination of hemoglobin in blood; and ^{131}I for the determination of bismuth in the presence of an excess of iodide ions using extraction separation [2].

8.5.1.2. Determination with Labeled Analyte

The first step involves labeling of the analyte. To this end, the radioactive indicator is added to the unknown sample and isotope exchange is allowed to take place; then the reagent is added in a substoichiometric but known quantity. After the reaction is complete, the phases are separated

and their respective activities determined. In those cases where the activity of the equilibrium solution is measured, the unknown concentration can be calculated from:

$$x + x_0 = \frac{A_{init}B}{(A_{init} - A_{equilib})}$$

where x is the quantity of analyte (equivalents), x_0 the amount of added radioactive indicator (equivalents), B the amount of unlabeled reagent (equivalents), and A_{init} and $A_{equilib}$ represent the total activity of the initial and equilibrium solutions, respectively. In this variant of the RRM, solvent extraction is frequently used.

8.5.1.3. Determination with Labeled Competing Substances

Analyte M_2 competes with the radioactively labeled substance $^*M_1$ in the formation of a compound with a reagent R, where the compound precipitates from the solution

$$^*M_1R \downarrow + M_2 \rightleftharpoons M_2R \downarrow + {}^*M_1$$
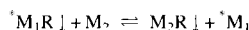
$^*M_1R$ and M_2R occur in a different phase than either substance M_1 or M_2. The equilibrium constant of the reaction, as a rule is > 1 so that a significant change takes place.

This method is used when the labeling of reagent R cannot be applied or when the reagent and its compound with the analyte are transferred into the same phase. Generally, a relatively soluble precipitate with one radioactively labeled component (e.g., $^{45}CaCO_3$) is brought into contact with a solution containing the analyte (Pb^{2+}), which displaces the radioactive component ($^{45}Ca^{2+}$) from the precipitate into the solution, and a less soluble precipitate ($PbCO_3$) is formed. The resulting radioactivity of the solution is proportional to the initial amount of lead.

8.5.2. Method of Concentration-Dependent Distribution

Concentration-dependent distribution (CDD) is based on the utilization of a calibration graph that shows the dependence of the distribution ratio of the substance to be determined in a two-phase system on the total concentration of the substance [2], [32], [87], [92], [95], [96]. Characteristic features of the method of CDD are as follows:

- The analysis is based on the distribution of a radioactive substance between two phases or between several parts of the systems used (paper chromatography, electrophoresis),
- The ratio of the activities in both phases or parts of the system depends strongly on the initial concentration of the unknown substance,
- The distribution is not given by the stoichiometric ratio of the reacting substances, i.e., not by total saturation of one phase or a nearly complete consumption of the reagent, but is given exclusively by the corresponding equilibrium or kinetic constants.

Example. Determination of chloride in water samples. The sample is shaken with an excess of labeled AgCl precipitate and the radioactivity of silver is measured in the equilibrium solution. The higher the chloride concentration the lower is the concentration of silver in solution (owing to the constancy of the AgCl solubility product). The Cl^- ions determined do not react with the AgCl precipitate, they only suppress its dissolution. The amount of chloride in the precipitate has no direct relationship to the concentration of Cl^- ions in the sample solution.

In CCD is useful to define two extreme types of determination: *saturation analysis* and *nonsaturation analysis*.

In *saturation analysis*, the cause of the change in the distribution of the radioactive substance between the phases is an increase in the total saturation of the reagent by the given substance. This term originally appeared in the literature in connection with the determination of biochemically important substances (steroids, hormones, vitamins) and is equivalent to the term radioimmunoassay.

In *nonsaturation analysis*, the cause of the change in the distribution coefficient is the shift of the chemical equilibrium due to an increase in the concentration of the substance to be determined, irrespective of the extent of saturation. The determination of some extractants can serve as a typical example. The number of procedures described that use nonsaturation analysis is limited.

8.5.3. Isotope Exchange Methods

Isotope exchange methods (IEM) are based on the exchange of isotopes between two different compounds of the element X, one of which (XA) is nonradioactive, the other (XB) being labeled with a radioactive isotope [2], [61], [87], [97]. After isotopic equilibrium is reached

$$XA + {}^*XB \rightleftharpoons {}^*XA + XB$$
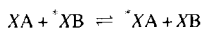
the specific activities of the element X in both compounds are equal

$$A_1/m = A_2/m_x$$

where A_1 and A_2 are the equilibrium radioactivities of XB and XA, respectively, and m_x and m are the amounts of X in XA and XB, respectively. The value of m_x can be computed from:

$$m_x = mA_2/A_1$$

or from the calibration graph of $m_x = f(A_2/A_1)$, $m_x = f(A_2)$ or $m_x = f(A_1)$.

Isotopic exchange can be carried out in either a heterogeneous or a homogeneous system.

Example. Methylmercury and phenylmercury are determined down to 10 pg/mL by liquid-liquid extraction from 3 mol/L aqueous HCl into benzene, and addition of $K^{131}I$ to the separated organic phase. Inactive chloride is completely displaced by ^{131}I:

$$(CH_3HgCl)_{org} + {}^{131}I^- \longrightarrow (CH_3Hg^{131}I)_{org} + Cl^-$$
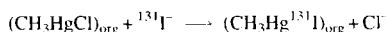
The method has been used for determination of methylmercury species in fish and in drinking water [61].

8.5.4. Radioimmunoassay

Radioimmunoassay is the widely used radio-reagent method.

8.5.5. Radiorelease Methods

In radiorelease methods the analyte substance A reacts with a radioactive reagent, so that radioactive R^* is released into a second phase, without being replaced by an inactive analyte [62]. Applications involve either release of radioreagents from solids or liquids into the gas phase, or release from solids into a liquid. Radiorelease methods may be classified according to the type of the radioactive reagents employed, i.e. (1) radioactive kryptonates, (2) radioactive metals, and (3) radioactive salts and other substances.

8.5.5.1. Radioactive Kryptonates [2], [87], [92], [94], [95]

The term radioactive kryptonates is used for substances into which atoms or ions of krypton-85 (^{85}Kr) are incorporated (by diffusion of ^{85}Kr; by

bombardment with accelerated krypton ions; by crystallization of the kryptonated substances from a melt, or by placing the solution in an atmosphere of ^{85}Kr, etc.). The radionuclide can be released from the solid lattice by any chemical or physical reaction that breaks down the lattice at the solid surface. The released inert gas is conveniently measured from its main β radiation (0.67 MeV). The half life of 10.27 a is suitable for long-term remote sensing devices.

Applications of Radioactive Kryptonates [32], [92], [94]–[96]. The determination of *oxygen* is performed by surface oxidation of copper or pyro-graphite kryptonate, at elevated temperature, resulting in destruction of the surface layer and release of ^{85}Kr proportional to the oxygen mass. Detection limits are at the 10 ng/m^3 level. Ozone oxidizes copper kryptonate at temperatures below 100 °C, whereas reaction with oxygen starts well above 200 °C, so this detector can detect O_3 and O_2 differentially. Determination of *ozone* in air is feasible over a concentration range of $10^{-7} - 10^{-3}$ g/m^3 with hydroquinone kryptonate:

$$[C_6H_4(OH)_2]_3[^{85}Kr] + O_3 \longrightarrow 3\,C_6H_4O_2 + 3\,H_2O + {}^{85}Kr\uparrow$$
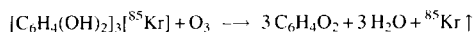
Sulfur dioxide has also been determined by a method based on the mechanism of double release. In the first stage, sulfur dioxide reacts with sodium chlorate to release chlorine dioxide which is a strong oxidizing agent. The chlorine dioxide then oxidizes radioactive hydroquinone kryptonate and gaseous ^{85}Kr is released. The following reactions are invoved:

$$SO_2 + 2\,NaClO_3 \longrightarrow 2\,ClO_2 + Na_2SO_4$$

$$ClO_2 + [C_6H_4(OH)_2]_3[^{85}Kr] \longrightarrow {}^{85}Kr\uparrow$$
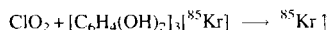
A radioactive kryptonate of silica has been suggested for the determination of *hydrogen fluoride* in air. *Oxygen dissolved in water* can be measured by the use of thallium kryptonate

$$4\,Tl[^{85}Kr] + O_2 + 2\,H_2O \longrightarrow Tl^+ + 4\,OH^- + {}^{85}Kr\uparrow$$
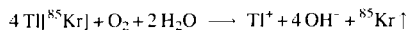
Thallium is oxidized by oxygen, and the amount of ^{85}Kr released or the decrease in activity of kryptonated thallium is proportional to the dissolved oxygen concentration in the sample (at constant pH) down to 0.3 µg/mL.

8.5.5.2. Radioactive Metals

Oxidizing agents in solution can react at the surface of a labeled metal, releasing radioactive ions into the solutions which are used for determination of the oxidants. The decisive factors in the choice of the metal are the following: the metal should not react with water, but with oxidizing agents to yield ions which do not form precipitates in aqueous media. The metal should have a radionuclide with suitable nuclear properties. These conditions are met by thallium (^{204}Tl) and silver (^{110m}Ag) [54], [60], [61].

Dissolved *oxygen* has been determined in seawater, drinking water, and wastewater down to the ng/g level with metallic thallium labeled with ^{204}Tl. Selectivity requires removal or masking of other oxidizing agents. The *vanadate* ion is assayed by acidifying the corresponding sample (pH ≈ 3) and passing it over a column containing radioactive metallic ^{110m}Ag. The labeled silver is oxidized, dissolved, eluted from the column, and detected. The measurement of *dichromate* ion concentration in natural waters can be carried out similarly.

8.5.5.3. Radioactive Salts and Other Radioactive Substances [2], [32], [92], [96]

Sulfur dioxid is determined on the basis of the reaction:

$$5 SO_2 + 2 K^{131}IO_3 + 4 H_2O \longrightarrow K_2SO_4 + 4 H_2SO_4 + {}^{131}I_2 \uparrow$$
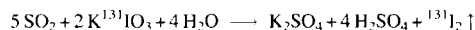
This reaction takes place in an alkaline solution through which air containing sulfur dioxide is bubbled. After completion of the reaction, the solution is acidified and the iodine released is extracted. *Active hydrogen* in organic substances may be determined by reaction with lithium aluminum hydride labeled with tritium (^{3}H). The activity of released tritium is measured using a proportional counter.

A number of determinations are based on the formation of a soluble complex between the analyte in solution with a radioactively labeled precipitate. In this way it is possible to determine anions forming soluble complexes (e.g., CN^-, $S_2O_3^-$, I^-, F^-). The principle of the determination are given in the following equations:

$$2 CN^- + {}^{203}Hg(IO_3)_2\downarrow \longrightarrow {}^{203}Hg(CN)_2 + 2 IO_3^-$$
$$(> 0.50\ \mu g/mL,\ \pm 55\ \%)$$

$$2 Cl^- + {}^{203}Hg(IO_3)_2\downarrow \longrightarrow {}^{203}HgCl_2 + 2 IO_3^-\ (> 1\ \mu g/mL,\ \pm 5\ \%)$$

$$n\ S_2O_3^{2-} + m\ {}^{110m}AgSCN\downarrow \longrightarrow {}^{110m}Ag_m(S_2O_3)_2^{(2n-m)-} + m\ SCN^-$$

$$2 I^- + {}^{203}HgI_2\downarrow \longrightarrow {}^{203}HgI_4^{2-}\ (> 25\ \mu g/mL,\ \pm 5\ \%)$$
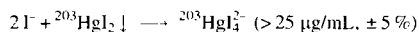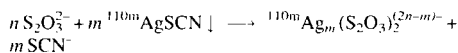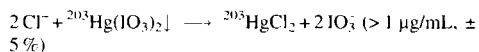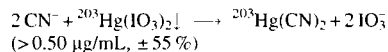
8.5.6. Radiometric Titration

Radiometric titrations follow the relation between the radioactivity of one component or phase of the solution under analysis and the volume of added titrant. The compound formed during the titration must be easily separable from the excess of unreacted ions. This separation is directly ensured only in the case of precipitation reactions. In other types of reactions, the separation can be accomplished using an additional procedure. The endpoint is determined from the change in the activity of the residual solution or of the other phase.

According to the type of chemical reaction used, methods based on the formation of precipitates and methods based on complex formation can be distinguished. Because of the necessity for handling precipitates, precipitation radiometric titrations are difficult to apply to less than milligram amounts and, therefore, have no special advantages over other volumetric methods. The sensitivity of complexometric titrations is limited by the sensitivity of the determination of the endpoint. However, the use of radiometric detection can substantially increase the sensitivity of this type of determination. For the separation of the product from the initial component, liquid-liquid distribution, ion-exchange, electrophoresis, or paper chromatography are most often used [2], [63], [88], [93], [97].

This application of radiometric titrations has declined over the past three decades [64]. Their main advantage is where classical methods for detection of the endpoint are either impossible or subject to interference from the titration medium.

Radiopolarography offers highly increased sensitivity and selectivity over polarographic current measurement, without interference from major components of the solution. It measures the amount of labeled ions deposited in single drops in a dropping mercury electrode as a function of potential [85].

8.6. References

General References

[1] IUPAC Commission on Radiochemistry, Nuclear Techniques, Anal. Chem. Div., IUPAC, *Pure Appl. Chem.* **63** (1991) 1269–1306.

[2] J. Tölgyessy, M. Kyrš: *Radioanalytical Chemistry*, vol. 1, Ellis Horwood, Chichester 1989, p. 354 ff.

[3] J. Tölgyessy, M. Kyrš: *Radioanalytical Chemistry*, vol. 2, Ellis Horwood, Chichester 1989, p. 498 ff.

[4] H. A. Das, A. Faanhof, H. A. van der Sloot: "Radioanalysis in Geochemistry," *Dev. Geochem.* **5** (1989) 482 pp.

[5] P. Bode, H. T. Wolterbeek, *J. Trace Microprobe Tech.* **81** (1990) 121–138.

[6] D. De Soete, R. Gijbels, J. Hoste: *Neutron Activation Analysis*, Wiley-Interscience, London 1972.

[7] W. D. Ehmann, D. E. Vance, *CRC Crit. Rev. in Anal. Chem.* **20** (1989) 405–443.

[8] N. M. Spyrou, *J. Trace Microprobe Tech.* **6** (1988) 603–619.

[9] J. Hoste, F. De Corte, W. Maenhaut in P. J. Elving, V. Krivan, E. Kolthoff (eds.): *Treatise on Analytical Chemistry*, 2nd ed., **part I, vol. 14**, J. Wiley & Sons, New York 1986, p. 645.

[10] V. Krivan in W. Fresenius et al. (eds.): *Analytiker Taschenbuch*, **vol. 5**, Springer-Verlag, Berlin 1985, p. 36.

[11] Y. Suzuki, *Radioisotopes* **36** (1987) 414–421.

[12] C. G. Clayton, M. R. Wormald, J. S. Schweitzer in IAEA (ed.): *Proc. Int. Sympos. Nucl. Techn. Explor. Exploit, Energy Miner. Resourc.*, IAEA, Vienna 1991, p. 107.

[13] G. F. Knoll: *Radiation Detection and Measurement*, 2nd ed., J. Wiley, New York 1989, p. 754.

[14] M. F. L'Annunziata: *Radionuclide Tracers: Their Detection and Measurement*, Academic Press, New York 1987, p. 505.

[15] L. I. Wiebe, *Radiat. Phys. Chem.* **24** (1984) 365–372.

[16] H. Filthuth, *Chem. Anal. (N.Y.)* **108** (1990) 167–183.

[17] H. J. M. Bowen: "Nuclear Activation and Radioisotopic Methods of Analysis," in P. J. Elving, V. Krivan, E. Kolthoff (eds.): *Treatise on Analytical Chemistry*, 2nd ed., part I, Sect. K, vol. 14, J. Wiley & Sons, New York 1986, p. 193.

[18] G. Tölg, *Analyst (London)* **112** (1987) 365–376.

[19] J. A. C. Broekaert, G. Tölg, *Fresenius' Z. Anal. Chem.* **326** (1987) 495–509.

[20] V. Krivan: "Activation and Radioisotopic Methods of Analysis," in P. J. Elving, V. Krivan, I. M. Kolthoff (eds.): *Treatise on Analytical Chemistry*, part I, Sect. K, 2nd ed., vol. 14, J. Wiley & Sons, New York 1986, p. 340.

[21] M. Linscheid, *Nachr. Chem. Tech. Lab.* **39** (1991) 132–137.

[22] W. R. Wolf, M. Stoeppler: "Proc. 4th Int. Symp. Biol. Environm. Reference Mater.," *Fresenius' J. Anal. Chem.* **338** (1990) 359–581.

[23] L. L. Jackson et al., *Anal. Chem.* **63** (1991) 33 R–48 R.

[24] J. W. Ferkany, *Life Sci.* **41** (1987) 881–884.

[25] W. D. Ehmann, J. D. Robertson, St. W. Yates, *Anal. Chem.* **62** (1990) 50 R–70 R;**64** (1992) 1 R–22 R.

[26] R. Naeumann, E. Steinnes, V. P. Guinn, *J. radioanal. Nucl. Chem.* **168** (1993) 61–68.

[27] R. R. Greenberg, *J. Radioanal. Nucl. Chem.* **113** (1987) 233–247.

[28] V. P. Guinn, H. S. Hsia, N. L. Turglio, *J. Radioanal. Nucl. Chem.* **123** (1988) 249–257.

[29] L. Grodzins, *Nucl. Instrum. Methods Phys. Res. Sect. B* **B 56–B 57** (1991) 829–833.

[30] F. De Corte, A. Simonits, *J. Radioanal. Nucl. Chem.* **133** (1989) 43–130.

[31] J. Tölgyessy, T. Braun, M. Kyrš: *Isotope Dilution Analysis*, Pergamon Press, Oxford 1972.

[32] J. Tölgyessy, E. H. Klehr: *Nuclear Environmental Chemical Analysis*, Ellis Horwood, Chichester 1987.

[33] V. Krivan, *Sci. Total Environ.* **64** (1987) 21–40.

[34] J. R. Jones (ed.): *Isotopes – Essential Chemistry and Applications*, 2nd ed., Royal Society of Chemistry, CRC Press, Boca Raton 1988, p. 270.

[35] G. R. Choppin, J. Rydberg: *Nuclear Chemistry: Theory and Applications*, Pergamon Press, Oxford 1980, p. 75.

[36] J. Mandel in I. M. Kolthoff, P. J. Elving (eds.): *Treatise on Analytical Chemistry*, **part I, vol. 1**, J. Wiley & Sons, New York 1978, p. 243.

[37] C. T. Peng: *Sample Preparation for Liquid Scintillation Counting*, Amersham International, United Kingdom, 1977.

[38] K. Robards, P. Starr, E. Patsalides, *Analyst (London)* **116** (1991) 1247–1273.

[39] A. W. Rogers: *Techniques of Autoradiography*, 3rd ed., Elsevier, Amsterdam 1979.

[40] A. Mizuike: *Enrichment Techniques for Inorganic Trace Analysis*, Springer-Verlag, Berlin 1983, p. 144.

[41] A. Dyer, *Analyst (London)* **114** (1989) 265–267.

[42] V. Krivan, *Talanta* **29** (1982) 1041–1050.

[43] Yu. A. Zolotow, M. Grasserbauer, *Pure Appl. Chem.* **57** (1985) 1133–1152.

[44] J. Versieck, R. Cornelis: *Trace Elements in Human Plasma or Serum*, CRC Press, Boca Raton 1989, p. 23.

[45] P. Tschöpel, G. Tölg, *J. Trace Microprobe Tech.* **1** (1982) 1–77.

[46] R. Cornelis, *Mikrochim. Acta* 1991 (1991) no. III, 37–44.

[47] *Ullmann*, 4th ed., **5**, 709.

[48] L. Melander, W. H. Saunders: *Reaction Rates of Isotopic Molecules*, Wiley-Interscience, New York 1980.

[49] K. H. Lieser: "Activation and Radioisotopic Methods of Analysis," in P. J. Elving, V. Krivan, I. M. Kolthoff (eds.): *Treatise on Analytical Chemistry*, part I, Sect. K, 2nd ed., vol. 14, J. Wiley & Sons, New York 1986, p. 1.

[50] W. Schmid, V. Krivan, *Anal. Chem.* **57** (1985) 30–34.

[51] R. Cornelis, J. Versieck: "Activation and Radioisotopic Methods of Analysis," in P. J. Elving, V. Krivan, I. M. Kolthoff (eds.): *Treatise on Analytical Chemistry*, part I, Sect. K, 2nd ed., vol. 14, J. Wiley & Sons, New York 1986, p. 665.

[52] *Ullmann*, 4th ed., **5**, 1.

[53] G. Tölg et al., *Pure Appl. Chem.* **60** (1988) 1417–1424.

[54] J. Starý, J. Růžička in G. Svehla (ed.): *Wilson and Wilson's Comprehensive Analytical Chemistry*, vol. 7, Elsevier, Amsterdam 1976.

[55] D. M. Wieland, M. C. Tobes, T. J. Mangner: *Analytical and Chromatographic Techniques in Radiopharmaceutical Chemistry*, Springer Verlag, New York 1986, p. 300.

[56] B. B. Damaskin, O. A. Petrii, V. V. Batrakov in R. Parsons (ed.): *Adsorption of Organic Compounds on Electrodes*, Plenum Press, New York 1971.

[57] J. Růžička, J. Starý: *Substoichiometry in Radiochemical Analysis*, Pergamon Press, Oxford 1968.

[58] F. Umland, A. Janssen, D. Thierig, G. Wünsch: *Theorie und praktische Anwendung von Komplexbildnern.*

Methoden der Analyse in der Chemie, vol. 9, Akademische Verlagsgesellschaft, Frankfurt 1971, p. 759.

[59] H. Yoshioka, T. Kambara, *Talanta* **31** (1984) 509 – 513.

[60] J. Starý: "Activation and Radioisotopic Methods of Analysis," in P. J. Elving, V. Krivan, I. M. Kolthoff (eds.): *Treatise on Analytical Chemistry*, part I, Sect. K, 2nd ed., vol. 14, J. Wiley & Sons, New York 1986, p. 241.

[61] J. Tölgyessy, S. Varga: *Nuclear Analytical Chemistry*, vol. 1 – 3, University Park Press, London 1972.

[62] V. Balek, J. Tölgyessy: *Emanation Thermal Analysis and Other Radiometric Emanation Methods*, Elsevier, Amsterdam 1984.

[63] T. Braun, J. Tölgyessy: *Radiometrische Titrationen*, Hirzel Verlag, Stuttgart 1968.

[64] G. Kraft, J. Fischer: *Indikation von Titrationen*, De Gruyter, Berlin 1972, p. 304.

[65] S. J. Parry: *Activation Spectrometry in Chemical Analysis*, Wiley & Sons, New York 1991, p. 243.

[66] W. D. Ehrmann, D. E. Vance: *Radiochemistry and Nuclear Methods of Analyses*, Wiley & Sons, New York 1991, p. 531.

Specific References

[67] G. Hevesy, F. Paneth, *Z. Anorg. Chem.* **82** (1913) 322 – 328.

[68] O. Hahn, *Z. Phys. Chem. (Leipzig)* **103** (1923) 461.

[69] Amersham International plc, Amersham Life Science Products 1993, Amersham/Little Chalfont, Amersham 1992, p. 184.

[70] Du Pont de Nemours Biotechnol. Systems, NEN Research Products 1991, Bad Homburg 1991, p. 164.

[71] J. T. Van Elteren, H. A. Das, C. L. De Ligny, J. Agterdenbos, *Anal. Chim. Acta* **222** (1989) 159 – 167.

[72] V. Hodge, M. Stalland, M. Koide, E. D. Goldberg, *Anal. Chem.* **58** (1986) 616 – 620.

[73] H. Yoshioka, K. Hasegawa, T. Kambara, *J. Radioanal. Nucl. Chem.* **117** (1987) 47 – 59.

[74] J. Klas, Z. Korenova, J. Tölgyessy, *J. Radioanal. Nucl. Chem.* **102** (1986) 111 – 120; **109** (1987) 337 – 351.

[75] S. Banerjee, *Anal. Chem.* **60** (1988) 1626 – 1629.

[76] R. H. Clarke: The 1990 Recommendations of ICRP, Supplement to the Radiological Protection Bulletin No. 119, NRPB, Chilton 1991.

[77] L. A. Currie, *Anal. Chem.* **40** (1968) 586 – 593.

[78] K. Hoppstock, R. P. H. Garten, P. Tschöpel, G. Tölg, *Fresenius J. Anal. Chem.* **343** (1992) 778 – 781.

[79] G. P. Westphal, *J. Trace Microprobe Tech.* **2** (1984/85) 217 – 235.

[80] D. Gawlik, K. Berthold, F. Chisela, P. Brätter, *J. Radioanal. Nucl. Chem.* **112** (1987) 309 – 320.

[81] W. Schmid, V. Krivan, *Anal. Chem.* **57** (1985) 30 – 34.

[82] K. G. Heumann, *Fresenius Z. Anal. Chem.* **324** (1986) 601 – 611.

[83] M. Schuster, *Fresenius J. Anal. Chem.* **342** (1992) 791 – 794.

[84] K.-H. Thiemer, V. Krivan, *Anal. Chem.* **62** (1990) 2722 – 2727.

[85] C. F. Miranda, R. Muxart, J. Vernois, G. Zuppiroli, *Radiochim. Acta* **19** (1973) 153.

[86] K. Hoppstock, R. P. H. Garten, P. Tschöpel, G. Tölg, *Anal. Chim. Acta*, (1994) in press.

[87] J. Tölgyessy, E. Bujdosó: *Handbook of Radioanalytical Chemistry*, **vols. 1 – 2**, CRC Press, Boca Raton 1991.

[88] T. Braun, J. Tölgyessy: *Radiometric Titrations*, Pergamon Press, Oxford 1967.

[89] J. Klas, J. Tölgyessy, J. Lesný: *Sub- superekvivalentová izotopová zried'ovacia analýza (Sub-superequivalent Isotope Dilution Analysis)* (in Slovak), Veda, Bratislava 1985.

[90] J. Tölgyessy, T. Braun, M. Kyrš: *Analiz metodom izotopnogo razbavlenija (Isotope Dilution Analysis)* (in Russian), Atomizdat, Moscow 1975.

[91] J. Klas, J. Tölgyessy, E. H. Klehr, *Radiochem. Radioanal. Lett.* **18** (1974) 83 – 88.

[92] J. Tölgyessy, M. Harangozó: "Radio-reagent Methods," in *Encyclopedia of Analytical Science, Radiochemical Methods*, Academic Press, London 1995 pp. 4317 – 4323.

[93] J. Tölgyessy: *Nuclear Radiation in Chemical Analysis* (in Hungarian), Müszaki Kiadó, Budapest 1965, p. 432.

[94] J. Tölgyessy: *Radioaktivnye kriptonaty v nauke i technike (Radioactive Kryptonates in Science and Technology), in Buduschee nauki (The Future of Science)* (in Russian), Izdatelstvo Znanie, Moscow 1977.

[95] M. Kyrš et al.: *Novye metody radioanaliticheskoi chimii (New Methods of Radioanalytical Chemistry)* (in Russian), Atomizdat, Moscow 1982.

[96] J. Tölgyessy, Ju. V. Jakovlev, G. N. Bilimovitch: *Diagnostika okruchajushej sredy radioanaliticheskimi metodami (Diagnostics of the Environments with Radioanalytical Methods)* (in Russian), Energoatomizdat, Moscow 1985.

[97] J. Tölgyessy: *Radioanaliticheskaja chimija (Radioanalytical Chemistry* (in Russian), Energoatomizdat, Moscow 1987.

[98] S. Aronoff: *Techniques of Radiochemistry*, The Iowa State College Press Building, Ames, IA 1956.

[99] V. R. S. Rao, Ch. P. Rao, G. Tataiah, *Radiochem. Radioanal. Lett.* **30** (1977) 365.

[100] J. Klas, J. Tölgyessy, E. H. Klehr, *Radiochem. Radioanal. Lett.* **18** (1974) 83.

[101] J. Klas, J. Tölgyessy, J. Lesný, *Radiochem. Radioanal. Lett.* **31** (1977) 171 – 179.

[102] J. Lesný, J. Klas, Z. Koreňová, *J. Radioanal. Nucl. Chem., Lett.* **155** (1991) 145 – 153.

[103] J. Lesný, Z. Koreňová, J. Klas, *J. Radioanal. Nucl. Chem., Lett.* **155** (1991) 155 – 168.

9. Enzyme and Immunoassays

JAMES N. MILLER, Department of Chemistry, Loughborough University, United Kingdom (Chap. 9)

REINHARD NIESSNER, Technische Universität München, München, Federal Republic of Germany (Chap. 9.2)

DIETMAR KNOPP, Technische Universität München, München, Federal Republic of Germany (Chap. 9.2)

9.1. Enzymatic Analysis Methods

9.1.1. Introduction

Many analyses in biological and clinical chemistry, and in the environmental, food and forensic sciences, involve the study of exceedingly complex multicomponent sample matrices. Moreover, the analytes under study are frequently present in trace amounts. Analytical techniques which combine exceptional selectivity with high sensitivity are thus required. Preeminent among these methods are those where the necessary selectivity is provided by biomolecules with very specific activities. This article surveys two groups of methods, those based on the activities of enzyme and of antibodies (the latter methods are normally called immunoassays).

The catalytic activity of enzymes provides for an enormous range of analytical techniques. Analytes are not restricted to conventional organic molecules, but include virtually all chemical species, including gases and metal ions. The ability of a single enzyme molecule to catalyze the reaction of numerous substrate molecules also provides an amplification effect which enhances the sensitivity

of the analyses. A further advantage is that most enzyme-catalyzed reactions can be followed by simple, widely available spectroscopic or electrochemical methods. Enzymes are normally active only in mild conditions — aqueous solutions at moderate temperatures and controlled pH — and this restricts the circumstances in which they can be used, but some enzymes have proved remarkably robust to, e.g., mixed aqueous – organic solvent systems or elevated temperature; immobilized (insolubilized) enzymes often show enhanced stability compared with their solution analogs (see Section 9.1.7).

Antibodies have been used in analysis for over 60 years, and offer an unexpectedly wide range of techniques and applications. In some cases, the specific combination of an antibody with the corresponding antigen or hapten can be detected directly (e.g., by nephelometry), but more often such reactions are monitored by a characteristic label such as a radioisotope, fluorophore, etc. Since antibody reactions do not have a built-in amplification effect, these labels are frequently necessary to provide sufficient analytical sensitivity. Antibodies of different classes vary greatly in stability, but some are relatively robust proteins, and this contributes significantly to the range of methods available.

Enzymes and antibodies are proteins, with the ability to bind appropriate ligands very strongly and specifically. Enzymes frequently need cofactors or metal ions for (full) activity. Antibodies are multifunctional molecules, whose ability to bind to cell surfaces and to other proteins, at sites distinct from the antigen/hapten-binding sites, extends the range of labeling and detection methods available in immunoassays.

The analytical potential of enzymes and antibodies can be combined. Enzyme immunoassays, in which enzymes act as label groups in antibody-based analyses, are very well established; they are probably the most widely used immunosassay methods. An important development is the production of single molecules which combine antibody and enzyme activities. A critical feature of enzyme activity is that the transition state of the substrate binds very tightly to the enzyme at a position near the amino acid side-chain groups that participate in the catalytic reaction. It is thus possible, by using haptens which are analogs of substrate transition states, to generate antibodies with catalytic activity. These "catalytic antibodies" are of great research interest, and their potential is emphasized by the possibility that they might be used to de-velop catalytic activities not present in living organisms.

This article can only attempt a general survey of enzyme and immunoassay methods, with references to more detailed reviews and books in specific areas. Many important clinical, veterinary, and forensic analyses in which enzymes themselves are the analytes are not discussed; this article considers only those methods where enzymes (and antibodies) are analytical reagents.

9.1.2. Enzymes: Basic Kinetics

The fundamental features of enzyme catalytic activity are described in detail in textbooks [1]. The simplest mechanism, outlined by MICHAELIS and MENTEN, proposes that the enzyme E reacts reversibly with the substrate S to form a complex, which subsequently decomposes to release the free enzyme and the product molecule P:

$$E + S \longrightarrow ES \longrightarrow E + P \qquad (1)$$

In this model, the combination reaction of E and S, with a rate constant k_1, occurs at the active site of the enzyme. This is a relatively small region of the E molecule; often a cleft or depression accessible from the surrounding aqueous solution. Its conformation, polarity, and charge distribution are complementary to those of the S molecule. The reverse process, dissociation of the ES complex, has a rate constant k_2, and the decomposition of ES to give E and P has a rate constant k_3. These three rate constants can be used to calculate the Michaelis constant K_M:

$$K_M = (k_2 + k_3)/k_1 \qquad (2)$$

The Michaelis constant clearly has units in mol/L, and in practice values between 10^{-1} and 10^{-7} mol/L. It is also convenient to define the turnover number of an enzyme, k_{cat}, which is the number of substrate molecules converted to product per unit time by a single enzyme molecule, in conditions where the enzyme is fully saturated with substrate. From the kinetic model $k_{cat} = k_3$, but if more complex models are used k_{cat} depends on several rate constants. Depending on the enzyme and the reaction conditions (see Section 9.1.3) k_{cat} varies between ca. 1 s^{-1} and 500 000 s^{-1}.

The kinetic model outlined above is clearly oversimplified (e.g., it should really include an
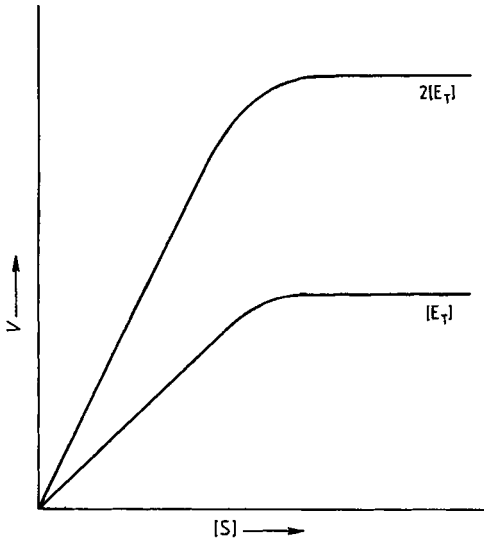
Figure 1. Relationship between reaction velocity V and substrate concentration [S] at different levels of enzyme concentration. [E$_T$] and 2 [E$_T$]

intermediate complex EP, in which the product molecule is bound to the enzyme), but it describes adequately the kinetic behavior of most enzyme-catalyzed reactions. If the steady-state hypothesis is applied to the ES complex in this model, the initial velocity of the enzyme-catalyzed reaction is:

$$V = k_3 [E_T][S]/([S] + K_M) \tag{3}$$

where squared brackets indicate concentrations, and [E$_T$] is the total enzyme concentration, irrespective of whether it is combined with S. When [S] >> K$_M$, i.e., when the enzyme active sites are saturated with substrate, V is simply k_3 [E$_T$]. This is commonly called V_{max}, so Equation (3) becomes:

$$V = V_{max}[S]/([S] + K_M) \tag{4}$$

Equations (3) and (4) show that, when V is plotted against [S] at constant [E$_T$] (Fig. 1), the reaction rate is approximately proportional to [S] at low substrate concentrations, but eventually reaches the plateau level V_{max}. So provided [S] << K$_M$, substrate concentrations can be determined directly from reaction rate measurements. Such analyses are widely used (Section 9.1.4). Equation (4) also shows that, when [S] = K$_M$, $V = V_{max}/2$. This provides a method of
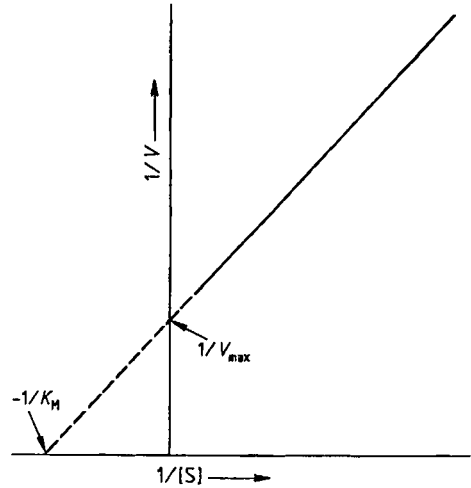
Figure 2. Double-reciprocal enzyme rate plot

estimating K_M. In practice, K_M and V_{max} are usually determined by one of several transformations of Equation (4); the most common is the Lineweaver–Burk double-reciprocal method, in which $1/V$ is plotted against $1/[S]$ (Fig. 2) to give a straight line of slope K_M/V_{max} and x- and y-axis intercepts $-1/K_M$ and $1/V_{max}$. This leads to the following:

$$(1/V) = (1/V_{max}) + (K_M/[S]V_{max}) \tag{5}$$

An important feature of enzymes is that their active sites can often be occupied by, or react with, molecules other than the substrate, leading to inhibition of enzyme activity. Several inhibition mechanisms are known, but it is necessary only to distinguish between irreversible and reversible inhibition. Irreversible inhibition arises when the inhibitor molecule I dissociates very slowly or not at all from the enzyme active site. The best-known examples occur when I reacts covalently with a critical residue in the active site. Inhibition of cholinesterase enzymes by the reaction of organophosphorus compounds with a serine residue is a case in point. This type of inhibition is said to be noncompetitive—enzyme activity cannot be restored by addition of excess substrate. So although addition of I reduces V_{max}, K_M is unaffected. The double-reciprocal plot in such cases has the same x-axis intercept as the plot for the uninhibited enzyme, but greater slope.

Reversible inhibition occurs when the EI complex can dissociate rapidly, just as the ES complex

does. The most common examples arise when S and I are chemically or sterically similar, and compete for the active site (competitive inhibition — but note that not all reversible inhibition mechanisms are competitive). Thus, alcohol dehydrogenase enzymes oxidize ethanol to acetaldehyde; but other alcohols, e.g., methanol or ethylene glycol, can act as competitive and reversible inhibitors. The extent of inhibition in such cases depends on the relative concentrations of I and S: if the latter is present in great excess, inhibition is negligible. In these cases, V_{max} is not affected, but K_M may be increased; the double-reciprocal plot has the same y-axis intercept, but greater slope than the plot for the uninhibited enzyme.

When steady-state theory is applied to these mechanisms, in both types of inhibition the initial rate of the enzyme-catalyzed reaction is inversely related to [I], provided that the latter is small. Again, the possibility of a simple quantitative analysis of I by rate measurements is apparent. Such analyses are not so numerous as determinations of substrates, but some are very important (Section 9.1.5).

A number of enzymes are inactive or weakly active in the absence of activators. The best known examples arise with metalloenzymes, whose active sites include a main-group or transition metal ion. When the active holoenzyme, i.e., the complete molecule, including the metal ion, is treated with a complexing agent to remove the metal ion, the resulting inactive apoenzyme may provide a sensitive reagent for determination of the metal, enzyme activity being directly proportional to the concentration of activator at low levels of the latter (Section 9.1.6).

9.1.3. Enzyme Assays: Practical Aspects

9.1.3.1. General Considerations

All enzymatic analyses of substrates, inhibitors, and activators involve the determination of reaction rates: since initial rates are required, experiments are usually quite short. It is possible in principle to mix the reactants and measure the rate directly, by recording an optical or electrochemical signal at regular intervals, the rate being the slope of the resulting signal plotted against time. When many samples are to be analyzed, such an approach is impracticable, and pseudo-rate methods are used. The reaction is allowed to proceed for a fixed time after the initial mixing; it is then

stopped abruptly by a large change in pH or temperature (see below), or by addition of a specific enzyme inhibitor, and the extent of substrate depletion or product formation is measured. The signal recorded is then compared with the results of matching experiments on standard analyte solutions, and the test concentrations calculated by interpolation. In simple enzymatic assays, the calibration graphs are usually linear (see above), but in complex systems such as enzyme immunoassays, curved calibration plots are usual, requiring more complex statistical evaluation.

9.1.3.2. Spectrometric Methods

UV – Visible absorption spectrometry is the most commonly used method in modern enzymatic analysis. Suitable spectrometers are found in every laboratory, and many analyses use simple, robust, relatively cheap instruments. The availability of disposable polystyrene or acrylic cuvettes is a further advantage for conventional solution studies, but many analyses now use microtiter plates, the individual wells of which can be examined in a simple spectrometer. Continuous flow systems are also simple to set up. Most spectrometers are interfaced with personal computers, which offer a range of control and data handling options.

Several generic applications of UV – visible spectrometry are common in enzymology (Section 9.1.4). Many oxidoreductase enzymes involve NAD/NADH or NADP/NADPH as cofactor systems. Such reactions are readily followed by absorption measurements at 340 nm, where the absorption of NADH and NADPH is strong, but that of NAD and NADP negligible (Fig. 3). This approach is so convenient that it is often applied in the form of a coupled enzyme reaction: if neither substrate nor product in the main reaction of interest is readily determined, one of them may be transformed via a second enzyme-catalyzed reaction involving NAD(P)/NAD(P)H, with detection at 340 nm. This principle has been extended so that three or more reactions occur in sequence, the final one either generating or consuming the readily determined NAD(P)H. Such indirect detection of the primary reaction is much less complex than it seems: often, all the enzymes, and any other reagents needed, can be included in the initial reaction mixture, the specificities of the of enzyme-catalyzed reactions is the sequence ensuring that they do not interfere. Extra sensitivity, and the opportunity to work in the visible region, is
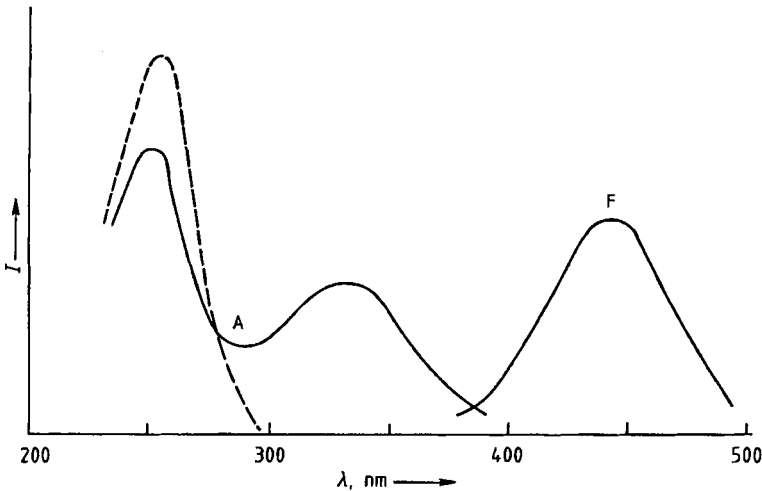
Figure 3. Absorption (A) and fluorescence (F) spectra NAD[P]H (———); and NAD[P] (- - -) (absorption only)

provided by using NAD(P)H to reduce colorless tetrazolium salts, in the presence of phenazine methasulfate (PMS), to intensely colored formazans, which absorb at ca. 500 nm. This further step is of particular value when a very rapid qualitative or semiquantitative result is required.

A further approach using UV – visible spectrometry involves colorigenic substrates. These molecules change color when hydrolyzed in the enzymatic reaction under study. A good example is provided by p-nitrophenol esters, which are colorless, but are hydrolzyed by appropriate enzymes to yellow p-nitrophenol, which can be determined at ca. 405 nm. Many substrates of this type are readily available, e.g., p-nitrophenyl phosphate as a phosphatase substrate, the corresponding sulfate as a substrate for aryl sulfatases, N-carbobenzoxy-L-tyrosine-p-nitrophenyl ester as a substrate for proteolytic enzymes such as chymotrypsin, etc.

Several analytically important enzymes catalyze reactions in which hydrogen peroxide is generated. Such reactions are easily followed colorimetrically, hydrogen peroxide being used in the presence of a second (peroxidase) enzyme to oxidize leuco-dyes to colored products. The same reactions can be used to monitor peroxidases themselves when they are used as labels in enzyme immunoassays; in modern systems, the enzyme-catalyzed reaction is halted by a large pH change, which simultaneously generates a distinctive color in the dye reaction product.

Fluorescence spectrometry is widely used in enzymology, usually because of its extra sensitiv-

ity compared with absorption methods. Fluorophores are characterized by two specific wavelengths, absorption and emission. The latter is the longer of the two, excited molecules having lost some energy between the processes of photon absorption and emission, and this allows the fluorescence to be determined (at 90° to the incident light beam in most instruments) against a dark background. This provides limits of detection unattainable by absorption spectrometry. A second important characteristic of fluorescence spectrometry is its versatility; it is just as easy to study concentrated solutions, suspensions, solid surfaces, flowing systems, etc., as it is to measure dilute solutions.

The generic approaches in absorption spectrometry are mirrored in fluorescence. Thus, NAD(P)H can be determined at an emission wavelength of ca. 460 nm with excitation at ca. 340 nm (Fig. 3). For work at longer wavelengths, NAD(P)H reduces nonfluorescent resazurin to the intensely yellow-fluorescent (ca. 580 nm) resorufin (this can also be determined colorimetrically at ca. 540 nm). Fluorigenic substrates are available for many hydrolytic enzymes, the best-known being those based on resorufin esters, and on esters of 4-methylumbelliferone; the esters are nonfluorescent, but are hydrolyzed to fluorescent products. Peroxidase activity can be monitored via the conversion of a resorcin derivative to a fluorescent product with hydrogen peroxide.

Recent years have seen an explosion of interest in chemiluminescence (CL) and bioluminescence

(BL) methods in enzymatic analyses. As in fluorescence, these methods measure photon emission from excited molecules, but here the latter are generated chemically, as reaction products. The reactions concerned are oxidations, often in mildly alkaline solutions in the presence of a catalyst. The advantages are:

1) No exciting light source is needed, so scattered light phenomena are minimized
2) It is usual to collect as many as possible of the emitted photons without the use of optical filters or monochromators. Screening devices using photographic detection have been developed. Sensitive Polaroid films allow the convenient qualitative or semiquantitative monitoring of chemiluminescent reactions.
3) Extreme sensitivities are possible
4) The reagents are generally free of hazard, despite offering limits of detection similar to those secured by radiochemical methods

Since the intensity of luminescence is proportional to the rate of generation of the excited reaction product, the principal precautions to be observed are those that apply when reaction rates are measured (control of temperature, pH, etc.), with the important additions that the order and the reproducibility of mixing of the reagents are critical, especially in conditions which yield short bright bursts of light. In such cases, fully automated CL/BL methods, such as those using flow injection analysis, are of obvious value. The common CL/BL techniques mirror those in UV – visible and fluorimetric measurements. Processes involving NAD(P) and NAD(P)H can be followed with the aid of coupled reactions (in which two or more enzymes catalyzing consecutive reactions are involved), and reactions in which hydrogen peroxide is generated are readily monitored by using the peroxide to oxidize a well-known luminescent compound, such as luminol. Luminogenic substrates, e.g., for phosphatase enzymes, have been developed, and these find particular use in CL immunoassays. Sensitive microtiter plate readers are now widely available for both fluorescence and chemiluminescence detection, so these spectroscopic methods are proving ever more popular in the monitoring of enzyme catalyzed reactions.

9.1.3.3. Other Methods

Although spectrometric methods still dominate the measurement systems used in biospecific anal-

yses, other methods are important. Some of these are applied in specialized areas. The optical activity of saccharides can be used to follow reactions such as the inversion (hydrolysis) of sucrose, catalyzed by invertase. The products (D-glucose and fructose) have a combined optical rotation (e.g., at 589 nm) quite different from that of sucrose, so a simple polarimeter provides a good method of determining sucrose, the enzyme's substrate, and of other species which inhibit invertase activity. Radiolabeled substrates are often employed in research (^{14}C and ^3H are the most common isotopes), but safety precautions preclude the routine use of this approach. Since the radioactivity of the label is not changed in the enzymatic reaction, physical separation is necessary before the radiolabeled reaction product (or residual substrate) can be determined (see Section 9.2.4).

Electrochemical methods of following enzyme-catalyzed reactions are rapidly growing in popularity, mirroring the growth of these methods across analytical science in general. Particularly important are electrochemical enzyme sensors, portable or disposable devices which combine the specificity of an enzyme reaction with the simplicity and compactness of an electrical transducer. Selective electrodes have been adapted to form such sensors: for example, oxygen electrodes can monitor any enzyme-catalyzed reaction in which oxygen is consumed or produced. Other potentiometric and amperometric methods have achieved impressive successes, exemplified by the variety of glucose sensors now available. These devices use the glucose oxidase reaction (Section 9.1.4), and a variety of coupled reaction systems which provide an electrochemical response. In early applications, the hydrogen peroxide generated in the oxidation of glucose was used to oxidize hexacyanoferrate(II) to hexacyanoferrate(III), this second reaction being followed amperometrically. More sophisticated methods are now applied, and are generally applicable to all reactions in which peroxide is produced.

The rates of enzyme-catalyzed reactions can be followed, not only by a wide range of instrumental techniques, but also with the aid of a variety of chemical principles. In favorable cases, reaction rates can be determined directly by observing consumption of the substrate or appearance of the product, but if these compounds lack distinctive physical properties, indirect approaches, especially coupled reactions, can be used (see Section 9.1.3.2).
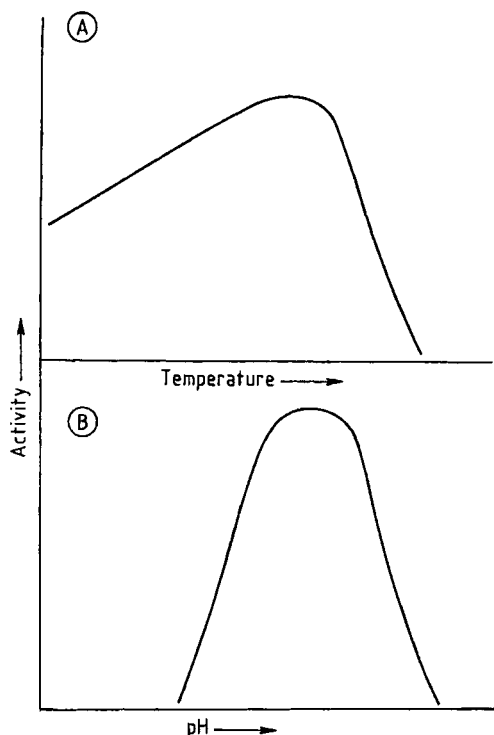
Figure 4. Effects of temperature (A), and pH (B) on the activity of a typical enzyme
The maxima and widths of the curves vary from one enzyme to another.

9.1.3.4. Effects of pH, Buffer, Composition, and Temperature

Virtually all enzymatic assays are carried out at 20–50 °C in aqueous buffers of known pH and controlled composition. Both temperature and buffer properties affect the rates of enzyme-catalyzed reactions markedly. The effects of temperature can usually be summarized by a bell-shaped curve (Fig. 4 A). At lower temperatures, reaction rates increase with temperature, but beyond a certain point, denaturation (unfolding) of the enzyme molecules begins, so they lose their ability to bind the substrate, and the reaction rate falls. The temperature giving maximum activity varies from one enzyme to another, according to the robustness of the molecule. In some cases, it may be convenient to use a temperature rather below this maximum, otherwise the rate becomes too high to measure precisely. The rates of many enzyme-catalyzed reactions increase by a factor of ca. 2 over a range of 10 °C in the region below the maximum of the

curve; it is thus necessary to control the temperature to within ca. ± 0.1 °C.

The pH of the buffer solution in which the analysis takes place must also be controlled. Most enzymes have a pH dependence of their activity of the type shown in Figure 4 B, and pH should be controlled to within ± 0.02. The optimum pH, and the range over which the enzyme is active, vary widely from one enzyme to another. These phenomena arise from the effects of pH on the structure of the enzyme itself, on the affinity constant between the enzyme and the substrate, on V_{max}, and also on any coupled indicator reaction that is used. The choice of buffer recipe for any given pH may be important. The ionic strength of the buffer and the salts contained in it can influence the rate and mechanism of the main and coupled reactions, sometimes with unpredictable results. It may be necessary to choose a buffer whose properties represent a compromise between the ideal conditions for the main reaction, and the ideal conditions for the indicator reaction(s). Immobilized enzymes (Section 9.1.7) often have pH (and temperature) dependence significantly different from their soluble counterparts. This all points to the value of establishing and maintaining a well-defined buffer system for enzymatic analyses.

9.1.3.5. Sources and Activity of Enzymes

A wide range of enzymes is commercially available from numerous suppliers, so it is rarely necessary for the analyst to prepare such materials. Enzymes with the same name, but isolated from different species, may be quite distinct in their chemical and biological properties, and their activity (e.g., their requirements for cofactors). Enzymes are normally supplied as freeze-dried or crystallized proteins, and should be stored carefully at 4 °C or −20 °C. Repeated freezing and thawing of protein solutions is not advisable, so it may be necessary to divide the dissolved enzyme into small aliquots, each of which is frozen and used just once.

The specificity of many enzyme-catalyzed reactions is a major reason for their use. Many organisms, or separate organs from a particular species, contain only one enzyme capable of catalyzing a given reaction. Thus, it may be unnecessary to use high-purity enzyme preparations; quite crudely purified material is often sufficient. In some cases, however, it is essential to ensure removal of a particular contaminant; enzyme preparations used in reactions involving hydrogen per-

oxide should be free of catalase, the enzyme which decomposes H_2O_2 to oxygen and water.

In practice many enzymatic analyses are performed with the aid of pre-packed kits. Such kits contain a number of vials which incorporate optimized mixtures of enzymes, cofactors, buffer salts, etc., as required, often in a dried form. Each vial provides a single analysis. The analyst needs only to reconstitute the reagents by the addition of water, add the sample, and perform the measurement after a given time period, recommended by the kit manufacturer.

Since an enzyme used as an analytical reagent may be quite impure, it is essential to know how much substrate a given weight of the enzyme preparation will convert to the corresponding product. This information can be derived from the activity of the material, given in International Units (IU). Unit enzyme activity converts one micromole of substrate to product per minute at 25 °C and optimal pH. Commercial enzyme preparations are described in terms of IU/mg or, in the case of an enzyme supplied in solution or suspension, IU/mL.

9.1.4. Enzyme Assays: Determination of Substrates

9.1.4.1. General Considerations

As shown in Section 9.1.2, the rate of an enzyme-catalyzed reaction is proportional to the concentration of substrate, if the latter is small compared with K_M. (The guideline $[S] < 0.2 K_M$ is often used.) This provides a simple approach to substrate determinations, which certainly represent the largest class of enzymatic analyses in modern practice, e.g., the measurement of blood glucose or blood cholesterol. In many instances, the analyses are highly selective, but not absolutely specific for the target substrate.

9.1.4.2. Determination of Glucose

Several different enzymes have been used to determine β-D-glucose, but by far the most popular approach uses glucose oxidase. Enzymes are often known by their Enzyme Commission numbers: glucose oxidase is E.C. 1.1.3.4. The first digit defines the enzyme as an oxidoreductase, the second as an oxidoreductase in which the hydrogen or electron donor is an alcohol, and the third as one in which the hydrogen or electron acceptor is molec-

ular oxygen. As there are several enzymes exhibiting these three characteristics, the fourth digit purely acts as a means of differentiating between similar substances. This enzyme, normally obtained from *Aspergillus niger,* has a molecular mass of ca. 80 000, uses flavin adenine dinucleotide as cofactor, and catalyzes, at pH 7 – 8, the reaction:

D-Glucose + O_2 ⟶ D-Gluconolactone + H_2O_2
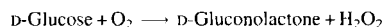
This reaction is highly specific (α-D-glucose is oxidized at only 1 % of the rate of the β-anomer), but not perfectly so, as 2-deoxy-D-glucose (which does not occur at significant levels in blood) is also readily oxidized. Scores of different methods have been used to follow the rate of this reaction, almost all based on the properties of the very reactive product, hydrogen peroxide. (Many of these methods can also be used to determine other substrates which are oxidized by oxidoreductase enzymes to yield, among other products, H_2O_2, e.g., cholesterol, xanthine, L- and D-amino acids, D-galactose etc.). Colorimetric methods are based on conversion of a chromogen [e.g., 2,2'-azino-di(3-ethylbenzothiazolinsulfonate) diammonium salt (ABTS)] to a colored product (for ABTS, $\lambda_{max} = 405$ nm) with the aid of a second enzyme, horseradish peroxidase. Analogous fluorimetric methods, such as the oxidation of nonfluorescent resorcin derivative to a fluorescent product, have also been used. The optimum pH for the coupled peroxidase step is ca. 10, but both enzymes show adequate activity at pH ca. 8.5, an example of a compromise pH. Glucose oxidase reactions can also be monitored by the chemiluminescence from luminol stimulated by peroxidase in the presence of hexacyanoferrate (III). Numerous electrochemical methods have also been developed. Many glucose tests are available which use the enzyme in immobilized form. A typical dipstick test for urine samples uses a cellulose matrix which contains immobilized glucose oxidase and peroxidase along with a chromogen. In the presence of glucose, the initially colorless matrix develops a color within a few seconds. (An exactly similar test is available using galactose oxidase to test for D-galactose.)

With such methods, glucose is routinely determined in blood (normal level ca. 5 mol/L) and foodstuffs. In all these analyses, it is assumed that molecular oxygen is present in excess, but it is also possible to use the enzyme in an excess of glucose to estimate levels of dissolved oxygen in aqueous or mixed aqueous – organic solvents. This unexpected application exemplifies the broad range of enzymatic analyses.

9.1.4.3. Determination of Ethanol

Ethanol can be determined with the aid of alcohol oxidase (E.C. 1.1.3.13), but a more usual approach uses alcohol dehydrogenase (ADH, E.C. 1.1.1.1.—the third digit indicates that this oxidoreductase has $NAD(P)^+$ as its hydrogen or electron acceptor). In neutral buffers, this zinc metalloenzyme, normally obtained from yeast, but also from mammalian liver, catalyzes the reaction:

$CH_3CH_2OH + NAD^+$ ⟶ $CH_3CHO + NADH + H^+$
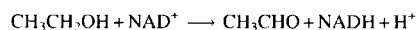
The oxidation can be followed by monitoring the appearance of NADH by UV-absorption spectrometry, colorimetry, fluorimetry, or CL/BL methods (Section 9.1.3.2). Common sample matrices are blood plasma or various foods and beverages; as in the case of glucose, their analyte levels are often quite high, so few sensitivity problems arise. The catalytic action of this enzyme is not very specific: n-propanol and n-butanol, isobutanol, allyl alcohol, ethylene glycol, glycerol, and methanol (oxidized to formaldehyde, hence its toxicity) are among other alcohols that may interfere. This lack of specificity is one reason for the interest shown in alcohol oxidase as an alternative. The importance of routine alcohol determinations has encouraged the development of many methods using immobilized ADH in the form of "alcohol electrodes" or continuous flow analyzers.

9.1.5. Enzyme Assays: Determination of Inhibitors

9.1.5.1. Inhibitors of Cholinesterase Enzymes

Enzyme assays based on inhibition effects are not as commonly employed as substrate determinations, but one or two are very important. Preeminent is the determination of organophosphorus compounds by using their inhibitory effect on cholinesterase enzymes (E.C. 3.1.1.8 — the first digit signifies a hydrolase enzyme, the second that the compounds hydrolyzed are esters, and the third that they are phosphoric monoesters). The latter catalyze the conversion of acylcholines to choline and the corresponding acid:

Acylcholine + H_2O $\longrightarrow$ Choline + Acid
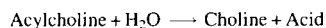
This reaction is crucial to many living systems, so the organophosphorus compounds (pesticides, chemical warfare agents) that inhibit it by irreversible binding to the active site are often highly toxic. A range of cholinesterase enzymes is available, the most common being the enzyme isolated from horse serum, which is a pseudocholinesterase or butyrylcholinesterase. The specificity of such an enzyme is broad, and a wide range of esters can be hydrolyzed. This is of value in inhibitor analyses as the "normal" substrate can be replaced by, e.g., a fluorigenic substrate to facilitate detection. Thus, resorufin butyrate, a nonfluorescent ester of resorufin, is hydrolyzed by the horse se-

rum enzyme to the intensely fluorescent resorufin. In the presence of an inhibitor, the expected growth in fluorescence intensity is reduced. Since a free acid is produced in the hydrolysis reaction, electrochemical methods provide viable alternatives, and enzyme reactors using immobilized cholinesterases are commercially available.

Organophosphorus compounds are often used as mixtures in pesticide formulations, and an ingenious application allowed the identification of such mixtures by combining the fluorigenic enzyme assay with thin layer chromatography (TLC). The sample was separated using TLC, and the plate treated sequentially with the enzyme and the fluorigenic substrate. The result was a mostly fluorescent TLC plate, with dark areas marking the positions of the enzyme inhibitors. In addition to organohosphorus compounds, carbamates (also used as pesticides) and the calabar bean alkaloid physostigmine act as inhibitors of cholinesterases, and can also be determined.

9.1.5.2. Other Inhibition Methods

The literature on enzyme inhibition assays is very large, but many of the methods are of limited value because they lack specificity, and/or do not compete with other techniques. Thus, many enzymes are inhibited to a greater or lesser extent by transition metal ions, but the lack of specificity for individual ions and the preference for alternative techniques, e.g., ion chromatography, has reduced the value of such approaches.

One area of great current importance is the search for enzyme inhibitors, especially inhibitors of proteolytic enzymes, in high-throughput screening programs for new drugs. Combinatorial chemistry methods can generate candidate inhibitors at the rate of thousands per working day, so rapid and efficient screening procedures are essential. Most currently used methods employ microtiter plate formats with fluorescence detection.

9.1.6. Enzyme Assays: Determination of Activators

9.1.6.1. Determination of Metal Ions by Metalloenzymes

Activator analyses are the least common application of enzymatic methods, but, as in the case of inhibitor analyses, a small number of methods are extremely valuable. Of great interest are methods

where incorporation of metal ions in the active site of an enzyme is critical. Removal of the metal from the holoenzyme (e.g., by mildly acidic solutions of EDTA) generates the apoenzyme, which lacks enzyme activity. This apoenzyme then becomes the reagent in the subsequent analysis, along with its natural substrate and any indicator reagents that may be required. In a sample containing the appropriate metal ion, the activity of the enzyme is wholly or partially restored. In contrast to the metal ion inhibition effects mentioned above, these analyses are often highly selective, as the active sites of metalloenzymes accommodate (with generation of activity) only a very few metal ions (ionic radius may be as important as ionic charge).

An example of such analyses is provided by isocitrate dehydrogenase (E.C. 1.1.1.42), normally isolated from pig heart. This enzyme catalyzes the oxidation of isocitrate ions to α-oxoglutarate ions; the reaction is NADP dependent, and results in a decarboxylation:

Isocitrate + NADP $\longrightarrow$ NADPH + CO_2 + H^+ + α-Oxoglutarate

The usual methods for following NAD(P)-dependent reactions are available here. The enzyme requires the presence of Mg^{2+} ions for activity, and extensive studies show that, apart from Mg^{2+}, the only other ions that combine with apo-ICDH (apo-isocitrate-dehydrogenase) to restore its activity are Mn^{2+} and to a lesser extent Zn^{2+} and Co^{2+}. This approach has been successfully applied to the determination of Mg^{2+} in blood and other samples at concentrations below 1 μmol/L, and to the analysis of Mn^{2+} at concentrations as low as 5 ppb. The advantage of this method of trace metal determination is that only a single oxidation state of the metal is determined — in the manganese analysis, only Mn(II), with no interference from Mn(VII) or other oxidation states.

9.1.6.2. Other Activator Analyses

The most obvious activator assays are those in which cofactors are analyzed: all the common cofactors can be determined in this way, and in many cases their own properties provide simple means of measuring reaction rates. The analyses are usually very selective. ATP in red blood cells and many other samples is often determined by using its specific participation in the well-known firefly bioluminescence reaction:

Luciferin + O_2 + ATP $\longrightarrow$ Oxyluciferin + ADP + PO_4^{3-} + Light
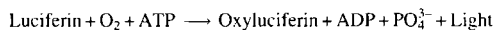
This reaction is catalyzed by firefly luciferase, which requires Mg^{2+}, and can also be used to determine this ion. Even in simple instruments, ppb levels of magnesium can be detected.

9.1.7. Immobilized Enzymes

9.1.7.1. Introduction

Analysts have attempted to use immobilized enzymes for over 30 years. The original motivations were two-fold: to conserve enzymes that were expensive and difficult to isolate, and to incorporate enzymes in reusable sensors, such as enzyme electrodes. During such researches, other advantages and applications of immobilized enzymes were found, and immobilization technologies were extended to other areas of biospecific analysis such as immunoassays (see below) and affinity chromatography. Immobilized enzymes have also been used extensively in manufacturing processes and for therapeutic purposes.

Several distinct approaches to the problem of immobilizing an enzyme in or on a solid or gel, with optimum retention of biological activity, are available. These include covalent binding, surface adsorption, gel entrapment, encapsulation within a semipermeable membrane, and chemical cross-linking. The first is by far the most popular, and many enzymes are commercially available in covalently immobilized form. Common solid phases include particles of well-known chromatographic media, e.g., agarose, glass and silica, polystyrene, polyacrylamide and cellulose, modern perfusion chromatography phases, and nylon in the form of flow tubing or membranes. The best-known covalent linkage methods include the use of cyanogen bromide (CNBr) to bind $-NH_2$ groups to hydroxylic matrices such as agarose or cellulose, chloro-*sym*-triazinyl derivatives to perform a similar function, and carbodiimides to link amine and carboxylic acid functional groups between enzymes and solid phases. Such covalent linkages are not infinitely stable; even the best immobilized enzyme reactors suffer slow leakage of the enzyme molecules into solution. However, many derivatives remain stable on storage and in use over periods of several months.

The idea of entrapping enzyme molecules within a polymeric matrix of known (average) pore size is attractive and has been frequently

studied, polyacrylamide gels being widely used. However the outcomes have often been disappointing. The polymerization process, carried out in the presence of a solution of the enzyme, may generate sufficient heat to denature some of the enzyme; leakage may occur from within the polymer matrix; enzyme activity with respect to large substrates may diminish sharply because of the molecular sieving effect of the polymeric gel; and even the activity to small molecules may be diminished because the enzyme molecules are less available than they are on the surface of a solid particle.

However, interest in this approach to enzyme immobilization has recently been renewed by the application of silica-based sol-gel entrapment matrices. These provide mild hydrophilic environments with controlled pore sizes. Moreover solgels can be formed into optically clear monoliths of varying dimensions, thin films, and fibers. Several applications including the determination of metal ions and of peroxidase have already been described.

Enzymes are readily adsorbed to the surfaces of ion exchange matrices, plastic microtiter plates, hydroxyapatite, even charcoal. Such immobilized molecules may be suitable for one-off assays (i.e., not for repeated use), but they are often readily desorbed by quite small changes in their environment, and enzymes immobilized in this way have not found great application in routine analyses, despite the ease with which adsorption is achieved. However, the adsorption of antibodies on plastic surfaces is routinely used in many immunoassays, including enzyme immunoassays (Section 9.2.4.3.1).

Bifunctional reagents, such as glutaraldehyde, can cross-link enzyme molecules to produce insoluble aggregates that retain at least some activity, but these cross-linked enzymes often have unsuitable mechanical properties, and they have found little practical use.

Microencapsulation of enzymes (and other biomolecules) has been extensively studied. The microcapsules have semipermeable walls that allow small substrate and product molecules to pass through freely, while large enzyme molecules are retained in solution within the capsules. However, the polymerization process used in forming the capsules generates heat that may damage the enzymes.

9.1.7.2. Properties of Immobilized Enzymes

Many experiments have shown that the advantages of (usually covalently) immobilized enzymes extend beyond their reusability; their physicochemical properties are often significant-ly different from those of the corresponding enzyme in free solution. The immobilized enzyme usually has somewhat less activity than the same mass of enzyme in solution. But unless activity losses are severe, this disadvantage is not serious, and can be compensated by using more of the immobilized preparation, knowing that it can be used repeatedly.

A significant advantage of immobilization is the increased thermal stability conferred on many enzymes, allowing their use for longer periods at higher temperatures than would be possible for the soluble molecule. An early experiment showed that immobilized papain (a protease derived from papaya latex) retained over half its room-temperature activity at 80 °C in conditions in which the soluble enzyme was almost entirely denatured.

Also interesting and valuable are the changes in the pH – activity curves that often accompany immobilization. A common result is broadening of the curve, i.e., the enzyme is active over a wider pH range than its soluble counterpart. This is normally ascribed to the range of microenvironments of different enzyme molecules in or on the solid matrix. Depending on the charge properties of this matrix, the optimum pH may undergo significant shifts. The optimum pH for an enzyme bound to a negatively charged carrier such as carboxymethyl-cellulose is shifted to higher values, while immobilization on a cationic matrix such as DEAE-cellulose (diethylaminoethyl-cellulose) has the opposite effect. These effects are ascribed to the change in the enzyme's microenvironment brought about by neighboring charged groups. Immobilization on a neutral carrier is not expected to change the pH optimum. Since these pH effects can be controlled to some degree, immobilization of an enzyme may confer significant advantages, e.g., using two or more enzymes with different pH optima in solution (Section 9.1.4.2) may be facilitated.

9.1.7.3. Application of Immobilized Enzymes

Since most of the enzymes routinely used in analytical science are now commercially available in immobilized form, it might be expected that many analyses conventionally carried out in solu-

tion are also performed with the immobilized products. In practice, many of the important applications of immobilized enzymes are those in which the enzyme preparation is incorporated into a more specialized analytical system, such as an enzyme electrode or continuous enzyme reactor.

One of the first applications of immobilized enzymes was the construction of an enzyme electrode for glucose, in which a thin layer of glucose oxidase entrapped in a polyacrylamide gel was placed over a conventional Clark oxygen electrode. When the enzyme electrode was immersed in glucose solution, the glucose penetrated the enzyme layer, consuming oxygen and reducing the potential of the Clark electrode. Apart from concerns over the long-term stability of the enzyme layer, the principal drawbacks of this and many other similar devices were their response times (often minutes rather than seconds), and their recovery times, i.e., the rates at which the electrodes could be used with different samples. Some modern devices have largely overcome these problems, but although enzyme electrodes (most based on platinum, oxygen, or other established sensors) have been developed for numerous analytes (amino acids, urea, sugars, simple alcohols, penicillin, cholesterol, carboxylic acids, inorganic anions, etc.), only a very few seem robust enough to survive routine laboratory use over long periods.

On the other hand disposable "one-shot" sensors based on immobilized enzymes are now of major importance in several biochemical and environmental application areas.

A second obvious area of application is in continuous flow analysis or flow injection analysis systems, in which the immobilized molecules form reactors that can be readily inserted and replaced in a flow analysis manifold. The physical form of the enzymes varies widely; packed-bed reactors are often used, but open-tube wall reactors and membrane reactors have also been investigated. A principal advantage of all such systems is that they can use all the optical or electrochemical detectors routinely used in flow analysis. However, the problems of producing stable and robust immobilized enzyme reactors have proved more intractable than many researchers hoped, and other advances (e.g., the use of more sensitive detectors, improved availability of low-cost soluble enzymes) have minimized the advantages of using solid phase enzymes.

Nonetheless, much research continues in this area, particularly work on thin layers or membranes to which enzymes are attached. In such cases, the enzyme kinetics may be favorable, and, of the available detection procedures, fluorescence spectroscopy is particularly suitable for the study of solid surfaces, and is highly sensitive.

9.2. Immunoassays in Analytical Chemistry

9.2.1. Introduction

In recent years, analytical chemistry has grown tremendously, especially in the fields of environmental and process analysis. Modern measurement techniques have led to lower limiting values in the areas of water, soil, and air analysis. This has been followed by legislation requiring environmentally compatible production processes under constant control. As a result, less expensive (and often faster) monitoring techniques are necessary. The search for "chemical sensors," which, similar to optical and electrical sensors, can be integrated into production monitoring and quality control, is a logical consequence of this ecological and economic pressure.

Immunoassays (IAs) are based on the formation of a thermodynamically stable antigen – antibody complex. These methods play an important role, especially in clinical chemistry, being used for the fast and safe detection of proteins, hormones, and pharmaceutical agents. These techniques promise to close the gap between the cheapest chemical sensors and conventional, expensive, slower analytical methods.

On the other hand, classical analytical chemistry is just starting to accept immunoassays; the reasons for this hesitancy are to be found in the necessity of:

1) Synthesizing an immunogen and coating antigen (hapten linked to different carrier proteins)
2) Isolating an antiserum after immunization, usually of a vertebrate
3) Development, optimization, and synthesis of a labeled analyte (hapten) derivative after isolation and characterization of the first antibodies (AB)
4) Validation of both the antibodies obtained, and the entire test

5) Possibly having to start the immunization all over again, if the antibodies do not fulfill the selectivity and affinity requirements

Every analytically usable method must fulfill the criteria of selectivity, sensitivity, calibration ability, and reproducibility.

The development of immunoassays is frequently based on empirical findings which are an obstacle to certification. In particular, despite the fact that polyclonal antibodies often permit better detection tests, they are not reproducible in any given laboratory. Even the production of monoclonal antibodies, which in principle is easily reproducible, does not aid worldwide acceptance and recognition of immunoassays because of their extremely complex production processes, which are increasingly being patented.

Nevertheless, a steady rapid development of immunotests can be observed, especially in the field of environmental monitoring [1]–[3], where the question of costs has become so important that the establishment of an immunotest can be worthwhile in spite of all the facts mentioned above. In the development of chemical sensors, efforts are also being made to utilize the biochemical recognition principle by coupling with optical, electrochemical, or other transducer (signal transfer) [4]–[6]. However, the slower kinetics involved in molecular biological processes require stricter maintenance of the experimental protocol, e.g., the time of reaction between the antibody and the tracer molecule or analyte (hapten) should be set up in an exactly reproducible manner by means of flow-injection analysis [7].

Immunoassays become important when:

1) Fast measurement and evaluation are required
2) Highest possible detection strength is required
3) Large numbers of samples are to be expected
4) Only complex and expensive analytical methods are otherwise available

The greatest potential for the use of immunoassays in environmental analytical chemistry is in screening, i.e., for the selection of contaminated and uncontaminated samples for further validation analysis.

9.2.2. Polyclonal, Monoclonal, and Recombinant Antibodies

The isolation of highly selective (specific), high-affinity antibodies is the primary requirement

for the successful development of an immune test [8], [9].

In analytical chemistry, especially in the environmental field, analytes of molecular mass < 1000 are often of interest, but molecules of this type (haptens) cannot alone induce the generation of antibodies.

For this reason, it is necessary to produce an immunogen by coupling the hapten to a high molecular carrier protein. The possibilities for immunogen synthesis are restricted by the structure of the hapten. The presence of amino or carboxyl groups, which can be directly coupled to the carrier protein by a peptide bond, is most favorable. If the hapten contains no reactive group, it must be introduced. An example is the reaction of haptens containing hydroxyl groups, which are converted to hemisuccinates with succinic anhydride or N-bromosuccinimide. Haptens with keto groups can be converted to oxime derivatives.

Modified haptens or haptens containing reactive groups are then coupled to a high molecular carrier protein. High yields and no interfering side products are usually obtainable from reactions at room temperature in aqueous media and at near-neutral pH. The carrier proteins commonly used are keyhole limpet hemocyanin human or bovine serum albumin, bovine gammaglobulin, or thyroglobulin. Coupling reactions comprise: (1) the carbodiimide-mediated conjugation, (2) the mixed anhydride procedure, (3) the NHS (N-Hydroxysuccinimide) ester methods, and (4) the glutaraldehyde condensation [10]. The degree of coupling, i.e., the number of hapten molecules per protein molecule, should be determined, using either radiolabeled hapten, UV and visible spectroscopy, matrix assisted laser desorption time-of-flight mass spectrometry (MALDI TOF MS), or indirect methods.

The choice of a suitable position of attachment to link the hapten to the carrier is important because it is responsible for recognition by the antibody produced. Especially in the case of parent substances which are rich in isomers, only some of which are of toxicological interest, e.g., PCBs or dioxins, the reactive positions used for conjugate synthesis must be carefully selected.

For the generation of polyclonal antibodies, the immunogen is applied to a vertebrate (rabbit, sheep, etc.), usually with the aid of adjuvants which enhance immunogenicity, by a mechanism that is not fully understood. The production of polyclonal antibodies is not very reproducible. For this reason, antibody concentration (titer),

specificity (selectivity of the antibodies), and affinity must be continually monitored, and the ability of the isolated antiserum to bind to the hapten must be determined. Further, the cross-reactivity of the antibodies should also be determined as early as possible. Cross-reactivity refers to the reaction of an antibody with molecules which are structurally very similar to the desired analyte (hapten); it determines the sensitivity to interfering substances in the final test.

Polyclonal immune sera from an immunization of this type consist of a spectrum of antibodies, which are active against the diverse epitopes of the same antigen and, therefore, possess various affinity constants.

The acceptance of immunoassays depends on the availability of antibodies. Since the amount of polyclonal antibodies is always limited, and even repeated immunizations of the same experimental animal species results in nonreproducible characteristics (specificity, affinity), the production of suitable monoclonal antibodies (they are product of a single cell alone) is most important. The method was developed by KÖHLER and MILSTEIN and encompasses the immortalization of a single-antibody producing cell (B-lymphocyte) by fusion to a myeloma cell [11]. The resulting hybridoma produces just one species of antibody which can be grown indefinitely and, therefore, offers a means to advise a consistent product in unlimited quantities. Up to now, compared to polyclonal antibodies monoclonal antibodies are less frequently applied in environmental analysis.

In recent years, genetic engineering processes have entered the field of antibody production. A number of functional recombinant single-chain antibody fragments (scAbs, recombinant antibodies) which recognizes and binds to small compounds has been expressed and characterized in microorganisms and transgenic plants [12]–[14]. However, these techniques are still at the development stage.

9.2.3. Sample Conditioning

Sample preparation is of tremendous importance in the application of an immunoassay. Apart from the usual criteria for representative sampling, special attention should be paid to the following influences:

1) pH value
2) Humic substances
3) Surfactants
4) Heavy metals
5) Organic solvents

If enzymatic tracers are used, the actual detection reaction is frequently disturbed by influences exerted on the reactive center of the enzyme, especially those due to interactions of the antibody with polyfunctional groups, e.g., with the humic substances in natural water samples [15]. This should be taken into account, especially in the use of immunoassays for soil screening. In competitive immunoassays, analytes that are sparingly soluble in water can be solubilized by the addition of surfactants [16]. On the other hand, the surfactant can alter the tertiary structure of the antibody, and possibly of the enzyme involved. The use of organic solvents is being intensively studied [17]. In some cases, immunoassays are known to tolerate more than 10 vol % of solvent.

The separation or complexation of humic substances and/or heavy metals is successful in favorable cases, e.g., by the addition of bovine serum albumin (BSA) in a trinitrotoluene enzyme-linked immunosorbent assay (ELISA) [18]. In the screening of soil samples for pesticides, persistent centrifugation at 17 000 rpm, and simple dilution with bidistilled water in the ratio 1 : 10 was sufficient [19].

In each application, the appropriate measurement must be established by independent validation analysis. Nonspecific interactions are very frequent in immunoassays, and usually lead to a slight overdetermination of the analyte.

9.2.4. Immunoassays

9.2.4.1. Radioimmunoassay

The development of radioimmunoassays (RIA), based on studies by YALOW and BERSON, has opened up an area of application especially in clinical chemistry [20], [21].

In RIAs, a radioactive label is used for detection of the formation of an antibody–antigen complex. Thus, a simple, specific, sensitive, and precise determination of the radioactive isotope is available. At present, however, immunoassays are increasingly carried out with nonisotopically labeled antigens or antibodies because of the legal restrictions on the use of radioactive material.

9.2.4.1.1. Isotopic Dilution Radioimmunoassay

In this classical RIA, a variable amount of unlabeled analyte (hapten) competes with a constant amount of a radiolabeled substance for a limited number of antibody binding sites:

$$Ag^* + Ag \nearrow\!\!\!\!\!\!\!\!\!\searrow \begin{matrix} Ag^* \, AB(s) + Ag(b) \\ \\ Ag \, AB(s) + Ag^*(b) \end{matrix}$$
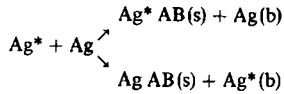
where Ag is the antigen (hapten or analyte); Ag* the radiolabeled antigen; AB the antibody; and (b) and (s) indicate species bound or in solution.

The more analyte in the sample, the more labeled antigen is found in solution. To quantify this amount, a separation step must be carried out, usually by subjecting the adsorptively bound antibodies (e.g., on a microtiter plate) to a washing step.

In principle, standard and analyte should have the same ability to displace a tracer molecule from the binding site on the antibody.

9.2.4.1.2. Immunoradiometric Assay (IRMA)

This process uses radioactively labeled antibody AB*. Two antibodies are required which recognize different binding sites on the antigen; the first is fixed on a microtiter plate or in a test tube, and the second is labeled, e.g., with ^{125}I. Sequential incubation is carried out, first with the sample or the standard analyte, followed by a separation washing step. Subsequent incubation with radiolabeled antibody AB* shows the antigen binding density through sandwich coupling:

$$AB(s) + Ag \longrightarrow AB - Ag(s)$$

$$AB - Ag(s) + AB^* \longrightarrow AB - Ag - AB^*$$
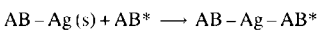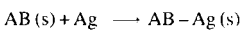
Small analyte molecules (haptens) cannot be detected by this technique because of the lack of several binding sites (epitopes), so it cannot be used for environmental analysis.

9.2.4.2. Nonisotopic Homogeneous Immunoassay

For diverse reasons, such as restrictions on the use of radioactive substances, nonisotopic immunoassays have gained general acceptance, especially in trace analysis and environmental analysis 1] – [3].

Consequently, a sensitive measuring principle other than radioactive decay is required. Furthermore, a distinction must be made between tests that require a separation step to determine the bound component and those that do not. Nonisotopic, homogeneous immunoassay allows the direct observation of the hapten bound to the antibody.

9.2.4.2.1. Latex Particle Agglutination Immunoassay

A large number of latex agglutination immunoassays have recently been adopted from clinical chemistry. These assays are based on the visualization of antigen – antibody complexes by the attachment of latex particles or gold colloids. Entities of this type with dimensions in the nanometer or micrometer range can be quantified by turbidimetry, nephelometry, light scattering techniques, and particle counters [22] – [25].

9.2.4.2.2. Enzyme-Multiplied Immunoassay Technique (EMIT)

In the EMIT [26], [27], the analyte is covalently bound to the enzyme in spatial proximity to the active site and, consequently, the formation of the antibody – antigen complex inactivates the enzyme; addition of hapten results in a reduction of this inactivation.

9.2.4.2.3. Apoenzyme Reconstitution Immunoassay System (ARIS)

If, however, the antigen is covalently bound to the prosthetic group of an enzyme such as glucose oxidase (E.C. 1.1.3.4) and an aliquot of the coupled antigen to flavin – adenine dinucleotide) is added to determine an analyte, free antibodies prevent the reconstitution of the enzyme. The concentration of the free antibody naturally depends on the analyte concentration in the sample. Similar to the EMIT technique, the ARIS is used in automatic analyzer systems in clinical chemistry [28].

9.2.4.2.4. Fluorophore-Labeled Homogeneous Immunoassay

At first glance, fluorescent labeling appears to have a much higher detection strength compared to colorimetric detection, but this is not the case. First, the affinity constant generally limits the detection strength of a process. Second, fluorophores

are exposed to many influences, such as quenching by impurities, or even adsorption of the fluorophore molecule. However, the fact that the detection can be repeated is advantageous, whereas a chemical reaction is irreversible [29].

The best-known variant is the substrate-labeled fluorescence immunoassay (SLFIA) test [30]. Galactose is linked to the antigen and to the fluorophore, methylumbelliferone. β-Galactosidase (E.C. 3.2.1.23) is capable of hydrolyzing and releasing the fluorophore, as long as the antigen linked to galactose is not stabilized in an antibody – antigen complex. If free antigen is present, the formation of galactose – antigen – fluorophore – antibody complex is reduced. The measuring signal is conveniently proportional to the concentration of free antigen.

In pesticide trace analysis, efforts have been made to observe directly the coupling of antigen – antibody in the quenching of the fluorophore bound to a tracer molecule, by the use of time-resolved, laser-induced fluorescence [31]. On formation of an Ag*–AB complex, a significant change in the fluorescence lifetime is observed. However, the fluorescence quantum yield suffers drastically, and this test has never been used.

9.2.4.2.5. Homogeneous Fluorescence Polarization Immunoassay (FPIA)

Direct observation of the formation of a hapten – (fluorescently labeled) antibody complex is also possibe in polarized light [32]. The presence of free hapten reduces the antibody-tracer complex concentration, and the degree of polarization is lowered. The detection strength of this test is in the µmol/L range and thus not yet high enough for environmental analysis.

9.2.4.2.6. Homogeneous Liposome Immunoassay

Owing to "cell wall formation" by multilayer membranes, liposomes may encapsulate up to several hundred fluorophore molecules.

The measuring principle is based on the fact that hundreds of hapten molecules are incorporated into the cell membrane of liposomes, e.g., by linkage of phosphatidylethanolamine (a functional part of the multilayer membrane). In the presence of an antibody for the formation of an immune complex, destabilization of the entire vesicle occurs. The contents of the liposome, e.g., fluorophores dyes, spin-labeled molecules, or

enzymes, are released. Liposome stability is a general problem associated with this method [33].

9.2.4.3. Nonisotopic Heterogeneous Immunoassay

The most common type of immunoassay is the nonisotopically labeled heterogeneous test [1]. Either the antibody or the hapten (analyte) to be detected is fixed to a solid interface via a covalent bond or by adsorption. The other significant difference lies in the fact that no radioactive labeling is used for signal production.

9.2.4.3.1. Enzyme-Linked Immunosorbent Assay (ELISA)

This assay is the most important in trace and environmental analysis. First, antibodies are immobilized on a solid carrier. For this, it may be necessary to take suitable measures, such as ^{60}Co radiation, to make the carrier sorption active. Standard 96-well microtiter plates, polystyrene beads, or test tubes can be used.

In a typical ELISA test on a microtiter plate (Fig. 5), the carrier is first coated with analyte-binding antibody. The sample or standard is added after washing off excess antibody with a surfactant solution. The volume added is typically 100 – 200 µL. After (optional) preincubation, a constant quantity of tracer, e.g., enzyme-labeled hapten, is added to the sample or standard. This initiates a competitive reaction, because only a limited number of antibodies are available for binding. After the tracer incubation period, sample (or standard) and excess reagents are washed away. The bound tracer concentration is inversely proportional to that of the analyte. The amount of bound tracer can be determined via an enzyme substrate reaction with a chromogenic substrate. After a reaction time long enough to produce sufficient dye, the enzyme is denatured (enzyme reaction stopped) by addition of acid. Subsequently, the depth of color formed in the individual cavities of the microtiter plate can be automatically determined with a "plate reader" (photometer).

The application of a competitive heterogeneous immunoassay is very common. Some important immunoassays for trace amounts of environmentally relevant compounds are presented in Table 1. Especially for agrochemicals, a large number of determination procedures have been established, because of legislation on drinking

Table 1. Selected ELISA tests in environmental analysis

Substance	Detection limit	Reference
Pesticides		
Alachlor	1 ppb	[34]
Aldicarb	300 ppb	[35]
Atrazine	50 ppt	[36]
Benomyl	250 ppb	[37]
Bentazon	2 ppb	[38]
Diflubenzuron	1 ppb	[39]
Endosulfan	3 ppb	[40]
Fenpropimorph	13 ppt	[41]
Hydroxyatrazine	50 ppt	[36]
Metazachlor	10 ppt	[42]
Methabenzthiazuron	50 ppt	[43]
Molinate	30 ppb	[44]
Norflurazon	1 ppb	[45]
Terbuthylazin	60 ppt	[46]
2.4-D	100 ppb	[47]
Aromatic compounds		
Benzo[a]pyrene	100 ppt	[48]
1-Nitropyrene	50 ppt	[49]
Pyrene	35 ppt	[50]
Trinitrotoluene	50 ppt	[51]
Halogenated aromatic hydrocarbons		
Pentachlorophenol	30 ppb	[52]
2.3.7.8-Tetrachlorodibenzodioxin	5 ppb	[53]
Heavy metals		
Mercury(II)	500 ppt	[54]
Surfactants		
Triton X-100	2 ppb	[55]

Figure 5. Principle of a direct competitive enzyme immunoassay (ELISA) with photometric detection

water which stipulates (as a precautionary measure) a limiting value of 100 ng/L for a single pesticide; this is not based on toxicological findings. The demand for sampling and pesticide analysis has given a sudden worldwide impetus to the development of immunoassays as a fast and cost-effective method of analysis.

The use of indirect competitive heterogeneous immunoassays (Fig. 6) is also common. Here, the hapten is bound to the surface of the carrier. Since hapten alone cannot be bound very reproducibly to a solid support by adsorption, a protein-linked hapten is used (e.g., to BSA). In principle, however, a conjugate (coating conjugate) different from the immunogen should be employed.

The advantage of this ELISA variant lies in the fact that (1) synthesis of a coating antigen is easier than that of a hapten-enzyme conjugate, (2) interfering matrix constituents are washed out before the addition of the enzyme-labeled second antibody, and (3) a prepared microtiter plate contains only slight amounts of a possibly toxic analyte [important when working with 2,3,7,8-tetrachlorodibenzodioxin (2,3,7,8-TCDD)]. After a washing step, sample and analyte-binding antibody are

added. If larger amounts of free analyte are present in the added sample, only a few adsorptively bound analyte molecules react with AB 1. The AB – Ag complexes in solution are removed by a further washing step. The AB 1 molecules bound through hapten – BSA, which contain the analytical information, are detected by means of an enzyme-labeled second antibody. For this purpose, it is possible to use AB_2 that recognizes antibodies (IgG) from rabbits, from which AB 1 originates; these antibodies are commercially available. Signal production is again inversely proportional to the concentration of analyte. Enzymes commonly used are horseradish peroxidase (E.C. 1.11.1.7) and alkaline phosphatase (E.C. 3.1.3.1).

For environmental analysis, enzyme-labeled tracers are predominantly used because increasingly lower limiting values demand maximum sensitivity. The reason for this choice is based on the enormous turnover rate in chromophore or fluorophore formation by enzymes. In the most common tests, photometric evaluation predominates.

There has been no lack of attempts to reduce the detection limit further by the use of liposomes filled with enzymes [56] or fluorophores [57]. Unfortunately, liposomes are not very stable, and
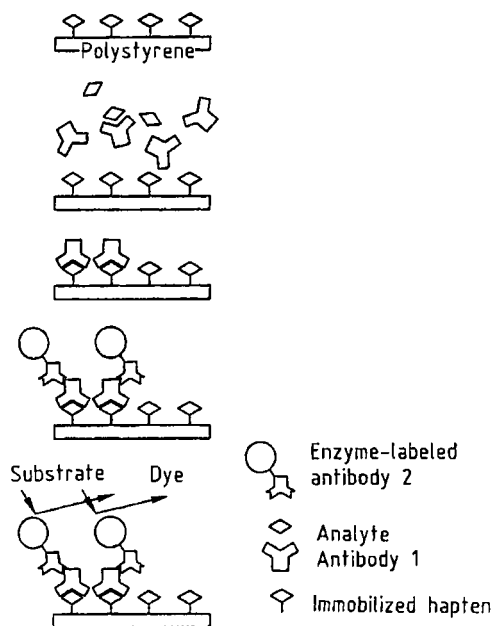
Figure 6. Principle of an indirect competitive enzyme immunoassay (ELISA)

The immobilized hapten corresponds to a hapten – BSA conjugate. AB 1 binds to the analyte, AB 2 binds to foreign antibodies and is covalently linked to an enzyme.

fluorophores with the highest quantum yields are extremely hydrophobic, and subject to quenching by adsorption to vessel walls.

Although they depend on special evaluation techniques, immunoassays based on signal production via chemi- and bioluminescence are successful [58]. In clinical chemistry, this type of signal production has already gained acceptance. This is, however, not the case in environmental analysis because of the often uncontrolled interaction of matrix components with the luminescence-producing enzyme.

Efforts are being made to improve the sensitivity of immunoassays by using long-lived fluorophore labeling and highly intense light sources (e.g., lasers). Labeling with rare earth chelates, which have fluorescence decay times in the range $600 - 1000 \, \mu s$, permits an especially easy separation of scattered light from the signal [59] – [62]. The high cost of UV laser systems is presently an obstacle to its wider acceptance. However, the first applications at 780 nm, using less expensive semiconductor diode lasers, have already been published [63].

The application of electrically produced time-resolved luminescence without light source represents an extremely interesting variant [64].

9.2.4.3.2. Sandwich Immunoassay

This type of noncompetitive heterogeneous immunoassay (Fig. 7) is encountered especially in the case of larger antigens with several epitopes [65]. The antibody, immobilized on the microtiter plate in the first step, has the function of binding the analyte (antigen) or standard in the solution. After a washing step, the second antibody, which recognizes the second epitope of the analyte, is added. After another washing step, a labeled (in this case with an enzyme) third antibody, which binds to AB 2, is added. Apart from the large number of steps involved, all of which hinder reproducibility, problems are also posed by the necessity of suppressing nonspecific adsorption in the presence of added reagents. Furthermore, the color formation is not inversely proportional to the analyte concentration. Sandwich immunoassay is not applicable to environmentally relevant haptens because it requires an analyte molecule of a certain minimum size.

9.2.4.3.3. Immunoassays using Immobilized Affinity Ligands

New immunoassay concepts were described in which immobilized affinity ligands are used in unique flow-through systems (flow-injection immunoassay, FIIA). These systems are exclusively heterogeneous. Both immobilization of a specific antibody or a hapten on the solid support is possible. In most cases the scheme of the assay used is based on a sequential competitive enzyme immunoassay procedure [66], [67].

9.2.4.3.4. Immunoassays with Biotin and Avidin

Avidin (an egg yolk protein) exhibits an extremely strong affinity (ca. 10^{15} L/mol) for biotin (vitamin H). This irreversible interaction can be utilized in the design of heterogeneous immunoassays with biotinylated, antigen-specific antibodies [68], [69]. Biotinylation can be performed without interfering with antibody function. The antibodies are applied in the usual manner in a competitive immunoassay (Fig. 8). In the last step, enzyme-labeled avidin is added in excess. Since there are four possibilities for the binding of avidin
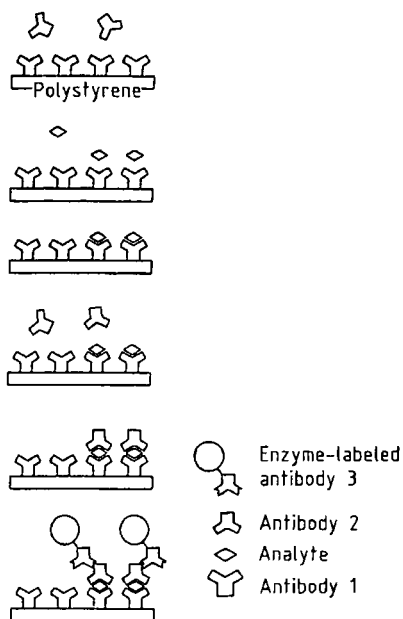
Figure 7. Principle of a noncompetitive immunoassay AB 1 is a catcher antibody, and binds to the analyte. AB 2 should be able to bind a second (identical or different) antigenic epitope. AB 3 binds to foreign antibodies (of type AB 2) and is covalently linked to an enzyme.

to biotin, labeled biotin can also be used to occupy binding sites of the bound avidin to amplify the recognition signal reaction. Streptavidin can be used instead of avidin.

9.2.4.3.5. Magnetic Particle Immunoassay

An especially elegant heterogeneous immunoassay format depends on coupling antibodies to magnetic particles. The assay is carried out in a test tube. The conjugate of analyte and peroxidase (tracer) is first added to the sample or standard and mixed well. An aliquot of an AB – magnetic particle suspension is then pipetted into the test tube and again mixed well. After a preselected incubation time, the particles are precipitated by a magnet. After pouring off the supernatant, the AB – magnetic particles are resuspended and re-precipitated (washing step). The depth of color, which is inversely proportional to the concentration of analyte in sample, is then determined in the usual manner via an enzyme – substrate reaction. Applications in the field of pesticide analysis have been published [70] – [72].

Instead of the tedious and destructive coupling of antibodies to magnetic particles, the utilization of natural bacterial magnetic particles is also possible [73].

9.2.5. Immunoaffinity Techniques

An immunosorbens represents an excellent agent for the specific preconcentration of analytes with the aid of antibodies (immunoextraction). Until now, this technique has been used mainly in clinical chemistry. Environmental applications started in 1994 [74]. The appropriate antibody is immobilized on a precolumn and, after a concentration phase, subjected to a desorption step. It is favorable if the eluent used for this purpose is the same as that employed in the subsequent liquid chromatographic separation.

The repeated use of the precolumn sometimes poses problems because desorption must be forcibly effected by denaturation of antibody, and is often reversible. The substantial band broadening in the desorption step represents another problem. The bands can be sharpened only if the analyte can be reconcentrated on a subsequent reversed phase column with a relatively weak eluent (low proportion of organic solvent).

Alternatively, immunoselective desorption can be carried out with a "displacer" having a still higher affinity for the immobilized antibody. In the subsequent chromatography, however, a reduction of the analytical "window" owing to the structural similarity between the displacer and the analyte is to be expected.

In recent years, a rather new technique — the sol-gel technology — was applied for the preparation of new immunosorbents (Fig. 9). Using this rather simple method, the antibodies are encapsulated in a silica network. They retain their activity and, in some cases gain even higher stability, and can react with ligands that diffuse into the highly porous matrix. Several papers were reported on the entrapment of polyclonal and monoclonal antibodies against environmental pollutants [75] – [78].

Preconcentration with immunosorbents, followed by quantification with an immunoassay, is not recommended because no increase in selectivity is achieved by using the same antibody in the preconcentration and in the test.

Immunosorbents are of interest in trace analytical problems which require optimum detection strength and selectivity.
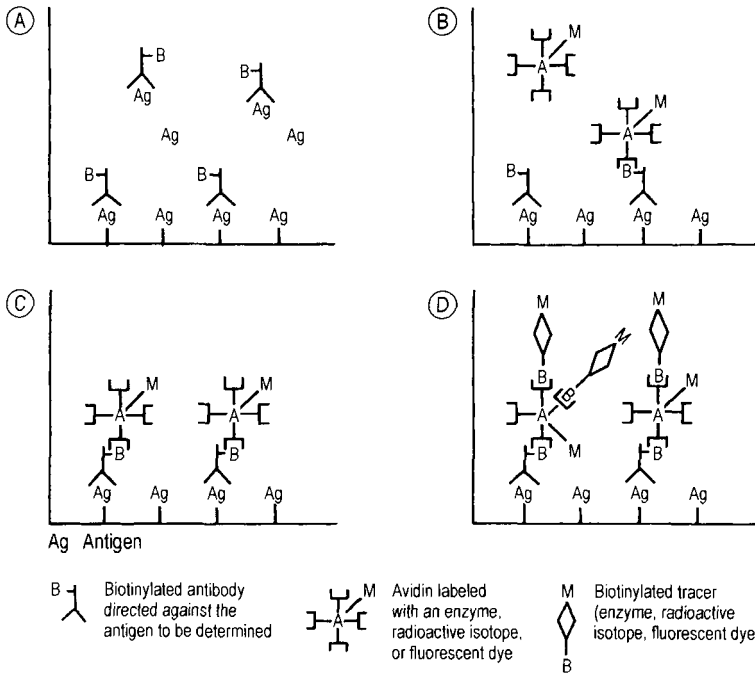
Figure 8. Principle of an immunoassay using the biotin–avidin system for labeling antibodies
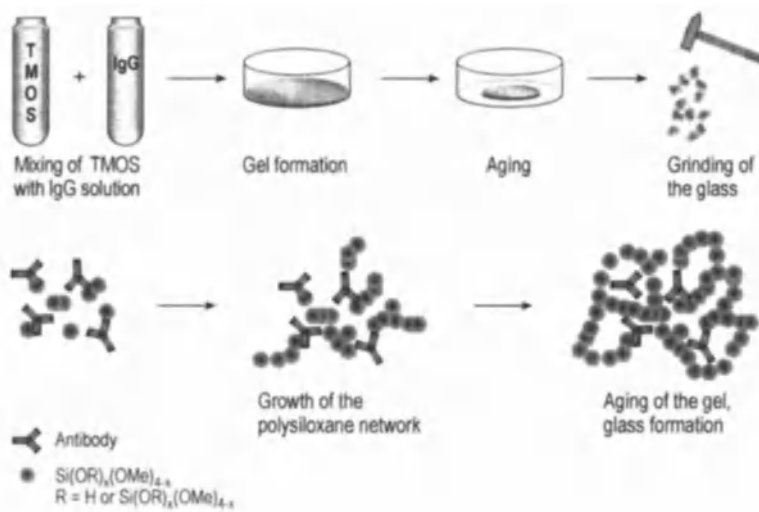
Figure 9. Encapsulation of antibodies in silicate glasses prepared by the sol–gel method. TMOS: tetramethoxysilane; IgG: immunoglobulin G

9.2.6. Immunoassays on Test Strips and other Planar Structures

Dry tests or test strips (also called lateral flow devices or immunomigration strips) are popular as

quick tests in clinical chemistry. Immunological tests are especially interesting because they can make use of the high selectivity of AB – Ag complex formation [79], [80]. The objetive is to accommodate all the reagents required for a quanti-

Immunochromatographic
strip with immobilized
antibody
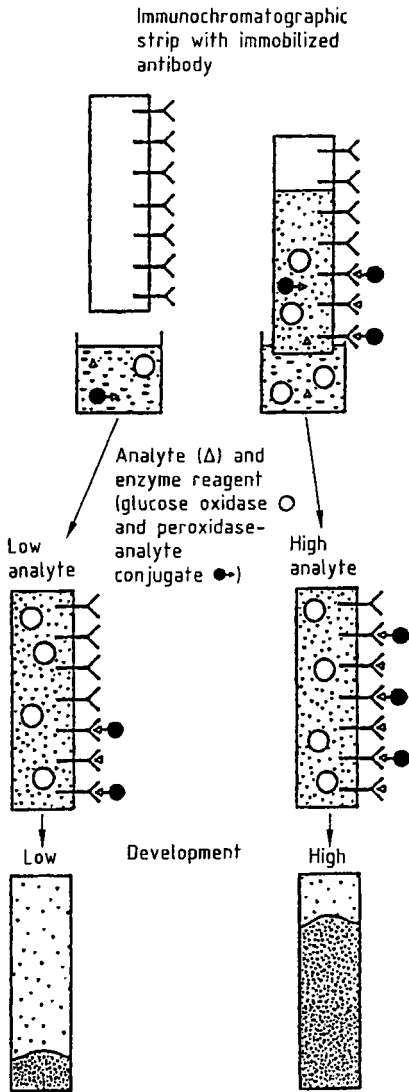
Figure 10. Principle of a test strip using enzyme immuno-
chromatography

fiable test on a simple strip, filter, or capillary. It
should be possible to dip the strip quickly into the
liquid sample or to place a drop of sample on the
carrier, and to determine the analyte concentration
from the resulting depth of color or length of a
colored band. Evaluation can be done visually or
spots can be read-out by a pocket reflectometer.
There should be tremendous possibilities for this
test, especially for on-the-spot environmental
analysis.

Two main variants are possible:

1) Accommodation of a homogeneous separation-
   free immunoassay on a dry matrix. This is
   possible for SLFIA and for ARIS. Other pos-
   sibilities are labeling with gold particles or
   liposomes. The art lies in contact-free accom-
   modation of all the reagents on the test strip, in
   such a way that no immune reaction takes
   place during storage.
2) The enzyme-channeling principle is especially
   suited to test strips. The test contains co-immo-
   bilized antibody and glucose oxidase. Two test
   strips must be used: one contains analyte-spe-
   cific antibody, and the other is impregnated
   with peroxidase-specific antibody. The strips
   are dipped first into the sample solution and
   then into a color developer. The color devel-
   oper contains a peroxidase tracer, glucose, and
   a chromogenic substrate (e.g., 4-chloro-
   1-naphthol). Channeling occurs by formation
   of H_2O_2 by glucose oxidase; this forms an
   insoluble dye on the surface of the test strip
   with 4-chloro-1-naphthol and peroxidase. In
   the case of the indicator test strip, the depth
   of color is influenced by competition with free
   analyte in the sample. The reference strip gives
   the standard value.

Enzyme immunochromatography (Fig. 10)
represents an especially attractive variant [81]. A
narrow strip of paper has the analyte-specific an-
tibody immobilized on it. The sample is mixed
with the tracer (analyte-labeled peroxidase and
glucose oxidase) and the end of the test strip is
dipped into this mixture. By capillary action, the
solution is sucked up into the strip matrix. The
sample is bound to the immobilized antibody as
it travels up the strip, resulting in competition
between the tracer and the analyte for the limited
binding sites along the strip. A zone is formed
which is proportional to the analyte concentration,
and occupied by tracer molecules. After this mini-
chromatography step, the entire strip is transferred
to a color developing solution which consists of
glucose and the peroxidase substrate, 4-chloro-
1-naphthol.

A strong color is developed, owing to enzyme
channeling only in places where the peroxidase
tracer is bound, in addition to the glucose that is
present everywhere. Thus, the occupied length,
which is proportional to the concentration of an-
alyte, can be visually quantified as in a thermom-
eter.
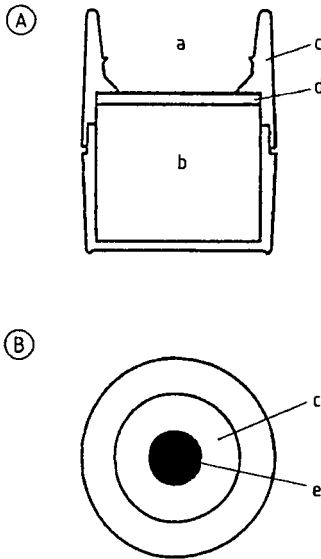
Figure 11. Cross section of the rapid immunofiltration test Hybritech ICON [72]
A) Side view; B) Top view
a) Fluid chamber; b) Absorbent reservoir; c) Membrane;
d) Support disk; e) Area of immobilized antibody

Figure 12. Typical calibration curve of a pyrene ELISA [50] The detection limit DL is calculated from the blank extinction value plus a three-fold standard deviation: 35 ppt. Test center point, $C=0.583$ µg/L; Margin of error: 1 Δ; $n=12$.

Another heterogeneous immunoassay, developed for field use [82], [83], employs immunofiltration (Fig. 11). Antibodies are immobilized in the middle of a nylon membrane. The first few drops placed on this membrane are immediately sucked down into a reservoir by capillary action. A few drops of washing buffer are then added, followed by AP- (alkaline phosphatase) labeled antibody to the analyte and, subsequently, a few drops of indoxyl substrate. As in the case of the sandwich immunoassay, the depth of color of the spot formed in the middle is linearly proportional to the analyte concentration; i.e., an ELISA is obtained for determination of aflatoxin [84], atrazin [85], [86], [87], carbaryl [88] or polychlorinated biphenyls [89].

9.2.7. Characterization and Interference

Immunochemical determination seems to be especially suited to the screening of large numbers of samples (e.g., water and soil analyses). Immunoassays are well suited for the analysis of the smallest sample volumes (e.g., in fog water analysis) because they allow analysis without sample preparation in a few microliters [90]. In critical

applications, immunoassays do not compete with, but rather supplement conventional trace analytical methods.

In contrast to applications in clinical chemistry, matrix interference must be expected when immunoassays are applied to environmental samples. Disturbances can be caused not only by the presence of high molecular, organic water components (humic acids, lignosulfonic acids), but also by inorganic ions (Al^{3+}, Fe^{3+}, other transition metals) and oxidizing agents (Cl_2, O_3, ClO_2, H_2O_2) used in water treatment.

An exceptional feature of competitive heterogeneous tests, which are at present the only tests that are sufficiently sensitive, is evaluation by photometrical readout of the microtiter plates.

In multiple determinations, the median is calculated for each calibration point (Fig. 12). The margins of error correspond to simple standard deviations. The medians, or the singly determined values, are fitted to a four-parameter function by a simplex process; each curve has seven calibration points. The zero value is not included in the mathematical evaluation, but is used only for control purposes:

$$Y = \frac{(A - D)}{\left[1 + \left(\frac{x}{C} \right)^R \right]}$$

where x is the concentration of analyte or hapten (µg/L); Y the extinction; A the maximal extinction (upper asymptote); D the minimal extinction
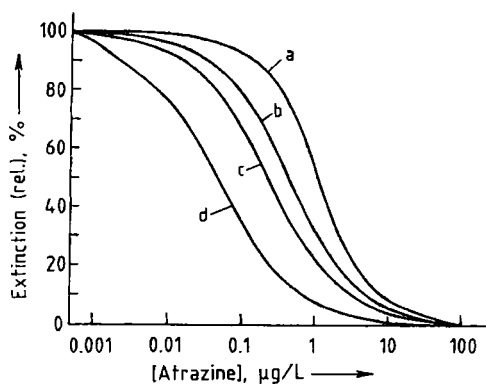
Figure 13. Calibration curves of an atrazine ELISA as a function of the incubation time of the tracer
a) 30 min; b) 10 min; c) 5 min;
Preincubation of unlabeled hapten: 60 min [91].

Table 2. Concentrations that cause 10 % inhibition of horse-radish peroxidase (N = 12)

Ion	Concentration, mg/L ($\pm 2\,s$)
CN^-	0.001 ± 0.0002
SCN^-	$0.02\ \pm 0.004$
ClO_3^-	$0.03\ \pm 0.004$
N_3^-	$0.04\ \pm 0.005$
ClO_4^-	$0.02\ \pm 0.003$
F^-	$0.15\ \pm 0.016$
Cu^{2+}	35.40 ± 4.04
Ni^{2+}	$8.70\ \pm 0.96$
Co^{2+}	17.30 ± 2.11
Al^{3+}	43.10 ± 2.97
Fe^{3+}	33.10 ± 4.74
Fe^{2+}	$1.30\ \pm 0.08$
$[Fe(CN)_6]^{3-}$	$0.02\ \pm 0.002$
$[Fe(CN)_6]^{4-}$	$0.17\ \pm 0.02$
Sn^{2+}	18.00 ± 2.11

(lower asymptote); C the test center point (μ g/L); and R a curvature parameter.

The curves are normalized with respect to both A and D:

$$Y_N = \frac{1}{\left[1 + \left(\frac{x}{C}\right)^B\right]}$$

where Y_N approximately corresponds to the signal B/B_0 normalized to the zero value (B represents the extinction of the sample and B_0 the blank reading determined with the plate reader) and enforces convergence of all calibration curves for [analyte] $\rightarrow 0$ and [analyte] $\rightarrow \infty$. This standardization has the advantage that, unlike B/B_0, neither a zero value (associated with a relatively large error) nor an excess value (difficult to obtain owing to solubility or contamination problems) is required.

In the case of an ELISA, the tracer incubation time is important, as illustrated in Figure 13 for a triazine herbicide test [91]. The test center point C can be lowered by a factor of 25 by reducing the incubation time from 30 to 1 min; if it is increased to 40 h, a marked increase in the test center point C is obtained. Thus, the test covers atrazine concentrations of more than five powers of ten by simple variation of the incubation time. It is not known whether polyclonality or a kinetic effect (imbalance between antibody, hapten, and labeled hapten) is responsible for the observed shift of the calibration curves.

In immunoassays, antibodies react to a large number of chemically closely related species with varying sensitivity. This cross-sensitivity, also called cross-reactivity, is an indication of high structural similarity between haptens. Since cross-sensitivities cannot be exactly predicted, they must be experimentally determined for each antibody and each test design. Crossreactivities CR_{50} are normally determined by forming ratios of the 50 % values of the substances to be compared.

However, the actual influence of cross-reacting compounds can depend on their concentrations and, therefore, a single value for the cross-reactivity cannot contain all the information. For unknown samples, the influence of cross-reacting substances as a function of concentration should be determined in advance.

Water components can influence either the tracer (the tracer enzyme in ELISA), or the antibody, or the hapten – AB binding. The concentrations of substances that cause a 10 % inhibition of the horseradish peroxidase normally used as label in enzyme immunoassays are listed in Table 2.

Strongly oxidizing substances, such as ClO_3^- or complexing agents (F^-, SCN^-, CN^-) exert an inhibiting effect. The azide ion, which is used as a bacteriostatic agent in immunology, can also cause inhibition. If horseradish peroxidase is used as a tracer enzyme, NaN_3 can be employed only for the preservation of antibody solutions. No effect is exerted by the following ions: Hg^{2+}, $Cr_2O_7^{2-}$, Na^+, K^+, Li^+, Sr^{2+}, Mg^{2+}, Ba^{2+}, Ca^{2+}, Ag^+, Pb^{2+}, NH_4^+, Cl^-, NO_2^-, NO_3^-, and SO_4^{2-}.

In the presence of Ca^{2+}, the activity of horseradish peroxidase can be increased by up to 40 %, depending on the contact time and the Ca^{2+} concentration [15].
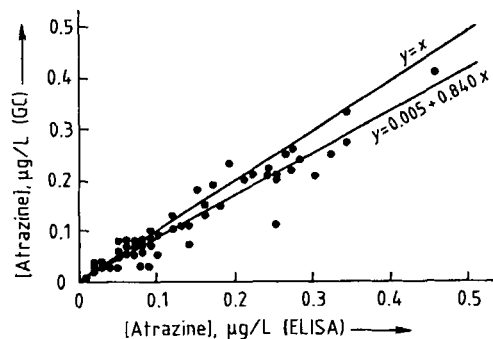
Figure 14. Correlation between two different analytical techniques (GC and ELISA) using atrazine as an example $n = 89$, $r = 0.956$ [15]

Sample components can interfere with antibodies in diverse ways. Evaluation is possible by means of calibration curves with the addition of various potentially interfering substances. Organic solvents (e.g., acetonitrile, methanol, and ethanol) in relatively low concentrations can affect the test. The maximal extinction falls increases on addition of solvents, the test center point is shifted by up to two powers of ten in the presence of 10 vol % of solvent [15]. In the immunological determination of trace substances (e.g., pesticides), the standard solutions should also be made up with the corresponding solvents.

An extremely important point in process characterization is validation with, e.g., classical methods, such as gas chromatography (GC) Fig. 14) or liquid chromatography. Atrazine in surface waters was measured by means of ELISA and GC, and the results compared. The results correlate well, and confirm the theory that immunological methods can give reliable values in practice. The results of both methods are equivalent with a probability of 95 %. No false negative values were obtained. The same applies to the use of immunoassays on soil extracts. In comparison with GC analyses, ELISA usually gives higher analyte concentrations, which is put down to the effects of humic acid.

9.2.8. Future Immunological Applications and Techniques

Apart from the extension of the tests already available to include analytes measurable only by expensive and tedious instrumental analytical techniques, the spectrum of applications is certain to widen.

Until now, water and body fluid (serum, urine) represent the most common test matrices. Soil analysis has already begun to profit tremendously from immunoassays. Very few applications exist in air monitoring, even though active and passive measuring procedures could profit increasingly from immunological techniques for the monitoring of diffuse sources.

The further development of immunolocalizing analytical techniques in conjunction with state-of-the-art optoelectronics and luminescence microscopy also seems worthwhile for the recognition of colloidal or cell-bound haptens.

In chemical sensors, the adaptation of immunological recognition techniques to optical and electrochemical transducers has been under investigation for several years.

An intense discussion presently goes on the usefulness of miniaturized analytical techniques, which may be considered part of the emerging nanotechnology complex. Immunological technology, after hyphenation with chromatographic and electrophoretic techniques, may be transferred to a chip format, obtaining a complete and very efficient "analytical-lab-on-a-chip". If a library of antibodies (an antibody array) is available, exhibiting a different pattern of affinity for a series of structurally related compounds, an appropriate statistical analysis (multivariate statistical techniques, parametric models) has the power to turn the problem of cross-reactivity into an advantage (multianalyte immunoassays). The breakthrough of this technology will mainly depend on the availability of improved solid supports and antibodies, and on the development of more simple instruments [92]–[95].

Increasing activities in validation and standardization of immunochemical tests, initiated by federal government and regulatory agencies of several countries and well-recognized national and international organizations will lead to more transparency and uniformity in method development and evaluation. This will help that the assays get rid of their image as "dubious biological tests", as they are sometimes called by analytical chemists [96]. Immunological methods cannot be assessed simply as "good" or "bad" but rather as suitable for an application or not.

9.3. References

[1] J. P. Sherry, *Crit. Rev. Anal. Chem.* **23** (1992) 217.

[2] J. P. Sherry, *Chemosphere* **34** (1997) 1011.

[3] E. P. Meulenberg et al., *Environ. Sci. Technol.* **29** (1995) 553.

[4] E. A. H. Hall (ed.): *Biosensoren.* Springer, Berlin 1990.

[5] B. Eggins (ed.): *Biosensors.* Wiley and Teubner, Chichester 1996.

[6] K. R. Rogers, A. Mulchandani (eds.): *Affinity Biosensors,* Humana Press, Totowa 1998.

[7] R. Schmid (ed.): *Flow Injection Analysis (FIA)* Based on Enzymes or Antibodies, VCH Verlagsgesellschaft, Weinheim 1991.

[8] P. Tijssen (ed.): *Practice and Theory of Enzyme Immunoassays.* Elsevier, Amsterdam 1985.

[9] J. Peters, H. Baumgarten (eds.): *Monoklonale Antikörper,* Springer, Berlin 1990.

[10] G. T. Hermanson (ed.): *Bioconjugate Techniques.* Academic Press, San Diego 1996.

[11] G. Köhler, C. Milstein, *Nature* **256** (1975) 52.

[12] H. A. Lee, M. R. A. Morgan, *Trends in Food Sci. Technol.* **4** (1993) 129.

[13] K. Kramer, *Anal. Letters* **31** (1998) 67.

[14] C. W. Bell et al., *ACS Symp. Ser.* **586** (1995) 413.

[15] T. Ruppert, L. Weil, R. Niessner, *Vom Wasser* **78** (1991) 387.

[16] P. Albro et al., *Toxicol. Appl. Pharmacol.* **50** (1979) 137.

[17] W. F. W. Stöcklein, A. Warsinke, F. W. Scheller, *ACS Symp. Ser.* **657** (1997) 373.

[18] C. Keuchel, L. Weil, R. Niessner, *SPIE* **1716** (1992) 44.

[19] R. Schneider, L. Weil, R. Niessner, *Int. J. Environ. Anal. Chem.* **46** (1992) 129.

[20] R. Yalow, S. Berson, *J. Clin. Invest.* **39** (1960) 1157.

[21] T. Chard (ed.): *An Introduction to Radioimmunoassay and Related Techniques,* Elsevier, Amsterdam 1990.

[22] P. Masson et al. in J. Langone, H. van Vunakis (eds.): *Immunochemical Techniques,* Part C, Academic Press, New York 1981, p. 106.

[23] Y. V. Lukin et al., *ACS Symp. Ser.* **657** (1997) 97.

[24] S. Obata et al., *Anal. Sci. Suppl.* **7** (1991) 1387.

[25] H. Weetall, A. Gaigalas, *Anal. Lett.* **25** (1992) 1039.

[26] E. Engvall in H. van Vunakis, J. Langone (eds.): *Methods in Enzymology,* **vol. 70,** Academic Press, New York 1970, p. 419.

[27] D. Morris et al., *Anal. Chem.* **53** (1981) 658.

[28] D. L. Morris, R. T. Buckler, *Meth. Enzymol.* **92** (1983) 413.

[29] E. Soini, I. Hemmilä, *Clin. Chem. (Winston-Salem, N.C.)* **25** (1979) 353.

[30] T. Ngo et al., *J. Immunol. Methods* **42** (1981) 93.

[31] R. Niessner et al.: "Application of Time-Resolved Laser-Induced Fluorescence for the Immunological Determination of Pesticides in the Water Cycle," in B. Hock, R. Niessner (eds.): *Immunological Detection of Pesticides and their Metabolites in the Water Cycle,* VCH Verlagsgesellschaft, Weinheim, in press.

[32] S. A. Eremin et al., *ACS Symp. Ser.* **586** (1995) 223.

[33] S. G. Reeves, S. T. A. Siebert, R. A. Durst, *ACS Symp. Ser.* **586** (1995) 210.

[34] J. Rittenburg et al. in M. Vanderlaan et al. (eds.): *Immunoassays for Trace Chemical Analysis,* **vol. 451,** American Chemical Society, Washington D.C., 1990, p. 28.

[35] J. Brady et al.: "Biochemical Monitoring for Pesticide Exposure," *ACS Symp. Ser.* **382** (1989) 262.

[36] J.-M. Schläppi, W. Föry, K. Ramsteiner, "Immunochemical Methods for Environmental Analysis," *ACS Symp. Ser.* **442** (1990) 199.

[37] H. Newsome, P. Collins, *J. Assoc. Off. Anal. Chem.* **70** (1987) 1025.

[38] Q. Li et al., *J. Agric. Food Chem.* **39** (1991) 1537.

[39] S. Wie, B. Hammock, *J. Agric. Food Chem.* **32** (1984) 1294.

[40] R. Dreher, B. Podratzki, *J. Agric. Food Chem.* **36** (1988) 1072.

[41] F. Jung et al., *J. Agric. Food Chem.* **37** (1989) 1183.

[42] H. Scholz, B. Hock, *Anal. Lett.* **24** (1991) 413.

[43] S. Kreissig et al., *Anal. Lett.* **24** (1991) 1729.

[44] Q. Li et al., *Anal. Chem.* **61** (1989) 819.

[45] B. Riggle, B. Dunbar, *J. Agric. Food Chem.* **38** (1990) 1922.

[46] P. Ulrich, R. Niessner, *Fres. Environ. Bull.* **1** (1992) 22.

[47] C. Hall et al., *J. Agric. Food. Chem.* **37** (1989) 981.

[48] A. Roda et al., *Environ. Technol. Lett.* **12** (1991) 1027.

[49] D. Knopp et al., *ACS Symp. Ser.* **657** (1997) 61.

[50] K. Meisenecker, D. Knopp, R. Niessner, *Analytical Methods and Instrumentation* **1** (1993) 114.

[51] C. Keuchel, L. Weil, R. Niessner, *Anal. Sci.* **8** (1992) 9.

[52] J. Van Emon, R. Gerlach, *Bull. Environ. Contam. Toxicol.* **48** (1992) 635.

[53] B. Watkins et al., *Chemosphere* **19** (1989) 267.

[54] D. Wylie et al., *Anal. Biochem.* **194** (1991) 381.

[55] S. Wie, B. Hammock, *Anal. Biochem.* **125** (1982) 168.

[56] T.-G. Wu, R. Durst, *Mikrochim. Acta* **1990**, 187.

[57] S. Choquette, L. Locascio-Brown, R. Durst, *Anal. Chem.* **64** (1992) 55.

[58] A. Tsuji, M. Maeda, H. Arakawa, *Anal. Sci.* **5** (1989) 497.

[59] E. Diamandis, *Clin. Biochem.* **21** (1988) 139.

[60] E. Diamandis, T. Christopoulos, *Anal. Chem.* **62** (1990) 1149 A.

[61] T. Christopoulos, E. Diamandis, *Anal. Chem.* **64** (1992) 342.

[62] Y.-Y. Xu et al., *Analyst (London)* **116** (1991) 1155.

[63] T. Imasaka et al., *Anal. Chem.* **62** (1990) 2405.

[64] J. Kanakare et al., *Anal. Chim. Acta* **266** (1992) 205.

[65] Hybritech, US 4 376 110, 1983 (G. S. David, H. E. Greene).

[66] B. B. Kim et al., *Anal. Chim. Acta* **280** (1993) 1991

[67] P. M. Kraemer et al., *ACS Symp. Ser.* **657** (1997) 153.

[68] M. Wilchek, E. A. Bayer, *Meth. Enzymol.* **184** (1990) 14.

[69] F. Eberhard: "Entwicklung von Immunoassays und Markierungsmethoden für die Bestimmung von Schimmelpilzgiften mit immunooptischen Biosensoren," Dissertation, Universität Hannover 1991.

[70] F. Rubio et al., *Food Agric. Immun.* **3** (1991) 113.

[71] T. Lawruk et al., *Bull. Environ. Contam. Toxicol.* **48** (1992) 643.

[72] J. A. Itak et al., *ACS Symp. Ser.* **657** (1997) 261.

[73] N. Nakamura, T. Matsunaga, *Anal. Sci.* **Suppl. 7** (1991) 899.

[74] G. S. Rule, A. V. Mordehai, J. Henion, *Anal. Chem.* **66** (1994) 230.

[75] J. Zuehlke, D. Knopp, R. Niessner, *Fresenius J. Anal. Chem.* **352** (1995) 654.

[76] A. Turniansky et al., *J. Sol – Gel Sci. Technol.* **7** (1996) 135.

[77] M. Cichna et al., *Chem. Mater.* **9** (1997) 2640.

[78] T. Scharnweber, D. Knopp, R. Niessner, *Field Anal. Chem. Technol.* **4** (2000) 43.

[79] D. Morris, D. Ledden, R. Boguslaski, *J. Clin. Lab. Anal.* **1** (1987) 243.

[80] D. Litman in T. Ngo, H. Lenhoff (eds.): *Enzyme-Mediated Immunoassay*, Plenum Press, New York 1985, p. 155.

[81] R. Zuck et al., *Clin. Chem. (Winston-Salem, N.C.)* **31** (1985) 1144.

[82] G. Valkiro, R. Barton, *Clin. Chem. (Winston-Salem, N.C.)* **31** (1985) 1427.

[83] J. Rittenburg et al.: "Immunochemical Methods for Environmental Analysis," *ACS Symp. Ser.* **442** (1990) 28.

[84] E. Schneider et al., *Food Agric. Immun.* **3** (1991) 185.

[85] C. Morber, L. Weil, R. Niessner, *Fres. Environ. Bull.* **2** (1993) 151.

[86] C. Keuchel, R. Niessner, *Fresenius J. Anal. Chem.* **350** (1994) 538.

[87] A. Dankwardt, B. Hock, *Biosens. Bioelectron.* **8** (1993) XX.

[88] S. Morias, A. Maquieira, R. Puchades, *Anal. Chem.* **71** (1999) 1905.

[89] M. Del Carlo et al. in B. Hock et al. (eds.): *Biosensors and Environmental Diagnostics*, Chap. 2, Teubner Verlag, Stuttgart 1998.

[90] F. Trautner, K. Huber, R. Niessner, *J. Aerosol Sci.* **23** (1992) S 999.

[91] M. Weller, L. Weil, R. Niessner, *Mikrochim. Acta* **108** (1992) 29.

[92] R. Ekins, F. Chu, E. Biggart, *Anal. Chim. Acta* **227** (1989) 73.

[93] A. Brecht, G. Gauglitz, *Anal. Chim. Acta* **347** (1997) 219.

[94] M. G. Weller et al., *Anal. Chim. Acta* **393** (1999) 29.

[95] H. Kido, A. Maquieira, B. D. Hammock, *Anal. Chim. Acta* **411** (2000) 1.

[96] J. Rittenburg, J. Dautlick, *ACS Symp. Ser.* **586** (1995) 301.

10. Basic Principles of Chromatography

GEORGES GUIOCHON, Department of Chemistry, University of Tennessee, Knoxville, TN, 37996-1600 , United States, Division of Chemical and Analytical Sciences, Oak Ridge National Laboratory, Oak Ridge, TN, 37831-6120, United States

Abbreviations

a	initial slope of the isotherm
C, C_m	mobile-phase concentration
C_s	stationary-phase concentration
d_f	film thickness
d_p	particle size (diameter)
D_a	axial dispersion coefficient
D_l	diffusion coefficient, stationary (liquid) phase
D_m	diffusion coefficient, mobile (gas) phase
F	phase ratio
F_v	mobile-phase flow rate
h	reduced plate height
H	height equivalent to a theoretical plate (HETP)
k_0	column permeability
k', k'_0	retention factor
K	Henry constant
L	column length
M_n	nth moment of a statistical distribution
N, N_c	number of plates
p	partial pressure
P	probability
P^0	vapor pressure
q	Langmuir isotherm
R	fraction of sample in the mobile phase; resolution
S_a	adsorbent surface area
S_i	detector response factor
t_R	retention time
u	mobile-phase velocity
v	reduced velocity
v_m	volume per plate of the mobile phase
v_s	volume per plate of the stationary phase
V_g	specific retention volume
V_G	volume of gas
V_l	volume of liquid
V_m	mobile-phase volume
V_N	correlated retention volume
V_R	retention volume
α	relative retention
γ^{∞}	activity coefficient at infinite dilution
ε_T	total packing porosity
η	mobile-phase viscosity
$\mu_n, \bar{\mu}_n$	absolute and central moments
Φ	fractional loss of efficiency
σ	standard deviation
τ	time variance

10.1. Introduction

Chromatography is the most powerful separation technique available. For reasons that will be made clear later, it is easy to implement chromatographic techniques with small samples, and to carry out separations on an analytical scale. Separated components are usually detected online, and sometimes they are also characterized, after which they are generally discarded. In preparative applications, the purified components are collected for further investigation, although new difficulties may be encountered in an attempt to scale-up the equipment and operate at high concentrations.

Chromatography is a separation method based on differences in equilibrium constants for the components of a mixture placed in a diphasic system. A chromatographic system is one in which a fluid mobile phase percolates through a stationary phase. The stationary phase is often a bed of nonconsolidated particles, but this is not essential. All that is in fact required is two phases in relative motion and excellent contact between these phases so that concentrations in the stationary phase are always very near their equilibrium values. For example, a tubular column through which the mobile phase flows and whose walls have been coated with a layer of stationary phase is a most suitable implementation of chromatography. In all cases the mobile phase is responsible for transporting the sample components through the stationary phase. The velocity of each component, hence its residence time, depends on the both mobile-phase velocity and the distribution equilibrium constant for that component. At least in principle, proper choice of the two phases constituting a chromatographic system permits selective adjustment of the relative migration rates of the mixture components and of the extent of their eventual separation.

The mobile phase is a fluid. The main criteria for its selection are:

1) Good solubility for the analytes
2) Low viscosity

All types of fluids have been used, but the chief types of chromatography are gas, dense-gas (more commonly called supercritical-fluid), and liquid chromatography. In *gas chromatography* (GC; → Gas Chromatography), a quasi-ideal gas is used; pressures are kept low, analyte solubility in the mobile phase depends only on its vapor pressure, and the role of the mobile phase is almost purely mechanical, with minor corrections due to nonideal behavior of the gas phase [1]. Retention adjustment is accomplished primarily by changing the temperature. With denser gases, an analyte's fugacity differs markedly from its vapor pressure, and analyte solubility depends strongly on the mobile-phase density as well as the temperature. Thus, retention adjustment in supercritical-fluid chromatography can be effected by changing not only the temperature but also the average pressure. In both gas and supercritical-fluid chromatography, the *selectivity* (i.e., the ability to achieve

a separation) is relatively insensitive to temperature and density. In *liquid chromatography* (LC), the mobile phase is a mixture of solvents ($\rightarrow$ Liquid Chromatography). Retention adjustments and selectivity changes are accomplished by changing the mobile-phase composition and introducing a variety of strong solvents or additives.

The stationary phase may be a solid (i.e., an adsorbent, with a relatively large specific surface area and very accessible pore channels) or a liquid coated on a solid support (to avoid stripping of the stationary phase by the streaming action of the mobile phase). Intermediate systems include immobilized polymers or surface-bonded shorter chemical species (e.g., $C_{18}H_{37}$) of widely varying chain length and molecular mass. Gas – solid, gas – liquid and liquid – solid chromatography are the most popular chromatographic methods. Ion-exchange resins in conjunction with ionic solvents, gels with pores comparable in size to the analyte molecules, and stationary phases containing suitable complexing agents or groups have also served as the basis for important methods, producing selectivities very different from those observed with more classical molecular interactions.

Another useful distinction involves the mechanism by which the mobile phase is transported through the stationary phase. The most popular approach is forced convection by pressurization, based on cylinders of compressed gases in gas chromatography, or a mechanical pump in liquid chromatography. In both cases the stationary phase is placed inside a tube, the *chromatographic column*. Other methods are available as well. For example, in thin layer chromatography ($\rightarrow$ Thin Layer Chromatography) capillary forces suffice to draw a liquid stream through a dry, porous bed in the form of a paper sheet or a thin layer of adsorbent coated on a glass, metal, or plastic sheet. More recently, electro-osmosis has been used to force a liquid stream through a column [2]. Although the method is quite similar to standard column liquid chromatography, the resulting separations may be markedly superior.

10.2. Historical Development

It is not possible to review here the entire history of chromatography and do justice to all the scientists who have contributed to designing and developing the dozens of major separation techniques based on the principle of chromatog-

raphy. The history of the origin of chromatography has been recounted by SAKODINSKY [3] and ETTRE [4]. Numerous other publications recall landmark contributions in the field.

Although a number of early publications introduced certain ideas related to chromatography, it is clear that the method was conceived and developed by the Russian chemist TSWETT at the beginning of this century [5]–[7]. TSWETT is best described as a biophysicist. He had a remarkably clear understanding of the interactions between the various phenomena involved in chromatography — some thermodynamic and others kinetic — and of the technical problems that would need to be solved to turn the chromatographic principle into a useful group of analytical methods. Unfortunately, living during the very early stages of the high-technology age, in a country soon to be ravaged by civil war, he was unable to succeed. His ideas remained untried until a group of German chemists who were informed of TSWETT's early work through WILLSTÄTTER [3] succeeded in using chromatography for the extraction and purification of plant pigments.

The original work of TSWETT is most interesting to read — at least the fragments available in the modern literature [3], [4], [8]. Many of the modern problems of high-performance liquid chromatography are already addressed in these early writings. TSWETT knew that small particles should be used for optimum performance, but he had no way to force a mobile-phase stream under pressure through a column. It is interesting also to note that the initial difficulties KUHN's group had to solve in developing their analytical schemes were due to an isomerization of carotenoid derivatives catalyzed by the silica gel used as stationary phase [3], [4]. TSWETT had deliberately used calcium carbonate for these separations, reserving silica gel for the separation of more stable compounds [7].

A. J. P. MARTIN has been associated with several critical developments in chromatography. His Nobel prize paper with SYNGE [9] represents the origin of both partition and thin layer chromatography (TLC). Originally carried out with paper sheets, TLC evolved later under the leadership of STAHL in the direction of consolidated layers of fine particles. Later, MARTIN, GORDON and CONSDEN [10] published the first two-dimensional chromatograms. Here a sample is introduced at one corner of a large, square paper sheet, and two successive developments are conducted in orthogonal directions, using two different solvent

mixtures as mobile phases. This method leads to a dramatic increase in separation power. Finally, in cooperation with JAMES, MARTIN invented gas–liquid chromatography [11], a technique ideally suited to the incorporation of many developments in modern instrumentation that occurred at that time. Gas–liquid chromatography led first to a major new instrumentation technique, then to the rejuvenation of conventional liquid chromatography, and finally to the more recent development of capillary-zone electrophoresis.

10.3. Chromatographic Systems

Chromatography is based on phase-equilibrium phenomena. The components of the analyte sample are caused to equilibrate between two phases, a mobile phase and a stationary phase. Because the mobile phase percolates through the stationary phase, rapid mass transfer takes place, and the mobile phase carries the components through the column to a detector. The velocity of this transfer is related to the equilibrium constant. Hence, only compatible combinations of mobile and stationary phases can be used in practice.

10.3.1. Phase Systems Used in Chromatography

The mobile phase must be a fluid: A gas, a dense gas, or a liquid, where the rate of mass transfer through the mobile phase, characterized by the diffusion coefficient, decreases in the order listed. To compensate for this marked decrease, finer and finer particles are used as the mobile phase becomes more dense. The practical lack of compressibility of liquids supports this trade-off, since small particles demand higher pressures. However, viscous liquids (e.g., glycerol, oligomers, or concentrated solutions of polymers) are excluded, because the associated pressure requirements would be unrealistic.

The stationary phase must be compatible with the mobile phase. Lack of solubility of the stationary phase in the mobile phase and a large interfacial area between the two phases are the two major requirements. Most adsorbents are applicable to any of the three types of mobile phase, giving rise, for example, to both gas–solid and liquid–solid chromatography. The specific sur-

face area of the absorbent must be large, not only to enhance column efficiency but also to ensure sufficient retention [12]. This surface must be highly homogeneous as well to avoid the presence of active sites on which the adsorption energy is exceptionally high leading to excessive retention, and a poor band profile. Most of the adsorbents for chromatography have been developed to such a point that they display the required properties: alumina, silica, molecular sieves, and activated or graphitized carbons are the most popular.

Silica comes in two broad types: *Pure* or *underivatized silicas* (all of which have very similar properties, but among which chromatographers can recognize more varieties than there are commercial brands) and the *chemically bonded silicas* [13]. Virtually every conceivable type of chemical group—from alkyl derivatives to sugars and peptides—has been bonded to silica, leading to a large number of phases, each with its specific properties and applications. In liquid–solid chromatography, two types of systems are distinguished: *Normal-phase* systems based on a polar stationary phase and a nonpolar mobile phase (e.g., silica and dichloromethane), and *reversed-phase* systems, in which the stationary phase is nonpolar and the mobile phase polar (e.g., C_{18} silica and water–methanol solutions).

Liquids can also be used as the stationary phase in chromatography, but they must be immiscible with the mobile phase. In gas–liquid chromatography this means that the stationary phase must have a low vapor pressure to avoid its premature depletion by the mobile phase. Convection would be highly detrimental to the effectiveness of a separation, and this is avoided by coating the liquid phase on an inert solid. The support is usually a pulverulent solid consisting of porous particles. To avoid significant contributions from adsorption (i.e., mixed retention mechanisms), the solid should have a low specific surface area, but it should also be highly porous to permit the use of a large amount of liquid. Alternatively, a stationary liquid can be coated on a nonporous support (e.g., glass beads, or the inner wall of an empty tube). Considerable effort has been devoted to the development of reproducible methods for coating the inner wall of a quartz tube with a homogeneous, stable, thin film of liquid.

While gas–liquid chromatography has become the preferred implementation of gas chromatography because of the excessive adsorption energy associated with gas–solid equilibria, for most compounds but gases, liquid–solid chromatog-

raphy is the preferred implementation of liquid chromatography. It is difficult to find a stationary liquid that is insoluble in the liquid mobile phase but at the same time not so different from this phase that the partition coefficients are either negligible or nearly infinite. The stability of a liquid coated on an inert support is low, and often the stationary phase is quickly stripped from the support in the form of a dilute emulsion.

Other important stationary phases for liquid chromatography are ion-exchange resins (*ion-exchange chromatography*, which uses anionic or cationic resins) and porous solids that do not adsorb the analytes but instead have pores so small that access to a fraction of the pore volume by some of the analyte molecules is more or less restricted depending on molecular size (*size-exclusion chromatography*).

Finally, chemical groups capable of forming charge transfer complexes or other types of labile complexes with specific components of a sample have also been used successfully as stationary phases. *Affinity chromatography* is one implementation of this concept, in which complex formation between a specific antibody and one analyte results in both exceptional selectivity for a particular component and an inherent lack of flexibility, two striking characteristics of this method.

10.3.2. Methods of Implementing Chromatography

There are three methods of implementing chromatography: Elution, displacement, and frontal analysis. An immense majority of analytical applications are carried out on the basis of elution. Equilibrium isotherms are often measured using frontal analysis. Finally, preparative separations, though most often accomplished by overloaded elution, also lend themselves to displacement. This has the advantage of producing more concentrated fractions, but it requires special development, choice of an appropriate displacer, and the extra step of column regeneration between the processing of successive batches.

In *elution*, a pulse of sample is injected into the mobile-phase stream before it enters the column. Under the linear conditions most often associated with analytical applications, each pulse of sample component behaves independently, migrating along the column at a velocity proportional to the mobile-phase velocity but dependent also on the equilibrium coefficient between the two phases. If the experimental conditions are properly chosen, pulses of the mixture components arrive separately at the column exit, and the chromatogram consists of a series of peaks, one for each component.

In *frontal analysis*, a stream of a solution of the sample in the mobile phase is abruptly substituted for the mobile-phase stream at the beginning of the experiment. If the components are present at very low concentrations, the front of each component "step" moves at the same velocity as the pulses in an elution experiment, resulting in a "staircase" in which the profile of each individual step can be regarded as an error function. However, in most applications of frontal analysis the component concentrations are in fact rather high, and the isotherms are no longer linear. Accordingly, they are not independent of each other, but are competitive; i.e., the concentration a given compound adsorbed at equilibrium depends on the concentrations of all the components present. The propagation of the various component steps results in breakthrough curves that are steep, and the concentrations of all the components in the eluent change when any one of them breaks through. As a consequence, frontal analysis can be used only to purify the first eluted component of a mixture, or to measure single-component or competitive (mainly binary) isotherms.

In *displacement* chromatography, a pulse of mixture is injected, and this is followed by a step of a single component called a *displacer*, which is adsorbed more strongly than any of the mixture components. After a certain period of time, during which the profiles of each pulse become reorganized, an isotachic pulse train is formed. In the isotachic train, each component forms its own concentration "boxcar," with a height that depends on the component and displacer isotherms as well as on the displacer concentration, and a width proportional to the amount of the corresponding component in the sample [14], [15].

10.4. Theory of Linear Chromatography

One can distinguish between two basic types of assumption common to all theoretical efforts to understand chromatography [16]. First, it may be assumed that the equilibrium isotherm is linear; alternatively, one might acknowledge that, under the experimental conditions selected, this isotherm

is not linear. One can thus study models of either linear or nonlinear chromatography. Analytical applications almost always involve small or dilute samples, and under these conditions any curvature in the isotherm near the origin can be ignored and the isotherm can be regarded as linear. In preparative applications — more generally, in any case of concentrated solutions — this assumption is no longer tenable, and it becomes necessary to consider nonlinear chromatography (see Chap. 10.10).

Similarly, a decision must be made whether or not to take into account the influence on band profiles of such phenomena as axial dispersion (dispersion in the direction of the concentration gradient in the column) and resistance to mass transfer (i.e., the fact that equilibration between mobile and stationary phases is not instantaneous). These phenomena are responsible for the finite efficiency of actual columns. Neglecting them and assuming the column to have infinite efficiency leads to a model of *ideal* chromatography. Taking them into account results in one of the models of *nonideal* chromatography.

The combination of these two types of consideration leads to four theories of chromatography. The first, *linear, ideal chromatography*, is trivial: each component pulse moves without changes in its profile, and its motion is independent of other pulses. In *linear, nonideal chromatography* one studies the influence of kinetic phenomena on band profiles, assuming each component pulse to move independently of all the others, and at a constant velocity. The chromatogram of a mixture in this case is the sum of the chromatograms for each independent component. This is no longer true in *ideal, nonlinear chromatography*. Here one studies the influence of nonlinear phase-equilibrium thermodynamics on the band profile. The result depends on the amount of each component present in the mixture, and individual elution profiles for the various components cannot be obtained independently (see Chap. 10.10). Nevertheless, both linear, nonideal and ideal, nonlinear chromatography provide relatively simple solutions. In *nonlinear, nonideal chromatography* it is necessary to take into account both finite column efficiency and a nonlinear, nonindependent behavior of the isotherms, and only numerical solutions are possible.

10.4.1. Plate Models

These models assume that the chromatographic column can be divided into a series of a finite number of identical plates. Each plate contains volumes v_m and v_s of the mobile and stationary phases, respectively. The sample is introduced as a solution of known concentration in the mobile phase used to fill the required number of plates. Plate models are essentially empirical, and cannot be related to first principles. Depending upon whether one assumes continuous or batch operation, two plate models can be considered: The Craig model and the Martin and Synge model.

10.4.1.1. The Craig Model

In this model [17], mobile phase is transported through the column by withdrawing it from the last plate, moving the remaining mobile phase from each plate to the next one in succession, and refilling the first plate. The process is repeated as often as necessary to elute the material introduced. It is easily shown that the fraction R of the sample in the mobile phase is expressed by

$$R = \frac{v_m C_m}{v_m C_m + v_s C_s} = \frac{1}{1 + Fa} = \frac{1}{1 + k_0'} \tag{1}$$

where C_m and C_s are the equilibrium concentrations in the mobile and stationary phases, respectively. F is the phase ratio, a is the initial slope of the isotherm, and k_0' is the retention factor. After the moving process has been repeated r times, the probability P of finding a molecule in the plate of rank l is

$$P_{l,r} = \frac{r!}{l!(r-l)!} R^l (1-R)^{r-l} \tag{2}$$

and the elution profile (at the exit of the last plate, of rank N_c) is

$$f(r) = R \frac{r!}{N_c!(r-N_c)!} R^{N_c} (1-R)^{r-N_c} \tag{3}$$

When N_c exceeds about 50 plates, this profile cannot be distinguished from a Gaussian profile with a retention volume

$$V_R = V_m(1 + k_0') \tag{4}$$

where $V_m = N_c v_m$, and a standard deviation

$$\sigma_v = V_R \sqrt{\frac{1-R}{N_c}} \tag{5}$$

10.4.1.2. The Martin–Synge Model

This is a continuous plate model [9]. The mobile phase is transferred from one plate to the next, and each plate is regarded as a perfect mixer. Integration of the differential mass balance in every plate leads to a Poisson distribution inside the column, and to the elution profile:

$$f(v) = \frac{N}{V_R} e^{-\frac{Nv}{V_R}} \left(\frac{Nv}{V_R}\right)^{(N-1)} \frac{1}{(N-1)!} \tag{6}$$

As the plate number N increases, just as in the case of the Craig model the profile tends rapidly toward a Gaussian profile:

$$f(v) = \frac{1}{\sigma_v \sqrt{2\pi}} \exp\left[-\frac{(v - V_R)^2}{2\sigma_v^2} \right] \tag{7}$$

with a retention volume and standard deviation given by

$$V_R = N(v_m + a\,v_s) \tag{8a}$$

$$\sigma_v = (v_m + a\,v_s)\sqrt{N} \tag{8b}$$

Note that, although both models lead to the same profile, the resulting relationships between number of plates in the column and standard deviation differ. The conventional plate number as defined in chromatography is equal to N for the Martin and Synge model, and to $N_c(1 + k'_0)/k'_0$ for the Craig model. In any discussion of column efficiency it is convenient to consider the *height equivalent to a theoretical plate* (or HETP).

$$H = \frac{L}{N} \tag{9}$$

This concept is developed further in Chapter 10.7.

10.4.2. Statistical Models

The *random walk* model has been used by GIDDINGS to relate various individual contributions in the mechanism of band broadening to the column efficiency [18]. This leads directly to a Gaussian band profile. The contributions of individual sources of band broadening to changes in

the profile are additive. Thus, the column HETP is obtained as the sum of individual contributions

$$H = \sum h_i \tag{10}$$

which can be calculated for each identifiable source of band broadening.

GIDDINGS and EYRING [19] have introduced a more sophisticated, stochastic model that treats the chromatographic process as a Poisson distribution process. Assuming molecules undergo a series of random adsorption and desorption steps during their migration along the column, it is possible to derive a distribution for the residence times of the molecules injected in a pulse. This model has been extended by McQUARRIE to the case of mixed mechanisms [20]. More recent investigations by DONDI et al. confirm that the model can account for both axial dispersion and mass-transfer resistance [21]. When the column efficiency is not too low, the resulting band profile tends toward a Gaussian distribution.

10.4.3. Mass-Balance Models

These models involve writing and integrating a differential mass-balance equation. The basic assumptions associated with writing such an equation are discussed in the literature [16]. There are two important models of this type, the equilibrium–dispersive model [22] and the lumped kinetic models, in addition to the general rate model. The differential mass balance equation is

$$\frac{\partial C}{\partial t} + F\frac{\partial q}{\partial t} + u\frac{\partial C}{\partial z} = D_a\frac{\partial^2 C}{\partial z^2} \tag{11}$$

where C and q are the mobile and stationary phase concentrations, respectively, F is the phase ratio [$F = (1 - \varepsilon_T)/\varepsilon_T$, with ε_T the total packing porosity], u is the mobile-phase velocity, and D_a is the axial dispersion coefficient. The first two terms on the left of this equation are the accumulation terms, the third is the convective term. The term on the right is the dispersive term. This equation contains two functions of the time (t) and space (z) coordinates, C and q. These must be related by a second equation to fully determine the problem.

10.4.3.1. The Equilibrium–Dispersive Model

In this model, one assumes that any departure from equilibrium between the two phases of the

chromatographic system is very small, and the stationary-phase concentration q is given practically everywhere by the equilibrium isotherm:

$$q = f(C) \tag{12}$$

For theoretical discussions, the Langmuir isotherm $q = a\,C/(1 + b\,C)$ is convenient. For a review of isotherms used in liquid chromatography, see [23]. With $D_a = 0$ this assumption would be equivalent to assuming that the column efficiency is infinite. However, the finite efficiency of an actual column can be taken into account by including in the axial dispersion coefficient the influences on band profiles due to both axial dispersion and the kinetics of mass transfer:

$$D_a = \frac{H\,u}{2} = \frac{L\,u}{2N} \tag{13}$$

where H is the column HEPT and L is the column length.

With a linear isotherm, Equation 11 becomes

$$\left(1 + k'_0\right)\frac{\partial C}{\partial t} + u\frac{\partial C}{\partial z} = D_a\frac{\partial^2 C}{\partial z^2} \tag{14}$$

where k'_0 is the retention factor. Solving this equation requires a choice of initial and boundary conditions [22]. In analytical chromatography, the initial condition is an empty column. Three types of boundary conditions have been studied:

1) An infinitely long column stretching from $z = -\infty$ to $z = +\infty$
2) An infinitely long column stretching from $z = 0$ to $z = +\infty$
3) A finite column from $z = 0$ to $z = L$

Depending on the boundary condition chosen, different solutions are obtained [16]. In practice, for values of N exceeding 30–50 plates there are no noticeable differences between the profiles. The first and second moments of these profiles are

1) $\left(1 + \frac{1}{N}\right)t_R$ and $\frac{t_R^2}{N}\left(1 + \frac{2}{N}\right)$
2) $\left(1 + \frac{1}{2N}\right)t_R$ and $\frac{t_R^2}{N}\left(1 + \frac{3}{4N}\right)$
3) t_R and $\frac{t_R^2}{N}\left[1 - \frac{1}{2N}\left(1 - e^{-2N}\right)\right]$

with $t_R = (1 + k'_0)\,t_0$, and $t_0 = L/u$. The differences relative to a Gaussian curve become rapidly negligible as N increases, ultimately reaching the values commonly achieved.

10.4.3.2. Lumped Kinetic Models

These models recognize that equilibrium between the mobile and stationary phases can never actually be achieved, and use a kinetic equation to relate $\partial q/\partial t$, the partial differential of q with respect to time, and the local concentrations, q and C. Several different kinetic models are possible depending on which step is assumed to be rate controlling. All are called "lumped kinetic" models because, for the sake of simplicity, the contributions of all other steps are lumped together with the one considered to be most important. These models are discussed and compared in detail in [24]. All such models are equivalent in linear chromatography, but not when the equilibrium isotherm is nonlinear [16]. The kinetic equation can be written as:

$$\frac{\partial q}{\partial t} = k_m(C_s - q) \tag{15}$$

where C_s is the stationary-phase concentration in equilibrium with the mobile-phase concentration C (and the isotherm is $C_s = a\,C$). A general solution of these models has been given by LAPIDUS and AMUNDSON [25]. Van DEEMTER et al. [26] have shown that the solution can be simplified considerably if one assumes reasonably rapid mass-transfer kinetics. The profile then becomes Gaussian, and the column HETP is given by:

$$H = \frac{2D_L}{u} + 2\left(\frac{k'_0}{1 + k'_0}\right)^2\frac{u}{k'_0 k_m} \tag{16}$$

This is the well-known Van Deemter equation.

10.4.3.3. The Golay Equation

In the particular case of a cylindrical column the wall of which is coated with a layer of stationary phase, the system of partial differential equations of a simplified kinetic model can be solved analytically [27]. For a noncompressible mobile phase, GOLAY derived the following equation:

$$H = \frac{2D_m}{u} + \frac{1 + 6k'_0 + 11k_0'^2}{96\left(1 + k'_0\right)^2}\frac{u\,d_p^2}{D_m} + \frac{k'_0}{24\left(1 + k'_0\right)^2}\frac{u\,d_f^2}{D_l} \tag{17}$$

where d_p is the particle size, d_f is the thickness of the liquid film, D_m is the diffusion coefficient in the carrier gas, and D_l is the diffusion coefficient in the liquid phase.

This equation is in excellent agreement with all experimental results so far reported.

10.4.4. The General Rate Model

In this model, each step of the chromatographic process is analyzed in detail [16]. Separate mass-balance equations are written for the mobile phase that flows through the bed and for the stagnant mobile phase inside the particles. Separate kinetic equations are then written for the kinetics of adsorption/desorption and for mass transfer. Again, if these kinetics are not unduly slow, the band profile tends toward a Gaussian shape (in which case a simpler model is actually more suitable).

10.4.5. Moment Analysis

The *moments* of a statistical distribution provide a convenient description of this distribution. Since a band profile is the distribution of residence times for molecules injected into a chromatographic system, the use of moments to characterize these profiles is a natural approach. Furthermore, analytical solutions of the general rate model can be obtained in the Laplace domain. The inverse transform to the time domain is impossible to solve in closed form, but the moments of the solution can easily be calculated. Equations are available for the first five moments [28], [29].

The nth moment of a distribution is

$$M_n = \int_0^{\infty} C(t) t^n \, dt \tag{18}$$

It is more practical to use normalized distributions, so the absolute moments, μ_n, $n \geq 1$, and the central moments, $\bar{\mu}_n$, $n > 1$ are most com-monly employed:

$$\mu_n = \frac{M_n}{M_0} = \frac{\int_0^{\prime} C(t) t^n \, dt}{\int_0^{\prime} C(t) \, dt} \tag{19a}$$

$$\bar{\mu}_n = \frac{\int_0^{\prime} C(t)(t - \mu_1)^n \, dt}{\int_0^{\prime} C(t) \, dt} \tag{19b}$$

These moments have been used to study band profiles. Interpretations are straightforward for the zeroth (peak area), first (retention of the peak mass center), and second ($H = \bar{\mu}_2 L / \mu_1^2$) moments. Higher moments are related to deviations of the band profile from a Gaussian profile [30]. The use of such moments is limited, however, in large part because the accuracy of determination decreases rapidly with increasing order of a moment. Use of the first moment rather than the retention time of a band maximum has been suggested as a way of relating chromatographic and thermodynamic data for an unsymmetrical peak [30], but this is justified only with slow mass-transfer kinetics.

10.4.6. Sources of Band Asymmetry and Tailing in Linear Chromatography

The theory of chromatography shows that there should be no noticeable difference between a recorded peak shape and a Gaussian profile so long as the number of theoretical plates of the column exceeds 100 [16]. The difference is small even for 25 plates, and careful experiments would be needed to demonstrate that difference. However, it is common to experience in practice peak profiles that are not truly Gaussian.

There are basically four possible causes of unsymmetrical band profiles in chromatography. First, the injection profile may exhibit some tailing. Second, the column may be overloaded such that the isotherm of equilibrium between the two phases is no longer linear. Third, the stationary phase may be inhomogeneous, or mixed retention mechanisms may be operative. Finally, the column packing may not be homogeneous. Slow mass transfer is usually not responsible, contrary to common wisdom, because this would lead to an efficiency much lower than that actually observed.

Injections are rarely carried out under experimental conditions that permit the achievement of narrow, rectangular sample pulses. More often, the injection profile tails markedly. If the length of this tail is significant compared to the standard deviation of the band, the result is tailing of the analytical band. For an exponential injection profile with a time constant τ, the ratio τ/σ should be < 0.10 [31]. Such tailing is usually restricted to high-efficiency bands eluted early, often as a result of poorly designed injection ports or connecting tubes, or an insufficiently high injection-port temperature in gas chromatography. Direct connection of the injection system to the detector

permits assessment of this particular source of tailing [32].

When the sample size is too large, the corresponding isotherm can no longer be considered identical to its initial tangent. If the isotherm is convex upward, the retention time decreases with increasing concentration, and the peak tails. Isotherms in gas chromatography are usually convex downward, and peaks from large samples tend to "front." However, peaks eluted early tend to have very sharp fronts and some tail. This is caused by a sorption effect, since the partial molar volume of the vapor is much larger than that of the dissolved solute [33].

In most cases, peak tailing can be explained by heterogeneity of the surface of the stationary phase [16]. Strongly adsorbing sites saturate before weak ones, and they have a much smaller saturation capacity. The efficiency increases when the sample size is reduced, but this observation is often of limited use, because detection limits establish a minimum for the acceptable sample size.

10.5. Flow Rate of the Mobile Phase

The velocity of the mobile phase is related to the pressure gradient, the mobile phase viscosity η, and the column permeability k_0 by the Darcy law [34]:

$$u = -\frac{k_0}{\eta}\frac{dp}{dz} \tag{20}$$

The minus sign here indicates that the flow is directed from high to low pressure. The differential equation can be integrated provided the relationship is known between local volume and pressure. Two equations of state can be used, one for gases (Boyle–Mariotte law), the other for liquids. For dense gases, the equation of state is far more complex [35].

Liquids can be considered as noncompressible for purposes of calculating a pressure gradient. In this case u is constant along the column, and the difference between the inlet and outlet pressures is

$$\Delta p = \frac{u \eta L}{k_0} \tag{21}$$

The retention time for an inert tracer is

$$t_0 = \frac{L}{u} \tag{22}$$

In gas chromatography the carrier gas behaves as an ideal gas from the standpoint of compressibility, so $p\,u = p_0\,u_0$, and the integration of Equation 20 gives

$$u_0 = \frac{k_0 p_0}{2\eta L}\,(p^2 - 1) \tag{23}$$

Because of gas-phase compressibility, the average gas velocity $\bar{u}$ is lower than the outlet velocity. It can be calculated that

$$\bar{u} = j\,u_0 = \frac{3\,(p^2 - 1)}{2\,(p^3 - 1)}\,u_0 \tag{24}$$

where j is the James and Martin compressibility factor [11]. The retention time for an inert tracer in this case is

$$t_0 = \frac{4\,\eta L^2\,(p_i^3 - p_0^3)}{3 k_0\,(p_i^2 - p_0^2)} \tag{25}$$

10.5.1. Permeability and Porosity of the Packing

10.5.1.1. Column Porosity

There are three different porosities that must be considered: The *interparticle* or *external* porosity, the *intra particle* or *internal* porosity, and the *total* porosity. The external porosity is the fraction of the overall geometrical volume of the column available to mobile phase that flows between the particles. Chromatographers always strive to achieve a dense column packing as the only known practical way of packing reproducible, stable, and efficient columns, so this porosity tends to be similar for all columns, with a value close to 0.41.

The internal porosity depends on the nature and structure of the particles used to pack the column. It varies considerably from one stationary phase to another, ranging from zero (nonporous beads) to ca. 60 % for certain porous silicas.

The total porosity is the fraction of the bed accessible to the mobile phase; i.e., it is the sum of the internal and external porosities.

10.5.1.2. Column Permeability

The *permeability* of a column depends on the characteristics of the particular packing material used, and especially on the column external po-

rosity, but it is the same regardless of whether the mobile phase is a gas, a liquid, or a dense gas. This permeability is inversely proportional to the square of the particle diameter:

$$k_0 = \frac{d_p^2}{\Phi} \tag{26}$$

For packed columns the numerical coefficient Φ is of the order of 1000, with reported values ranging between 500 and 1200. It tends to be smaller for spherical than for irregular particles. For empty cylindrical tubes, such as those used for open tubular columns, it is equal to 32.

10.5.2. Viscosity of the Mobile Phase

The viscous properties of gases and liquids differ significantly [36].

10.5.2.1. Viscosity of Gases

The viscosity of a gas increases slowly with increasing temperature, and is nearly independent of the pressure below 2 MPa. It is approximately proportional to $T^{5/6}$. The rare gases (He, Ar) have higher viscosities than diatomic gases (N_2, H_2) and polyatomic molecules. Hydrogen has the lowest viscosity of all possible carrier gases, and is often selected for this reason. The viscosity of the sample vapor is unlikely to exceed that of the carrier gas.

10.5.2.2. Viscosity of Liquids

The viscosity of a liquid decreases with increasing temperature, following an Antoine-type equation:

$$\log \eta_{liq} = A + \frac{B}{T + C} \tag{27}$$

In most cases, however, C is negligible, and if T is expressed in kelvins B is of the order of 1 to 3×10^{-3}. For example, the viscosity of benzene decreases by one-half with an increase in temperature of 120 °C. The viscosity increases slowly and linearly with increasing pressure below 200 MPa, and may increase by $10-25\%$ when the pressure rises from 0 to 20 MPa. The viscosity of a liquid also increases with the complexity of the molecule and its ability to engage in hydrogen bonding. The viscosity of a sample may be much higher than that of the corresponding mobile phase, which

presents a serious potential problem in preparative chromatography [37].

10.6. The Thermodynamics of Phase Equilibria and Retention

Chromatography separates substances on the basis of their equilibrium distributions between the two phases constituting the system. Without retention there can be no separation. Retention data are directly related to the nature of the phase equilibrium and to its isotherm. Although the details depend very much on the nature of the mobile and stationary phases used, a few generalizations are possible.

10.6.1. Retention Data

It is easy to relate retention times to the thermodynamics of phase equilibria. In practice, however, it is difficult to measure many of the relevant thermodynamic parameters with sufficient accuracy, while retention times can be measured with great precision. Furthermore, it should be emphasized at the outset that even though very precise measurements can indeed be made in chromatography [38], the available stationary phases are often so poorly reproducible that such precision is unwarranted. Column-to-column reproducibility for silica-based phases is reasonable only for columns packed with material from a given supplier. For this reason, chromatographers prefer whenever possible to use relative retention data.

10.6.1.1. Absolute Data

The experimental data acquired directly in chromatography are retention time, column temperature, and the mobile-phase flow rate. All other data are derived from these primary parameters. The retention time is inversely proportional to the flow rate, so it is often preferable instead to report the retention volume and the retention factor (both of which are independent of the flow rate).

The *retention factor* k' is the ratio of the fraction of component molecules in the stationary phase n_s to the fraction in the mobile phase n_m at equilibrium. The retention factor is thus

$$k' = \frac{n_s}{n_m} = \frac{1 - R}{R} \tag{28}$$

This also corresponds to the ratio of the average times spent by a typical molecule in the two phases:

$$k' = \frac{t_R - t_0}{t_0} \tag{29}$$

Accordingly, k' is proportional to the thermodynamic constant for the phase equilibrium at infinite dilution.

The *retention volume* V_R (see Eq. 4) is the volume of mobile phase required to elute a component peak. For compressible phases it must be referred to a standard pressure:

$$V_R = t_R F_v f(P) \tag{30}$$

where F_v is the mobile-phase flow rate and $f(P)$ is derived from the equation of state for the mobile phase. For the ideal gases used in GC,

$$f(P) = j = \frac{3(P^2 - 1)}{2(P^3 - 1)}$$

(Eq. 24); for liquids, $f(P) = 1$; for dense gases it is a complex function [35].

10.6.1.2. Relative Data

Because absolute retention data depend on so many controlled (temperature, pressure) and uncontrollable (activity of the stationary phase) factors, and because their quantification requires independent determinations that are time-consuming (e.g., density of the stationary liquid phase in the case of GC) or not very accurate (flow-rate measurements), and also because chromatographers are more often concerned with practical separation problems than with phase thermodynamics, it is more convenient to work with relative retention data. The parameter of choice is the *relative retention* α, which is also the *separation factor*:

$$\alpha_{2,1} = \frac{t_{R,2} - t_0}{t_{R,1} - t_0} = \frac{k_2'}{k_1'} \tag{31}$$

Relative retention data are more convenient than absolute data, but the choice of an appropriate reference may prove difficult. An ideal reference compound

1) Should not be present in mixtures subject to routine analysis
2) Should be introduced into the sample as a standard (though it might also be used as an external standard in quantitative analysis)
3) Should be nontoxic and inexpensive
4) Should elute near the middle of the chromatogram so the relative retention data will be neither too large nor too small
5) Should have chemical properties similar to those of most of the components in the sample mixture, thereby ensuring that relative retention volumes will be insensitive to changes in the level of activation of the stationary phase

For complex mixtures, especially those analyzed with temperature programming (in GC) or gradient elution (in LC), it is impossible to select a single reference compound. The Kovats retention-index system [39], based on homologous series of reference compounds, provides an elegant solution in GC, one which has been widely accepted. In LC, the absence of suitable homologous series, and the fact that retention depends more than in the case of GC on the polarities of compounds and less on molecular weight makes the use of an index system impractical.

Finally, because of its definition, the separation factor for two compounds is related to the difference between the corresponding Gibbs free energies of transfer from one phase to the other:

$$RT \ln \alpha = \Delta(\Delta G^0) \tag{32}$$

When α is close to unity [40],

$$\alpha - 1 = \frac{\Delta(\Delta G^0)}{RT}.$$

10.6.2. Partition Coefficients and Isotherms

From Equation 29 it can be shown that for gas–liquid chromatography:

$$k' = \frac{RT \rho}{\gamma^x P^0 M_s} \frac{V_L}{V_G} = K \frac{V_L}{V_G} \tag{33}$$

where P^0 is the vapor pressure of the analyte, γ^x is the activity coefficient at infinite dilution, ρ is the density, and K is the thermodynamic equilibrium constant, or the initial slope of the isotherm. Alternately, the constant K can be derived from

measurement of the specific retention volume. The retention volume, $V_R = F_v t_R$, is the volume of carrier gas required to elute a particular component peak. The *corrected* retention volume V_N is the retention volume corrected for gas compressibility (see Eq. 24) and for the hold-up time in the mobile phase:

$$V_N = j(t_R - t_0)F_v = j k' t_0 F_v \qquad (34)$$

The *specific* retention volume is the corrected retention volume under STP conditions, reported for a unit mass of liquid phase:

$$V_g = \frac{V_N}{m_L}\frac{273}{T_c}\frac{p_0}{P_n} = \frac{K}{\varrho}\frac{273}{T_c} \qquad (35)$$

where T_c is the column temperature. In practice, derivation of the constant K from k', V_L, and V_G or from V_g requires the same measurements, and the accuracy and precision are the same.

Since V_g is proportional to a thermodynamic constant of equilibrium, it varies with the temperature according to

$$\frac{d(\ln V_g)}{d(1/T)} = -\frac{\Delta H_S}{R} \qquad (36)$$

Since the molar enthalpy of vaporization of solute ΔH_S from an infinitely dilute solution is negative, retention volumes decrease with increasing temperature.

In liquid–liquid chromatography the retention volume depends on the ratio of activities of the component in the two phases at equilibrium; i.e., on the ratio of the activity coefficients at infinite dilution. As liquid–liquid chromatography has been little used so far (because of instability of the chromatographic system), and activity coefficients are not easily accessible, this relationship has not been extensively explored.

10.6.3. Gas – Solid Adsorption Equilibria

Gas – solid chromatograhy is used mainly for the separation of gases. In spite of major advantages, such as a high degree of selectivity for geometrical isomers and the high thermal stability of many adsorbents, it is not used for the analysis of organic mixtures.

Two conditions are required for successful analysis by gas – solid chromatography: The isotherm must be close to its initial tangent so that peaks migrate under conditions of linear equilib-

rium, and the surface of the adsorbent must be highly homogeneous. Gases are analyzed at temperatures much above their boiling points, so their partial pressures relative to their vapor pressures are very low, and equilibrium isotherms deviate little from their initial tangents. This is impractical with heavier organic vapors that decompose thermally at moderate temperatures. Graphitized carbon black, with a highly homogeneous surface, has been used successfully for the separation of many terpenes and sesquiterpenes, especially closely related geometrical isomers [41]. A number of adsorbents developed recently for reversed-phase LC have never been tried in gas chromatography despite their apparent potential advantages.

At low partial pressure (more precisely: low values of the ratio p/P^0 of the partial pressure of the analyte to its vapor pressure) the amount of a substance adsorbed on a surface at equilibrium with the vapor is proportional to the partial pressure:

$$m = K_H p \qquad (37)$$

where the constant K_H is the Henry constant [41]. Retention data are sometimes reported in terms of mass of adsorbent in the column, analogous to the specific retention volume, although the adsorbent mass itself is not actually related to the adsorption equilibrium. Retention data should rather be reported as a function of area S_a of adsorbent:

$$K_H = \frac{V_R}{m_a S_a} \qquad (38)$$

However, the determination of S_a is complex and relatively inaccurate. The Henry constant of adsorption depends on the geometrical structure of the adsorbent as well as on the interaction energy between the surface and chemical groups in the adsorbate molecule. The Henry constant of adsorption decreases rapidly with increasing temperature. Equation 36 applies in gas – solid chromatography as well.

10.6.4. Liquid – Solid Adsorption Equilibria

Liquid – solid chromatography is the most widely applied method of liquid chromatography. Other modes such as hydrophobic-interaction

chromatography are in fact variants of liquid–solid chromatography based on adsorption from complex solutions. The two most important methods in liquid–solid chromatography are *normal phase*, which uses a polar adsorbent (e.g., silica) and a low-polarity mobile phase (e.g., methylene chloride), and *reversed phase*, which uses a nonpolar adsorbent (e.g., chemically bonded C_{18} silica) and a polar eluent (e.g., methanol – water mixtures).

10.6.4.1. Normal-Phase Chromatography

Retention volumes in normal-phase chromatography depend much on the polarity and polarizability of the analyte molecules, and little on their molecular masses [42]. Thus, all alkyl benzenes, or all alkyl esters, will be eluted in a narrow range of retention volumes. On the other hand, structural changes affecting the polarity of a molecule are easily seen. The amount of water adsorbed by the stationary phase controls its activity, hence the resolution between analytes, and this is in turn a sensitive function of water dissolved in the mobile phase. Thus, control of the water content of the eluent is a primary concern. The influence of the composition of the mobile phase on retention has been discussed by SNYDER, who introduced the concept of *eluotropic strength* [42].

Strong solvents (i.e., those that are retained in a pure weak solvent) compete with components of the sample for adsorption. For this reason, a pulse of a weak solvent generates as many peaks as there are components in the mobile phase [43], [44]. These peaks are called *system peaks* [44]. This property has been used for detecting compounds that themselves produce no response from the detector. Thus, if a compound that does produce a response with the detector is introduced at a constant concentration in the mobile phase, injection of a sample is equivalent to injection of a negative amount of the background compound, leading to a negative peak in the chromatogram.

10.6.4.2. Reversed-Phase Chromatography

In reversed-phase chromatography the retention volume depends very much on the molecular mass of the analyte. Methylene selectivity (i.e., the relative retention of two successive homologues) varies over a much wider range in reversed-phase LC than in GC, and can be quite large [45].

Water is the weakest of all solvents in reversed-phase chromatography. The mechanism by which a strong solvent reduces the retention volume of an analyte is quite different in reversed-phase chromatography from that in normal-phase mode. In reversed-phase mode there is no competition. Methanol, acetonitrile, and other organic solvents miscible in water are poorly retained on reversed-phase columns eluted with pure water. Addition of organic solvent instead increases analyte solubility in the mobile phase, thereby decreasing retention. The phenomenon of system peaks is nearly negligible.

10.7. Band Broadening

As discussed previously (Chap. 10.4), the band profile in linear chromatography is Gaussian. The relative standard deviation of this Gaussian profile is used as a way of characterizing column efficiency. By definition

$$N = \frac{t_R}{\sigma_t} = 2k^2 L n 2 \frac{t_R}{w_{1/k}} = 5.64 \frac{t_R}{w_{1/2}} \tag{39}$$

where $w_{1/k}$ is the peak width at the fraction $1/k$ of its height, and L is the length of the column. The experimental height equivalent to a theoretical plate is

$$\bar{H} = \frac{L}{N} \tag{40}$$

where N and $\bar{H}$ depend on the experimental conditions. Experimental conditions must be so established that N is at a maximum and H at a minimum to ensure that the peaks are as narrow as possible, thereby maximizing the resolution between successive bands.

10.7.1. Sources of Band Broadening

Various phenomena contribute to the broadening of peaks during their migration along a column [18]. The most important are:

1) Axial diffusion along the concentration gradient of the band profile
2) Dispersion due to inhomogeneous flow distribution between the packing particles
3) Resistance to mass transfer back and forth from the mobile phase stream passing between the particles to the stagnant mobile phase inside the particles

4) Resistance to mass transfer through pores inside the particles

5) The kinetics of the retention mechanism (e.g., of adsorption – desorption)

Calculations of the contributions from these several phenomena have been accomplished by integration of the mass-balance equation (e.g., Eq. 17), or by a statistical approach.

The simplest approach involves the random-walk model, derived first for the study of Brownian motion. The trajectory of a given molecule through the column can be represented as a succession of steps in which the molecule either moves forward at the mobile-phase velocity (u) or stays in the stationary phase, where it is immobile (i.e., it moves backward with respect to the mass-center of the peak, which moves at the velocity $R u$). If one can identify the nature of the steps made by the molecule, and then calculate their average length l and number n, it is possible to derive the contribution to the variance from this particular source of band broadening; the result is in fact $n l^2$ [18].

10.7.2. The Plate-Height Equation

Because the plate number is related to particle size, and contributions from axial dispersion and resistance to mass transfer in the mobile phase are related to both the velocity of this phase and the molecular diffusion coefficient, it is convenient to express plate height on the basis of reduced, dimensionless parameters, in the form of a *reduced plate height*, $h = H/d_p$ together with a *reduced velocity*, $v = u \, d_p/D_m$ [46].

Contributions from all the various sources of band broadening listed in the previous section have been calculated [18]. If we assume that these sources are independent, the band variance is the sum of each contribution to the variance, and the column HETP is given by the equation [46]

$$h = \frac{B}{v} + A v^{1/3} + C v \tag{41}$$

where the coefficients A, B, and C are dimensionless functions of the experimental conditions, but velocity independent. The use of reduced plate height and reduced velocity has become standard practice in LC, but not in GC, probably because the concept arose only in the mid 1970s — although this is not a valid reason for ignoring it.

Depending on the type of chromatography under consideration, the plate-height coefficients are obtained from different relationships. The Golay equation [27] is typical. This describes the plate height for an open tubular column. It is characterized by $A = 0$ and C a simple function of k'. In a packed GC column, anastomosis of the flow channels causes the axial dispersion to be more complex, and results in the $A v^{1/3}$ contribution in Equation 41 as well as a more complicated equation for the C coefficient than in the Golay equation. It is striking, however, that GC produces $h = f\,(v)$ (i.e., HETP versus velocity) curves that are quite similar to those from LC, with minimum values of h between 2 and 3, and corresponding values of v between 3 and 5. These values are excellent approximations for order-of-magnitude calculations. The only marked difference between experimental results reflects the distinction between open-tubular and packed columns. For the former, $A = 0$, and the curve is a hyperbola. For the latter, the curve has an inflection point; a tangent drawn from the origin has its contact point at a finite velocity.

Identifying those sources of band broadening that contribute significantly to increased peak width, and then calculating their contributions, makes it possible to derive the value of the local plate height; i.e., the coefficient relating an incremental increase in band variance to an incremental distance of migration [12]:

$$d(\sigma^2) = H \, dz \tag{42}$$

The *apparent plate height*, $\bar{H}$, given by Equation 40, is the column length average of the local plate heights

$$\bar{H} = \frac{\displaystyle\int_0^L H \, dz}{L} \tag{43}$$

This equation can be used to calculate apparent HETPs in both GC and TLC, which are types of chromatography in which the mobile-phase velocity varies quite markedly during the course of peak migration [47].

10.7.3. Resolution Between Peaks

The *resolution* (R) between two successive Gaussian bands is defined as

$$R = 2\frac{t_{R,2} - t_{R,1}}{w_1 + w_2} \tag{44}$$

Assuming the two bands are close (since otherwise the concept of resolution loses most of its relevance), the relationship $t_{R,2} + t_{R,1} \cong 2\, t_{R,2}$ represents a good approximation, and

$$R = \frac{\sqrt{N}}{4}\frac{\alpha - 1}{\alpha}\frac{k'}{1 + k'} \tag{45}$$

This equation [48] shows, in the first place, that there is no separation if there is no retention. Separations become very difficult when k' tends toward 0 even if α is large. A case in point is the separation of nitrogen from argon by gas chromatography on molecular sieves (0.5 nm) at room temperature.

Furthermore, solving Equation (45) for N gives the column efficiency required for separating two compounds with a given degree of resolution. This shows that the required column length increases rapidly as α approaches 1. With a 100-plate column it is possible to separate with unit resolution two compounds with $\alpha = 2$ and $k' = 3$; a 1 000 000-plate column is required to separate two compounds with $\alpha = 1.005$.

10.7.4. Optimization of Experimental Conditions

It is quite easy to find the conditions corresponding to maximum column efficiency. The column must be operated at the velocity producing the minimum plate height, and if the efficiency is still insufficient, a longer column must be prepared. In practice, certain constraints are applicable, however. First, there is a maximum pressure at which any column can be operated, a pressure that depends on the equipment available. Furthermore, analysts are often more interested in performing separations rapidly than in achieving the highest possible resolution. It should be emphasized that the less time a band spends in a column, the narrower it will be, hence the higher its maximum concentration and the lower the detection limit. Optimization for minimum analysis time becomes an important consideration. It is easily shown that minimum analysis time is achieved by operating a column at the reduced velocity for which the ratio h/v is a minimum, with a particle size such that this velocity is achieved at the maximum acceptable column inlet pressure [49]–[51].

A considerable amount of work has been published on optimizing the experimental conditions for minimum analysis time under various constraints [52]. One complication arises from the definition of reduced mobile-phase velocity. The actual mobile-phase velocity depends largely on the molecular-diffusion coefficient of the analyte. Thus, very small particles can be used for the analysis of high molecular mass compounds, which have low D_m values. The actual flow rate required will remain compatible with pressure constraints despite the resulting high pneumatic or hydraulic resistance. Detailed results obviously depend greatly on the mode of chromatography used.

10.7.5. Instrumental Requirements

A column alone cannot give the information the analyst requires. A complete instrument is needed, one with a mobile-phase delivery system, an injection device, a detector, and a data station. At all stages the requirements and properties of the modules force the analyst to accept compromises and lose resolution. Contributions of the injector and detector to band broadening should be of special concern. The lucid analysis of these problems with respect to GC by STERNBERG [53] has become a classic. Most of the conclusions are easily extended to LC.

Band broadening in the equipment results from axial dispersion in valves and connecting tubes, as well as from time delays associated with injection of the sample and detection of the separated peaks. The contributions to band broadening from these sources can be expressed as additional terms in the equation for apparent column HETP:

$$H_c = \frac{\sum \tau_j^2 u^2}{(1 + k')^2 L} \tag{46}$$

where τ^2 is the time variance caused by any of the band-broadening sources indicated above.

10.7.5.1. Injection Systems

There is a dearth of data regarding actual injection profiles for every type of column chromatography. Dependence of the instrument variance on the mobile-phase flow velocity has been studied in liquid chromatography [32]. Probably in part because of the onset of eddy turbulence at various

flow rates in different parts of the instrument, this variation is not always simple.

If accurate efficiency data are required, the variance of the injection profile should be measured as a function of the flow rate, with a correction applied by subtracting the first moment of the injection profile from the peak retention time and the variance of the injection profile from the band variance. In actual analytical practice it is often sufficient to minimize this contribution by making sure that the volume of the injection device is much smaller than the volume of the column, that the device is properly swept by the mobile phase, and that actual injection is rapid. The use of a bypass at the column inlet permits a higher flow rate through the injection device than through the column, effectively reducing the band-broadening contribution of the injection system.

10.7.5.2. Connecting Tubes

STERNBERG [53] and GOLAY and ATWOOD [32], [54] have studied this particular contribution, showing that for short tubes it is smaller than predicted by application of the Golay equation (Eq. 17) to an empty tube. On the other hand, dead volumes, spaces unswept by the mobile-phase stream but to which the analytes have access by diffusion, act as mixing chambers, and they may have a highly detrimental effect on efficiency. The relative importance of the problem increases with decreasing column diameter.

10.7.5.3. Detectors

The detector is a concentration sensor appropriate to the analysis in question. As in the case of all sensors, detectors do not react immediately, but instead have a time constant. Thus, they contribute to band broadening through both their cell volume and their response time. The cell-volume contribution can be handled in the same way as the volume of the injection device. The response-time contribution is given by Equation 46, and it is proportional to the square of the mobile-phase velocity. The response time should be less than one-tenth of the standard deviation of a Gaussian peak in order to have a negligible effect on the retention time and profile of the peak [31].

10.7.5.4. Instrument Specifications

If a fractional loss of efficiency equal to Φ is acceptable, it follows that:

$$\left(\frac{\sigma_{eq}}{\sigma_{col}}\right)^2 \leq \Phi \tag{47}$$

where σ_{eq}^2 and σ_{col}^2 are the variance contributions of the equipment and the column, respectively. Application of this relationship permits easy derivation of equipment specifications as a function of the particular system used and the analysis of interest [52].

10.8. Qualitative Analysis

Although chromatography is a most powerful method of separation, it affords few clues regarding the identity of the separated eluates. Only a retention time, volume, or factor is available for each peak, and the method cannot even establish whether a given peak corresponds to a pure compound or to a mixture. Supplying information that is both meager and not very accurate, chromatography alone is poorly suited to the identification of single components, although previously identified patterns simplify the matter (e.g., alkane series in petroleum derivatives, or phytane – pristane in crude oils). Innumerable attempts at processing retention data for identification purposes have been reported, including the use of the most refined chemometric methods, but success has been limited. In practice, a combination of chromatographic separation and spectrometric identification has become the preferred method of identification for unknown chemicals.

10.8.1. Comparisons of Retention Data

The most efficient identification procedure involves the determination in rapid succession of retention data (retention volumes, or preferably retention factors) for the components of a mixture and a series of authentic compounds. This method permits straightforward identification of components known to be present in the mixture, and it can also rule out the presence of other possible components. However, it cannot positively identify the presence of a given, isolated component. It is recommended that such determinations be carried out using similar amounts of compounds, and that a number of spike analyses be performed, where small amounts of authentic compounds are added to the mixture in order to approximately double the sizes of unknown peaks. Comparison of

the efficiencies of an unknown peak in the original sample and in the spiked mixtures increases the probability of recognizing differences between the analyte and an authentic compound.

GIDDINGS and DAVIS [55] have shown that, for complex mixtures, analyte retention data tend to reflect a Poisson distribution. This result permits a simple calculation of the probability of finding a certain number of singlets, doublets, and higher-order multiplets in the course of analyzing a mixture of m components on a column with a peak capacity n. If the ratio m/n is not small, the probability is low. This confirms that there is little chance of separating a complex mixture on the first phase selected [56]. Method development is a long and onerous process because the probability of random success is low.

In practice, qualitative analysis is made easier by the use of retention indices I [39]:

$$I = 100z + \frac{\log\left(t_{R,X} - t_0\right) - \log\left(t_{R,P_z} - t_0\right)}{\log\left(t_{R,P_z+1} - t_0\right)} \qquad (48)$$

where $t_{R,X}$ is the retention time of component X, and P_z is the n-alkane containing z carbon atoms. The retention index varies little with temperature, usually following an Antoine-type equation. The retention index increment

$$\Delta I = I^P - I^A \qquad (49)$$

is the difference between the retention indices for a compound on two stationary phases: A polar phase, P, and a nonpolar phase, A. Squalane, Apiezon L (a heavy hydrocarbon grease), and various polymethyl siloxanes have been used as nonpolar phases. KOVATS and WEHRLI [57] have shown that the retention index increment for a compound can be related to its structure.

10.8.2. Precision and Accuracy in the Measurement of Retention Data

The *precision* of retention data can be quite high [38], [58]. This is related to the precision with which the peak maximum can be located; i.e., it depends on column efficiency and on the signal-to-noise ratio. *Reproducibility* depends on the stability of the experimental parameters (i.e., the mobile-phase flow rate and the temperature). Reproducibility also depends on the stability of the stationary phase, which is, unfortunately, much less satisfactory than often assumed. In gas–liquid

chromatography, the stationary liquid phase tends to oxidize, decompose, and become more polar unless extreme precautions are taken to avoid oxygen, which diffuses into the carrier gas through septa and flow-rate controller membranes. In liquid–solid chromatography, impurities in the mobile phase (e.g., water in the case of normal phase LC) are adsorbed, and these slowly modify the activity of the stationary phase; alternatively, the aqueous mobile phase may slowly react with a chemically bonded phase and hydrolyze it.

The *accuracy* of retention data is still more questionable than their reproducibility. Poor batch-to-batch reproducibility for silica-based phases has plagued LC for a generation. For this reason, retention data cannot be compiled in the same way as spectra. Such data collections have limited usefulness, providing orders of magnitude for relative retention rather than accurate information on which an analyst can rely for the identification of unknowns. Precision and accuracy with respect to absolute and relative retention data have been discussed in great detail [58].

10.8.3. Retention and Chemical Properties

Considerable effort has been invested in elucidating the relationships between chemical structure and retention data in all forms of chromatography. A wide compendium of results has been published in the literature. These studies are much more detailed for GC than LC, probably because work was begun earlier in the former case, before the sterility of the approach for purposes of qualitative analysis was recognized.

A plot of the logarithm of the retention volumes for homologous compounds versus the number of carbon atoms in their chains produces a nearly straight line. Exceptions are found for the first few members of a series and in other rare cases. Over a wide range of carbon-atom numbers, it also appears that such a plot exhibits definite curvature [59]. This result has been observed in all modes of chromatography capable of separating homologues. It should be noted, however, that the approximate linearity of the plots depends on a proper estimate of the hold-up volume, which is not readily defined in liquid chromatography [60]. Lines observed for different functional groups are parallel, or nearly so.

More generally, the Gibbs free energy of retention can be written as a sum of group

contributions [61], and increments can be ascribed to different groups or substituents. This permits the approximate calculation of retention data from tables of increments. Unfortunately, multifunctional compounds are subject to interactions between the respective groups, and interaction coefficients must be introduced, so the system rapidly becomes too complex for practical application.

In GC, the retention volumes of nonpolar or slightly polar compounds obey the following correlation [62]:

$$\log V_g = \log \frac{273R}{M_S} + 0.22k\left(\frac{T_{Eb}}{T} - 1\right) - 2.9 - \log \gamma^\lambda$$

(50)

where M_S is the molecular mass of the stationary phase, k is Trouton's constant (ca. 88 J mol^{-1} K^{-1}), and T_{Eb} is the boiling point of the compound. More detailed relationships linking retention-index increments in gas chromatography with the structures of the corresponding components have been suggested [57]. These correlations apply well within groups of related, monofunctional compounds, but difficulties arise in the attempt to extend the relations to multifunctional compounds. However, the identification of complex compounds is the only real challenge remaining in this area. Applying such relationships to a series of standards has facilitated the classification of gas-chromatographic stationary phases [63].

10.8.4. Selective Detectors

The use of selective detectors is the most promising route to the identification of unknown components in a complex mixture. Chromatography can separate complex mixtures into individual components, but it cannot positively identify them. Most spectrometric methods, on the other hand, supply sufficient information to identify complex compounds, but only if they are so pure that all the spectral clues can be ascribed safely to the same compound. Among the possible combinations, coupling with mass spectrometry, and to a lesser degree IR spectrometry, have been the most fruitful, because the amount of sample easily handled in classical chromatography is well suited to these techniques. Less sophisticated detectors have also been used to advantage.

10.8.4.1. Simple Selective Detectors

Paradoxically, the gas-density detector, which can supply the molecular mass of an unknown, can be considered to be a selective detector in gas chromatography [58]. This comment effectively illustrates the dearth of convenient selective detectors. At best such detectors help to identify compounds that contain halogen atoms (electron-capture or thermoionic detectors), sulfur, nitrogen, or phosphorus (flame-photometric or thermoionic detectors). Physiological detectors have also been used in certain rare cases (insects that react to sexual pheromones, for example, or the chemist's nose, a dangerous and hazardous application).

The choice of selective detectors in liquid chromatography is still more limited. Electrochemical detectors and the UV-spectrophotometric detector can provide clues, but many types of compounds resist oxidation or reduction in an electrochemical cell, and the paucity of clues provided by a UV spectrum has been a source of disappointment to many chromatographers.

Selective detectors are much more useful for the selective quantitation of a small number of chemicals known to be present in a complex matrix than they are for identification purposes. The use of on-line chemical reactions can be very selective for certain chemical families (e.g., amino acids, sugars). This technique has become widespread in column chromatography, both gas and liquid, and is even the principle underlying many detection schemes in thin layer chromatography [64]. The identification of an unknown presupposes methods supplying far more information and consistent with the ready association of elements of this information with structural details of the corresponding compound.

10.8.4.2. Chromatography – Mass Spectrometry (→ Mass Spectrometry)

The coupling of gas chromatography and mass spectrometry (GC/MS) was developed during the 1970s, and it has become the generally accepted method for identifying the components of complex mixtures. Its use has considerably influenced the development of modern environmental, clinical, pharmacological, and toxicological analysis. Preferably implemented with open tubular columns and associated with selective derivatization of the sample components, gas chromatography permits the separation and elution of all compounds

that have significant vapor pressures at 400 °C and do not decompose, or have derivatives fulfilling these conditions. Study of the resulting mass spectra in conjunction with a rapid computer search in a large spectrum library permits the identification of separated compounds. Quantitation can be accomplished with the aid of the total ion current, currents for selected masses, or the signal from an associated chromatographic detector.

The coupling of liquid chromatography with mass spectrometry (LC/MS) has also been intensively investigated since the early 1980s, but so far without complete success. The new ionization techniques, especially ion spray, made the coupling of HPLC with MS to a routine technique. The decreasing prices(e.g. of ion trap MS instruments open the use of these instruments in routine HPLC analysis. For more details see → Mass Spectrometry. In GC/MS it is relatively easy to remove the mobile phase, or to limit and control its effects on the ionization source, but this has proven to be much more difficult in LC/MS. A variety of interfaces has been developed for that purpose:

1) A direct liquid interface, where the mobile phase is nebulized into the source, vaporized, and condensed onto a cold surface
2) Thermo-spray, in which the mobile phase is rapidly heated and vaporized into a spray of fast-moving droplets passing the mass spectrometer inlet, which allows ions to enter the analyzer, but not the unionized solvent
3) Various spraying methods
4) Fast-atom bombardment of liquid flowing out of the column
5) Electro-spray

The variety of these methods, as well as the fact that they produce widely differing spectra and that, for reasons still unclear, one method often appears better suited to a particular group of chemicals, demonstrates that the method is still not a fully developed analytical technique. It nevertheless holds great promise, and has already proven quite useful despite difficulties more serious than those encountered in GC/MS.

10.8.4.3. Chromatography – IR Spectrometry
(→ Infrared and Raman Spectroscopy)

Infrared (IR) spectrometry is less sensitive than MS, and it therefore requires cells with a long optical path length, but IR spectrometry has also proven quite useful as an identification method in

chromatography. Fourier-transform IR (FTIR) spectrometry has become the method of choice because of its sensitivity and the possibility of subtracting most of the contributions due to solvent (in liquid chromatography). It is possible to obtain useful spectra with as little as a few nanograms of a compound, provided it has been chromatographically separated from interfering components. The only remaining source of technical difficulty in coupling a chromatographic separation with an FTIR spectrometer is the size of the sample cell, which is generally excessively large relative to the volume of a band. As a consequence, the resolution between separated components must be greater than in the case of coupling to a mass spectrometer. Columns wider than the conventional 4.6 mm i.d., packed columns could be used with advantage.

10.9. Quantitative Analysis

Once a mixture has been separated into its components it becomes possible to determine the amount of each [58]. This requires a determination of the size of the concentration profile. Two possible methods are available: Use of either the *peak height* or the *peak area*. The former is more sensitive in the case of GC to fluctuations in details of the injection procedure and the column temperature, and in the mobile-phase composition and column temperature in LC. Peak areas are more sensitive to flow-rate fluctuations in both GC and LC, at least so long as the detector is concentration sensitive. In practice, however, only the peak-area method has been implemented in data packages currently available for chromatographic computers.

With respect to the most frequent case of a concentration sensitive detector, the mass of a particular compound (i) eluted from a chromatographic column is given by [58]

$$m_i = \int 0 C_i \, dV \simeq \int_{t_i}^{t_f} y_i S_i F_v \, dt \simeq F_v S_i \int_{t_i}^{t_f} y_i \, dt \qquad (51a)$$

where t_i and t_f are the initial and final times between which integration of the detector profile, $y(t)$, is calculated, and S_i is the appropriate response factor. Thus, the peak area $\int_{t_i}^{t_f} y_i \, dt$ is inversely proportional to the mobile-phase flow rate. For a mass-flow sensitive detector (e.g., mass

spectrometer, flame-ionization detector), the result is simpler:

$$m_i = \int\limits_0^x \frac{dm_i}{dt}\,dt \simeq \int\limits_{t_1}^{t_1} dm_i = S_i \int\limits_{t_1}^{t_1} v_i\,dt \qquad (51b)$$

and the area is independent of the flow rate.

The main problems encountered in quantitative analysis are associated with sampling and sample introduction, and with measurement of the peak area and the calibration factors [58]. Chromatographic measurements are usually carried out with a precision of a few percent, but can be much more precise when the sources of error are properly recognized and controlled [65].

10.9.1. Sampling Problems in Chromatography

As in all analytical techniques, the accuracy of a chromatographic analysis depends on the quality of the sample studied. The techniques involved in ensuring that a representative sample is selected from the system under investigation are discussed elsewhere (→ Sampling). The sample is usually introduced into the mobile-phase stream ahead of the column through a sampling valve, or with a syringe into an injection port. Calibration of the volume of sample injected is difficult, so most analytical work is accomplished by determining only relative concentrations from the actual experimental chromatograms, with internal standards added to permit absolute determinations. A discussion of the repeatability associated with sampling valves can be found in [58]. The approximate volume of an injected sample is most easily measured by injecting a standard acid solution into a flask with the value to be calibrated and titrating.

10.9.2. Measurement of Peak Area

Peak areas are now measured with the aid of automatic integrators or appropriately programmed computers. These devices incorporate A/D converters and noise filters, and make logical decisions based on the value of the observed signal. One of the major problems is the fact that the devices are closed, and often not transparent to the end user. They are typically programmed by electronic engineers or programmers, who are often poorly informed with respect to the analysts' problems. It is necessary that chemists always have access to raw data when needed, and they should demand that. The nature of the algorithms used to correct for baseline drift, to detect the beginning and end of a peak, and to allocate peak areas between incompletely resolved bands should be clearly indicated, and it should be easy to select the particular algorithm required. Various methods used in automatic integration of chromatographic signals and the related problems have been discussed in detail in [58].

Of special concern is the algorithm used for peak-area allocation when two or more bands are not sufficiently well resolved [58]. In most cases, and if the two peaks are comparable in size, a vertical line is dropped from the bottom of the valley between the peaks. This may cause an error ranging from one to a few percent. When the relative concentration is very different from unity, as in the case of trace analysis, and if the minor component is eluted second in the tail of a major component, an artificial baseline is assumed by establishing a tangent to the major-component profile before and after the minor peak, or the profile of the tail may be interpolated and subtracted on the assumption of an exponential decay. Such a procedure can never be as satisfactory as achieving complete resolution of the peaks, and can in fact cause quite significant errors [58].

10.9.3. Calibration Techniques

Quantitative analysis requires the determination of individual response factors for the various components of interest in the analyzed mixture [58]. No detector gives the same response factor for two different compounds unless they are enantiomers. In practice, quantitative analysis requires that the detector be calibrated by injection of mixtures of known compositions containing one or two internal standards. Response factors with respect to the internal standards are then determined in the concentration range expected for the analytes. Response factors should never be extrapolated, even if there are good reasons to believe that the detector provides a linear response. The concentration of the analyte C_{an} is

$$C_{an} = C_{is} S_{an/is} \frac{A_{an}}{A_{is}} \qquad (52)$$

where $S_{an/is}$ is the relative response factor for the analyte and A_{an} and A_{is} are the observed peak areas for analyte and internal standard. The need to calibrate the detector means that no quantitative analysis is possible for compounds that have not been identified. If no source of authentic compound of sufficient purity is available, calibration becomes difficult, but it is not impossible [58].

Use of an external standard, injected separately from the sample after a predetermined delay selected so as to avoid interferences between peaks of the standard and those of the mixture components, represents an extremely useful calibration tool, and it also serves as a control for monitoring proper performance of the analytical instrument, especially in process-control analysis [58], [66]. The peak area of the external, deferred standard should remain constant for as long as the chromatograph is in good working order.

10.9.4. Sources of Error in Chromatography

Any fluctuation of an operating parameter of the chromatograph, including those that control the retention times and band widths of peaks and those that determine response factors, constitutes a source of error in chromatographic analysis [58]. Such parameters include the column temperature and the mobile-phase flow rate, as well as various detector settings [67].

10.10. Theory of Nonlinear Chromatography

Equilibrium isotherms are not linear as they are assumed to be in classical treatments of analytical chromatography. In the low concentration range used for analysis, observed band profiles are in good agreement with predictions made on the basis of replacing an isotherm by its initial tangent. When large samples are used, however, and the concentrations of mixture components are high, the nonlinear character of the thermodynamics of phase equilibria causes band profiles to deviate significantly from the classical Gaussian shape. Furthermore, the equilibrium composition of the stationary phase at high concentrations depends on the concentrations of all the components. In other words, the concentration of one particular compound in the stationary phase at equilibrium is

a function of the mobile-phase concentrations of all the components. Thus, the elution profiles interact. In preparative chromatography, unlike analytical chromatography, the chromatogram obtained for a mixture is not simply the sum of those obtained when equivalent amounts of each component are injected separately [23].

The prediction of such band profiles requires the solution of several problems:

1) How does the nonlinear behavior of isotherms influence band profiles?
2) How do bands of incompletely resolved components interact?
3) What is the relative importance of thermodynamic effects compared to the kinetic sources of band broadening already known to exist from studies in linear chromatography?
4) How should one model competitive isotherms?

The answers to all these questions are now well known. The ideal model of chromatography responds to the first and second questions [68]. The equilibrium – dispersive model speaks to the first three questions, at least so long as the column efficiency under linear conditions exceeds 30 – 50 plates [22]. The last question is outside the scope of this review, but has been abundantly discussed in the literature [23].

Band profiles in chromatography are obtained as solutions to the differential mass-balance equations for the mixture components (Eq. 11). The models differ in the way such balance equations are simplified and solved.

10.10.1. The Ideal Model of Chromatography

In this model we assume that the column has infinite efficiency [69]. Thus, this model makes it possible to investigate the influence of the nonlinear thermodynamics of phase equilibria independently of the blurring, dispersive effects of axial dispersion and the finite rate of mass transfer. Neglecting one of the two major phenomena controlling band profiles in chromatography, it is as "theoretical" as linear chromatography. The results must be convoluted with those of linear chromatography to lead to a realistic description of actual chromatograms. Note that in linear, ideal chromatography, band profiles are identical to injection profiles, and the bands move at the same velocity as in linear chromatography.
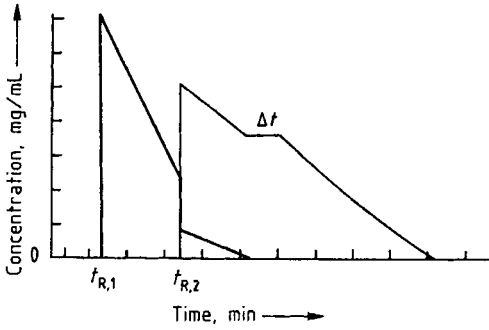
Figure 1. Solution of the ideal model for the elution of a narrow rectangular pulse of a binary mixture
Component 1; — Component 2

For single-component bands it can be shown that, in the ideal model, a particular migration velocity can be associated with each concentration [22], [69]. However, this assigned velocity is a function of the concentration. Therefore, each concentration leads to a well-defined retention time, which is a linear function of the differential of the isotherm q:

$$t_{R,C} = t_0 \left(1 + F\frac{dq}{dC} \right) \tag{53}$$

where F is the phase ratio [see Eq. (11)].

If the isotherm is convex upward, as in most cases found in liquid chromatography, the retention time decreases with increasing concentration. High concentrations cannot pass lower ones, however, as there can be only one concentration at a time at any given point. The concentrations thus pile up at the front, creating a concentration discontinuity or shock. This shock migrates at its own velocity, which is a function of its height. It can be shown [22] that the maximum concentration of the shock is the solution of the equation

$$\left| q(C) - C\frac{\partial q}{\partial C} \right| = \frac{N}{Ft_0F_c} \tag{54}$$

The retention time is obtained by introducing the maximum concentration into Equation 53.

In gas chromatography the isotherms, and especially the gas–liquid isotherms, tend to be convex downward. In this case the concentration shock forms at the rear of the band, which appears to be fronting [33]. However, another effect becomes important. The partial molar volumes of the sample components in the two phases, which are

nearly equal in LC, are very different in GC. A sorption effect arises from the much higher velocity of the mobile phase in the band, where the component vapor adds to the mobile phase, relative to the region outside the band [33]. The higher the partial pressure of the vapor, the faster the band travels. Thus, bands eluted early display a front shock, later ones a rear shock. An intermediate band has a shock on both sides [70].

In the case of a binary mixture, an analytical solution exists only if the competitive equilibrium behavior is accounted for by the Langmuir isotherm

$$q_i = \frac{a_i C_i}{1 + b_1 C_1 + b_2 C_2} \tag{55}$$

where a_i and b_i are numerical coefficients characterizing the phase system and the compound.

This solution has been discussed in detail [71]. An example is provided in Figure 1. Characteristics of this solution include:

1) The existence of two concentration shocks at $t_{R,1}$ and $t_{R,2}$, one for each of the two components

2) The fact that the positive shock of the second component concentration is accompanied by a partial, negative shock from the first component concentration (at $t_{R,2}$)

3) A concentration plateau of length Δt for the second component following the end of the elution profile of the first component

4) The fact that the profile for the end of the second component is the same as the profile of a pure band of that compound

The front of the second component displaces the first component (*displacement effect*), while the tail of the first component drags forward the front part of the second-component profile (*tag-along effect*). These effects have important consequences in preparative chromatography [22], [23].

10.10.2. The Equilibrium – Dispersive Model

In this model, constant equilibrium is still assumed between the two phases of the system, but finite column efficiency is accounted for by introducing a dispersion term in the mass-balance equation [22]. This term remains independent of the concentration. Solutions of the model must be obtained by numerical calculation. Several possi-

ble algorithms have been discussed. Excellent agreement has been reported between such solutions and experimental band profiles [72].

The solutions of the equilibrium – dispersive model exhibit the same features as those of the ideal model at high concentrations, where thermodynamic effects are dominant and dispersion due to finite column efficiency merely smoothes the edges. When concentrations decrease, the solutions tend toward the Gaussian profiles of linear chromatography.

10.10.3. Moderate Column Overloading

At moderate sample sizes, band profiles depend on the initial curvature of the isotherm, while retention times are functions of both initial slope and curvature [73]. Accordingly, retention varies linearly with increasing sample size [74]. The widespread belief that there is a range of sample sizes over which retention time is independent of sample size is erroneous. What happens is simply that, below a certain size, variation in the retention time is less than the precision of measurement. As the retention time decreases the band front or tail becomes steeper and the band becomes unsymmetrical.

If an isotherm can be reduced to the first two terms of a power expansion (i.e., to a parabolic isotherm), an approximate solution can be obtained by the method of HAARHOFF and VAN DER LINDE [75] in the case of a single-component band.

10.10.4. Preparative Chromatography

Preparative chromatography consists in using chromatography to purify a substance, or to extract one component from a mixture. Unlike analytical chromatography it is not devoted to the direct collection of information. This difference in objective results in marked differences in operation [23]. Production of finite amounts of purified compounds requires the injection of large samples, hence column operation under overloaded conditions. Resolution between individual bands in the chromatogram is no longer necessary provided mixed bands between the components are narrow. Instead, product recovery is the goal, demanding efficient condensers in the case of GC and rotary evaporators for LC. However, equivalent column performance is required in both analytical and preparative applications, and the packing materials must be of comparable quality.

10.11. Reference Material

10.11.1. Journals

The literature of chromatography is enormous. Several journals are dedicated entirely to this field, including the *Journal of Chromatography*, *Journal of Chromatographic Sciences*, and *Chromatographia*, while a larger number specialize in various subfields of chromatography (e.g., *Journal of Liquid Chromatography*, *Journal of Planar Chromatography*, *Journal of Microcolumn Separation*, *High Resolution Chromatography*). Several other journals publish one or several papers on chromatography in every issue. Important papers on chromatography are found in journals important in other scientific fields as well, ranging from *Biochemistry Journal* to *Chemical Engineering Science*.

One journal provides only abstracts of chromatographic papers appearing in other journals: *Chromatography Abstracts*, published by Elsevier for the Chromatographic Society. Chemical Abstracts also issues several *CA Selects* dedicated to specific aspects of chromatography.

10.11.2. Books

Probably the oldest general reference book still available is *Theoretische Grundlagen der Chromatographie* by G. SCHAY [76]. In the theoretical area, *Dynamics of Chromatography* by J. C. GIDDINGS [12] is still a mine of useful discussions, comments, and ideas. A number of excellent books have been written about gas chromatography. *Gas Chromatography* by A. B. LITTLEWOOD [77], though obsolete in many respects, contains excellent fundamental discussions on GC and GC detectors. More practical, *The Practice of Gas Chromatography* by L. S. ETTRE and A. ZLATKIS [78] supplies much valuable information. Similarly, *Modern Practice of Liquid Chromatography* by J. J. KIRKLAND and L. R. SNYDER [79] has long been the primary reference book for LC.

Among recent books, the latest editions of *Chromatography*, edited by E. HEFTMAN [80], and *Chromatography Today* by POOLE and

POOLE [81] contain excellent practical discussions, covering almost all aspects of implementing chromatographic separations. However, two important areas were virtually ignored: quantitative measurement and preparative separations. With respect to the former, *Quantitative Gas Chromatography* [58] is the only book that discusses in detail the sources of errors, their control, and problems of data acquisition and handling, which should not be left exclusively in the hands of computer specialists. For the latter, *Fundamentals of Preparative and Nonlinear Chromatography* [82] presents all aspects of the theory of chromatography, and discusses the optimization of preparative separations.

10.12. References

[1] D. H. Everett, *Trans. Faraday Soc.* **61** (1965) 1637.

[2] T. Tsuda, K. Nomura, G. Nakagawa, *J. Chromatogr.* **248** (1982) 241.

[3] K. Sakodynski, *J. Chromatogr.* **73** (1972) 303.

[4] L. S. Ettre in Cs. Horváth (ed.): *High Performance Liquid Chromatography: Advances and Perspectives*, vol. 1, Academic Press, New York 1980, p. 25.

[5] M. S. Tswett, *Ber. Dtsch. Bot. Ges.* **24** (1906) 316.

[6] M. S. Tswett, *Ber. Dtsch. Bot. Ges.* **24** (1906) 384.

[7] M. S. Tswett: *Khromofilly v Rastitel'nom Zhivotnom Mire (Chromophylls in the Plant and Animal World)*, Izd. Karbasnikov, Warzaw, Poland, 1910 (reprinted in part in 1946 by the publishing house of the Soviet Academy of Science, A. A. Rikhter, T. A. Krasnosel'skaya, eds.).

[8] V. Berezkin: *M. S. Tswett*, Academic Press, New York 1992.

[9] A. J. P. Martin, R. L. M. Synge, *Biochem. J.* **35** (1941) 1358.

[10] R. Consden, A. M. Gordon, A. J. P. Martin, *Biochem. J.* **38** (1944) 224.

[11] A. J. P. Martin, A. T. James, *Biochem. J.* **50** (1952) 679.

[12] J. C. Giddings: *Dynamics of Chromatography*, Marcel Dekker, New York 1965.

[13] K. K. Unger: *Porous Silica, Its Properties and Use as Support in Column Liquid Chromatography*, Elsevier, Amsterdam 1979.

[14] A. Tiselius, S. Claeson, *Ark. Kemi Mineral. Geol.* **16** (1943) 18.

[15] J. Frenz, Cs. Horváth in Cs. Horváth (ed.): *High-Performance Liquid Chromatography, Advances and Perspectives*, vol. 5, Academic Press, New York 1988.

[16] S. Golshan-Shirazi, G. Guiochon in F. Dondi, G. Guiochon (eds.): "Theoretical Advancement in Chromatography and Related Separation Techniques," *NATO Adv. Study Inst. Ser. Ser. C* **383** (1992) 61.

[17] L. C. Craig, *J. Biol. Chem.* **155** (1944) 519.

[18] J. C. Giddings in E. Heftmann (ed.): *Chromatography*, Van Nostrand, New York 1975, p. 27.

[19] J. C. Giddings, H. Eyring, *J. Phys. Chem.* **59** (1955) 416.

[20] D. A. McQuarrie, *J. Chem. Phys.* **38** (1963) 437.

[21] F. Dondi, M. Remelli, *J. Phys. Chem.* **90** (1986) 1885.

[22] S. Golshan-Shirazi, G. Guiochon: "Theoretical Advancement in Chromatography and Related Separation Techniques," in F. Dondi, G. Guiochon (eds.): *NATO Adv. Study Inst. Ser. Ser. C* **383** (1992) 35.

[23] A. M. Katti, G. Guiochon in J. C. Giddings, E. Grushka, P. R. Brown (eds.), *Adv. Chromatogr. (N.Y.)* **31** (1992) 1.

[24] S. Golshan-Shirazi, G. Guiochon, *J. Chromatogr.* **603** (1992) 1.

[25] L. Lapidus, N. L. Amundson, *J. Phys. Chem.* **56** (1952) 984.

[26] J. J. Van Deemter, F. J. Zuiderweg, A. Klinkenberg, *Chem. Eng. Sci.* **5** (1956) 271.

[27] M. J. E. Golay in D. H. Desty (ed.): *Gas Chromatography 1958*, Butterworths, London 1959, p. 36.

[28] M. Kubin, *Collect. Czech. Chem. Commun.* **30** (1965) 2900.

[29] E. Kucera, *J. Chromatogr.* **19** (1965) 237.

[30] E. Grushka, *J. Phys. Chem.* **76** (1972) 2586.

[31] L. J. Schmauch, *Anal. Chem.* **31** (1959) 225.

[32] J. G. Atwood, M. J. E. Golay, *J. Chromatogr.* **218** (1981) 97.

[33] G. Guiochon, L. Jacob, *Chromatogr. Rev.* **14** (1971) 77.

[34] H. Darcy: *Les Fontaines Publiques de la Ville de Dijon*, Dalmont, Paris 1856.

[35] D. E. Martire, R. E. Boehm, *J. Phys. Chem.* **91** (1987) 2433.

[36] R. C. Reid, J. M. Prausnitz, B. E. Poling: *The Properties of Gases and Liquids*, 4th ed., McGraw-Hill, New York 1987.

[37] A. Felinger, G. Guiochon, *Biotechnol. Progr.* **9** (1993) 421.

[38] C. Vidal-Madjar, M. F. Gonnord, M. Goedert, G. Guiochon, *J. Phys. Chem.* **79** (1975) 732.

[39] E. sz Kovats, *Helv. Chim. Acta* **41** (1958) 1915.

[40] B. L. Karger, *Anal. Chem.* **39** (1967) no. 8, 24 A.

[41] A. V. Kiselev, Ya. I. Yashin: *Gas Adsorption Chromatography*, Plenum Press, New York 1970.

[42] L. R. Snyder in C. Horváth (ed.): *High Performance Liquid Chromatography, Advances and Perspectives*, vol. 1, Academic Press, New York 1980, p. 280.

[43] M. Denkert, L. Hackzell, G. Schill, E. Sjögren, *J. Chromatogr.* **218** (1981) 31.

[44] S. Golshan-Shirazi, G. Guiochon, *Anal. Chem.* **62** (1990) 923.

[45] W. Melander, C. Horváth in C. Horváth (ed.): *High Performance Liquid Chromatography, Advances and Perspectives*, vol. 2, Wiley, New York 1980, p. 114.

[46] J. H. Knox, *J. Chromatogr. Sci.* **15** (1977) 352.

[47] G. Guiochon, A. Siouffi, *J. Chromatogr. Sci.* **16** (1978) 470.

[48] J. H. Purnell, *J. Chem. Soc.* 1960, 1268.

[49] B. L. Karger, W. D. Cooke, *Anal. Chem.* **36** (1964) 985.

[50] G. Guiochon, *Anal. Chem.* **38** (1966) 1020.

[51] G. Guiochon, E. Grushka, *J. Chromatogr. Sci.* **10** (1972) 649.

[52] G. Guiochon in C. Horváth (ed.): *High Performance Liquid Chromatography, Advances and Perspectives*, vol. 2, Wiley, New York 1980, p. 1.

[53] J. C. Sternberg in J. C. Giddings, R. A. Keller (eds.), *Adv. Chromatogr. (N.Y.)* **2** (1966) 205.

[54] M. J. E. Golay, J. G. Atwood, *J. Chromatogr.* **186** (1979) 353.

[55] J. M. Davis, J. C. Giddings, *Anal. Chem.* **57** (1985) 2178.

[56] J. C. Giddings: *Unified Separation Science*, Wiley-Interscience, New York 1991.

[57] A. Wehrli, E. sz Kovats, *Helv. Chim. Acta* **42** (1959) 2709.

[58] G. Guiochon, C. Guillemin: *Quantitative Gas Chromatography*, Elsevier, Amsterdam 1988.

[59] R. V. Golovnya, D. N. Grigoryeva, *Chromatographia* **17** (1983) 613.

[60] H. Colin, E. Grushka, G. Guiochon, *J. Liq. Chromatogr.* **5** (1982) 1391.

[61] A. J. P. Martin, *Discuss. Faraday Soc.* **7** (1949) 332.

[62] D. A. Leathard, B. C. Shurlock: *Identification Techniques in Gas Chromatography*, Wiley-Interscience, New York 1970.

[63] L. Rohrschneider, *J. Gas Chromatogr.* **6** (1968) 5.

[64] R. W. Frei, J. F. Lawrence: *Chemical Derivatization in Analytical Chemistry, 1. Chromatography*, Plenum Press, New York 1981.

[65] M. Goedert, G. Guiochon, *J. Chromatogr. Sci.* **7** (1969) 323.

[66] C. L. Guillemin et al., *J. Chromatogr. Sci.* **9** (1971) 155.

[67] G. Guiochon, M. Goedert, L. Jacob in R. Stock (ed.): *Gas Chromatography 1970*, The Institute of Petroleum, London 1971, p. 160.

[68] S. Golshan-Shirazi, G. Guiochon: "Theoretical Advancement in Chromatography and Related Separation Techniques," in F. Dondi, G. Guiochon (eds.): *NATO Adv. Study Inst. Ser. Ser. C* **383** (1992) 1.

[69] D. DeVault, *J. Am. Chem. Soc.* **65** (1943) 532.

[70] P. Valentin, G. Guiochon, *J. Chromatogr. Sci.* **10** (1975) 271.

[71] S. Golshan-Shirazi, G. Guiochon, *J. Phys. Chem.* **93** (1989) 4143.

[72] G. Guiochon et al., *Acc. Chem. Res.* **25** (1992) 366.

[73] S. Golshan-Shirazi, G. Guiochon, *J. Chromatogr.* **506** (1990) 495.

[74] A. Jaulmes, C. Vidal-Madjar, M. Gaspar, G. Guiochon, *J. Phys. Chem.* **88** (1984) 5385.

[75] P. C. Haarhof, H. J. Van der Linde, *Anal. Chem.* **38** (1966) 573.

[76] G. Schay: *Grundlagen der Chromatographie*, Springer Verlag, Berlin.

[77] A. B. Littlewood: *Gas Chromatography*, 2nd ed., Academic Press, New York 1970.

[78] L. S. Ettre, A. Zlatkis: *The Practice of Gas Chromatography*, Interscience, New York 1967.

[79] J. J. Kirkland, L. R. Snyder: *Modern Practice of Liquid Chromatography*, Wiley-Interscience, New York 1971.

[80] E. Heftman (ed.): *Chromatography*, 5th ed., Elsevier, Amsterdam 1993.

[81] C. Poole, S. K. Poole: *Chromatography Today*, Elsevier, Amsterdam 1992.

[82] G. Guiochon, S. Golshan-Shirazi, A. Katti: *Fundamentals of Preparative and Nonlinear Chromatography*, Academic Press, Boston 1994.

11. Gas Chromatography

PAT J. F. SANDRA, University of Stellenbosch, Matieland, South Africa, Universiteit Gent, Vakgroep
Organische Chemie, Gent, Belgium

Abbreviations

AED   atomic emission detection
CGC   capillary gas chromatography
CI   chemical ionization
CpSi   cyanopropylsilicone
d_f film thickness
D_M diffusion coefficient of the mobile phase
d_p particle diameter
ECD   electron capture detector
EI   electron ionization
ELCD   Hall eletrolytic conductivity detector
EPC   electronic pressure control
FFAP   free fatty acid phase
FID   flame ionization detector
FPD   flame photometric detector
FSOT   fused silica open tubular
GALP   good automated laboratory practice
GLC   gas – liquid chromatography
GLP   good laboratory practice
GSC   gas – solid chromatography
H column efficiency, plate height
HPLC   high performance liquid chromatography
HS   head space
i.d.   internal diameter
K Kelvin temperature
k retention factor
L column length
MAOT   maximum allowable operating temperature
MDCGC   multidimensional capillary chromatography
MeSi   methylsilicone
MiAOT   minimum allowable operating temperature
N plate number
NPD   nitrogen phosphorus detector
PCGC   packed column gas chromatography
PID   photoionization detector
PLOT   porous layer open tubular
PT   purge and trap
PTV   programmed-temperature vaporization
r column radius
RCD   redox chemiluminescence detector
R_s peak resolution
SFC   supercritical fluid chromatography
SFE   supercritical fluid extraction
SIM   selected ion monitoring
TCD   thermal conductivity detector
TD   thermal desorption
TEA   thermal energy analyzer
TID   thermionic detector
t_M retention time of solvent
t_R retention time of analyte
u mobile phase velocity
WCOT   wall-coated open tubular
α separation factor

11.1. Introduction

Gas chromatography (GC) is the separation technique that is based on the multiplicative distribution of the compounds to be separated between the two-phase system solid or liquid (*sta-tionary phase*) and gas (*mobile phase*). Gas chromatography by definition thus comprises all separation methods in which the moving phase is gaseous. Contrary to the other chromatographic techniques, i.e., high-performance liquid chromatography (HPLC) and supercritical fluid chromatography (SFC), the role of the gaseous mobile phase — quasi-ideal inert gases such as nitrogen, helium, or hydrogen — is purely mechanical: they just serve for the transport of solutes along the column axis. The residence time (retention) of solutes is affected only by their vapor pressure, which depends on the temperature and on the intermolecular interaction between the solutes and the stationary phase.

Currently, gas chromatography is one of the most important and definitely the most economic of all separation methods. Its applicability ranges from the analysis of permanent gases and natural gas to heavy petroleum products (up to 130 carbon atoms, simulated distillation), oligosaccharides, lipids, etc. Moreover, as far as chromatographic efficiency and GC system selectivity is concerned, no other separation technique can compete with gas chromatography. The dictum "If the separation problem can be solved by gas chromatography, no other technique has to be tried out" is now generally accepted. As illustration, more than 80 % of the priority pollutants on the lists of the EC and of the EPA in the United States, are amenable to GC analysis.

Gas chromatography was developed in the early 1950s [8]. Tens of thousands of publications and more than 250 textbooks on theory and practice have been published. The purpose of this article is not to write a new textbook on GC but rather to help newcomers in the field to find their way in the present state of the art of modern GC. For the basic aspects of GC, see → Basic Principles of Chromatography. Fundamental equations are given or repeated only to clarify some important statements. The terms, symbols, and nomenclature used are those that have been advised by IUPAC [9].

At the end of the references, some general and specialized textbooks are listed, together with the most important journals dealing with gas chromatography. The reader is advised to consult these books and journals for a more detailed study.
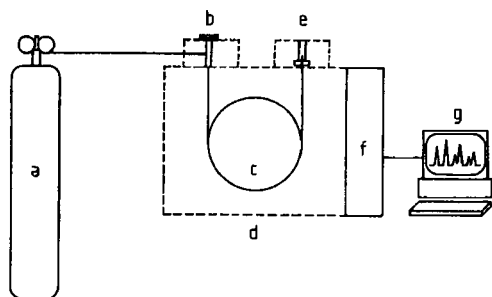
Figure 1. Basic components of a modern GC system
a) Carrier gas supply; b) Injector; c) Column; d) Column
oven; e) Detector; f) Controller; g) Recorder

11.2. Instrumental Modules

A schematic drawing of a modern gas chromatographic system is shown in Figure 1. The basic parts are the carrier gas supply (a), the injector (b), the column (c), and the detector (e). The dashed lines indicate thermostatted regions. The carrier gas is usually supplied from a high-pressure cylinder equipped with a two-stage pressure regulator. Via the controller of the GC instrument (f), the gas flow can be fine-tuned, and preheated carrier gas is delivered to the column in the constant-pressure, constant-flow, or pressure-programmed mode. The sample is introduced into the carrier gas stream via the injector, and vaporized components are transported into the column, the heart of the system, where separation occurs. The column is placed in an oven, the temperature of which can be kept constant (*isothermal operation*) or programmed (*temperature-programmed operation*). Some applications are performed at subambient temperature and instruments can be equipped with cryogenic units. After separation, the component bands leave the column and are recorded (g) as a function of time (chromatogram) by a detection device (e). The chromatogram provides two types of data. The residence time or retention time of the components in the column is characteristic for the solute – stationary phase interaction and can be used for qualitative interpretation or component identification. The detector's response is proportional to the amount of separated sample component and gives quantitative information on the composition of the mixture.

11.3. The Separation System

11.3.1. Modes of Gas Chromatography

Two types of gas chromatography exist: gas – liquid chromatography (GLC) and gas – solid chromatography (GSC). Other classification schemes such as GSC, GLC plus capillary gas chromatography (CGC) are outdated because nowadays GLC and GSC can be performed both in packed columns and in capillary or open tubular columns.

11.3.1.1. Gas – Liquid Chromatography

In GLC the stationary phase is a liquid acting as a solvent for the substances (solutes) to be separated. The liquid can be distributed in the form of a thin film on the surface of a solid support, which is then packed in a tube (packed column GC, PCGC) or on the wall of an open tube or capillary column (capillary GC, CGC). The term WCOT (*wall-coated open tubular*) is used only for capillary columns coated with a thin film of a liquid. In GLC the separation is based on partition of the components between the two phases, the stationary phase and the mobile phase, hence the term partition chromatography. This mode of GC is the most popular and most powerful one.

11.3.1.2. Gas – Solid Chromatography

Gas – solid chromatography comprises the techniques with an active solid as the stationary phase. Separation depends on differences in adsorption of the sample components on inorganic adsorbents (i.e., silica, alumina, carbon black) or on organic adsorbents such as styrene – divinylbenzene copolymers. Separation can also occur by a size exclusion mechanism, such as the separation of gases on synthetic zeolites or molecular sieves. GSC is performed on packed columns or on open tubular columns on the walls of which a thin layer of the porous material is deposited [*porous layer open tubular* (PLOT) columns]. GSC nowadays is used only for special separation problems, and GSC columns are, therefore, referred to as tailor-made columns.

11.3.2. Selection of the Carrier Gas

Common carrier gases are nitrogen, helium, and hydrogen. The choice of carrier gas depends
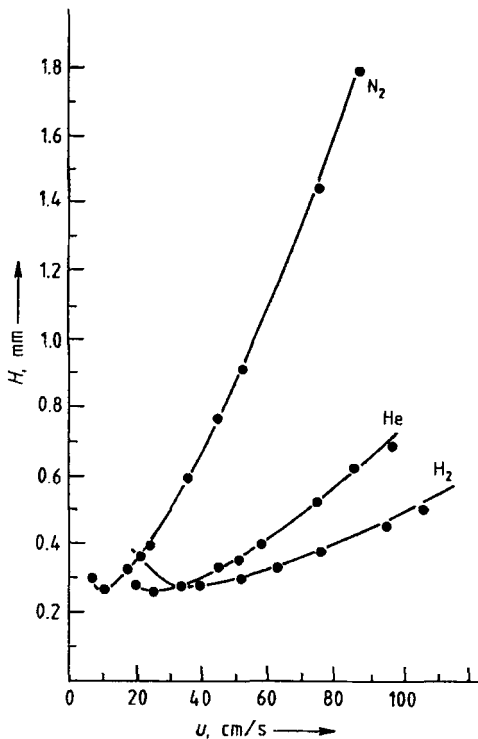
Figure 2. Experimental $H - u$ curves as function of carrier gas

on the column and the detector used but, above all, on the required speed of analysis and detectability. The influence of the mobile-phase velocity u on column efficiency H has been detailed in → Basic Principles of Chromatography. The relation $H - u$ was derived for packed column GC by VAN DEEMTER, ZUIDERWEG, and KLINKENBERG [10] and adapted to capillary columns by GOLAY [11]. Practical consequences of the carrier gas selection in capillary GC have been described in [12], [13]. Due to the relatively high density of nitrogen compared to helium and hydrogen, longitudinal diffusional spreading of the solutes in nitrogen is small, but the resistance to mass transfer in the mobile phase is high. In contrast, for low-density gases such as helium and especially hydrogen, longitudinal diffusional spreading is large but resistance to mass transfer small. For columns with small amounts of stationary phases, the resistance to mass transfer in the stationary phase can be neglected, and the minimum plate height H_{min} is nearly independent of the nature of the carrier gas. The optimum mobile-phase velocity u_{opt}, corresponding to H_{min} is proportional to the diffusion

coefficient of the mobile phase D_M; u_{opt} is 10 cm/s for nitrogen, 25 cm/s for helium, and 40 cm/s for hydrogen. In practical GC on thin-film columns, hydrogen is the best choice because the analysis time is reduced by a factor of 4 compared to nitrogen, while the detectability is increased by the same order. Some experimental $H - u$ curves for a *capillary column* 20 m in length and 0.27 mm internal diameter (*i.d.*) coated with a 0.22-μm film of methylsilicone (MeSi) are shown in Figure 2. The experimental data fit the theory perfectly. The fact that the slope (i.e., the increase of plate height with velocity) for nitrogen is much steeper than for the two other gases is noteworthy. Working close to u_{opt} is, therefore, a must with nitrogen, whereas for helium and hydrogen the velocity can be increased to values greater than u_{opt} without losing too much resolution power. For thin-film columns, hydrogen is definitely the best choice. When hydrogen is not allowed for safety reasons — most GC systems nowadays are equipped with hydrogen sensors — helium should be chosen. Because of the important contribution of the resistance to mass transfer in *thick-film columns*, this situation is more complicated. The selection depends on what is considered most important, efficiency or speed. This is dictated by the separation problem at hand. For a given column, the selection of carrier gas and optimum flow rate is best determined experimentally by making $H - u$ plots. This is not difficult and can be carried out in 3–4 h in the following way:

1) A solution of a hydrocarbon (i.e., tridecane) in a highly volatile solvent (i.e., hexane) is injected at an inlet pressure P_x of carrier gas A and a column temperature is selected to give a retention factor k of ca. 5.
2) The plate number N is calculated according to

$$N = 5.54 \, (t_R/w_h)^2 \tag{3.1}$$

where t_R is the retention time of the hydrocarbon and w_h is the peak width at half height.
3) The H value is calculated from

$$H = L/N \tag{3.2}$$

where L denotes the column length.
4) The corresponding u value is given by

$$u = L/t_M \tag{3.3}$$
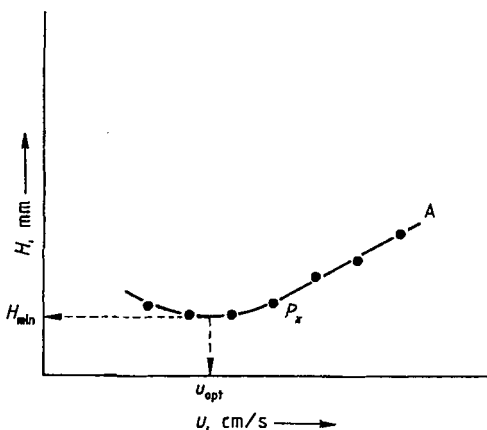
Figure 3. Making a *H–u* plot

where t_M is the retention time of the solvent; for thick-film columns, methane is injected instead of hexane to give t_M.

5) The H and u values corresponding to P_x of carrier gas A are marked in a $H–u$ plot (Fig. 3).
6) The inlet pressure P is set at values P_x^{-3}, P_x^{-2}, P_x^{-1}, P_x^{+1}, P_x^{+2}, P_x^{+3}, and the corresponding H and u values are drawn in the plot.
7) H_{min} and u_{opt} are deduced from the plot.
8) The same procedure is repeated for carrier gases B and C.

High-purity gases should be used in GC. Most gas suppliers have special GC-grade gases. If lower-quality carrier gases are used, purification through a molecular sieve trap to remove moisture and low molecular mass hydrocarbons and through an oxy-trap to remove oxygen is recommended.

11.3.3. Selection of the Gas Chromatographic Column

Two column types are in use, the packed and the capillary column. The selection is often dictated by the nature and complexity of the sample. The present trend is to replace PCGC by CGC whenever possible, because data obtained with the latter technique are much more reliable. The fundamental difference between the two methods is reflected in the resolution equation, the key equation for separation optimization:

$$R_S = \frac{\sqrt{N}}{4} \frac{\alpha - 1}{\alpha} \frac{k}{k + 1} \qquad (3.4)$$

Anything that increases the column efficiency N, the column selectivity α, or the retention factor k will enhance the separation power of the column. Packed columns are characterized by low plate numbers and PCGC is therefore a low-resolution technique. The lower efficiency is compensated by the high selectivity α of the stationary phase, and this is the main reason why so many different stationary phases have been developed for PCGC. Capillary columns on the other hand have very high plate numbers and, therefore, the number of stationary phases can be restricted because the selectivity is less important. In fact, most separation problems can be handled with four basic stationary phases and half a dozen tailor-made stationary phases. Other important features of capillary columns are their inertness and compatibility with spectroscopic detectors. In the framework of this discussion, emphasis is, therefore, on capillary columns.

11.3.3.1. Packed Columns

A packed column is defined by the material of which the tube is made, by its length L and internal diameter *i.d.*, by the nature and diameter d_p of the particles, and of course by the film thickness d_f and the nature of the stationary phase. The latter is most important because the stationary phase controls column selectivity (see Section 11.3.4). All other column parameters represent column efficiency. In packed column GC, H_{min} is controlled mainly by the particle diameter and is always $>2d_p$. A column 1 m long filled with particles of 100–120 mesh (150–125 µm) can never generate more than 3000 plates. Because of the packed bed, the resistance to the mobile flow is high, and columns longer than 3 m exhibit low permeability and are not practical. This means that a total plate number of 10000 is the maximum that can be reached with packed columns. Most of the columns for routine analysis in fact have only 3000 to 5000 plates. The column efficiency is independent of column diameter. The column tube can be made of copper, stainless steel, nickel, glass, and polytetrafluoroethylene. Glass and polytetrafluoroethylene are the most inert materials and are preferred for the analysis of polar and thermolabile compounds. In *gas – liquid chromatography*, the support material has a silicium oxide network, such as diatomaceous earth, synthetic spherical silica, or glass beads. The support materials are purified by acid washing and deactivated by silanization, which caps the active sites. However, even

the most highly deactivated surface will still ex-
hibit adsorption of polar compounds such as or-
ganic acids and bases. A variety of support mate-
rials are available commercially, the catalogues of
the column manufacturers (Agilent, Chrompack,
Supelco, Alltech, etc.) should be consulted for
detailed information. The stationary-phase film
is homogeneously coated on the particles by dis-
solving the phase in a suitable solvent and adding
the support material to the mixture. The solvent is
then removed on a rotavapor (rotary evaporator),
and after complete drying under a purified ni-
trogen flow, the column is packed tightly with
the coated support. In PCGC, the contents of liq-
uid coatings are expressed as a percentage of the
support material: 1, 3, 5, and 10 wt% are com-
monly used.

The dimensions of columns used in *gas – solid
chromatography* are very similar to those of gas–
liquid chromatography. Packing materials (silica,
alumina, carbon blacks, zeolites, porous polymers)
in proper mesh size are available commercially to
fill the columns. GSC columns give lower plate
numbers than GLC columns but possess very high
selectivities for some typical applications. Pre-
packed GLC and GSC columns can be purchased
commercially and are ready to be installed in the
GC instrument.

11.3.3.2. Capillary Columns

Major progress in gas chromatography, partic-
ularly in the area of applications, has been made
by the introduction of capillary columns. Invented
by M. GOLAY in 1957 [11], the real breakthrough
came only in 1979 with the development of flex-
ible fused silica as tubing material for capillary
columns [14] (*fused silica open tubular columns,*
FSOT). The inherent strength and flexibility of
fused silica made use of capillary GC easy, com-
pared to the glass capillary columns generally used
previously. In addition, no other column material
can offer the same inertness. A capillary column is
characterized by its dimensions (*L, i.d.*) and by the
nature of the stationary phase. For *GLC,* a rel-
atively thin stationary (liquid) phase film, with
film thickness d_f, is coated directly on the inner
wall (WCOT columns). The film thickness con-
trols the retention factor k and the sample capacity
of a capillary column. In the case of GSC, the
adsorbent with particle diameter d_p is deposited
in a thin layer d_l on the capillary wall (PLOT
columns). The main advantage of PLOT columns
compared to packed GSC columns is the speed of

analysis (see Section 11.3.4.2). Most of the appli-
cations (>90%) in capillary GC are performed on
WCOT columns. These columns can be sub-
divided in three types according to their internal
diameter. Columns with *i.d.* of 0.18 – 0.32 mm are
called *conventional capillary columns* because this
is the column type most often used for the analysis
of complex samples. Columns with an *i.d.* of
0.53 mm are known as *widebore* or *megabore col-
umns.* They have been introduced as an alternative
to packed columns for the analysis of less complex
mixtures. *Narrowbore columns* with *i.d.*'s ranging
from 0.1 to 0.05 mm provide very high efficien-
cies or short analysis times. This column type is
not often used nowadays, but this situation is ex-
pected to change in the next few years (see Sec-
tion 11.10.1).

The selection of a capillary column depends on
the complexity of the sample to be analyzed. The
column length, the internal diameter, the station-
ary phase, and its film thickness determine the
separation power (resolution), the sample capaci-
ty, the speed of analysis, and the detectability or
sensitivity. Theoretical considerations [12], [13]
indicate that for capillary columns with thin films
(<1 µm), the H_{min} value is roughly equal to the
column diameter. This is illustrated in Figure 4,
which shows experimental $H-u$ curves for col-
umns varying in internal diameter. H was calcu-
lated for dodecane at 100 °C with hydrogen as
carrier gas. H_{min} measured experimentally is in-
deed very close to H_{min} deduced theoretically. By
knowing this, the maximum plate number that a
capillary column can provide may be calculated
without performing any analysis:

$$N = L/H = L/H_{min} = L/d_c \qquad (3.5)$$

The H_{min} values, the maximum number of plates
per meter, and the minimum column length
needed for 100 000 plates for different *i.d.*'s of
commercially available FSOT columns are listed
in Table 1. In contrast to packed columns, the
length of a capillary column is not restricted be-
cause the permeability of an open tube is very
high. To a great extent, the length and internal
diameter of a capillary column define the efficien-
cy. The selection of the film thickness for a given
stationary phase is also of upmost importance. The
resolution equation (Eq. 3.4) shows that low k
values result in poor resolution. For the same col-
umn efficiency N and selectivity α, increasing the
k value from 0.25 to 5 corresponds to a resolution
gain of 378% (R_s increases from 0.37 to 1.40).

Table 1. Characteristics of FSOT columns coated with thin films

d_c, mm	H_{min}, mm	N, m^{-1}	L for 100 000 plates, m
0.05	0.05	20 000	5
0.10	0.10	10 000	10
0.18	0.18	5 556	18
0.22	0.22	4 545	22
0.25	0.25	4 000	25
0.32	0.32	3 125	32
0.53	0.53	1 887	53

Increasing the retention factor can be accomplished by increasing the film thickness

$$k = \frac{K\,2\,d_f}{r_c} \tag{3.6}$$

For a fixed column radius (r_c = constant), the retention factor of a solute at a given temperature (K = constant), doubles when the film thickness is increased by a factor of 2. Guidelines for the selection of film thickness for the stationary phase methylsilicone are listed below (values depend on column length and stationary phase):

≈ 0.1 µm	high molecular mass compounds 500 – 1500
≈ 0.2 µm	$M_r \approx 300 – 500$
≈ 0.3 µm	$M_r \approx 150 – 400$
≈ 0.5 µm	$M_r \approx 50 – 200$
≈ 1 µm	$M_r \approx 30 – 150$
$2 – 5$ µm	widebore – high-capacity narrowbore – high volatiles

Conventional Capillary Columns. Capillary columns with *i.d.*'s of 0.18 – 0.32 mm and lengths of 10 – 60 m are the "workhorses" for high-resolution separation. Long columns are preferred for the analysis of very complex samples. The less complex the sample, the shorter the column can be. Figure 5 shows the analysis of the oxygen fraction of the essential oil of black pepper [15]. The profile is extremely complex (more than 300 compounds), and the column efficiency must be high to unravel all solutes. Long columns, of course, imply long analysis times. Figure 6 shows the carbon number separation of the neutral lipids in palm oil. The separation is less challenging and the column length can be reduced to 5 m, which results in an analysis time of only 3 min [16]. For samples of unknown complexity, column lengths of 25 – 30 m offer a good compromise. Column dimensions are often adapted to the requirements
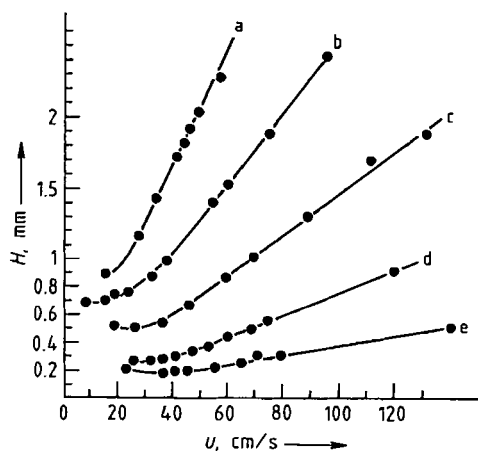
Figure 4. Experimental $H - u$ curves for columns with different *i.d.*'s
a) 0.88 mm; b) 0.70 mm; c) 0.512 mm; d) 0.27 mm; e) 0.18 mm

of pre- or postcolumn devices. As an example, for the combination with mass spectroscopy, an internal diameter of 0.18 – 0.25 mm is preferred because in this case the flow of the mobile phase is of the order of 1 mL/min. This flow rate fits well with the pump capacity of the spectrometer so that special interfacing is not required.

Widebore Capillary Columns. FSOT capillary columns with *i.d.*'s of 0.53 mm were introduced as alternatives to packed columns, with the inertness of fused silica as their main advantage. Widebore or megabore columns have higher sample capacities and are applicable for simple packed column-like separations. A serious limitation of small *i.d.* capillary columns indeed is the low sample capacity (25 m $L\times0.25$ mm *i.d.*; $d_f = 0.2$ µm results in a capacity of roughly 20 ng per compound). Sample capacity is the ability of a column to tolerate high concentrations of solutes. If the capacity is exceeded, the chromatographic performance is degraded as evidenced by leading (overloaded) peaks. Sample capacity is related to the film thickness and column radius. A column of 10 m$\times$0.53 mm, $d_f = 1$ µm, has a sample capacity >200 ng per compound. Simple inlet devices can be used with widebore columns in the low-resolution mode. In this mode, the carrier gas velocity is far above optimum values (flows of 8 – 12 mL/min), and a 10-m column generates only 5000 – 8000 plates. For simple mixtures, these efficiencies are sufficient. Megabore columns are
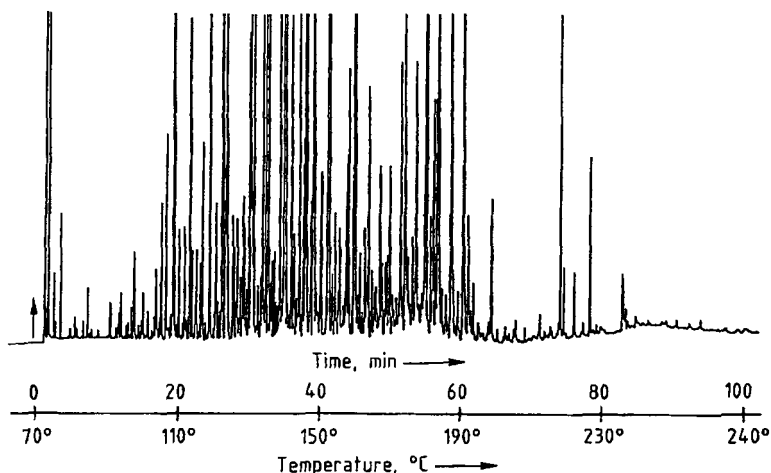
Figure 5. Analysis of oxygen-containing compounds of essential oil of pepper
Column: 40 m L×0.25 mm i.d.; 0.25 µm d$_f$ high molecular mass poly(ethylene glycol); temperature as indicated; split injection 1/50; carrier gas hydrogen at 15 psi (≈ 1 bar); FID
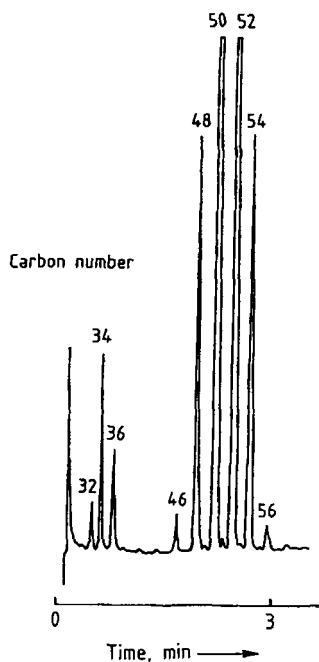
Figure 6. Fast separation of palm oil
Column: 5 m L×0.25 mm i.d.; 0.1 µm d$_f$ methylsilicone; temperature: 290–350 °C at 30 °C/min; cool on-column injection; carrier gas hydrogen at 15 psi (≈ 1 bar); FIDPeaks: 32–36 diglycerides, 46–56 triglycerides

also available in lengths of 30–60 m. At optimum velocities (flows of 1–3 mL/min) they offer 50 000 and 100 000 plates, respectively, which is considered high-resolution GC.

Narrowbore Capillary Columns. Narrowbore columns offer very high efficiencies (100 m L×0.1 mm i.d. = 1 000 000 plates) or ultrahigh speed at 100 000 plates but at the expense of sample capacity. This is illustrated in Figure 7 for the analysis of fatty acid methyl esters on a 10 m L×0.1 mm i.d., d$_f$ 0.1 µm, column of cyanopropylsilicone (CpSi). Total analysis time is only 90 s. Sample injection is critical because the sample capacity is < 1 ng. This is the reason for the slow acceptance of narrowbore columns. Normal sample size injection is expected to be possible with the development of new sample inlets such as programmed temperature vaporization in the solvent venting mode.

11.3.4. Stationary Phases in Gas Chromatography

11.3.4.1. General Considerations

The importance of the stationary phase in GC can, once again, be deduced from the resolution equation (Eq. 3.4). In Figure 8, the influence of the plate number N, the separation or selectivity factor α, and the retention factor k on peak resolution R_s is represented graphically. The peak resolution of two solutes ($k = 5$) with $\alpha = 1.05$ on a column with a plate number of 20 000 (e.g., 10 m×0.5 mm coated with 0.2 µm methylsilicone) is 1.4. This value, the crossing point of the curves, is taken as reference. The curves are obtained by changing
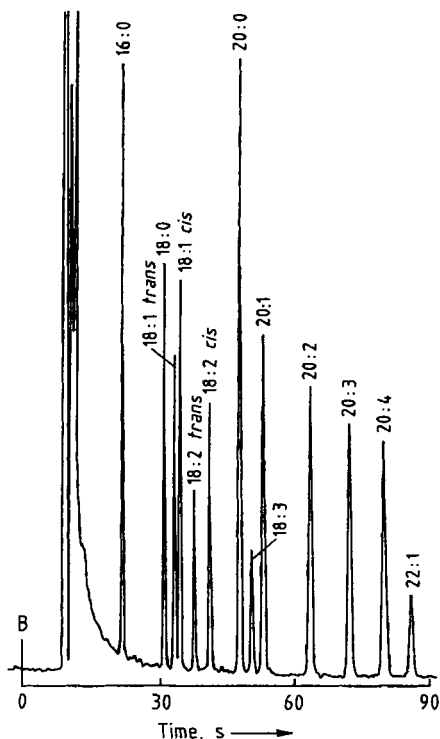
Figure 7. Fast separation of fatty acid methyl esters
Column: 10 m L×0.1 mm *i.d.*; 0.1 μm *d*₁ cyanopropylsilicone; temperature: 180 °C; split injection 1/500; carrier gas hydrogen at 60 psi (≈ 4 bar); FID
Peaks: as indicated (the figure before the colon denotes the number of carbon atoms; that after the colon, the number of double bonds per molecule)

Figure 8. Effect of N, α, and k on R_s

one of the variables while the other two remain constant. From the curves, the following conclusions can be drawn:

Plate Number N. Peak resolution is proportional to the square root of the plate number. Increasing the plate number by a factor of 2 ($N = 20\,000$ to $N = 40\,000$) improves peak resolution by only 1.4 ($\sqrt{2}$ gives $R_s = 1.40 \rightarrow 1.96$).

Retention Factor k. For values >5, the influence of k on R_s is small. Increasing k from 5 to 20 increases resolution by only 14 % ($R_s = 1.40 \rightarrow 1.60$).

Separation Factor α. Optimization of the separation factor has the most important impact on peak resolution. Increasing α from 1.05 to 1.10 nearly doubles the peak resolution ($R_s = 1.40$
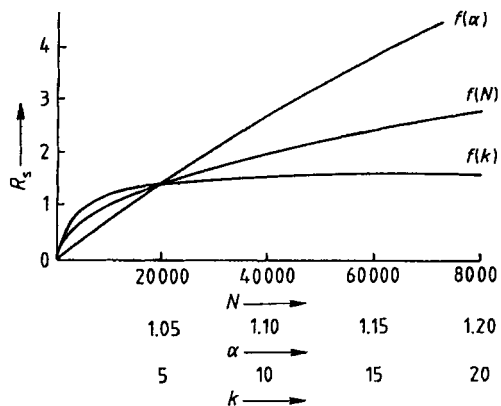
$\rightarrow 2.68$), while a slight decrease ($\alpha = 1.03$) causes a drastic loss of resolution ($R_s = 1.40 \rightarrow 0.86$). The selection of a "liquid phase" that maximizes the relative retention of solutes is thus of upmost importance. The x-scales in Figure 8 have not been selected arbitrarily. Varying plate number and retention factor, respectively, from $N = 20\,000$ to $40\,000$ and from $k = 5$ to 20 is not so difficult to realize. In contrast, increasing the separation factor is not easy to achieve, especially in gas chromatography. First of all, as already mentioned, the mobile phase is an inert gas and its role is purely mechanical. Second, optimization of the selectivity of the stationary phase for a particular sample means that the nature of the solutes must be known. Optimization then requires good chemical insight and intuition. Several separations have been achieved more by luck than by skill. In 1991 a book was published with very detailed discussion of practically every stationary phase used since the inception of gas chromatography [17]. The reader is referred to this book for more information.

The selectivity of a stationary phase is commonly expressed in terms of the relative retention of a compound pair:

$$\alpha = k_2/k_1 \tag{3.7}$$

The thermodynamic description of Equation (3.7) is

$$\alpha = p^\circ_1/p^\circ_2 \cdot \gamma^\circ_1/\gamma^\circ_2 \tag{3.8}$$

where p° refers to the saturated vapor pressure of the analyte and γ° is the activity of the analyte at

Table 2. Resolution R_s for different α values

α	R_s	
	GC	CGC
1.2	2.9	12.5
1.1	1.6	6.8
1.05	0.8	3.6
1.02	↓	1.5
1.01		0.8
	other stationary phase	↓
		other stationary phase or increase of column length

infinite dilution in the stationary phase. Equation (3.8) is split in two contributions: the first term is related to the solutes and depends only on temperature, whereas the second term depends on the selective properties of the stationary phases. Solute – stationary phase interactions are very complex and may include nonspecific interactions (dispersion interactions); specific interactions such as dipole – dipole, dipole – induced dipole, and hydrogen bonds; and chemical interactions such as acid – base, proton donor – proton acceptor, electron donor – electron acceptor, and inclusion. The terms *selectivity* and *polarity* are commonly used to describe a stationary phase. A polarity and selectivity scale of more than 200 stationary phases has been established by McREYNOLDS [18]. The Kovats retention indices (see Chap. 11.7) of benzene, butanol, 2-pentanone, nitropropane, and pyridine on phase X are measured, and the differences in indices in phase X, relative to the stationary phase squalane as typical apolar phase, are listed individually (selectivity) and as a sum (polarity). Although the polarity scale is discussed in all textbooks on GC, it has little practical value. The following are by far the most common methods used for selection of a stationary phase: trying a phase with good chromatographic properties, looking for a similar application that has already been published, or asking a colleague with experience (trial and error) for advice. At this stage the difference between PCGC and CGC should be restressed. Consider a packed column and a capillary column coated with the same stationary phase of the following characteristics: *packed column*, 2 m $L\times4$ mm *i.d.*, packed with 100 – 120 mesh diatomaceous earth particles, 3 % methylsilicone, plate number 5000; *capillary column*, 25 m $L\times0.25$ mm *i.d.*, coated with 0.25 µm methylsilicone, plate number 100 000. By neglect-

ing the k contribution in the resolution equation, the resolution for two solutes with different α values can be calculated (Table 2). For an α value of 1.05, the resolution on the packed column is insufficient and another stationary phase must be selected. On the capillary column for the same solute – stationary phase interaction the resolution is still 3.6 because of the very high efficiency of a capillary column. Even for α values of 1.02, baseline separation is obtained. In the case of α value 1.01, the resolution is only 0.8, but due to the high permeability of a capillary column, its length can be increased to provide the required plate number N_{req} on the same stationary phase. From the resolution equation follows:

$$N_{req} = 16 R_s^2 \, (\alpha/\alpha - 1)^2 = 160\,000 \text{ plates} \qquad (3.9)$$

and as $N = L/d_c$, the required length for $R_s = 1$ is 40 m. The analysis of a fatty acid methyl ester mixture on a *packed column* with a highly selective phase for these solutes (ethylene glycol succinate) is shown in Figure 9; several peaks are incompletely separated. The same analysis was performed on a capillary column coated with poly(ethylene glycol), a stationary phase with no particular selectivity for the solutes (Fig. 10): all peaks are baseline separated.

11.3.4.2. Solid Phases

For certain applications, GSC enjoys some advantages over GLC. Adsorbents are stable over a wide temperature range, and column bleed is virtually nonexistent especially for inorganic adsorbents and molecular sieves. *Silica* and *alumina* adsorbents give excellent separation of saturated and unsaturated hydrocarbons with low molar mass. Medium-polarity and polar solutes interact too strongly with the highly adsorptive surface of these adsorbents. *Graphitized carbon blacks* are well suited for separation of structural and geometric isomers. The disadvantage of carbon blacks is that their retention characteristics differ from batch to batch and they seem difficult to prepare in a reproducible way. Inorganic adsorption columns can be modified by coating the material with a small amount of nonvolatile liquid or inorganic salt. The most adsorptive sites will bind preferentially to these modifiers. As a result, the sites are unavailable to participate in retention, and retention times generally decrease with increased reproducibility. *Molecular sieves* (zeolites) are aluminosilicates of alkali metals. The most common
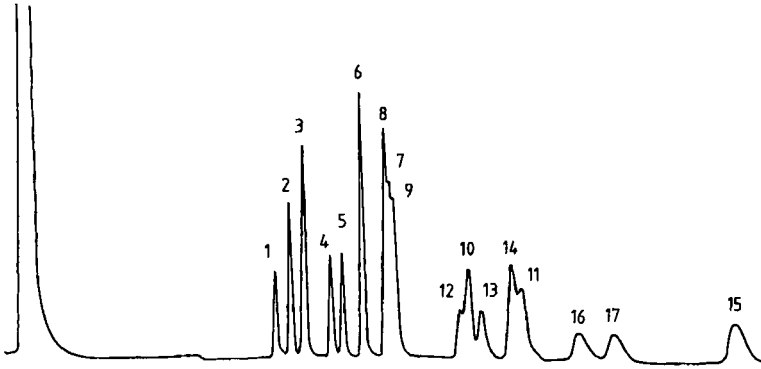
Figure 9. Isothermal analysis (200 °C) of fatty acid methyl esters on a packed column with stationary phase ethylene glycol succinate (EGS-X)
1. $C_{16}:0$; 2. $C_{16}:1$; 3. $C_{17}:0$; 4. $C_{18}:0$; 5. $C_{18}:1$; 6. $C_{18}:2$; 7. $C_{18}:3$; 8. $C_{20}:0$; 9. $C_{20}:1$; 10. $C_{20}:4$; 11. $C_{20}:5$; 12. $C_{22}:0$; 13. $C_{22}:1$; 14. $C_{23}:0$; 15. $C_{22}:6$; 16. $C_{24}:0$; 17. $C_{24}:1$

types for GC are molecular sieves 5A and 13X. Separation is based mainly on the size of the solutes and the porous structure of the molecular sieve, although adsorptive interactions inside and outside the pores may also contribute. Molecular sieves are used primarily for the separation of permanent and inert gases including low molar mass hydrocarbons. *Porous polymer beads* of different properties have found many applications in the analysis of volatile inorganic and organic compounds. The properties of the commercially available materials vary by chemical composition, pore structure, and surface area. Compared to inorganic adsorbents, polymer beads are stable to water injections, and on some of the materials, polar solutes give perfect peak shapes in reasonable analysis times. The best-known adsorbents are the Porapak series, the Chromosorb 101-108 series, and Tenax. The range of applications on those polymers can be found in catalogues of the column and instrument manufacturers. Tenax (a linear polymer of *para*-2,6-diphenylphenylene oxide) is somewhat unique because of its thermal stability up to 380 °C. Relatively high molecular mass polar compounds such as diols, phenols, or ethanolamines can be analyzed successfully. Some doubt exists as to the mechanism of retention on Tenax. At low temperature, adsorption appears to be the principal retention mechanism, but at higher temperature the surface structure becomes liquid-like and partitioning does occur.

Solid Phases for PLOT Columns. At the moment, GSC is enjoying a revival because of the introduction and commercialization of PLOT columns. The advantages of these columns are higher efficiency, speed of analysis, and better stability and reproducibility over time [19]. To date, PLOT columns are available in different lengths and internal diameters, with aluminum oxide – KCl for the separation of saturated and unsaturated hydrocarbons including low-volatile aromatics; with molecular sieve 5A for the analysis of permanent and inert gases; with porous polymers of the Porapak-type (Q,S,U) (PoraPlot columns) for the analysis of polar solutes and water; and with carbon black (CarboPlot columns) for volatile apolar solutes in petroleum products and environmental samples. Figure 11 shows the determination of benzene and toluene in natural gas on an Al_2O_3 – KCl PLOT column. The elution of benzene from the C_6 group and of toluene from the C_7 group illustrates the high selectivity.

11.3.4.3. Liquid Phases

In the past, numerous phases were used in PCGC, leading to what has been called "stationary-phase pollution" [20]. In the beginning, the many phases were a blessing; later, a curse. Today most separations are performed on a dozen preferred phases, which are all specially synthesized for GLC purposes. Some popular PCGC phases, listed according to increasing polarity, are

Methylsilicone (gum)
Methylsilicone (fluid)
Methylphenyl(5 %)silicone
Methylphenyl(50 %)silicone
Methyltrifluoropropyl(50 %)silicone
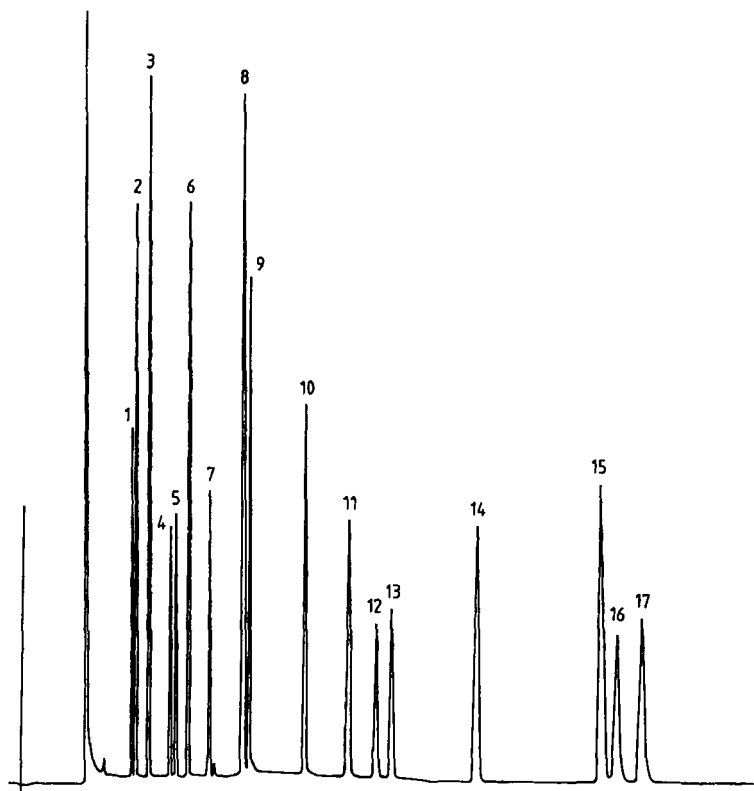
Figure 10. Isothermal analysis (200 °C) of fatty acid methyl esters on a capillary column with high-M_r poly(ethylene glycol) as stationary phase (for designation of peaks, see Fig. 9)

Methylphenyl(25 %)cyanopropyl(25 %)silicone
Poly(ethylene glycol), $M_r > 40\,000$
Cyanopropyl(50 %)phenylsilicone
Poly(ethylene glycol) esterified with 2-nitroterephthalic acid
Diethylene glycol succinate
Cyanopropyl(100 %)silicone
Ethylene glycol succinate
1,2,3-Tris(2-cyanoethoxy)propane

Of this list, by far the most important group are the silicones. The reasons can be summarized as follows: Their diffusion coefficients are high, and thus resistance to mass transfer is low, silicones have a high thermal stability (methylsilicone gum up to 400 °C) and are liquids or gums at low temperature, the structure is well defined, and the polymers are very pure. For practical GC this results in a broad temperature range, which is defined by the minimum allowable operating temperature range (MiAOT) and by the maximum allowable operating temperature range (MAOT),

and low column bleeding. Poly(ethylene glycol) with high molecular mass, which has a unique selectivity, is a solid below 60 °C (MiAOT = 60 °C) and bleeds at 240–250 °C as a result of the higher vapor pressure compared to silicones.

11.3.4.3.1. Basic Phases for Capillary GC

As already mentioned, the number of phases in CGC can be reduced further due to the high efficiency those columns offer. STARK et al. [21] studied stationary-phase characteristics in depth and concluded that the order of optimized phases for CGC, all higher molecular mass crosslinkable gums, is

Methylsilicone
Methylphenyl(50–70 %)silicone
Methylcyanopropylsilicones with medium (25–50 %) and high (70–90 %) cyanopropyl content
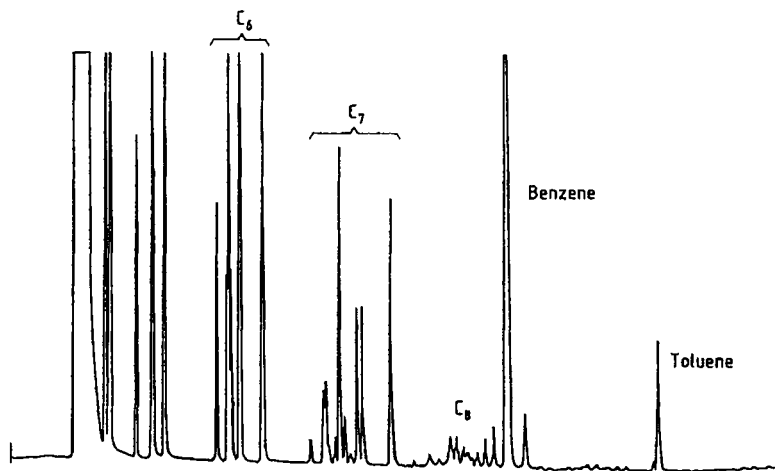
Figure 11. Natural gas analysis on an Al_2O_3 PLOT column
Column: 25 m $L\times0.25$ mm *i.d.*; Al_2O_3 – KCl; temperature: 75 – 200 °C at 3 °C/min; split injection; carrier gas nitrogen at 5 psi ($\approx 3.5\times10^4$ Pa); FID

Methyltrifluoropropylsilicone

High molecular mass poly(ethylene glycol) or a silicone substitute having the same polarity and selectivity

The functionalities in those phases are the methyl, phenyl, cyano, trifluoro, and hydroxy–ether group, respectively. The similarity with HPLC phases is remarkable. Based on the above, SANDRA et al. [22] proposed five basic phases for capillary GC. The formulas are given in Figure 12. For the majority of high-resolution separations, these stationary phases provide more than adequate performance. For some applications, the functionalities are combined on the same siloxane backbone, or columns are coupled (selectivity tuning). Functionalities can also be modified to provide specific interactions. Optical phases and liquid crystals complete the set of preferred CGC phases. These tailor-made phases are discussed in Section 11.3.4.3.3.

The most important phase is *methylsilicone*. Separations occur according to the boiling point differences of the solutes. A typical chromatogram is shown in Figure 13. Fatty acid methyl esters are separated according to their boiling point or molar mass. Hardly any separation occurs among oleic (C_{18}: 1 *cis*), elaidic (C_{18}: 1 *trans*), linoleic (C_{18}: 2) and linolenic (C_{18}: 3) acids. *Methylphenylsilicone* is the second most widely used stationary phase. Selectivity is increased due to the polarizable phenyl group. *Methyltrifluoropropylsilicone* possess rather unique selectivity for electron donor

solutes such as ketones and nitro groups. Incorporating the *cyano group* with its large dipole moment into the silicone backbone provides strong interactions with dipolar solutes and unsaturated bonds. Cyano phases are widely used for the separation of unsaturated fatty acid methylesters and for the resolution of dioxins and furans. Figure 14 represents the analysis of a sample whose composition is very similar to that in Figure 13. Separation according to the number of double bonds in the alkyl chains is due to $\pi-\pi$ interactions. Methyl oleate (*cis* C_{18}: 1) and methyl elaidate (*trans* C_{18}: 1) differing only in configuration are baseline separated as well. High *molecular mass poly(ethylene glycol)* has a unique selectivity and polarity for the separation of medium-polarity and polar compounds. The most popular of these phases is Carbowax 20M with average M_r of 20 000.

11.3.4.3.2. Selectivity Tuning

A solvent with a selectivity intermediate to the selectivities of the five basic phases can, for some applications, offer a better separation or a shorter analysis time. This is illustrated by the following example. On stationary phase X, compounds A and B can be separated from C, but they cannot be separated from each other. On another stationary phase Y, A and B are separated, but one of them coelutes with C. By using a chromatographic system with selectivity between X and Y, an opti-

Methylsilicone
(MeSi)

Methylphenylsilicone
(MePhSi)

Cyanopropylsilicone
(CPSi)

Methyltrifluoropropylsilicone
(MeTFSi)

High molecular mass
poly(ethylene glycol) $HO(CH_2CH_2O)_xH$
(PEG HMW)
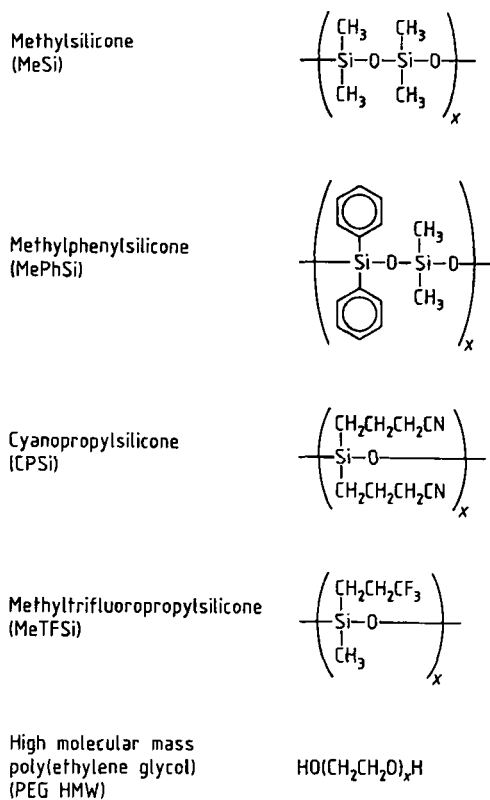
Figure 12. Structures of the basic phases for capillary GC

mum can be found to separate the three compounds. Intermediate selectivities and polarities can be obtained by selectivity tuning [23]. Selectivity tuning means that the selectivity is adapted to the analytical need by creating a selectivity between two (or more) extreme (basic phase) selectivities. In principle, this can be performed in three different ways: (1) by synthesizing a tuned phase with predetermined amounts of monomers containing the required functional groups; (2) by mixing, in different ratios, two or three basic phases in one column; and (3) by coupling two or more columns of extreme selectivities. Approach (1) and (2) have been commercially used, and columns are available for specific applications. Columns tuned for the analysis of classes of priority pollutants are best known. Method (3) is detailed in Section 11.8.1.

11.3.4.3.3. Tailor-Made Phases

The most widely used tailor-made phases are the free fatty acid phase (FFAP), liquid crystals, and the optical phases Chirasil-Val and derivatized cyclodextrins. *FFAP* is produced by condensing 2-nitroterephthalic acid with Carbowax 20M. The phase is recommended for the analysis of acidic compounds such as organic acids and phenols. The phase is not suitable for the analysis of alkaline compounds and aldehydes with which it reacts. *Liquid crystals* can separate isomers as a function of the solute length-to-breadth ratios. High selectivities have been noted for the separation of polycyclic aromatic compounds and dioxin isomers [24]. The most powerful involves the use of a liquid-crystal column as second column in a multidimensional CGC system. Optical isomers (enantiomers) can be separated principally by two different approaches. Enantiomeric mixtures containing a reactive group can be derivatized with an optically pure reagent to produce a mixture of diastereomers that can be separated on nonoptically active stationary phases such as methylsilicone. The direct separation on optically active stationary phases is, however, the current method of choice because it has certain advantages. Derivatization agents of 100% enantiomeric purity are not needed, and the method can be applied to enantiomers lacking reactive functional groups. In Chirasil-Val a diamide is incorporated into a silicone backbone, and separation occurs via a triple hydrogen interaction. A number of derivatized α-, β-, and γ-cyclodextrin stationary phases have recently been introduced for the separation of enantiomeric pairs. Separation is based on the formation of inclusion complexes with the chiral cavity of the cyclodextrins. However, none of the phases is universally applicable, and enantiomer separation is still based on trial and error. The separation of γ- and δ-lactones on a 2,3-O-acetyl-6-tert-butyldimethylsilyl-β-cyclodextrin stationary phase is shown in Figure 15. Full details on the separation of enantiomers by capillary GC can be found in [25], [26].

11.4. Choice of Conditions of Analysis

Under isothermal and constant-flow conditions, for a homologous series of hydrocarbons, an exponential correlation exists between the residence time in the stationary phase (adjusted re-
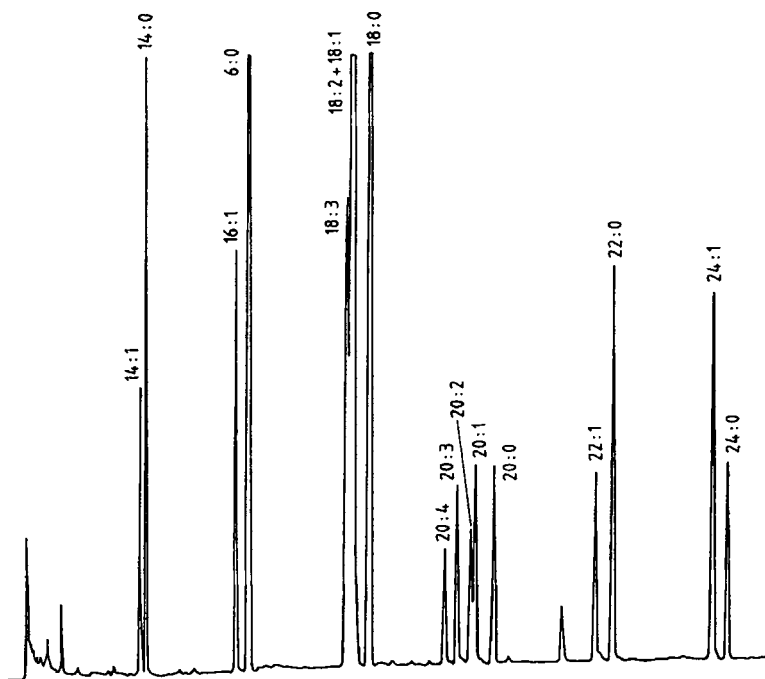
Figure 13. Analysis of fatty acid methyl esters on methylsilicone
Column: 25 m $L \times 0.25$ mm *i.d.*; 0.25 μm d_f MeSi; temperature: 100–220 °C at 3 °C/min; split injection 1/100; carrier gas hydrogen at 8 psi ($\approx 5.6 \times 10^4$ Pa); FID
(For designation of peaks, see Fig. 7)

tention time) and the solute vapor pressure, boiling point, or number of methylene groups (carbon number).

$$\log t'_R = a\, C_n + b \tag{4.1}$$

where t'_R is the adjusted retention time; C_n is the number of carbons in the individual homologues; and a, b are constants. For a mixture with boiling point range of the solutes less than 100 °C, an optimum isothermal temperature can be selected. For samples in which the boiling point difference of the solutes exceeds ca. 100 °C, working in the isothermal mode is impractical. Usually, early eluting peaks will show poor resolution (low k values), whereas late eluting peaks will be broad when a compromise temperature has been chosen. The problem can be solved by temperature or flow programming.

11.4.1. Temperature-Programmed Analysis

In temperature-programmed analysis, a controlled change of the column temperature occurs

as a function of time. The initial temperature, heating rate, and terminal temperature must be adjusted to the particular separation problem. The initial temperature is chosen so that the low-boiling compounds are optimally separated (k value >3). Selection of the program rate depends on the nature of the solutes and the complexity of the sample. The final temperature is adjusted to give reasonable total analysis times. Temperature programming has been applied since the beginning of gas chromatography (1960), but only since the end of the 1980s have the possibilities been fully exploited. This might be, in part, because of the time-consuming nature of temperature optimization by trial and error, but above all because of ignorance of the importance of the temperature programming rate to selectivity. This is especially so for polar columns where $\Delta t_R/\Delta T$ is much more pronounced than on apolar columns. Today, software is available to optimize a temperature program for a particular separation on the basis of only two experimental programmed runs. The theory of temperature-programmed optimization is treated in detail in [27], [28]. Figure 16 shows the computer predicted (A) and experimental (B)
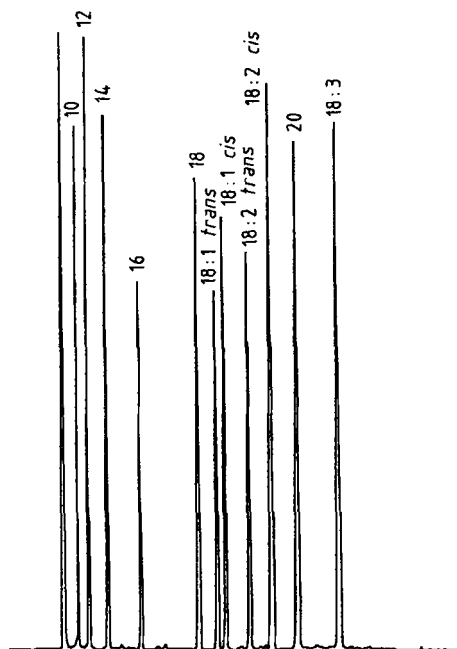
Figure 14. Analysis of fatty acid methyl esters on a cyano-propylsilicone
Column: 25 m $L \times 0.25$ mm *i.d.*; 0.25 µm d_f CPSi; temperature: 187 °C; split injection 1/100; carrier gas hydrogen at 8 psi ($\approx 5.6 \times 10^4$ Pa); FID

chromatograms for a complex fatty acid methyl ester mixture (Table 3) at a programmed temperature increase of 40 °C/min [29]. The column was 55 m in length, and a remarkably good separation was achieved in < 8 min. Without the aid of computer simulation, such fast programming rates would never have been considered for this separation.

11.4.2. Constant Pressure, Constant Flow, and Pressure Programming

Flow programming offers advantages such as reduced separation time, less column bleed, and lower eluting temperature for labile compounds but has not been widely used in gas chromatography because temperature programming is experimentally easier to perform. In 1990, electronic pressure control (EPC) [30] was developed, which allows operation in the constant-pressure, constant-flow, and pressure-programming modes. Whether temperature programming will be re-

placed by flow programming is questionable, although the features of EPC open new perspectives. Gases play a vital role in GC and are even more important in CGC. The carrier gas is the force moving the solutes through the column, and its velocity or flow controls the chromatographic band broadening and thus the efficiency. Keeping the velocity or flow under control is a prerequisite for good GC practice. Besides this, the carrier gas is often the transport medium of solutes from injection devices such as split, splitless, programmed-temperature vaporizing inlets, purge and trap, thermal desorption units, and pyrolysis units. Often the flows required in the precolumn devices do not correspond to the optimal carrier gas flows, and compromises must be accepted (i.e., split in purge and trap, long residence times of solutes in a splitless liner). On the other hand, optimization of total GC performance also requires precise control of the flows of the detector gases responsible for detector sensitivity and baseline stability. The quality of data generated by the hyphenated spectroscopic methods (see Section 11.6.4) also depends strongly on the stability of the flows of reagent gases (e.g., in chemical ionization mass spectroscopy and atomic emission detection). In the past, very little attention was paid to the mode of operation of the different gases in a GC system, and all gas supplies were more or less operated in the constant-pressure mode. In the constant-flow mode, all parts of a GC system run under optimum conditions. EPC is likely to become a standard mode of operation in the near future. The CGC applications with EPC published until now all show improved chromatographic performance going from reduced discrimination and large volume introduction in splitless injection by a pressure pulse shortening the residence time in the inlet, to increased sensitivity by EPC control of detector gases independently or in concert with the carrier gas flow. These are just a few examples, and more features are detailed in [31]. Yet more is to be expected because the possibilities of EPC have not yet been fully evaluated and exploited. A typical example can be found in multidimensional CGC (see Section 11.8.1). One of the reasons for slow acceptance of this extremely powerful technique is the necessary adjustment of the midpoint pressures, which can require several days of fine tuning with conventional pressure control devices. Flow control at the midpoint should solve this shortcoming. Nevertheless, EPC introduces a new era in good laboratory practice (GLP) and good automated laboratory practice (GALP). An-
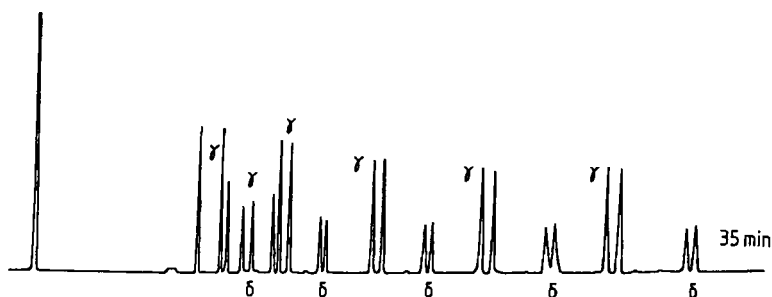
Figure 15. Enantioselective separation of γ- and δ-lactones on 2,3-*O*-acetyl-6-*tert*-butyldimethylsilyl-β-cyclodextrin
Column: 25 m *L*×0.25 mm *i.d.*; 0.25 μm *d*f; temperature: 150–185 °C at 1 °C/min; split injection 1/100; carrier gas nitrogen at 18 psi (≈ 1.3 bar); FID
Peaks: γ (6–11) and δ (8–12) lactones

Table 3. Retention times (n = 5) in initial runs, and predicted and experimental retention times at the optimum programming rate

| Compound | | Symbol | Initial runs | | Optimized program | |
No.	Name		1 °C/min, min	12 °C/min, min	Predicted, min	Experimental, min
1	Methyl butanoate	$C_4:0$	2.95	2.49	2.07	2.21
2	Methyl hexanoate	$C_6:0$	6.30	3.74	2.50	2.73
3	Methyl octadecanoate	$C_8:0$	14.65	5.38	3.01	3.32
4	Methyl decanoate	$C_{10}:0$	28.64	7.09	3.52	3.88
5	Methyl dodecanoate	$C_{12}:0$	45.02	8.69	4.00	4.41
6	Methyl phenylacetate	I.S.*	49.51	9.68	4.37	4.85
7	Methyl tetradecanoate	$C_{14}:0$	61.15	10.14	4.44	4.88
8	Methyl hexadecanoate	$C_{16}:0$	76.20	11.45	4.83	5.32
9	Methyl *cis*-9-hexadecanoate	$C_{16}:1$ *cis* 9	79.38	11.85	4.96	5.47
10	Methyl octadecanoate	$C_{18}:0$	90.08	12.65	5.19	5.72
11	Methyl *cis*-9-octadecanoate	$C_{18}:1$ *cis* 9	92.43	12.97	5.31	5.85
12	Methyl *cis*-9,12-octadecadienoate	$C_{18}:2$	97.23	13.47	5.48	6.04
13	Methyl *cis*-6,9,12-octadecatrienoate	$C_{18}:3$ γ	100.56	13.82	5.61	6.19
14	Methyl eicosanoate	$C_{20}:0$	102.95	13.75	5.55	6.11
15	Methyl *cis*-9,12,15-octadecatrienoate	$C_{18}:3$ α	102.95	14.03	5.68	6.27
16	Methyl docosanoate	$C_{22}:0$	114.64	14.77	5.92	6.49
17	Methyl *cis*-5,8,11,14-eicosatetraenoate	$C_{20}:4$	114.90	15.10	6.11	6.71
18	Methyl *cis*-4,7,10,13,16-eicosa-pentaenoate	$C_{20}:5$	120.20	15.64	6.39	6.89
19	Methyl tetracosanoate	$C_{24}:0$	126.08	15.72	6.34	6.97
20	Methyl *cis*-4,7,10,13,16,19-docosa-hexaenoate	$C_{22}:6$	133.60	16.83	7.15	7.64

* I.S. = Internal standard.

alytical methods can now be stored completely, providing exactly the same flow characteristics of all gases and tremendously enhancing repeatability and accuracy.

11.5. Sample Inlet Systems

11.5.1. General Considerations

"If the column is described as the heart of chromatography, then sample introduction may,
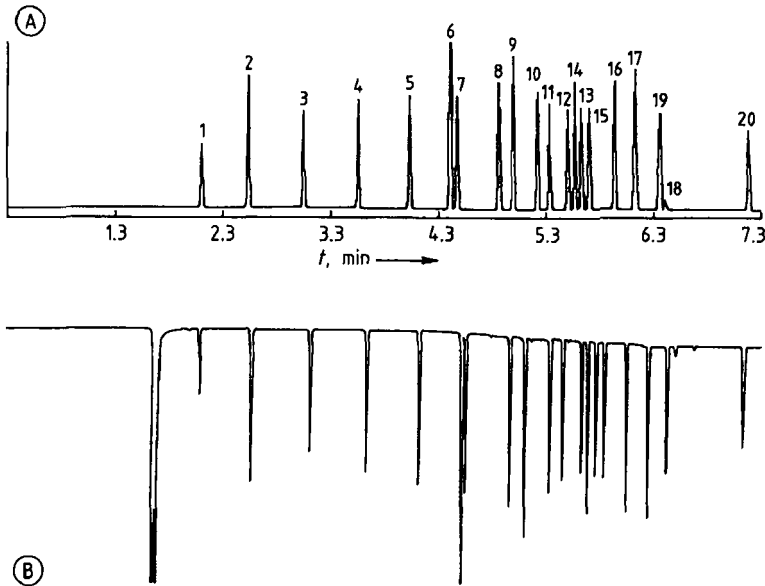
Figure 16. Predicted (A) and experimental (B) chromatograms for fatty acid methyl ester mixture with temperature-programmed analysis at 40 °C/min

with some justification, be referred to as the Achilles heel." This statement [32] clearly illustrates that sample introduction is of primary importance in GC and especially capillary GC. The performance of the sample introduction system is crucial for overall chromatographic performance of the complete system. Much progress has been made in recent years, and the understanding of injection phenomena has increased tremendously. Because of the diversity of samples that can be analyzed with modern GC (wide range of component concentrations; from highly volatile to less volatile components; different thermal stabilities; etc.), several injection methods have been developed. A final universal answer to sample introduction, however, has not been provided yet, and whether such an answer can be provided at all is questionable: "There does not now and probably never will exist injector hardware or methodology that is suitable for all samples, under all conditions" [33]. "There is still no such thing as a universal injection system and there probably never will be" [34]. This does not mean that precise and accurate results cannot be generated with the different injection systems that have been optimized over the years. The possibilities and the limitations of available systems, however, must be known exactly. In combination with a sufficient knowledge of the composition of the sample to be

analyzed, proper injector choice will guarantee good results.

A qualitative and quantitative analysis means that the chromatographically determined composition of the sample corresponds to its real composition. Possible difficulties encountered in obtaining such data have to be attributed to the sampling technique, various column effects, or a combination of both. A sample introduction system can be "discriminative," meaning that certain components are not introduced quantitatively into the column. However, the column itself can also act to discriminate (reversible and/or irreversible adsorption), the degree of which can, moreover, be injection dependent. The characteristics of the column in use, therefore, must be well established before any conclusion can be drawn with regard to the sampling system.

The basic prerequisite of a sampling system for gas chromatography is that the sample must be introduced into the column as a narrow band, whose composition is identical to the original composition of the sample. The width of the inlet band must be such that its variance does not contribute significantly, or contributes only slightly, to the chromatographic broadening process. The total measured variance σ_{total} is the sum of all contributions to the peak variance:
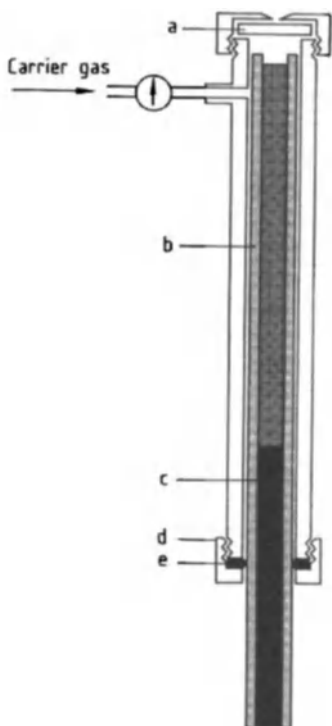
Figure 17. Schematic of a packed column injector
a) Septum; b) Glass wool; c) Packing; d) Nut; e) Ferrule

$$\sigma^2_{total} = \sigma^2_{column} + \sigma^2_{injection} + \sigma^2_{extra} \qquad (5.1)$$

With modern GC instrumentation, the extra peak variance contribution (electrometer, data system, recording) approaches zero. Because of the occurrence of the squared variances in this equation, the contribution of the initial bandwidth to the total peak bandwidth is less dramatic than might be expected, at least for GC on packed columns. For capillary columns, injection bandwidths can be too large compared to the chromatographic band broadening, and narrow initial bands have to be obtained by:

1) Introducing very small sample volumes (1–50 nL) into the column. A sample stream splitting device (splitter) accomplishes this task. The sample vapor, formed on injection at elevated temperature, is split into two sample streams of different flow rates.

2) The entire sample is introduced into the column, but the broad initial band is immediately focused into a narrow initial band by sharp-

ening mechanisms such as solvent, thermal, or stationary-phase focusing. In practice, these effects are used with great advantage to focus the solute bandwidths produced by splitless and on-column injection.

In subsequent sections, the most important sample introduction systems will be discussed in some detail with emphasis on capillary inlets. Because of the much larger σ_{column} in packed columns, injection is less stringent. Inlets are also divided into universal and selective inlets. The aim of universal injection is introduction of the complete sample into the column, whereas selective injection means that only a well-defined fraction enters the column.

11.5.2. Universal Inlets

11.5.2.1. Packed Column Inlets

Packed column inlets are simple to design and use [35]. Injection is performed directly in the column "on-column" (Fig. 17) or in an in-sert placed before the head of the column. Different inserts are used, depending on the type of packed column (1/8" = 0.32 cm or 1/4" = 0.63 cm). Polar and thermolabile compounds are most effectively analyzed by on-column injection on a glass column. The injector zone is heated, and upon manual or automated injection, usually with a microliter syringe, hot mobile-phase gas carries the vaporized sample down to the packing material. Packed column inlets are vaporizing so they can cause sample degradation and needle discrimination. This is discussed in detail for split injection. Sampling valves are often used to sample gases or volatile liquids in constant-flowing streams. The valve is substituted for the syringe and can be used both with packed and with capillary columns. The construction of GC valves is similar to HPLC valves.

11.5.2.2. Capillary Column Inlets

In the context of this contribution, it is impossible to treat every aspect of sample introduction in capillary GC in depth. Full details on the different sampling methods can be found in [36], [37]. A survey of sampling systems is also given in [34], [38].

11.5.2.2.1. Split Injection

Split sampling was the first sample introduction system developed for capillary GC [39]. The conventional split injector is a flash vaporization device. The liquid plug, introduced with a syringe, is volatilized immediately, and a small fraction enters the column, while the major portion is vented to waste. This technique guarantees narrow inlet bands. A schematic diagram of a split injector is shown in Figure 18. Preheated carrier gas enters the injector and the flow is divided in two streams. One stream of carrier gas flows upward and purges the septum (a). The septum purge flow is controlled by a needle valve (b). Septum purge flow rates are usually between 3 and 5 mL/min. A high flow of carrier gas enters the vaporization chamber (c), which is a glass or quartz liner, where the vaporized sample is mixed with carrier gas. The mixed stream is split at the column inlet, and only a small fraction enters the column (d). A needle valve controls the split ratio. Split ratios (i.e., column flow : inlet flow) range typically from $1:50$ to $1:500$ for conventional capillary columns ($0.22-0.32$ mm *i.d.*). Lower split ratios can be used in combination with focusing effects. For columns with a high sample capacity such as widebore or thick-film columns, low split ratios ($1:5-1:50$) are common. In high-speed capillary gas chromatography, applying 0.1 to 0.05 mm *i.d.* columns, split ratios can exceed $1:1000$. In split injection, initial bandwidths are only milliseconds. Because split injection is a flash vaporization technique, sample discrimination is difficult to avoid. This is especially the case if the sample contains components in different concentrations and with different volatilities and polarities. Sample discrimination in split injection is caused by inlet-related parameters as well as by operational parameters such as syringe handling. *Inlet-related discrimination* is often referred to as nonlinearity of the splitter device. Linearity in this respect means that the split ratio at the point of splitting is equal to the preset split ratio and equal for all the components in the sample. Linear splitting against varying sample components, be it in concentration, volatility, or polarity, is impossible to achieve, even when a sample is introduced in a nondiscriminative manner into the vaporization chamber. Different mechanisms can cause nonlinear splitting: different diffusion speed of the sample components, incomplete evaporation, and fluctuating split ratio. The different mechanisms and their respective contribution are discussed in

[36]. Nonlinearity can be minimized by complete vaporization of the sample, followed by homogeneous mixing with the carrier gas before the sample enters the column. As evident as this may seem, the sample introduced by split injection often arrives at the point of splitting as a mixture of vapor and a nonuniform mist of droplets. Two approaches can be used to minimize this phenomenon: (1) increased injection temperature and (2) optimization of the inlet configuration and glass liners. Different glass liners have been proposed (i.e., empty tube, short glass-wool plug in the splitting region, long and tight glass-wool plug, packing with chromatographic support or glass beads, deformation of cross section, Jennings tube, etc.) with the aim of enhancing efficient heat transfer to the injected sample and thoroughly mixing the vaporized sample. Still, whereas such modifications demonstrate an improvement in linearity for some applications, the same setup may give bad results for others. As a general rule, packed liners should not be used unless bad results have been obtained with an open liner. When discussing nonlinearity, it was taken for granted that sample introduction with a syringe into the vaporization chamber occurs without any alteration to the sample — in other words, no discrimination is caused by the syringe introduction. However, most of the discrimination problems encountered with vaporizing injectors are related to *syringe needle effects*. Upon introducing the syringe needle through the septum, volatiles immediately start to evaporate inside the needle itself, which is heated by the injector. Also after pushing down the plunger, solvent and volatile solutes are evaporated more readily than high-boiling solute material, which partly remains on the needle wall. On removing the needle from the injector body, the nonvolatile components are taken out of the vaporization chamber as well, resulting in a mass discrimination according to volatility. Different methods of syringe manipulation have been studied (e.g., filled needle, cold needle, hot needle, solvent flush, air flush, sandwich method) in combination with fast or slow injection. These different procedures are not discussed here, but rather the best method — "manual hot-needle, fast-sample introduction" — is described. This method guarantees minimal syringe discrimination, although complete avoidance is impossible when dealing with solutes having a large volatility difference. In the hot-needle method, the sample is taken into the syringe barrel, for example up to the 5-μL mark when a 10-μL syringe is used, without
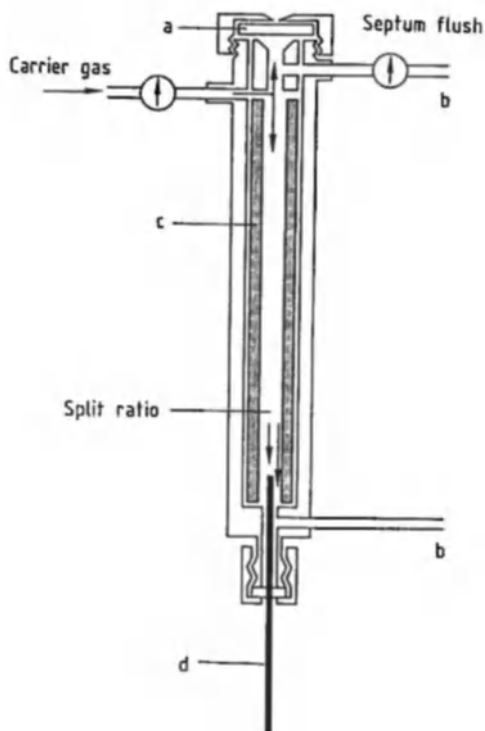
Figure 18. Schematic of a split injector
a) Septum; b) Needle valve; c) Vaporization chamber;
d) Capillary column

Figure 19. Effect of needle dwell time on sample discrimination

leaving an air plug between sample and plunger. After insertion into the injection zone, the needle is allowed to heat up for 3 – 5 s. This time is sufficient for the needle to be heated to the injector temperature. The sample is then injected by rapidly pushing the plunger down (fast injection), after which the needle is withdrawn from the injector within 1 s. Syringe discrimination is related to the warming up of the needle in the vaporization chamber. Systems can be worked out in which syringe discrimination is reduced or avoided. The best technique is very fast injection. A prerequisite here is *automated cold-needle* sample introduction since all steps of the injection sequence are identical for each injection. The objective of very fast injection is to get the needle into the injection port, inject the sample, and withdraw the needle so quickly that it has no time to warm up, in this way avoiding selective sample evaporation. Fast injection, moreover, implies that the delivered volume of sample equals the preset volume. The effect of the needle *dwell time* in the injection port on sample discrimination has been

studied [40]. Dwell time is defined as the interval between the needle tip piercing the septum on the way in and reaching the same point on the way out. Figure 19 shows a plot of $C_X : C_{20}$ area ratios ($X = 10 - 40$) as a function of C-number for different dwell times, with hexane as the solvent. Although these data have been obtained for direct injection, they are also valid for split injection. A dwell time of 500 ms or less shows no noticeable fractionation. Modern autosamplers inject in 100 ms. The previous discussion indicates that obtaining quantitative data by using a split injector can be problematical, but not impossible. For samples in which the solutes do not differ too much in volatility, precision and accuracy can be acceptable if the injection technique has been optimized. Some summarizing guidelines are given. In quantitative analysis, the standard addition or the internal standard method is preferred for manual injections. External standardization in which absolute peak areas are compared can be applied only with fast automated injection. Reproducibility will be enhanced by not varying the injected volume, which typically should be 0.5 – 2 µL. Injector temperature should be adapted to the problem at hand (i.e., close to the boiling point of the last eluting compound). Excessively high injector temperatures should be avoided. The hot-needle, fast-injection method is preferred for manual injection. The use of highly volatile solvents should be avoided whenever possible. If an open liner does not do the job, lose packing with deactivated glass wool or glass beads can provide a solution. However, adsorption and decomposition can occur. One of the main problems associated with split injection is syringe handling. The application of

an autosampling system can overcome this difficulty, enhancing both precision and repeatability. For several applications, however, split injection will not give high quantitative accuracy.

11.5.2.2.2. Splitless Injection

In splitless injection a conventional split injector is operated in a nonsplitting mode by closing the split valve during injection. The sample is flash vaporized in the vaporizing chamber from which the sample vapors are carried to the column by the flow of the mobile phase. Since this transfer takes several tens of seconds, broad initial bandwidths would be anticipated. However, through optimal use of focusing effects, such as solvent, thermal, and stationary-phase focusing, initial peak broadening can be suppressed. The main feature of splitless injection lies in the fact that the total injected sample is introduced in the column, thus resulting in much higher sensitivity compared to split injection. For a long time, splitless injection was the only sample introduction technique in capillary GC for trace analysis. The pneumatic configuration is similar to the classical split injector (see Fig. 18). The septum is purged continuously (flow rate 2 mL/min) to maintain the system free of contamination, while a flow of 20–50 mL/min passes the split outlet. Prior to injection, a solenoid valve is activated, so that the splitline is closed off while the septum purge is maintained. After waiting a sufficient time for the solvent–solute vapor to be transferred onto the column (i.e., 30–80 s), the solenoid valve is deactivated. Residual vapors in the vaporizing chamber are vented to waste via the split line. For this reason, in splitless injection the split line is often referred to as the purge line. The time interval between the point of injection and the activation of the split (purge) line is a function of the characteristics of solvent and solutes, the volume of the vaporizing chamber, the sample size, the injection speed, and the carrier gas velocity. The sample transfer from the vaporizing chamber into the column is a slow process. Solvent vapors especially tend to remain in the inlet for a long time. Purging the insert after injection removes the last traces of vapor from the vaporizing chamber. An important aspect in the design of a splitless injector is the dimension of the vaporizing chamber. Long and narrow inserts are preferred to obtain minimal sample dilution. Internal volumes vary between 0.5 and 1 mL. The column is installed

0.5 cm in the insert, and syringes with long needles are used, creating a distance of 1–1.5 cm between the needle tip and the column inlet. Overfilling of the vaporizing chamber is suppressed, and fast sample transfer is achieved. If sample volumes larger than 2 μL are injected, a widebore liner should be used. Because of the relatively long residence time of the solutes in the vaporizing chamber, lower injection temperatures can be used compared to split injection. This can be helpful in minimizing sample degradation. For the same reason, the use of packed inserts should be avoided in splitless injection. An illustration of the effect of glass or fused silica wool in the liner is presented in Figure 20. The organochloropesticide endrin (peak 1) tends to decompose into endrin aldehyde (peak 2) and endrin ketone (peak 3) when active sites are present in the liner. The first chromatogram shows that with an empty liner, 92 % of the endrin is detected, compared to cool on-column injection, which provides 100 % elution. Filling the liner with a 1-cm plug of glass wool or fused silica wool results in the elution of only 83 % and 48 % endrin, respectively. Taking into account the fact that endrin is not the most thermolabile compound, what will happen if polar compounds (i.e., carbamates, barbiturates, etc.) are injected with filled liners can easily be imagined. Recently, electronic pressure control was introduced to reduce the residence time of the solutes in the liner. With EPC, the carrier gas pressure or flow can be programmed. During injection a very high flow rate is applied, which speeds up the transfer liner–column. After injection, the flow is reset to the optimal value for the column in use [41]. Syringe discrimination, which is one of the largest sources of error when applying split injection, also occurs for splitless injection because both are vaporizing inlets. The performance of a splitless injector for a particular application depends on the optimization of experimental variables, the most important of which are sample size, injection speed, purge time, injection temperature, initial column temperature, carrier gas selection, and flow rate. General guidelines cannot be advanced, and some of the variables can be only tuned by trial and error. Whatever the application, however, refocusing of the solutes in the inlet section of the capillary column is necessary. In splitless injection, the initial bandwidths are broadened by two mechanisms: band broadening in time and band broadening in space. *Band broadening in time* is caused by the slow transfer from the vaporizing chamber to the column inlet section, which takes several tens of sec-
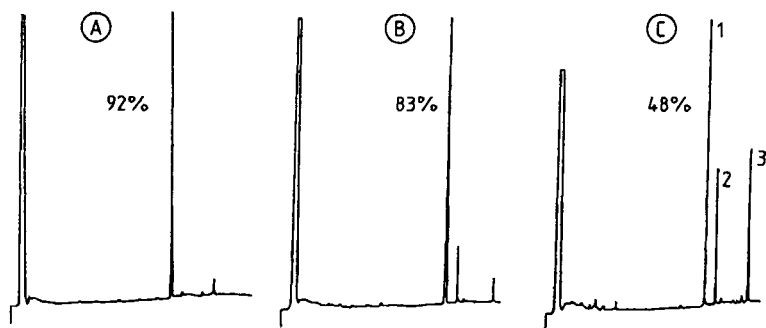
Figure 20. Decomposition of endrin on different splitless liners
A) Empty liner; B) 1-cm plug of glass wool; C) 1-cm plug of fused silica wool

onds. Upon their arrival in the column, the solutes spread over a certain column length, mainly by flooding of the sample liquid, thus causing *band broadening in space*. The fundamental difference between band broadening in time and band broadening in space is that, in the first case, solutes are spread equally with respect to gas chromatographic retention time, whereas in the second case, solutes are spread equally with respect to column length. Both phenomena cause distorted elution profiles if the solutes are not refocused before starting the chromatographic distribution process.

Solvent Focusing, Thermal Focusing, and Stationary-Phase Focusing. Splitless injection most often is combined with column temperature programming, starting with an oven temperature 25 – 30 °C below the boiling point of the solvent. Upon condensation of solvent and solutes, the droplet formed becomes too thick to be stable. The carrier gas pushes the plug further into the column, creating a "flooded zone" (Fig. 21). The solutes now are spread over the full length of the flooded zone, thus creating a solute bandwidth that equals the length of the flooded zone. For 1-µL injections, the length of the flooded zone roughly is 20 cm, provided the stationary phase is perfectly wettable by the solvent [i.e., isooctane solutions on apolar methylsilicone phases, ethyl acetate solutions on poly(ethylene glycol) phases]. On conventional capillary columns, 25 – 30 m in length, *i.d.* 0.32 mm, band broadening will hardly be observed for sample volumes of 1 µL. Only for sample sizes larger than 2 – 3 µL will peak deformation be noted. On the other hand, for polar solvents on apolar stationary phases, the length of the flooded zone can be as large as 1 m, resulting in distorted peaks. Stationary phase focusing

then becomes necessary. After creation of the flooded zone, the temperature is increased and the solvent starts to evaporate from the rear of the flooded zone (Fig. 22). Highly volatile solutes evaporate as well and the last traces of solvent act as a barrier since the residual solvent behaves as a thick stationary-phase film for those solutes. High volatiles are thus focused by the *solvent effect*. The low-volatility solutes are cold trapped since the column temperature is still too low to evaporate them. They are *thermally focused* over the length of the flooded zone. When the temperature is high enough to evaporate the low-volatility solutes, they start the chromatographic process at different points. If the flooded zone is homogeneous and short, peak distortion is hardly observed, but when the flooded zone is inhomogeneous and long, which is the case when polar solvents are injected on apolar capillary columns, broad distorted peaks are detected. The only reason for this is that the chromatographic process starts with the solutes distributed over the length of the flooded zone. This can be avoided by removing the stationary phase over the length of the flooded zone or, in other words, by creating a retention gap. Band broadening in space can thus be suppressed by *stationary-phase focusing* via a retention gap (Fig. 23). A retention gap is a definite length of the separation system that is uncoated. All solutes that are spread over the flooded zone are carried with no retention onto the stationary phase where they are retained. In practice, a retention gap is a separate piece of deactivated fused silica, connected to the analytical column through a coupling device such as a press fit connector. The length of the retention gap is a function of the length of the flooded zone, and depends on the sample volume and the nature of the solvent in use. Typical
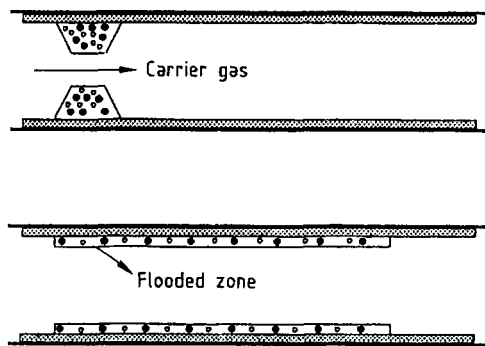
Figure 21. Creation of the flooded zone
o = High-volatile solute ; • = Low-volatile solute
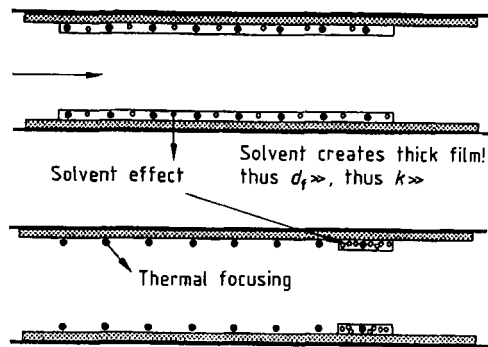
Figure 22. The solvent and thermal focusing effect
o = High-volatile solute ; • = Low-volatile solute

Figure 23. Stationary-phase focusing and the principle of a retention gap
• = Low-volatile solute

lengths are 0.5–1 m for injections of 1 to 2 µL. Longer lengths, however, do not harm the separation, and pieces of 3 m are standard and commercially available. Band broadening in space and the use of a retention gap are also required for cool on-column injection (see Section 11.5.2.2.3). In recent years, splitless injection has been overshadowed by cool on-column injection. Without any doubt, the most accurate and precise data are now provided with on-column injection. Notwithstanding this, splitless injection is still used for many routine determinations (i.e., in environmental analysis, pesticide monitoring, drug screening, etc.). In these fields, sample preparation is of primary concern, and cleaning up a sample to such an extent that it is accessible to on-column injection is not always possible or economically justifiable. Traces of nonvolatile or high-boiling components often remain in the sample. Splitless injection is an easy solution to such problems. The "dirty" components remain mainly in the vaporizing inlet,

which is easily accessible for cleaning. An illustration of the determination of polychlorinated biphenyls (PCBs) in waste oil via splitless injection is shown in Figure 24. 5 µL of an isooctane solution of the PCBs extracted from waste oil, was injected in the splitless mode on an apolar column equipped with a retention gap of 5 m. A pressure pulse was applied to minimize the residence time of the solutes in the liner [42]. Without a retention gap the flooded zone would have been too large to guarantee good peak shapes. When properly performed, reproducibility in splitless injection can be as high as 1–2% relative standard deviation (RSD). Standard addition or internal standard quantitation is the method of choice for manual injection. The external standard method can be adapted for automated splitless injection. Many capillary GC applications need the best attainable precision and accuracy. Cool on-column injection (see Section 11.5.2.2.4) then is the method of choice.

11.5.2.2.3. Direct Injection

The nomenclature of different injection techiques has been created as new developments occur. This often causes confusion with regard to the interpretation of some terms, which do not always stand for what they really mean. Direct injection is an example of such an expression, often confused or even identified with on-column injection. The understanding of the terminology given in this article, based on the chronological order in which techniques were developed, is as
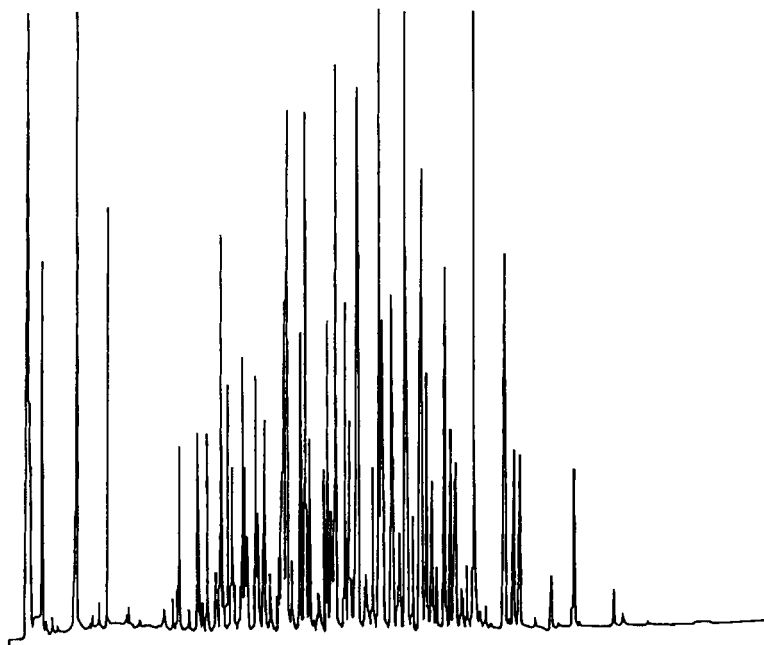
Figure 24. Optimized splitless injection of a PCB sample
Column: 25 m L×0.25 mm *i.d.*; 0.25 μm *d*$_f$ MeSi; retention gap: 5 m L×0.25 mm *i.d.*; temperature: 150-280 °C at 5 °C/min; splitless injection with pressure pulse; carrier gas hydrogen; FID

follows: *Direct injection* is a flash vaporizing injection method. The inlet system is heated separately and independently from the column oven, and evaporation occurs in the inlet. This inlet can be a glass liner (out-column evaporation) or a part of the column (in-column evaporation). Different direct injection devices are shown in Figure 25. *On-column injection* is a "cold" injection technique. The sample is injected as a liquid, directly on the column. During injection, the injection zone is cooled to avoid needle discrimination. Injection peak broadening is caused only by band broadening in space. The solutes are focused in the inlet section of the column, from where evaporation gradually starts as the oven temperature is raised. Whereas on-column injection can be applied for all conventional capillary columns (*i.d.* 0.25, 0.32, 0.53 mm), direct injection is restricted to widebore or megabore columns.

Direct injection into widebore columns was described in 1959 [43]. At that time, column efficiency was all important. Through the introduction of widebore fused-silica capillary columns with *i.d.*'s of 0.53 mm, the interest in direct sampling has been renewed. Injection on a 0.53-mm *i.d.*

FSOT column, operated in the high-resolution (flow rates 1 – 3 mL/min) or the low-resolution (flow rates 8 – 12 mL/min) mode, as a packed column alternative, is said to be as easy as installing a simple glass liner in the packed column inlet and connecting the column to it. This view is slightly oversimplified since the direct injection technique is based on the evaporation of the sample at elevated injection temperature, all previous statements about syringe discrimination and band broadening effects remain valid. Direct injection must definitely be optimized according to the problem at hand. Some experience-based guidelines on direct injection into widebore columns are advanced. However, when widebore capillary columns are used in the high-resolution mode, they should be connected to conventional capillary injectors (split, splitless, on-column, programmed temperature vaporization). For direct injection on widebore columns operated at high flow rates, the sample size should be kept as small as possible, and the sample should preferably be injected with an autosampler. For manual injections, the fast hot-needle injection technique is preferred to minimize syringe discrimination. To avoid sample
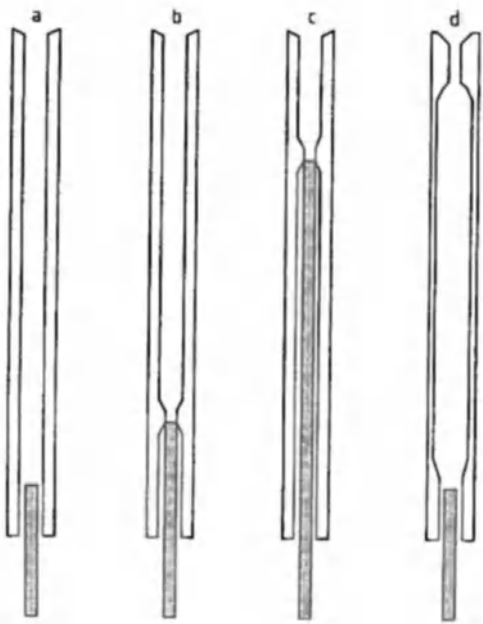
Figure 25. Different injection devices for direct injection
a) Open liner; b) Open liner with conical contraction;
c) Direct on-column liner; d) Open liner with expansion volume

losses due to overfilling the insert with sample vapors, the point of injection should be close to the column entrance. Long and narrow inserts are preferred to obtain minimum sample dilution (band broadening in time). Band broadening should be compensated by refocusing the solutes in the inlet section of the column. The principle of the solvent effect and of thermal focusing have been discussed in Section 11.5.2.2.2. Injections of microliter volumes at oven temperatures (far) above the boiling point of the solvent should be avoided, unless the sample is known to contain heavy compounds only, which will be refocused by cold trapping at the chosen column temperature. Figure 26 shows the analysis of the oxygenated fraction of hop oil dissolved in dichloromethane on a narrowbore column with split injection (Fig. 26 A) and on a widebore column with direct injection (Fig. 26 B) [44]. Both columns (25 m×0.25 mm and 50 m× 0.50 mm) offer the same efficiency. Peak broadening did not occur on the widebore column with direct injection, for which only 0.2 µL was injected.

11.5.2.2.4. Cool On-Column Injection

Syringe on-column injection with small-diameter capillary columns was first described in 1978 by the GROBs [45], [46]. Special on-column devices (a micro and a macro version) were introduced in 1977 by SCHOMBURG et al. [47], but the technical requirements were stringent and the devices lacked practical flexibility. Since this pioneering work, all instrument manufacturers have introduced cool on-column injectors. As an example, the cool on-column injector developed by Hewlett–Packard is shown in Figure 27. It can be used for manual and automated injection. The injector has a low thermal mass, which facilitates cooling. A key part of the injector for manual operation is the duck bill valve (b). The duck bill valve, made out of a soft elastomer, is a passive element in that it has no moving parts. It consists simply of two surfaces pressed together by the column inlet pressure. During injection, the syringe needle slips between the two surfaces to maintain the seal. For automated cool on-column injection, the system is modified with a disk septum (h). Today, cool on-column injection systems can be operated in the constant-pressure, constant-flow, or pressure-programmed mode. These features allow reduction of analysis time and elution of high molar mass compounds. The advent of cool on-column injection has extended the range of applicability of capillary GC to many classes of compounds that heretofore were difficult if not impossible to analyze. The technique of introducing the sample directly on the column without prior vaporization offers many advantages, such as elimination of sample discrimination and sample alteration, high analytical precision, and data operator independence. As with all other sampling techniques, the operational parameters strongly affect the chromatographic data. Parameters such as initial column temperature, solvent nature, injection rate, injected volume, and boiling point range of the sample components are interrelated. Once again, all aspects cannot be treated in detail in this article. *On-column Injection in Capillary GC* by K. GROB contains 590 pages [37]! The characteristics of the technique are highlighted here, and some important aspects for daily practical use are discussed. For sample sizes of the order of 0.5–2 µL, injection should be performed as quickly as possible, with the column oven temperature below or equal to the boiling point of the solvent. The oven temperature in on-column injection may be higher than in splitless injection. In
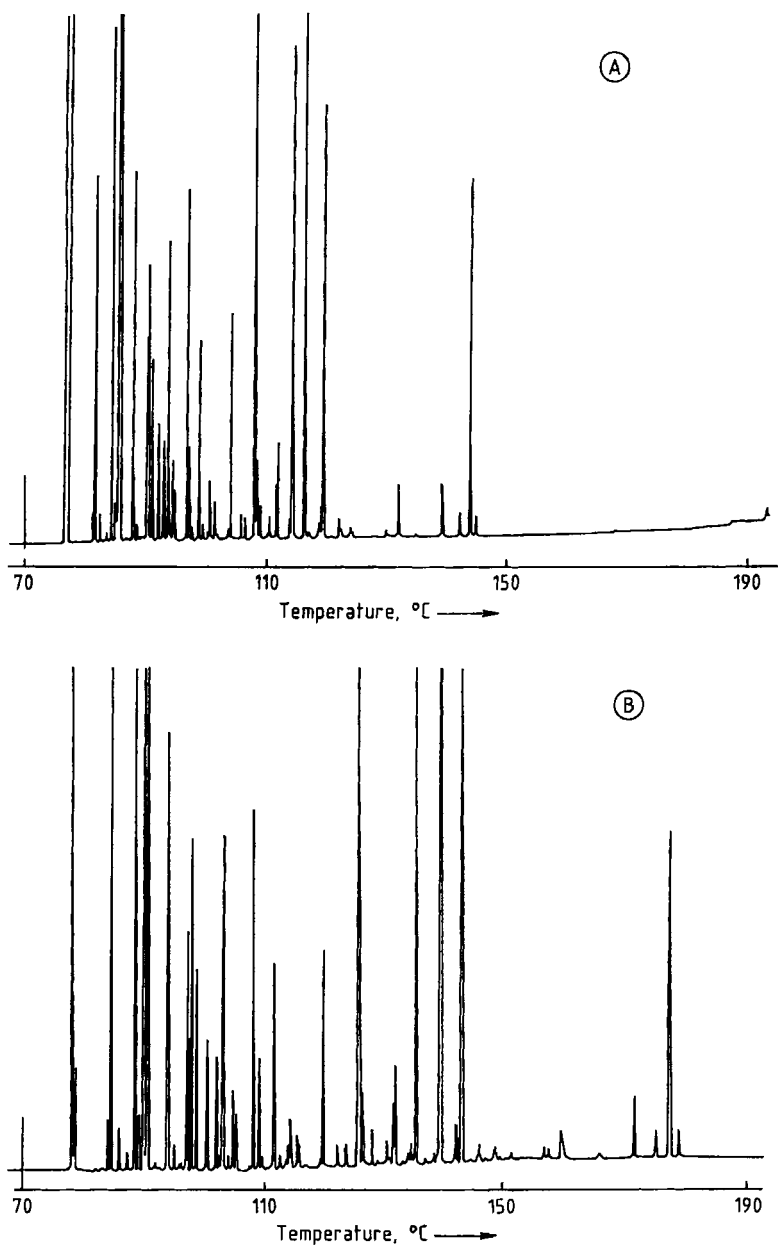
Figure 26. Analysis of the oxygenated fraction of hop essential oil on a conventional and a widebore capillary column
A) Column: 25 m $L\times$0.25 mm *i.d.* : 0.25 μm d_f high-M_r poly(ethylene glycol); temperature: 70 – 190 °C at 2 °C/min; split injection 1/20; carrier gas hydrogen; FID; B) Column: 50 m $L\times$0.5 mm *i.d.* : 0.25 μm d_f high-M_r poly(ethylene glycol); temperature: 70 – 190 °C at 2 °C/min; direct injection; carrier gas hydrogen; FID

splitless injection, the solvent has to recondense; whereas in on-column injection, the liquid is introduced directly into the column. As in splitless injection, the length of the flooded zone must be reduced because solutes are distributed over its entire distance. The width of the initial band equals the length of the flooded zone, i.e., band broadening in space (band broadening in time does not occur in on-column injection). The length of the flooded zone is a function of the column di-
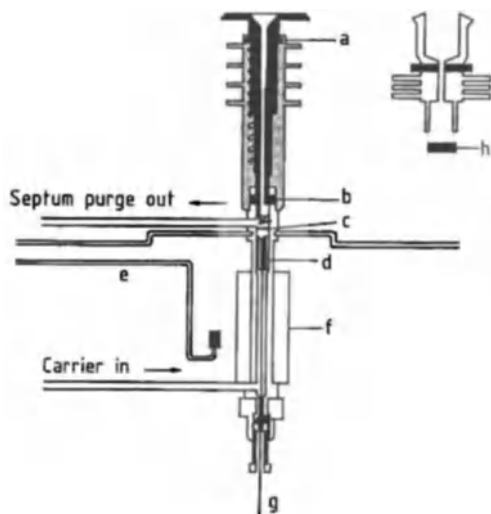
Figure 27. Schematic of a cool on-column injector
a) Cool tower, needle guide; b) Duck bill valve (isolation valve); c) Spring; d) Insert; e) Cryogenic cooling (optional); f) Heater block; g) Column; h) Septum

ameter, the volume injected, the actual inlet temperature, the stationary-phase thickness, and, most important, the affinity of the solvent for the stationary phase. For perfect wettability (i.e., apolar solvents on apolar silicone phases), the length of the flooded zone is about 20 cm per microliter injected. This length easily increases by a factor of five or ten if the wettability is poor, such as for methanol injections on apolar silicone phases. If the solutes are not refocused, band broadening in space makes qualitative and quantitative analysis impossible if large sample volumes are injected or if the solvent does not properly wet the stationary phase. The latter phenomenon is illustrated in Figure 28 with the analysis of a fraction collected from a step-elution HPLC analysis of the essential oil of *Valeriana celtica* L., with hexane and methanol. The fraction of the polar compounds was contained in 1 mL of hexane and 6 mL of methanol in a two-phase system. Both phases were injected on-column (1 µL) onto a 20 m×0.3 mm *i.d.* methylsilicone column with a film thickness of 0.3 µm. Figure 28 A shows the chromatogram of the hexane layer and Figure 28 B that of the methanol layer. Methanol causes peak splitting, whereas hexane does not [48]. The poor wettability (polar – apolar) in the case of methanol results in a long, nonhomogeneous flooded zone. This artifact can be bypassed via a retention gap

(see Section 11.5.2.2.2). On-column injection provides the most accurate and precise results because syringe discrimination, which is one of the main sources of error in the quantitative analysis of samples covering a wide range of molar masses, is completely avoided. Moreover, inlet-related discrimination does not occur either, because the liquid is introduced directly onto the column. The automated on-column analysis of free fatty acids from acetic to decanoic acid is shown in Figure 29. Relative standard deviations on absolute peak areas are < 1% and on relative peak areas < 0.4%, for 20 injections ($n = 20$). The analysis of polymer additives covering a broad range of functionalities and molecular masses is shown in Figure 30 [49]. To elute Irganox 1010 (M_r 1176) in a reasonable time, cool on-column injection was performed in the constant-flow mode. Both the external standard and the internal standard method can be applied for quantitation. Current literature data contain overwhelming evidence for the superior features of cool on-column injection in quantitative work. A second important feature of cold on-column injection is the elimination of sample alterations. Thermally labile components are not exposed to thermal stress: they begin the chromatographic process at relatively low temperature. Decomposition and rearrangement reactions are nearly completely eliminated. This feature allows analyses that heretofore were impossible with gas chromatography. Figure 31 shows the analysis of some retinoids [50]. Quantitative analysis in plasma samples was possible only with cool on-colum injection. The only disadvantage of cool on-column injection is that since the sample is introduced directly onto the column, relatively "clean" samples have to be prepared. Nonvolatile and less volatile material collects at the head of the column. The presence of this material causes a loss of separation efficiency or introduces adsorptive sites in the column inlet. Proper sample cleanup is thus very important. The use of a precolumn of 3 m, which at the same time acts as a retention gap, is therefore recommended. When the chromatographic performance decreases, 5 to 10 cm of the inlet section of the pre-column is cut off, restoring the original performance of the system.

11.5.2.2.5. Programmed-Temperature Injection

In 1979, VOGT and coworkers [51] described an injection method for the introduction of sample
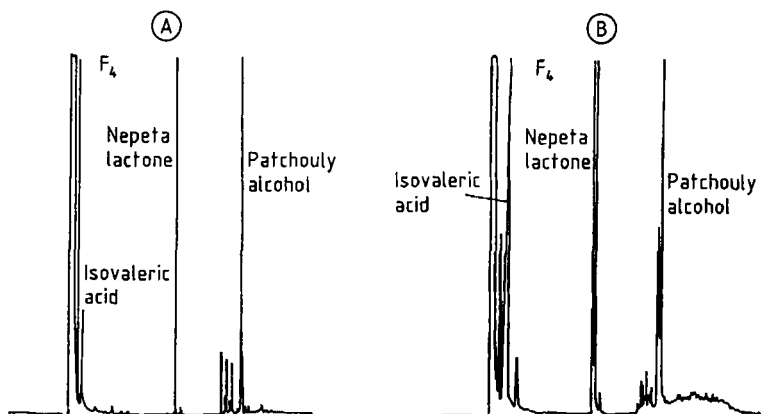
Figure 28. Illustration of peak splitting in cool on-column injection
A) Hexane as solvent; B) Methanol as solvent
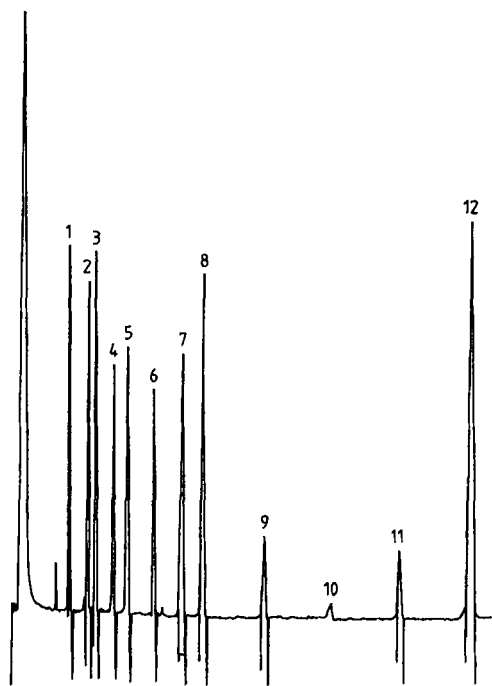
Figure 29. Analysis of free fatty acids on a widebore capillary column
Column: 10 m L×0.53 mm i.d.; 0.25 µm d_f FFAP; temperature: 60 - 100 °C balistically, to 140 °C at 2 °C/min; cool on-column injection; carrier gas hydrogen 4 mL/min; FID-Peaks: (1) acetic acid, (2) propionic acid, (3) isobutyric acid, (4) butyric acid, (5) isovaleric acid, (6) valeric acid, (7) isocaproic acid, (8) caproic acid, (9) heptanoic acid, (10) octanoic acid, (11) nonanoic acid, (12) decanoic acid

volumes up to 250 µL. The sample was introduced slowly into a cold glass insert packed with glass wool. The low-boiling solvent was evaporated continuously and vented through the split exit. The solutes, which remained in the inlet, were transferred to the capillary column by rapidly heating (30 °C/min) the inlet. During this transfer, the split line was closed (splitless injection). Based on this idea, the groups of SCHOMBURG [52] and POY [53] almost simultaneously developed programmed-temperature injection. Their aim, however, was not the injection of large sample volumes but rather the elimination of syringe needle discrimination, which at that time was the subject of interest. Different programmed-temperature vaporization (PTV) injection devices are now commercially available, offering a broad range of possibilities (i.e., hot or cold split injection, hot or cold splitless injection, cool on-column injection, direct injection, etc.). Through these possibilities, together with features such as the injection of large sample volumes, concentration by multiple injection, solvent venting, etc., PTV injection is claimed to be the most universal sample introduction system. It is, however, not the best choice for all applications. The upper part of a PTV inlet resembles a classical split—splitless device, including carrier gas inlet and septum flush. The sample is injected into a glass liner with a low thermal mass while the injector body is cold. After withdrawal of the syringe needle, the vaporizing tube is heated rapidly to volatilize the solvent and the solutes. The heat can be provided electrically or by means of preheated compressed air. Depend-
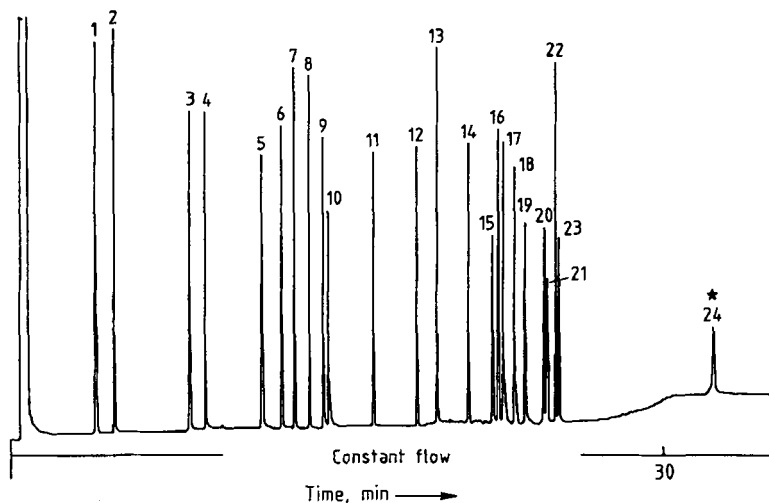
Figure 30. Analysis of polymer additives by high-temperature capillary GC
Column: 25 m L×0.32 mm i.d.; 0.17 μm d_f MeSi; temperature: 80 – 380 °C at 10 °C/min; cool on-column injection; carrier gas hydrogen 50 kPa constant flow; FID
Peaks: (1) butyl-4-methoxyphenol, (2) diethyl phthalate, (3) dibutyl phthalate, (4) Tinuvin P, (5) triphenyl phosphate, (6) dicyclohexyl phthalate, (7) dioctyl phthalate, (8) Tinuvin 327, (9) benzophenone UV, (10) erucamide, (11) Tinuvin 770, (12) Irgaphos 168, (13) Irganox 1076, (14) Tinuvin 144, (15) Irganox 245, (16) Irganox 259, (17) Irganox 1035, (18) Irganox 565, (19) crodamide, (20) Irganox 1098, (21) Irganox 3114, (22) Irganox 1330, (23) Irganox PS 802, (24) Irganox 1010

ing on its construction, heating of the device can be performed ballistically or linearly at selected rates (for example 2 – 12 °C/s) and even to consecutive temperature levels. These facilities allow optimization of conditions for the analysis of thermally labile compounds, operating in the solvent vent mode when working with specific detection such as electron capture detection (ECD) or mass spectrometry, concentration through multiple injection, etc. Split or splitless injection can be achieved through regulation of the split valve. During or after the chromatographic run, the vaporizing chamber is cooled by air or carbon dioxide, in preparation for the next injection. In an interesting mode of operation (i.e., *solvent elimination injection*), the sample is introduced into the cold injector with the split valve open. Conditions are selected such that only the solvent evaporates while the solutes of interest remain in the liner. The applicability of this technique is, therefore, restricted to the analysis of relatively high-boiling compounds. In this way, large sample volumes can be introduced to enrich traces. The principle has been applied to detect dioxins and furans down to the 100-fg/μL level with capillary GC – low-resolution mass spectroscopy [54].

11.5.3. Selective Inlets

A number of selective inlets are available in GC. Best known are static and dynamic head space (HS) injection (selectivity based on vapor pressure), purge and trap (PT) samplers (selectivity based on solubility of the solutes in water), thermal desorption (TD) units (selectivity based on characteristics of the adsorbent), and pyrolysis units (selectivity based on the pyrolysis or cracking temperature). These inlets are discussed in Sections 11.5.3.1 – 11.5.3.4. More sophisticated systems, which can also be considered as sample introduction systems, such as solid-phase extraction (SPE; selectivity based on the polarity of both adsorbent and desorbing liquid), supercritical fluid extraction (SFE; selectivity based on the solubility in the supercritical medium, the variables being density, temperature, pressure, and polarity), multidimensional capillary GC (MDCGC), the combinations of capillary GC with liquid chromatography (HPLC – CGC) or with supercritical fluid chromatography (SFC – CGC) allowing preseparation according to molar mass (size-exclusion chromatography, SEC), functionality (normal phase LC), hydrophobicity (reversed phase LC), or solubility (SFC) are treated in Chapter 11.8. In addition, *on-line sample preparation units* became

24 min
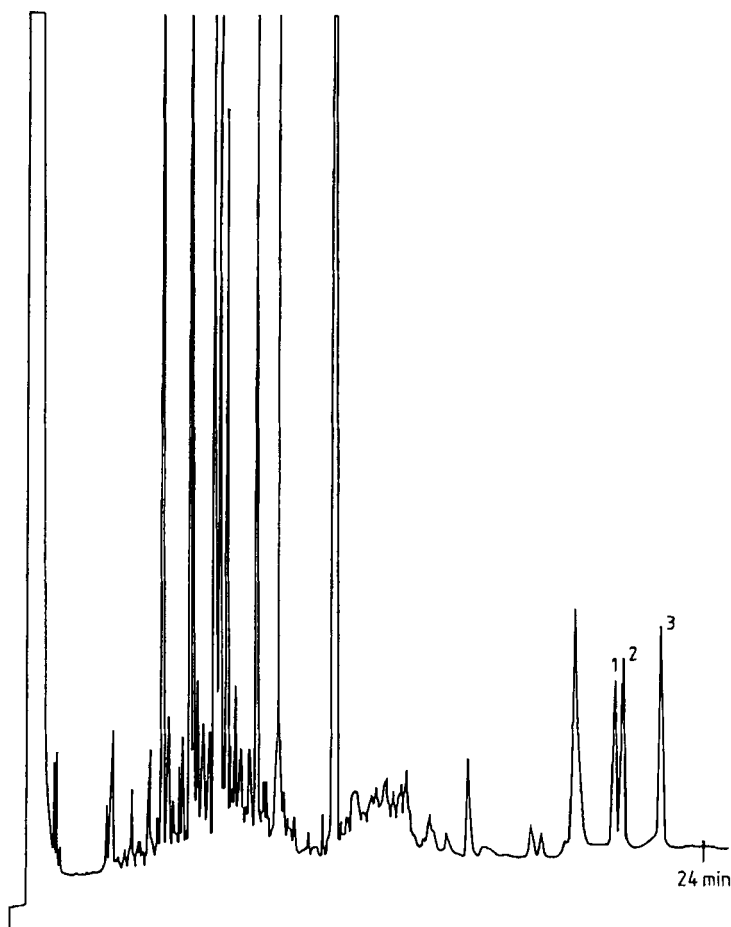
Figure 31. Illustration of the application of cool on-column injection to the analysis of thermolabile compounds
Column: 12.5 m L×0.25 mm i.d.; 0.25 µm d_f CPSi; temperature: 60–190 °C at 17 °C/min, to 220 °C at 6 °C/min and to 235 °C at 1.5 °C/min; cool on-column injection; carrier gas hydrogen; FIDPeaks: (1) trans-acitretin methyl ester, (2) cis-acitretin methyl ester, (3) cis-acitretin propyl ester

available in 1993, providing functions such as dispensing (addition of reagents or internal standards), diluting (concentration adjustments and multipoint calibration standards), heating (derivatization, solubilization), evaporating (solvent exchange, concentration), solid-phase extraction, and filtering.

11.5.3.1. Static and Dynamic Head Space Analysis

Head space analysis is used to analyze volatiles in samples for which the matrix is of no interest (i.e., water, soil, polymers, etc.). Typical applications include monitoring of volatiles in soil and water (environmental); determination of

monomers in polymers, aromas in food and beverages, residual solvents in drugs (pharmaceuticals), etc. Different head space autosamplers are commercially available, and they are based on the principle of static or dynamic head space. The difference between these modes is depicted in Figure 32. In *static head space* (Fig. 32 A), the sample (soil, water, solid) is transferred in a head space vial that is sealed and placed in a thermostat to drive the desirable components into the head space for sampling. Via a gas-tight syringe (a) or a sample loop of a gas sampling valve, an aliquot of the vapor phase is introduced in the GC system, which can be equipped with a packed or a capillary column. Because of the much lower sample loadability of a capillary column, split injection is
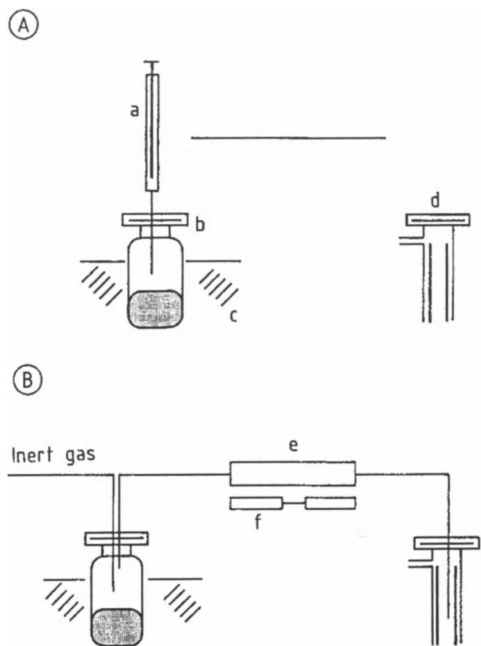
Figure 32. Principle of static (A) and dynamic head space (B)

a) Syringe; b) Head-space bottle; c) Thermostat; d) Injector; e) Cold trap; f) Solid-phase extraction

usually applied in this case. Static head space implies that the sample is taken from one phase equilibrium. To increase detectability, *dynamic head space analysis* has been developed. The phase equilibrium is continuously displaced by driving the head space out of the vial via an inert gas. The solutes are focused by cold trap (e) or an adsorbent such as Tenax and after enrichment the trap is heated and the solutes are introduced into the GC system. Whereas static head space is applied to analyze ppm amounts (1 part in 10^6), dynamic headspace allows determination of ppb amounts (1 part in 10^9) of volatiles. Sample pretreatment can often help to increase sensitivity and enhance reproducibility. The best-known methods are salting out of water samples with sodium sulfate, adding water to solid samples, and adjusting the pH to drive organic acids or bases into the head space.

11.5.3.2. Purge and Trap Systems

Purge and trap samplers have been developed for analysis of apolar and medium-polarity pollu-

tants in water samples. The commercially available systems are all based on the same principle. Helium is purged through the sample placed in a sealed system, and the volatiles are swept continuously through an adsorbent trap where they are concentrated. After a selected time (10–20 min), purging is stopped, the carrier gas is directed through the trap via a six-way valve, and the trap is heated rapidly to desorb the solutes. The method has proven to be highly accurate for environmental monitoring of ppb amounts of volatile aromatics, trihalomethanes, etc.

11.5.3.3. Thermal Desorption Units

Thermal desorption is the method of choice to monitor air pollution. A known amount of air is drawn through an adsorbent tube filled with activated carbon, Tenax, silicagel, or mixtures of these. Analytes of interest are trapped and concentrated on the adsorbent. The adsorbent tube is sealed and transported to the laboratory where it is installed in the thermal desorption unit. An example of a thermal desorption unit is shown in Figure 33. After the adsorbent tube is placed in the desorption module, the carrier gas is directed over the adsorbent, which is heated from room temperature to 250–300 °C. Since thermal desorption is a slow process, the solutes are focused by cold trapping on a fused silica trap. The cryogenic trap is then heated rapidly to transfer the sample to the column. The method works well for apolar and medium-polarity components in air. Highly polar solutes are very hard to desorb, and other sampling methods must be selected.

11.5.3.4. Pyrolysis Gas Chromatography

Pyrolysis involves the thermal cleavage (cracking) of large molecules, which are not amenable to GC, into small volatile fragments. Pyrolysis GC is an excellent technique for identifying polymers and microorganisms because the pyrolysis profiles are very typical fingerprints. Most important in this respect is the pyrolysis temperature, typically between 400 and 1000 °C. The higher the temperature, the greater is the degree of fragmentation. Several types of analytical pyrolyzers are available. The most common types are the *platinum resistively heated* and the *Curie point pyrolyzers*. Figure 34 shows an example of a resistively heated pyrolyzer. The pyroprobe (c) consists of an injector cap (a), a quartz tube (d), and a platinum wire (f). The sample is placed in the quartz tube, and
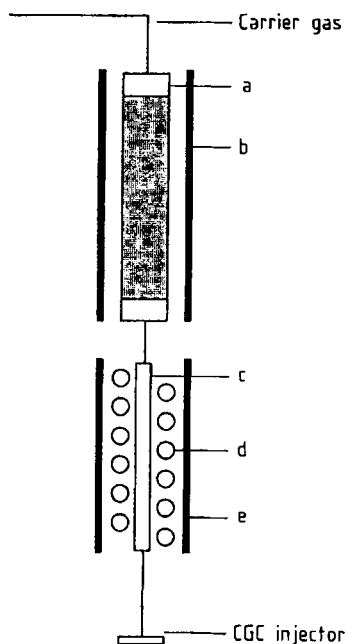
Figure 33. Schematic of a thermal desorption unit
a) Adsorbent tube; b) Heating device; c) FSOT trap;
d) Cooling; e) Heating

Figure 34. Principle of pyrolysis GC
a) Injector cap; b) Septum; c) Pyroprobe; d) Quartz tube;
e) Glass wool; f) Platinum wire; g) Inlet liner

the pyroprobe is introduced in the injection port of the GC. The carrier gas enters the system, and after equilibration, the platinum wire is heated to a preselected temperature. The material decomposes and the volatiles are analyzed.

11.6. Detectors

Detectors in gas chromatography sense the presence of analytes in the carrier gas and convert this information in an electrical signal, which is recorded. Numerous detection devices have been developed in GC [55]; here, only the currently most common detectors are dealt with. An excellent book describing detectors for capillary chromatography has been published [56].

11.6.1. Classification

GC detectors are classified mainly on the basis of response or of detector selectivity. Detectors whose response is proportional to the concentration of a sample component in the mobile phase (g/mL) are called *concentration-sensitive detec-*

tors, whereas detectors whose response is proportional to the amount of sample component reaching the detector in unit time (g/s) are called *mass-flow detectors*. *Universal detectors* respond to every component in the mobile phase; *selective detectors* respond only to a related group of substances; and *specific detectors* respond to a single sample component or to a limited number of components with similar chemical characteristics. The differentiation of selective versus specific is confusing, and therefore in this article, selective is used for both to distinguish them from universal. Another classification is destructive versus nondestructive detectors. Any detector can be coupled in series with a nondestructive detector. The most important characteristics for practical work are sensitivity, dynamic range, linear range, detector response factors, and selectivity.

Sensitivity is the signal output per unit concentration or unit mass of a substance in the mobile phase and is best expressed by the minimum detectable quantity, which is defined as the concentration or the amount of analyte for which the peak height is four times the intrinsic noise height of the detector (signal-to-noise ratio $S/N = 4$). The *dynamic range* of a detector is that range of concentration or quantity over which an incremental change produces an incremental change in detector signal. Figure 35 represents a plot used for the determination of the dynamic range of a detector. That part of the curve for which the response

Table 4. Characteristics of common GC detectors

Name	Type	Selectivity	Minimum detectability	Linear range
Flame ionization detector (FID)	upiversal[c]	[a]	10 pg C/s	10^7
Thermal conductivity detector (TCD)	universal[c]	[b]	1 ng/mL mobile phase	10^6
Electron capture detector (ECD)	selective	compounds capturing electrons, e.g., halogens	0.2 pg Cl/s	10^4
Nitrogen phosporus detector (NPD)	selective	N and P	1 pg N/s 0.5 pg P/s	10^4
Flame photometric detector (FPD)	selective	P and S	50 pg S/s	10^3
			2 pg P/s	10^4
Photoionization detector (PID)	selective	aromatics	5 pg C/s	10^7
Electrolytic conductivity detector (ELCD)	selective	halogens and S	1 pg Cl/s	10^6
			5 pg S/s	10^4
Fourier transform infrared spectroscopy (FTIR)	universal[c]	molecular vibrations	1 ng (strong absorber)	10^3
Mass spectroscopy (MS)	universal[c]	characteristic ions	1 ng full-scan mode	10^5
			1 pg ion-monitoring mode	
Atomic emission detection (AED)	universal[c]	any element	0.2 – 50 pg/s depending on element	10^4

[a] FID responds to all organic compounds that ionize in a flame.
[b] TCD responds if thermal conductivity is different from carrier gas.
[c] Spectroscopic detectors that can be operated in a universal or a selective mode.
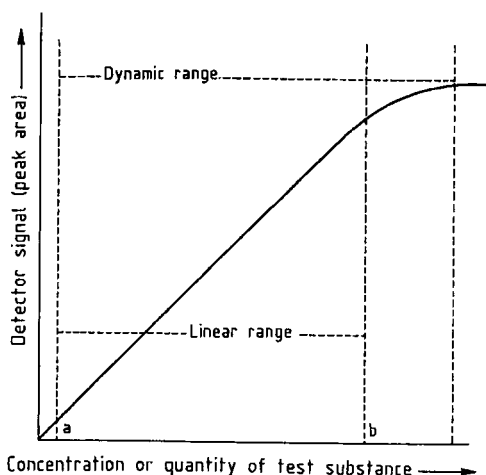
Figure 35. Dynamic range of a GC detector
a) Minimum detectability; b) Upper limit of linearity

increases proportionally with increased concentration or quantity is called the *linear range*. For quantitation, work in the linear range is preferable, it is expressed numerically as the ratio of the upper limit of linearity (b) to the minimum detectability (a), both measured for the same substance. The *detector response factor* f_i is a relative term expressing the sensitivity of a detector to a given compound i, relative to its sensitivity to a standard compound:

$$f_i = f_{st} (A_i / A_{st}) \tag{6.1}$$

where A_i and A_{st} are the respective peak areas and f_{st} is the detector response factor for the standard; f_{st} is usually assigned a value of 1.000.

Detector selectivity is the ratio of the signal obtained for a compound to which the detector has special selectivity versus the signal obtained for a compound to which the detector is not particularly sensitive. Both signals must be measured for equal amounts. A compilation of selectivity, minimum detectability, and linear range of the most common GC detectors is given in Table 4.

11.6.2. Universal Detectors

11.6.2.1. Flame Ionization Detector

The flame ionization detector (FID) is the most widely used GC detector. Its operation is simple, sensitivity is of the order of 10 pg carbon per second with a linear range of 10^7, the response is fast, and detector stability is excellent. The carrier gas is mixed with hydrogen, and this mixture is combusted in air at the exit of a flame jet. Ions are formed that are collected at an electrode producing a current that is proportional to the amount of sample compound in the flame. Analytes such as permanent gases, nitrogen oxides, sulfur oxides, carbon oxides, carbon disulfide, water, formic acid, formaldehyde, etc., do not provide a significant FID response. The flows of carrier and combustion gas should be set properly for optimal FID operation. Typical flow rates are 1:1:12

(30:30:360 mL/min), respectively, for nitrogen as carrier gas, and the combustion gases hydrogen and air. If hydrogen is applied as carrier gas in packed column GC, the hydrogen combustion gas is replaced by nitrogen. For capillary GC where the carrier gas flows are much lower (e.g., 1–3 mL/min), hydrogen combustion gas is adjusted to a total hydrogen flow of 30 mL/min. Nitrogen can be added as make-up gas, which increases sensitivity and stability slightly. Detector temperature must always be kept above 150 °C to prevent condensation of water produced in the combustion process. Flame tip and electrodes should be cleaned regularly because certain compounds can cause deposits that must be removed mechanically to prevent loss in sensitivity and detector stability. Bleeding of stationary phases and the use of derivatization agents are important sources of sensitivity loss and spikes. The flame shape for capillary columns should be optimized, and therefore flame tips with different bores are available. Small-bore jets produce the greatest signal for capillary GC.

11.6.2.2. Thermal Conductivity Detector

The thermal conductivity detector (TCD), also called the *hot-wire detector* or *katharometer*, measures the change in the thermal conductivity of the mobile-phase stream. Helium is the recommended carrier gas because of its high thermal conductivity. When an analyte is present in the carrier gas, the thermal conductivity will drop and less heat is lost to the cavity wall. A filament in the detector cell operated under constant voltage will heat up and its resistance will increase. This resistance is measured and recorded.

In the *original TCD* design the resistance is incorporated in a Wheatstone bridge circuit. The difference in thermal conductivity between the column effluent R_1 and a reference flow R_2 is recorded. This original design is still used in packed column GC (e.g., in process control). Nowadays, *single-filament*, flow-modulated thermal conductivity detectors are available for capillary work. The cell size is reduced to 3.5 µL to minimize detector band broadening, and the single filament measures the reduction in voltage required to maintain a constant filament temperature. A diagram of a commercially available TCD cell is presented in Figure 36. In Figure 36 A, the switching flow pushes the column effluent through the filament channel. On changing the switching flow (Fig. 36 B), the column effluent

passes through the empty channel while the filament channel fills with the pure reference gas. Switching between column effluent and reference gas occurs every 100 ms. An important parameter in practical work is the detector temperature. The lower the temperature, the higher is the sensitivity and the longer the lifetime of the filament. The temperature must, however, be high enough to avoid condensation of the solutes in the detector cell. Oxygen-contaminated carrier gas, column bleeding, and active components such as strong acids and bases all decrease sensitivity and reduce the lifetime of the filament. Note also that if helium is the carrier gas, and hydrogen together with some other solutes must be measured, the hydrogen peak will be negative because its thermal conductivity is higher than that of helium, while the other solutes give positive peaks.

11.6.3. Selective Detectors

11.6.3.1. Electron Capture Detector

One of the most popular selective GC detectors in use today is the electron capture detector (ECD). The ECD is an extremely powerful tool in environmental and biomedical studies. The principle is based on the fact that electronegative organic species (RX) react with low-energy thermal electrons to form negatively charged ions. The loss of electrons from the electron stream that is measured is related to the quantity of analyte in the carrier gas stream. The capturable electrons originate from ionization of the carrier gas or a make-up gas (usually nitrogen or an argon–methane mixture 95:5 or 90:10) by β particles from a radioactive source such as ^{63}Ni placed in the cell. The electron flow results in a current that is measured. Electronegative species in the carrier gas capture electrons and decrease the current. The highest sensitivity is noted for organic polyhalides such as trihalomethanes, organochloropesticides, PCBs, dioxins, and furans. Sensitivity increases in the order F < Cl < Br < I. The selectivity depends strongly on the compound's affinity for electrons and for polyhalides it is as high as 10^6 compared to hydrocarbons, and 10^4 compared to ketones, anhydrides, amines, etc. By chemical derivatization, the applicability of ECD can be expanded for trace analysis of species that do not or only weakly capture electrons. The best-known reagents are pentafluorobenzoic acid anhydride and pentafluorobenzoyl chloride for the derivati-
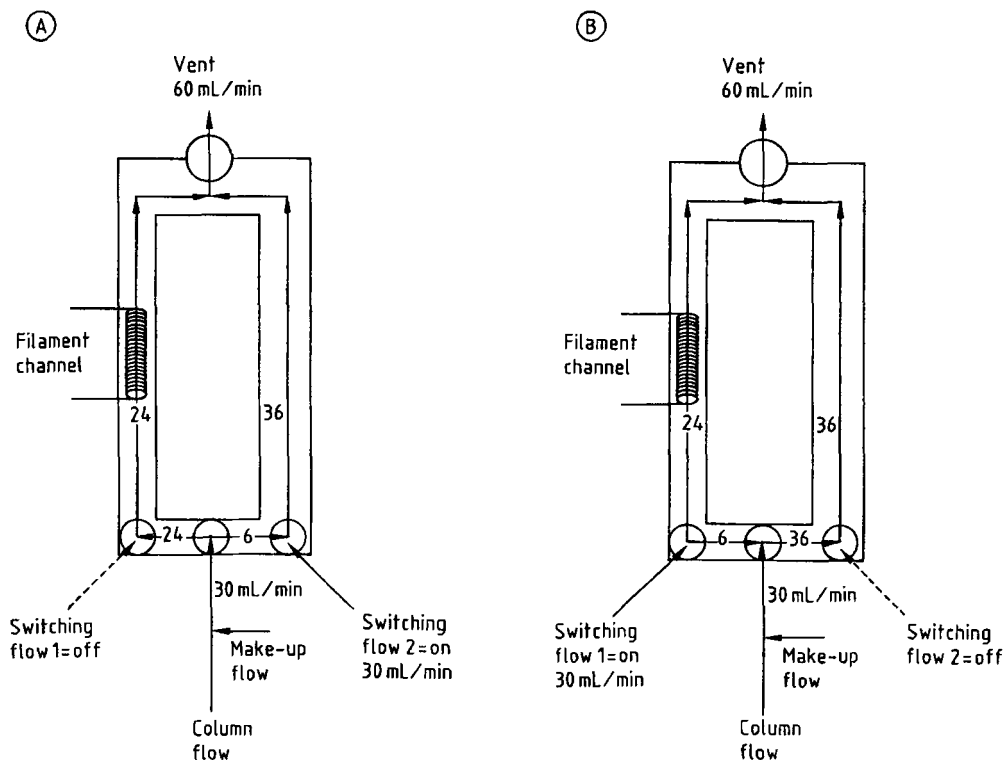
Figure 36. Diagram of a miniaturized TCD cell
A) Measurement of column effluent flow; B) Measurement of reference flow

zation of alcohols, amines, and phenols. (The fluorocompounds are chosen because of their high volatility, even though the detector is less sensitive for fluorine.) The derivatives formed are generally stable and have good chromatographic properties. As an example, the EPA method for the determination of chlorophenols in water samples includes a derivatization step with pentafluorbenzoyl chloride followed by ECD detection, in this way enhancing both sensitivity and selectivity. For proper use of the ECD detector, the carrier gas and the make-up gas must be very clean and dry. In packed column GC, nitrogen or argon – methane is used as carrier gas and no make-up gas is required at the column outlet. In capillary GC, helium or hydrogen is the preferred carrier gas because of the speed of analysis, and 25 – 30 mL/min nitrogen or argon – methane is added as make-up gas to produce the thermal electrons. The column should also be preconditioned before connection to the detector. High stability and sensitivity are guaranteed only with the detector in constant operation. Contamination is normally corrected for by baking

out the detector at 350 °C over a 24-h period. Some typical ECD chromatograms are shown in Figure 37 for the direct analysis of trihalomethanes in drinking water [57] and in Figure 38 for the determination of organochloropesticides.

11.6.3.2. Nitrogen Phosphorus Detector

The nitrogen phosphorus detector (NPD), also called the *thermionic detector* (TID) is similar in construction to the FID. The detector is made element specific for N and P because a source of an alkali salt — salts of potassium and rubidium are more commonly used — is positioned above the flame jet. By careful control of the combustion gases hydrogen and air, whose flow rates are less than those for the normal FID, the ionization of compounds containing no nitrogen or phosphorus is minimized, while the ionization of N- and P-containing compounds is increased. The exact mechanism is still unknown. Small changes in detector gas flows have a significant effect on sensitivity and selectivity. Optimal gas flows can
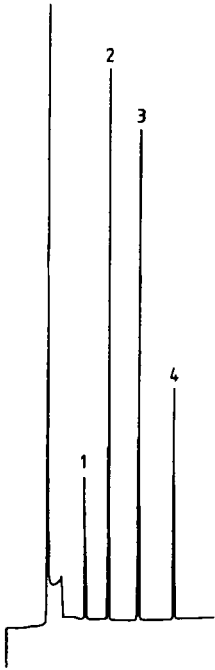
Figure 37. Direct analysis of trihalomethanes in drinking water
Column: 30 m $L \times 0.53$ mm *i.d.*; 2.65 μm d_f MeSi; retention gap 1.5 m $L \times 0.53$ mm *i.d.*; temperature: 50–150 °C at 10 °C/min; cool on-column injection; carrier gas 10 kPa hydrogen; ECD
Peaks: (1) chloroform, (2) dichlorobromoform, (3) dibromochloroform, (4) bromoform

be found by analyzing a mixture containing azobenzene, heptadecane, and malathion in the ratio 1 : 1000 : 2 (100 pg/μL, 100 ng/μL, and 200 pg/μL, respectively). Prominent responses should be obtained for azobenzene and malathion, whereas the hydrocarbon should hardly be detected. Optimal flow rates on the order of 4 – 10 mL/min for hydrogen and 60 – 100 mL/min for air are typical. For some versions of the NPD detector the use of a make-up gas such as helium increases the N or P response. The NPD detector is a valuable tool in clinical and pharmaceutical analysis, and especially in environmental monitoring of organophosphorus- and organonitrogen- containing pesticides.

11.6.3.3. Flame Photometric Detector

The flame photometric detector (FPD) uses the principle that sulfur- and phosphorus-containing compounds emit light by chemiluminescence when burned in a hydrogen-rich flame. For sulfur,
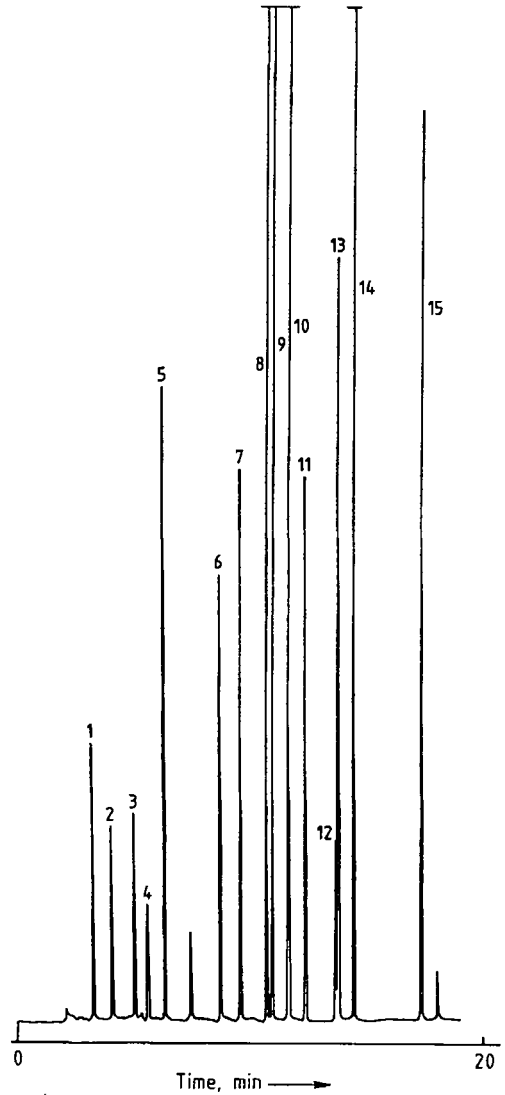
Figure 38. Analysis of organochloropesticides with ECD detection
Column: 20 m $L \times 0.32$ mm *i.d.*; 0.25 μm d_f 7 % CPSi, 7 % PhSi, 84 % MeSi (tailor-made called 1701); temperature: 180 – 240 °C at 4 °C/min; carrier gas hydrogen 8 psi ($\approx 5.6 \times 10^4$ Pa); ECD
Peaks: (1) HCB, (2) hexachlorocyclohexane, (3) lindane, (4) heptachlor, (5) aldrin, (6) heptachlorepoxide, (7) α-endosulfan, (8) 1,1-dichloroethenylidenbis(4-chlorobenzene), (9) dieldrin, (10) endrin, (11) 2,4-DDT, (12) β-endosulfan, (13)1,1'-(2,2-dichloroethylidene)bis(4-chlorobenzene), (14) 4,4'-DDT, (15) methoxychlor

for example, electronically excited diatomic sulfur (S^*_2) is formed, which upon relaxation emits light over the wavelength region 320 – 450 nm, with an emission maximum at 394 nm. An optical filter at
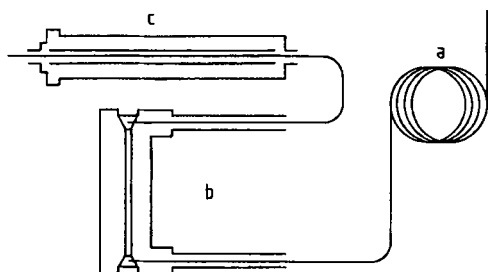
Figure 39. Diagram of a capillary GC – FTIR – MS system
a) Capillary GC; b) FTIR; c) MS

394 nm permits the light to enter a photomultiplier, and a signal is produced. For phosphorus the interference filter is 526 nm. Although widely used mainly as a sulfur-selective detector in the petroleum industry, for the analysis of flavors and fragrances, in environmental analysis, etc., the FPD detector has far from ideal characteristics as a GC detector. It has many shortcomings; the most important one is that the intensity of light emitted is not linear with concentration, but approximately proportional to the square of the sulfur atom concentration. Careful calibration is a must when quantitation is aimed at. Other negative aspects are that the flame extinguishes when high amounts of solvents are introduced (this problem is eliminated when a dual-flame burner is used), hydrocarbon quenching occurs, and self-quenching occurs at high concentrations of the heteroatom species. Moreover, optimization of the different flows is very critical for optimal sensitivity and selectivity. Therefore, for the analysis of sulfur-containing compounds the FPD detector is often replaced by the atomic emission detector, the Hall electrolytic conductivity detector, or the chemiluminescence detector based on ozone reactions, whereas for organophosphorus compounds the NPD is preferred.

11.6.3.4. Overview of Other Selective Detectors

Two of the more recently developed detectors, namely, the Hall electrolytic conductivity detector (ELCD) and the photoionization detector (PID) are recommended by the EPA for the analysis of volatile and semivolatile halogenated organic compounds and low molar mass aromatics. Chemical emission based detectors, such as the thermal energy analyzer (TEA) for its determination of

nitrosamines and the redox chemiluminescence detector (RCD) for the quantitation of oxygen- and sulfur-containing compounds, are becoming more and more important for some typical applications.

11.6.4. Detectors Allowing Selective Recognition

The selective detectors discussed in the previous sections often do not provide enough information to elucidate with 100% probability the nature of the eluting solutes. For this reason, data with selective detectors can be erratic. The future in this respect definitely belongs to the spectroscopic detectors that allow selective recognition of the separated compounds. Today, the hyphenated techniques CGC – mass spectroscopy (CGC – MS), CGC – Fourier transform infrared spectroscopy (CGC – FTIR), and CGC – atomic emission detection (CGC – AED) are the most powerful analytical techniques available. They provide sensitive and selective quantitation of target compounds and structural elucidation or identification of unknowns. The applicability and ease of use of the hyphenated techniques were greatly increased by the introduction of fused silica wall coated open tubular columns. The main reason for this is that because of the low flows of capillary columns, no special interfaces are required and columns are connected directly to the different spectrometers. The introduction of relatively inexpensive benchtop hyphenated systems has enabled many laboratories to acquire such instrumentation, which in turn has expanded their applicability ever further.

The hyphenated techniques CGC – MS, CGC – FTIR, and CGC – AED are generally used as stand-alone units. Due to the nondestructive character of FTIR, CGC – FTIR – MS units (Fig. 39) are possible and have been commercialized. The software then allows simultaneous recording of the infrared and mass spectra of the eluting compounds. In principle, CGC – FTIR – MS – AED is also possible if an open split interface is applied for the CGC – FTIR – MS combination and the split-line is directed into the AED detector. The fundamental aspects of CGC – MS, CGC – FTIR and CGC – AED are discussed in [58], [59], and [60].
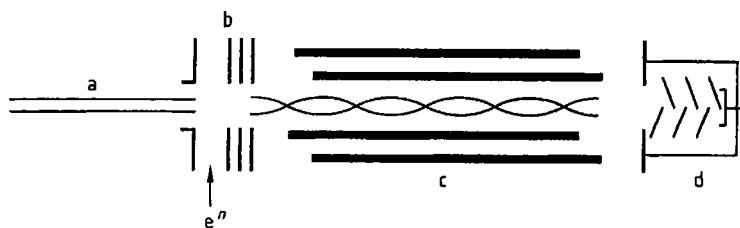
Figure 40. Schematic of a capillary GC – MS system
a) Capillary column; b) Ion source; c) Quadrupole mass filter; d) Multiplier

11.6.4.1. Mass Spectroscopy

The principle of CGC – MS is illustrated with a quadrupole mass analyzer as an example. As sample molecules exit the chromatographic column, they are introduced in the ion source of the mass spectrometer and ionized. Depending on the ionization method, molecular ions or fragment ions are formed, which are accelerated and separated from each other according to their mass-to-charge ratios (m/z) by a mass analyzer such as a quadrupole (c) (Fig. 40). Other mass analyzers are magnetic sector instruments, ion trap detectors, and time-of-flight mass spectrometers. The separated ions are then detected by a electron multiplier (d), and the resulting mass spectrum is a line spectrum of intensity of ions (y-axis) versus their m/z ratio (x-axis).

Two different ionization methods are available in CGC – MS, namely, electron ionization (EI) and chemical ionization (CI). In *electron ionization*, the molecules are bombarded with electrons of 70 eV emitted from a rhenium or tungsten filament. A molecular radical ion is formed ($M^+\cdot$) with sufficient amount of energy accumulated in its bonds to dissociate into typical fragment ions, radicals, and neutral species.

$$ABCD \longrightarrow ABCD^{+\cdot}(M^{+\cdot}) + e^-$$
$$\begin{array}{l} \longmapsto AB^{+\cdot} + CD \\ \longmapsto AB^+ + CD^\cdot \\ \longmapsto ABC^+ + D^\cdot \\ \qquad \longmapsto AB^+ + C \\ \qquad \longmapsto A^+ + BC \end{array}$$
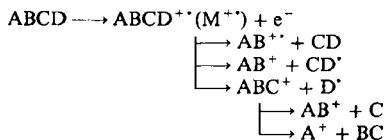
Since low ion source pressures are employed (10^{-5} torr $\approx 1.33 \times 10^{-2}$ Pa), reactions in the ion source are unimolecular, and association between molecule and fragment ions does not occur. A disadvantage of electron ionization is that energy in the molecule is so large that complete fragmentation can occur and the molecular ion is absent.

Less fragmentation is obtained with *chemical ionization*, which is a soft technique of ionizing molecules through gas-phase ion – molecule reactions. In chemical ionization, a reagent gas such as methane is introduced into the ion source and a relatively high source pressure is maintained (1 torr $\approx$ 133 Pa). The reagent gas is ionized by the electrons, and the ions formed interact with the sample molecules. Since these reactions are low in energy, abundant quasi-molecular ions, most often $M^+\cdot + 1$ are observed. Figures 41 and 42 show the electron-impact and the chemical-ionization spectra of trimethylsilylated clenbuterol. Chemical ionization is very useful to elucidate the molecular ion of unknowns and is often the ionization method of choice for quantitative mass spectroscopy via ion monitoring. After ionization, the positive ions are accelerated and separated in the mass analyzer by tuning the voltage on the quadrupole rods. An ion of unit mass will pass the rods and reach the detector only if the voltages are properly adjusted. To detect all ions, the voltage is varied in time, and the time to detect all ions is called a scan. This means that if all ions between mass 20 and 400 are to be detected in 1 s, the voltages are on the exact value for each ion to pass during 1/380 of a second. This mode, called the *full-scan mode*, is necessary to identify compounds via their fragmentation pattern. Data are acquired by continuous repetitive scanning of the column eluate. The reconstructed chromatogram of a mass spectrometer is a plot of the total ion current as a function of time or scan number. All recorded spectra or scans are stored in the computer and can be recalled for identification. Nowadays this is performed mostly by comparison with a spectral library. The information in the stored mass scans can provide, in a simple way, valuable information on specific compounds or compound classes. This is illustrated in Figure 43, which shows the recon-
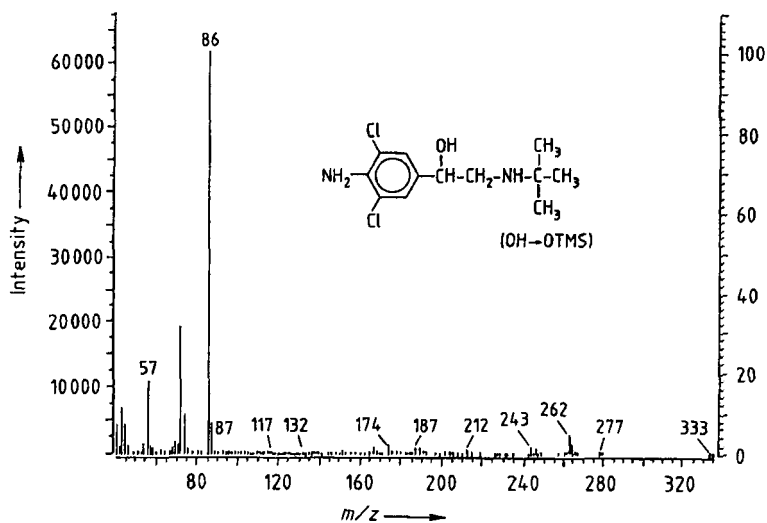
Figure 41. Electron ionization spectrum of TMS – clenbuterol

structed total ion chromatogram of a solvent mixture (Fig. 43 A) and the reconstructed chromatogram of three specific ions typical for C_3 (trimethyl or methylethyl), C_4 (tetramethyl, dimethylethyl, diethyl, etc.), C_5 (pentamethyl, trimethylethyl, dimethylpropyl, etc.) benzenes (Fig. 43 B – D). This is called *mass chromatography*, and only data stored in the computer of the full-scan mode are used to provide selectivity. *Mass fragmentography* or *selected ion monitoring* (SIM), on the other hand, is a method to quantify target compounds with high selectivity and sensitivity. In this mode of operation, the voltages on the quadrupole rods are adjusted stepwise to detect only two, three, or four ions. The time the rods are on a particular voltage to let an ion pass is much longer than in the case of the full-scan mode, which results in enhanced sensitivity. As an illustration, in the previous case for the full-scan mode, each ion was seen for only 1/380 of a second. Suppose clenbuterol has to be quantified. By selecting the ions 259 and 349 only (Fig. 42), each ion is detected in a 1-s scan for 1/2 s. Compared to the full-scan mode, the residence time is 190 times higher; thus the sensitivity increases by a factor 190. With modern CGC – MS systems, picogram amounts can easily be quantified when operated in the ion-monitoring mode. An interesting feature of SIM is that internal standards can be selected that chemically, physically, and chromatographically behave exactly the same as the compound to be measured. In the case of clenbuterol, for example,

the best internal standard is D_6-clenbuterol with typical ions at m/z 265 and 355. In routine analysis, D_6-clenbuterol is added to the sample (i.e., urine) before any pretreatment. Sample preparation and cleanup are then performed, and the four ions 259 and 349 for clenbuterol and 265 and 355 for the internal standard D_6-clenbuterol are recorded. Quantitation is very easy.

11.6.4.2. Fourier Transform Infrared Spectroscopy

In the combination CGC – FTIR, IR spectra of the eluting peaks are recorded as they emerge sequentially from the column outlet. The eluate is introduced in a light pipe where compounds absorb radiation at well-defined frequencies (i.e., frequencies characteristic of the bonds within a molecule). The absorption is sensed and recorded. The sensitivity of FTIR depends on the functionalities within a molecule, and for strong IR absorbers, 1 ng yields a good spectrum. Current FTIRs include computers, and the recorded spectra can be compared with a library of spectra to aid in compound identification. Monitoring of specific wavelengths allows the elucidation of classes of compounds (e.g., aldehydes, ketones, alcohols, ethers). On-line coupling of CGC–FTIR and CGC – MS can be realized in series since the FTIR is nondestructive. The IR spectra are complementary in nature to MS spectra, especially for the elucidation of isomers for which MS is not very
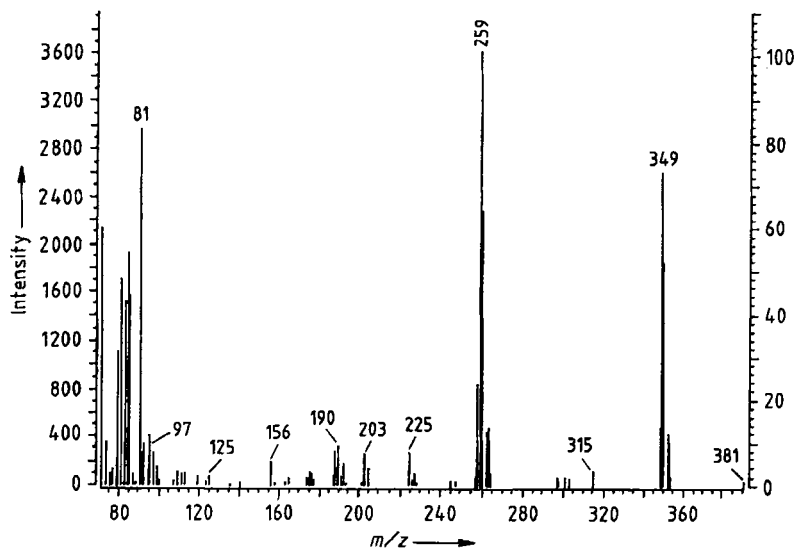
Figure 42. Chemical ionization spectrum (reagent gas isobutane) of TMS – clenbuterol

informative. A typical example is the determination of impurities in α-angelica lactone. Figure 44 shows the total response chromatogram obtained with FTIR and the total ion chromatogram, recorded by the mass spectrometer. Both chromatograms correspond very well, and at any time MS and IR data can be recalled and compared. The mass spectra of peak 1 and 2 are very similar and do not allow the differentiation of structural differences. The infrared spectra of peak 1 and 2 are compared in Figure 45. Some important differences are noted. The shift of the carbonyl band from 1833 cm^{-1} (peak 1) to 1806 cm^{-1} (peak 2) indicates that peak 2 corresponds to β-angelica lactone (α, β-unsaturated instead of β, γ-unsaturated).

11.6.4.3. Atomic Emission Detection

One of the latest developments in hyphenated techniques is the atomic emission detector. The availability of a bench-top model CGC –AED enables the use of this powerful technique in routine analysis. Detectabilities of this element-specific detector are of the order of 0.1 pg/s for organometallic compounds, 0.2 pg/s for carbon (more sensitive than FID), 1 pg/s for sulfur, and 15 pg/s for nitrogen, to mention only a few. The power of the technique lies in its supreme selectivity: all elements can be detected selectively. As opposed to ECD, AED allows differentiation between various halogenated organic compounds (e.g., fluoro, chloro, and bromo components). Multielement analysis can be carried out by simple preselection of the atoms to be monitored. In AED, the solutes eluting from the column are atomized in a high-energy source; the resulting atoms are in the excited state and emit light as they return to the ground state. The various wavelengths of the emitted light are then dispersed in a spectrometer and measured by using a photodiode array. Each element has a typical emission spectrum, with emission lines commonly occurring in clusters. The relative intensities of lines within these clusters are constant. Figure 46 shows the emission spectrum of a sulfur-containing compound, and the 180.7-, 182.0-, and 182.6-nm cluster serves as conclusive proof of the presence of sulfur in a peak.

Multiwavelength detection became possible through the introduction of diode array technology. Quantitative treatment of the data obtained in a multielement analysis moreover allows the calculation of empirical formulas, yielding information that is complementary to and helpful in the interpretation of mass spectral data. Quantitation is simplified considerably, since calibration of the detector no longer depends on the type of components to be quantified. Nontoxic compounds can thus be used to quantify harmful chemicals. Figure 47 illustrates the sensitivity of the detector for organometallic compounds. Organolead halides
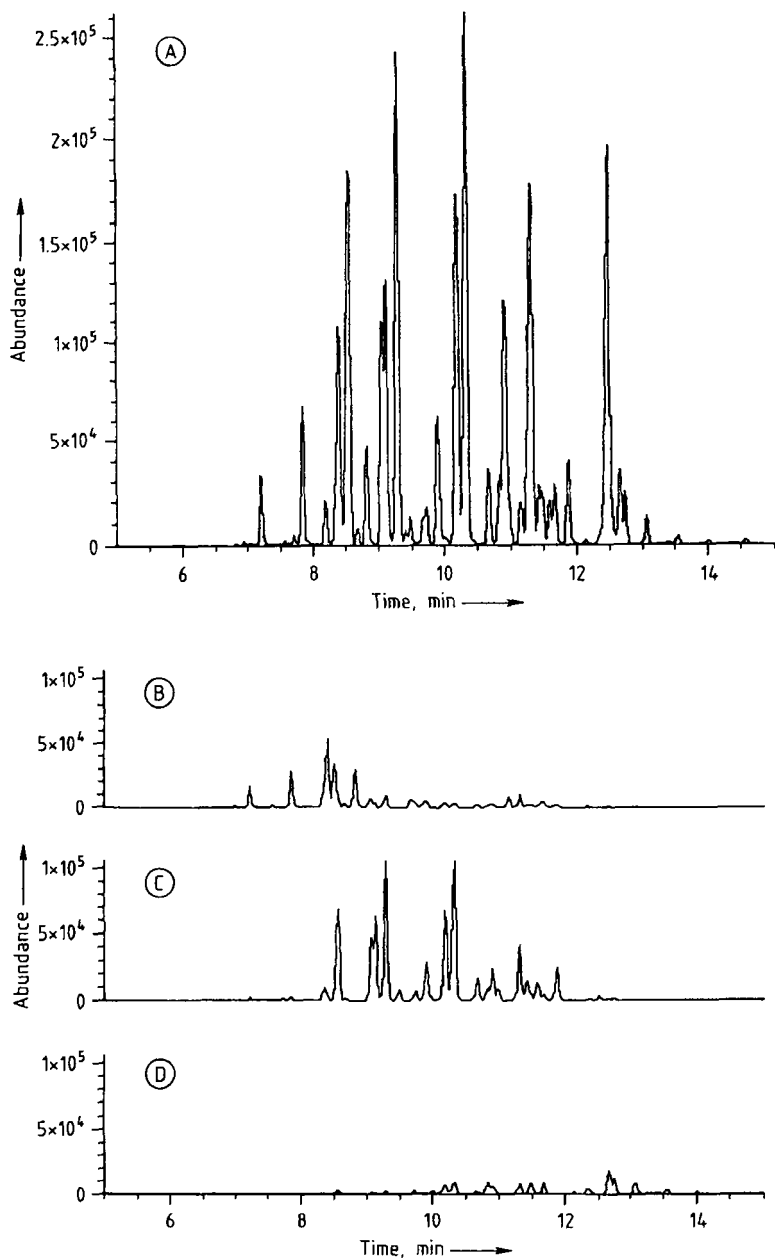
Figure 43. Illustration of mass chromatography
A) Solvent mixture; B) C_3-Benzene; C) C_4-Benzene; D) C_5-Benzene

are derivatized with butylmagnesium bromide into tetraalkylated solutes, which are amenable to CGC analysis. By selecting a wavelength of 406 nm, high selectivity and sensitivity (10 pg per compound) are guaranteed. Another important application concerns direct monitoring of oxygenates in petroleum products. Figure 48 shows the direct analysis of a light naphtha containing low concentrations (ppm) of oxygenated compounds. The concentrations for methanol and ethanol are 100 and 50 ppm, respectively.
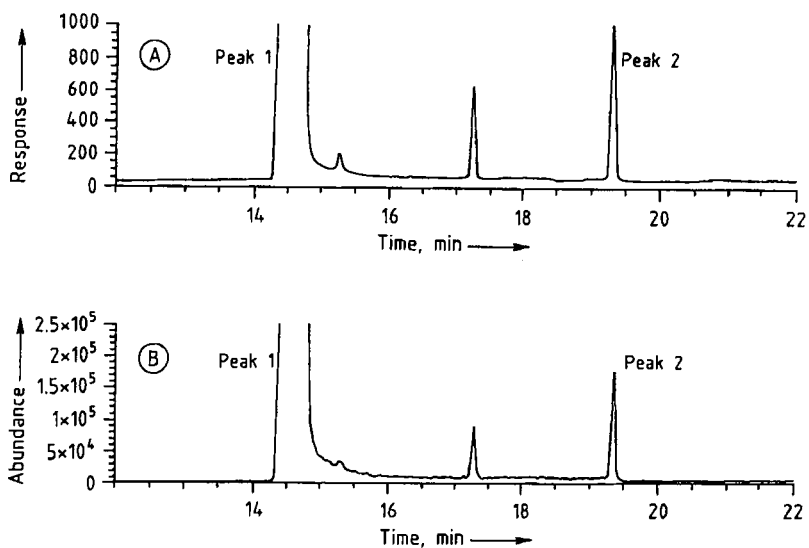
Figure 44. Capillary GC–FTIR–MS chromatograms of α-angelica lactone
A) Total response chromatogram (FTIR); B) Total ion chromatogram (MS)
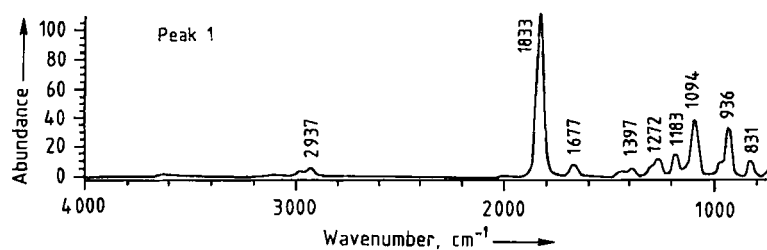
Figure 45. Infrared spectra of peak 1 and 2 of chromatogram in Figure 44
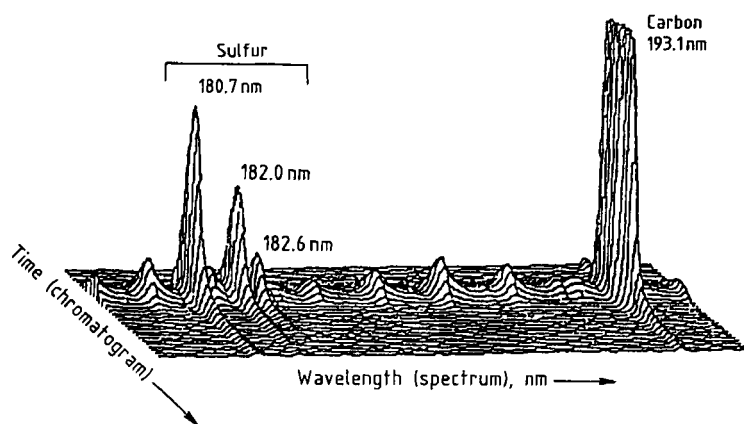
Figure 46. Emission spectrum of a sulfur-containing compound
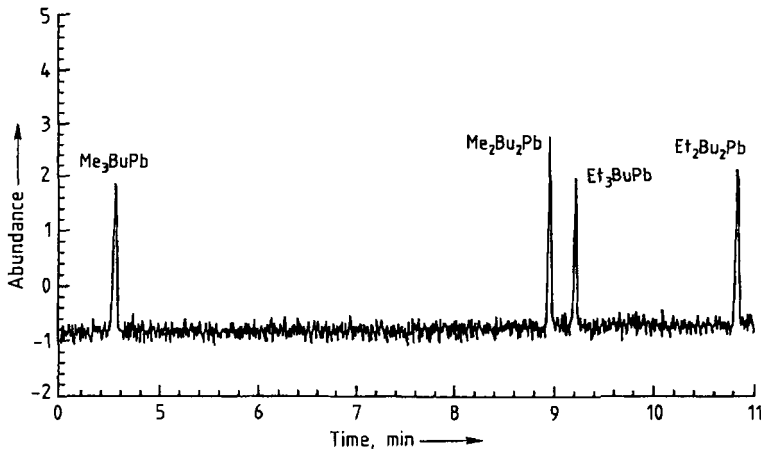
Figure 47. AED analysis of organolead compounds
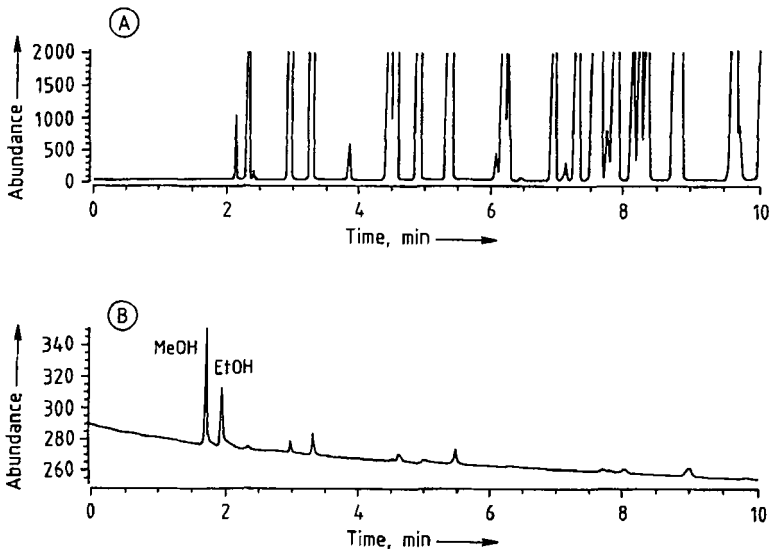
Figure 48. AED analysis of light naphtha in the C (A) and the O (B) emission line

11.7. Practical Considerations in Qualitative and Quantitative Analysis

A chromatogram provides information on the complexity (number of peaks), identity (retention time), and quantity (peak area or height) of the components in a mixture. This information can be considered suspect if the quality of separation is not optimal. Capillary columns have greatly enhanced the use of gas chromatography as a qualitative and quantitative tool.

11.7.1. Qualitative Analysis

Modern capillary GC is characterized by very good precision in retention time, and this allows the use of retention indices for peak identification. In a retention index system, the retention behavior of a particular solute is expressed in a uniform scale determined by a series of closely related standard substances. In the retention index scale developed by KOVATS [61] for isothermal separations and by VAN DEN DOOL and KRATZ [62] for temperature-programmed analyses, the fixed
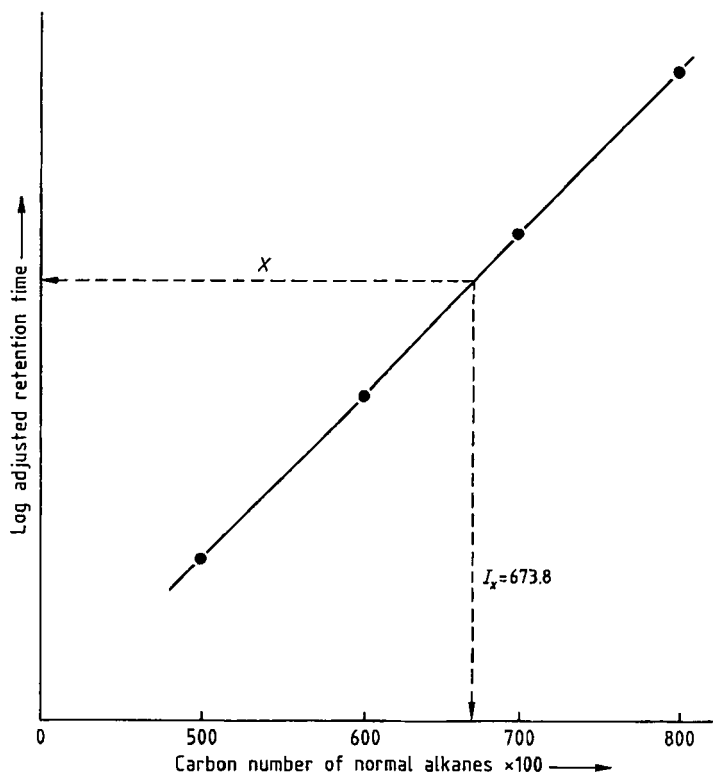
Figure 49. The Kovats retention index system

points of the scale are normal alkanes. The Kovats retention index system is illustrated in Figure 49. Compound X elutes between the hydrocarbons hexane and heptane with carbon numbers 6 and 7, respectively. For convenience, the carbon numbers are multiplied by 100, thus 600 and 700. The retention index I_x of compound X is 673.8. Indices are not graphically measured but calculated by Equation (7.1)

$$I_x = 100z + 100 \frac{\log t'_{R_x} - \log t'_{R_z}}{\log t'_{R_{z+1}} - \log t'_{R_z}} \qquad (7:1)$$

where z is the carbon number of the n-alkane eluting before X; $z+1$ is the carbon number of the n-alkane eluting after X; and t'_R is the adjusted retention time.

To perform the interpolation, the adjusted retention times of the n-alkanes must be known. This is often done by analyzing a separate mixture containing n-alkanes. This requires, however, automated injection. For manual injection, n-alkanes are added to the sample. In Equation (7.1), logarithmic interpolation is necessary because the

adjusted retention times increase exponentially with carbon number (Eq. 4.1). In linear temperature-programmed runs, the retention times of a homologous series (i.e., the n-alkanes) increase linearly with carbon number, and the equation for temperature-programmed analysis is simplified to

$$I_x = 100z + 100 \frac{t_{R_x} - t_{R_z}}{t_{R_{z+1}} - t_{R_z}} \qquad (:2)$$

Unadjusted retention times may be used. The isothermal indices depend on the stationary phase and film thickness and on the temperature. Specification of those variables is a must. The programmed indices, often called linear retention indices, require specification of the temperature profile as well. Once determined, the indices can be used to identify compounds from libraries with tabulated indices [63], [64]. The general utility of those libraries for identification is questionable. Indices are best used within a laboratory for specific samples (i.e., petroleum products, essential oils, steroids, etc.). Greater confidence in identification is

obtained by measuring indices on more than one stationary phase. Unequivocal compound elucidation requires spectroscopic characterization as provided by mass spectroscopy, Fourier transform infrared spectroscopy, and atomic emission detection.

Recently the concept of retention time locking was introduced and this opens new perspectives in capillary GC for solute elucidation (see Section 11.10.2).

11.7.2. Quantitative Analysis

Four techniques are commonly used in quantitative analysis: the normalization method, the external standard method, the internal standard method, and the standard addition method. Whatever method is used, the accuracy often depends on the sample preparation and on the injection technique. Nowadays these are two main sources of error in quantitative analysis. The quantitative results produced by PCGC and CGC are comparable.

11.8. Coupled Systems

Coupled systems include multidimensional and multimodal systems. *Multidimensional chromatography* involves two columns in series preferably two capillary columns, with different selectivity or sample capacity, to optimize the selectivity of some compounds of interest in complex profiles or to provide an enrichment of relevant fractions. In *multimodal systems*, two chromatographic methods or eventually a sample preparation unit and a chromatographic method are coupled in series. Coupled systems that received much interest in recent years are multidimensional CGC (MDCGC), the combination of high-performance liquid chromatography with CGC (HPLC – CGC) and the on- or off-line combination of supercritical fluid extraction with CGC (SFE – CGC). Multidimensional and multimodal techniques in chromatography are described in detail in [65].

11.8.1. Multidimensional Capillary GC

For highly complex samples, the separation power of a single capillary column is insufficient to achieve complete resolution for the compounds of interest. Even after optimization of the selec-

tivity, important compounds will still coelute, since the better separation of one pair of compounds is likely to be counteracted by the overlapping of another pair of compounds present in the sample. The only solution is the combination of more than one column system. In multidimensional CGC, a group of compounds that has not been separated on a first column is transferred (heartcut) to a second column, where complete resolution is achieved. Two operational modes can be distinguished. If the transferred fraction is coldtrapped before analysis on the second column (intermediate trapping), the selectivities of both columns are decoupled. The elution of the transferred solutes will be independent of the selectivity of the first column and will be the same as if the pure sample compounds were analyzed on the second column only. Alternatively, when no intermediate trapping is applied, the elution pattern observed after analysis on the second column will be the same as the elution pattern for analysis on a mixed-phase column or a coupled column system with an intermediate selectivity between the selectivities of both columns. Both principles are illustrated in Figure 50. Peppermint oil was analyzed on a methylsilicone capillary column, and the separation of the menthone – menthol fraction (peak 1 – 7) is incomplete (Fig. 50 A). This fraction was heartcut (Fig. 50 B) and further analyzed on a poly(ethylene glycol) capillary column. With intermediate trapping (Fig. 50 C) the separation is drastically improved, but the highest resolution, in this case, is without intermediate trapping (Fig. 50 D). This is an illustration of the power of multidimensional CGC in combination with selectivity tuning.

Developments in multidimensional CGC diverge in two main directions: the coupling of narrow to narrowbore capillaries to enhance resolution, on the one hand, and the coupling of widebore precolumns to narrowbore capillaries as a selective sample introduction method, on the other. Some inlet devices, such as the programmed temperature vaporization injector operated in the solvent vent mode, can also be classified as multidimensional systems. The solvent is removed through the split vent, while the solutes of interest remain in the liner. The multidimensional approach with selective sampling is of considerable interest in routine analysis. It allows preseparation of relatively large amounts of sample on the precolumn operated under constant-flow conditions. Fine-tuning of the separation can be achieved on the analytical column, operated in the constant-
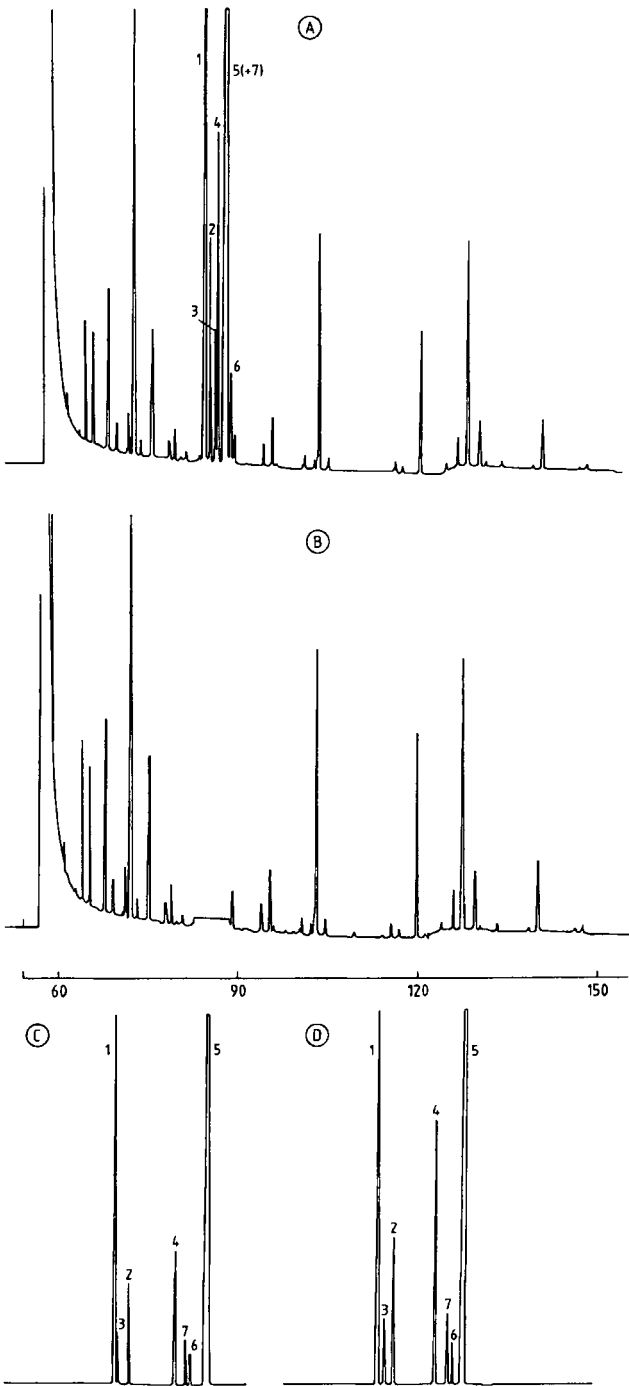
Figure 50. Multidimensional CGC analysis of pepermint oil, two column system constant pressure – constant pressure
A) Separation on MeSi; B) Analysis on MeSi after heartcut ; C) Analysis of heartcut fraction on column B with intermediate trapping; D) Analysis of heartcut fraction on column B without intermediate trapping (selectivity tuning)
Column A: 20 m $L \times 0.32$ mm $i.d.$; 0.24 μm d_f MeSi; Column B: 20 m $L \times 0.32$ mm $i.d.$; 0.24 μm d_f high-M_r PEG
Temperature: 60 °C (2 min) to 180 °C at 3 °C/min; split injection 1/50; carrier gas hydrogen; FID
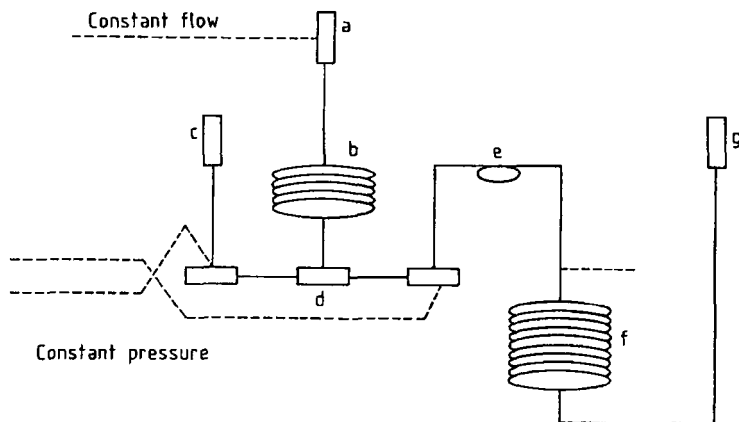
Figure 51. MDCGC system constant flow – constant pressure
a) Injector; b) Precolumn; c) Detector 1; d) Switching device; e) Cold trap; f) Analytical column; g) Detector 2

pressure mode. The system is easy to operate and can be installed without problems in every GC instrument. Systems based on constant pressure – constant pressure and constant flow – constant flow offer more flexibility because column dimensions are less critical than in the constant flow – constant pressure system; however, they are expensive and require time-consuming optimization procedures when new applications are carried out. The possibilities of multidimensional capillary GC as a sampling method are illustrated with the analysis of volatiles emitted by living plants [66]. The atmosphere surrounding a plant (headspace) is sucked in an open tubular trap (3 m in length, 530 µm *i.d.*) coated with a 15-µm film of methylsilicone where the volatiles are retained by the normal chromatographic partitioning process at ambient temperature. The advantages over traps filled with adsorbents are that breakthrough volumes can be calculated, the inertness is high, and thermal desorption of polar compounds is feasible. After sampling, the trap is installed as precolumn in a multidimensional CGC system (Fig. 51). The volatiles are thermally desorbed, reconcentrated in a cold trap (e), and then analyzed on-line on the analytical capillary column (f). Figure 52 shows the headspace CGC–MS analyses of a healthy and a fungus-infected *Mentha arvensis* plant. Menthol (peak 15) is the main compound in the healthy plant (Fig. 52 A), and the highly volatile fraction is relatively poor. In the infected plant (Fig. 52 B), menthyl acetate (peak 21) is the main compound, while the relative percentage of menthone (peak 13) is increased and the highly volatile fraction is more abundant. This hyphenated

combination live headspace sampling – multidimensional capillary GC – MS provides interesting results in phytochemical research. In 1992, a more sophisticated system, PTV injection — MDCGC with three columns—MS was described for the analysis of dioxins and furans [54].

11.8.2. Multimodal High-Performance Liquid Chromatography – Capillary GC

The on-line coupling of HPLC to capillary GC has received much attention since the mid-1980s. Pioneering work was performed by K. Grob, Jr. [67]. The coupling of HPLC to capillary GC can be considered as an on-line sample preparation and enrichment step. Preseparation can be carried out according to molecular mass (SEC), functionality (straight-phase LC), or hydrophobicity (reversed-phase LC). Similar results are obtained with off-line techniques (i.e., solid-phase extraction or LC off-line fractionation), but advantages of the on-line approach include easier automation, fewer artifacts, and higher sensitivity because the total elution volume of the fraction of interest is transferred to the capillary GC column.

Recently a fully automated and flexible on-line LC–GC system has been described [68]. The system is based on standard LC and capillary GC instrumentation. The on-line LC–GC system was evaluated for the determination of pesticides in orange oil. The essential oil was fractionated by normal phase LC, resulting in a separation according to polarity. The fraction containing pesticides was transferred and analyzed by capillary GC –
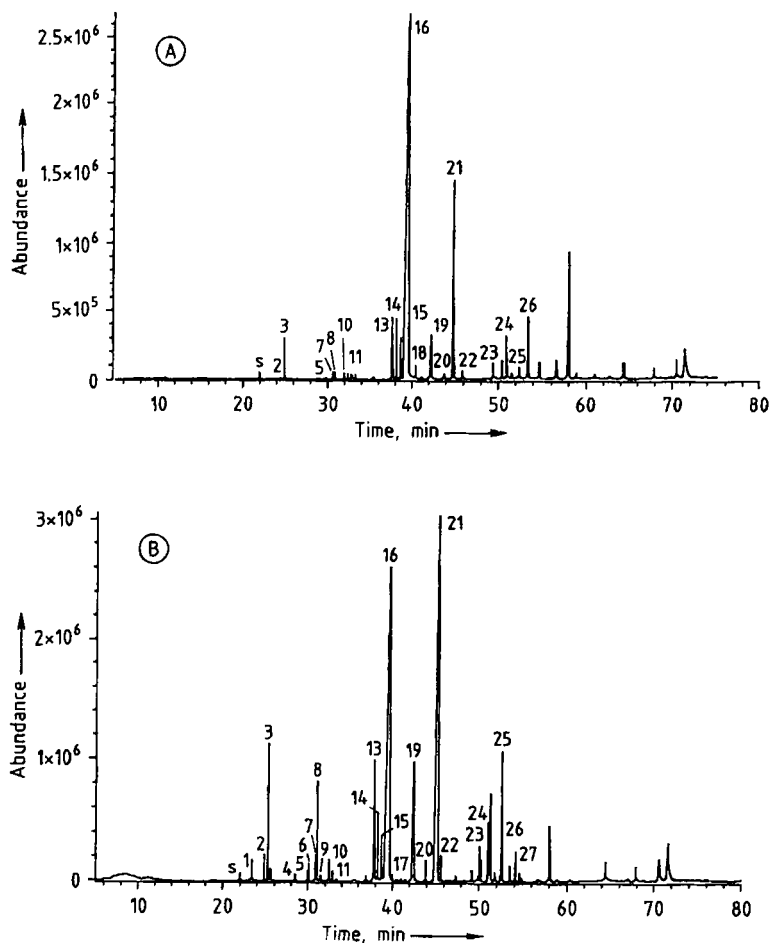
Figure 52. CGC pattern of the head space of a healthy (A) and an infected (B) *Mentha arvensis* plant
Thermal desorption and cryofocusing

NPD. The LC–GC interface is a modified large volume sampler (Gerstel, Mulheim a/d Ruhr, Germany). The heart of the system is a specially designed flow cell that replaces a normal vial. The mobile phase leaving the LC detector is directed via a capillary tube with well-defined dimensions, in a T-shaped flow cell. The cell is equipped with a septumless sampling head through which a syringe needle can be introduced. The sampler is completely computer controlled. To transfer a selected LC fraction, the transfer start and stop times, measured on the LC chromatogram, are introduced into the controller software. The time delay between the LC detector and the flow cell is automatically calculated from the connecting tubing dimensions and the LC flow rate. At the time the LC fraction of interest passes through the flow cell, the syringe needle penetrates the septumless head and samples the LC fraction at a speed equal to the LC flow rate. Volumes of up to 2 mL can be collected in the syringe. After collection, the needle is withdrawn from the flow cell which rotates away from the PTV inlet and a large volume injection is made using the PTV in the solvent vent mode. Depending on the fraction volume and solvent type, the sample introduction parameters (inlet temperature, vent flow, vent time, injection speed,...) are calculated by the PTV calculator program. For fractionation of the orange oil sample, the following HPLC parameters were used: column 250 mm × 4.6 mm i.d. × 10 µm Lichrospher 100 DIOL, injection volume 20 µL, mobile phase in a gradient from 100 % hexane for 10 min, to 40 % isopropanol at 20 min (2 min hold) at a
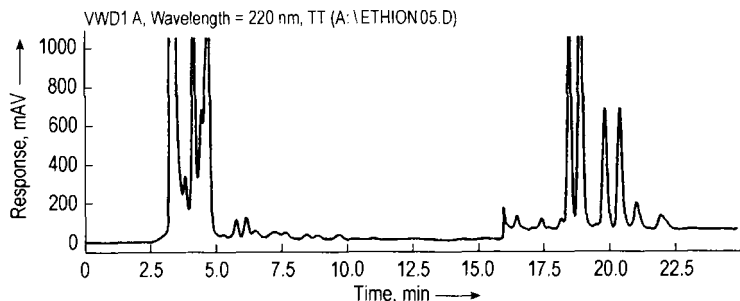
Figure 53. LC fractionation of orange oil on Lichrosorb 100 DIOL
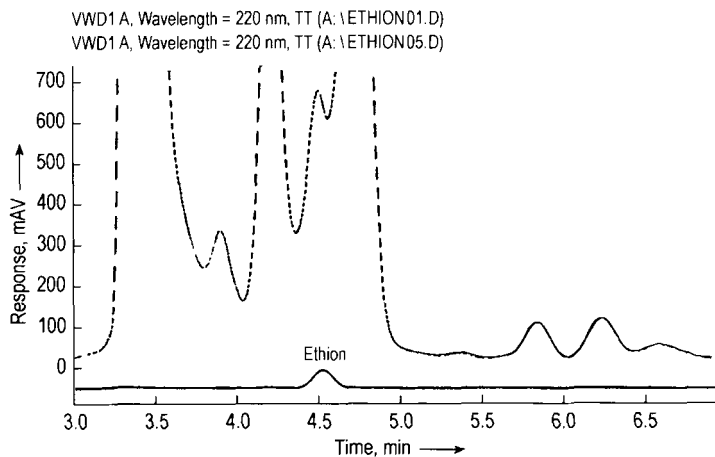
Figure 54. Overlay of relevant parts of LC chromatograms of a 1 ppm ethion standard (bottom trace) and orange oil (top trace)

flow rate of 1 mL/min and UV detection at 220 nm. The fraction eluting from 4.4 – 4.9 min (volume = 0.5 mL) was automatically transferred to the GC inlet. The LC – GC interface was programmed to take the sample at a 1000 µL/min sampling speed (the same as the LC flow rate). The complete fraction of 500 µL was injected in the PTV inlet at 250 µL/min. This injection speed corresponds to the injection speed calculated by the PTV software program. Capillary GC – NPD analysis was performed on a 30 m × 0.25 mm i.d. × 0.25 µm MeSi column and the oven was programmed from 70 °C (2 min) to 150 °C at 25 °C/min and then to 280 °C at 8 °C/min (10 min). The detector was set at 320 °C with 3 mL/min hydrogen, 80 mL/min air, and 30 mL/min helium make-up flow.

The LC profile for the orange essential is shown in Figure 53. The apolar mono- and sesquiterpenes elute first, followed by the terpenoids and

after 16 min also the flavanoids are eluted. These last compounds in particular give most interference in GC analysis because they have similar molecular weights to the pesticides. Using the same conditions, ethion and chlorpyriphos elute at 4.6 – 4.7 min. Figure 54 shows a comparison of the analysis of a 1 ppm ethion reference sample and the essential oil analysis.

The fraction eluting from 4.4 to 4.9 min was transferred to the CGC-NPD and analyzed. Figure 55 shows the resulting GC – NPD chromatogram and both ethion and chlorpyriphos (t_R 15.79 min) are detected. The chromatograms are very clean and no interferences are present. This demonstrates the excellent selectivity of the LC – GC combination. The concentrations of ethion and chlorpyriphos were calculated by external standard analysis and are 1.9 ppm for chlorpyriphos and 0.8 ppm for ethion.

NPD1 A, (A:\LCGC0006.D)
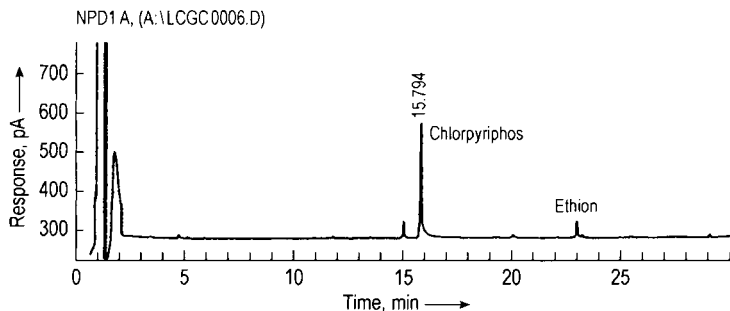
Figure 55. LC – Capillary GC – NPD analysis of orange oil
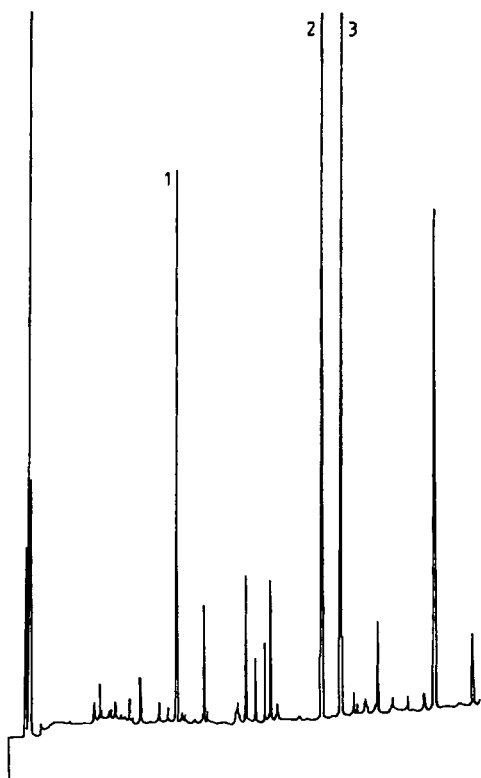
Figure 56. Analysis of SFE extract of a sediment with ECD detection
Peaks: (1) aldrin, (2) dieldrin, (3) endrin

11.8.3. Multimodal Supercritical Fluid Extraction – Capillary GC

Because supercritical fluids have excellent characteristics for the rapid extraction of organic compounds, combining supercritical fluid extrac-

tion with CGC is an interesting approach to the enrichment of organic solvents [69], [70]. Since the density of supercritical fluids—usually carbon dioxide—can be varied (pressure, temperature), a high degree of selectivity can be introduced in the sample preparation procedure (i.e., prior to the analytical system). The selectivity can be further enhanced by making use of modifiers that may be polar (e.g., $CO_2 - MeOH$) or nonpolar ($CO_2 - $hexane). The main applications of SFE – CGC are related to solid samples, such as soil or sediment samples, and plant materials. For aqueous solutions (e.g., wastewater, urine), samples are enriched by solid-phase extraction. The extraction cartridges are then submitted to SFE – CGC, thus providing all the advantages of SFE (i.e., selective extraction and ease of automation). There is some controversy in the literature whether on-line SFE – CGC is the best approach to the application of analytical supercritical fluid extraction. Indeed off-line SFE has some advantages, and the use of robotics allows automated transfer of the sample to the capillary GC instrument. Selection of on-line or off-line depends mainly on the specific problem. Off-line is preferred for method development and on-line for routine analysis. On-line can, however, be recommended only when the matrix from which the solutes have to be extracted is always the same (e.g., in chemical composition, particle size, pore size). Moreover, the literature on SFE extraction times and recoveries is often incorrect because spiked samples are analyzed. Applied to real contaminated samples, SFE is found to be much more difficult. Figure 56 shows the optimized off-line SFE – CGC analysis of pesticides from sediment: 500 mg sediment was placed in the extraction vessel, and 250 µL of methanol was added on top of the sediment. SF extraction was performed at 60 °C with pure CO_2
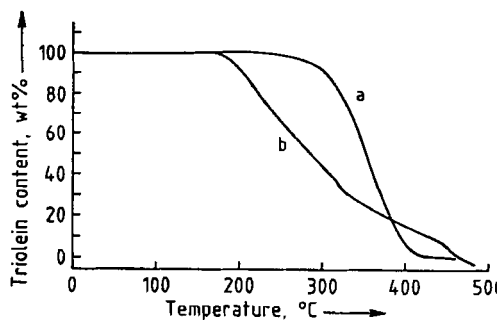
Figure 57. Thermogravimetric analysis of triolein
Ramp from 40–500 °C at 5 °C/min
a) Pure nitrogen; b) Air

at 35 MPa for 30 min. The concentration of the main compound dieldrin is 20 ppb. Classical Soxhlet extraction gave the same concentration.

11.9. Applicability

11.9.1. Solute Thermal Stability

Solutes subjected to GC must be thermally stable. The thermostability of organic compounds depends strongly on their nature, on the activity of the environment in which thermal stability is measured, and on the thermal stress given to the solutes. Capillary GC nowadays is performed in a completely inert system, i.e., highly pure carrier gas, purified stationary phases, fused silica with less than 0.1 ppm metal traces and specially deactivated, etc. Moreover, thermal stress can be avoided by applying cool on-column injection. Lipids serve as a good illustration. When oils or fats are used in food preparation, they decompose (formation of aldehydes etc.) or polymerize (formation of dimers, trimers, etc.) as a function of time, which make them no longer useful for cooking. These alterations are caused by the presence of water and oxygen. When heated under inert conditions, fats and oils are stable and evaporate. Figure 57 compares the thermogravimetric profiles recorded for triolein ($M_r = 886$) under a stream of pure nitrogen gas (a) and air (b). In present state-of-the-art GC, the systems are operated under circumstances shown in curve a.

11.9.2. Solute Volatility

Volatility is related to the vapor pressure (boiling point) of the compounds. Polydimethylsiloxanes with molecular masses as high as 5000 are volatile enough to be analyzed by GC, whereas poly(ethylene glycols) with M_r of ca. 1500 are not volatile at all. Volatility is thus related not only to the molecular mass, but to the polarity of the functional groups. Polydimethylsiloxane (M_r 5000) has approximately 68 apolar dimethylsiloxane units, whereas poly(ethylene glycol) (M_r 1500) has 114 polar ethoxy units besides the terminating hydroxyl groups. Some thermolabile compounds, although able to be volatilized, undergo partial decomposition at high temperature. Derivatization can be employed to impart volatility and to yield a more stable product, thereby improving chromatographic performance and peak shape. Three general types of analytical derivatization reactions are used in GC: alkylation, acylation, and silylation [71], [72]. The ideal derivatization procedure must be simple, fast, and complete. Alkylation, acylation, and silylation are used to modify compounds containing functional groups with active hydrogens such as –COOH, –OH, and –NH$_2$. The tendency of these functional groups to form intermolecular hydrogen bonds affects the inherent volatility of compounds containing them, their interaction with the stationary phase, and their stability. Figure 58 shows the analysis of sucrose in molasses with capillary GC. Trehalose is used as internal standard, and the sample is derivatized into the oxime – trimethylsilyl derivatives.

11.9.3. Comparison of Gas Chromatography, Liquid Chromatography, and Supercritical Fluid Chromatography

When a scientist is faced with deciding which separation technique to use for a specific analytical problem, the primary question is, which chromatographic technique best suits this particular separation. The initial answer is dictated by the chemical nature of the analytes themselves. For the analysis of volatile components, gas chromatography is preferred. For analyzing strongly polar or ionic compounds, liquid chromatography may be the best choice. In some cases, supercritical fluid chromatography has distinct advantages over GC and LC. However, for many applications —

Table 5. Calculated percent composition of steroid esters *

Compound	HPLC		CSFC	CGC on-column	Composition in mixture, %
	UV	MD **			
Testosterone propionate	15.13	11.93	10.98	11.98	12.00
Testosterone isocaproate	27.77	23.68	23.57	25.20	24.00
Testosterone decanoate	33.25	40.43	44.33	39.99	40.00
Testosterone phenylpropionate	23.38	23.96	21.11	23.83	24.00
Mean RSD, %	1.43	1.08	2.58	0.58	
Analysis time, min	22.50	25.00	21.50	21.50	

* All values calculated from uncorrected peak areas with response factors = 1.0; mean values of five experiments.
** MD = mass detector.
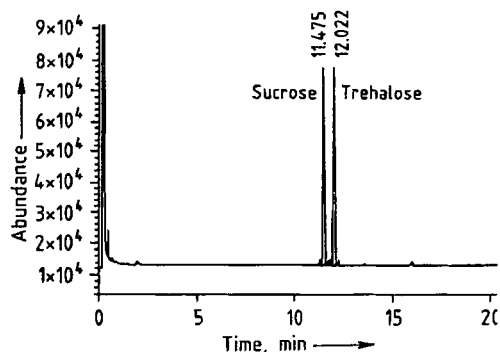
Figure 58. Capillary GC analysis of sugar derivatives (oxime – TMS)
Column: 10 m $L\times0.53$ mm *i.d.*: 0.1 µm d_f MeSi; temperature: 80–300 °C at 10 °C/min; cool on-column injection; carrier gas hydrogen 4 mL/min; FID

such as analyses of polycyclic aromatic hydrocarbons, polychlorinated biphenyls, lipids, fatty acids, steroids, fat-soluble vitamins, mycotoxins, etc., — several chromatographic techniques can be considered.

In such cases, strange as it may seem, the choice of a chromatographic technique is often biased on secondary considerations such as the analyst's know-how and skill or the availability of a particular instrument. This choice is usually justified with supplementary arguments, especially those concerning the simplicity of a technique. Recent technical developments in nearly all branches of chromatography, however, have largely overcome any "advantage of simplicity" (to some, the "disadvantage of complexity") of various techniques. The extended possibilities of automation further weaken arguments of this kind.

More important, especially when the analytical method must be applied routinely, is its cost. This will be governed mainly by the price of the in-

strument, columns, and mobile phases, as well as maintenance costs. In this respect, the chromatographic techniques can be ranked in increasing order of the prices of analysis: GC < LC < SFC.

From a practical viewpoint that disregards costs, the choice of method must be based on three fundamental aspects: separation efficiency, accuracy of the analytical method, and total required analysis time. For some applications, the specificity and sensitivity (detectability) of a technique are also important.

A study was conducted in which the performance of three techniques — high-performance LC (HPLC), capillary GC (CGC), and capillary SFC (CSFC) — was compared in the analysis of the percent composition of steroid esters in a pharmaceutical hormone formulation [73]. The study used standard chromatographic conditions, applicable in every laboratory, and the conditions were optimized so as to yield comparable analysis times of < 30 min.

High-Performance Liquid Chromatography. In the pharmaceutical industry, analysis of steroid ester formulations is performed routinely by using HPLC. Figure 59 A shows the HPLC chromatogram of the sample mixture applying UV detection at 254 nm. The steroid esters are baseline separated in 24 min. The column plate number, calculated for testosterone decanoate, is 4500, which corresponds to 45 000 plates per meter. With UV detection, allowance must be made for different response factors. Calibration graphs, however, can be recorded only if pure compounds are used, which was not the case in this study. A response factor of 1.0 was, therefore, applied for the various steroid esters, giving the quantitative data listed in Table 5. Very often, in practice, unknown compounds are present in HPLC chromatograms, and quantitation is impossible using UV detection.
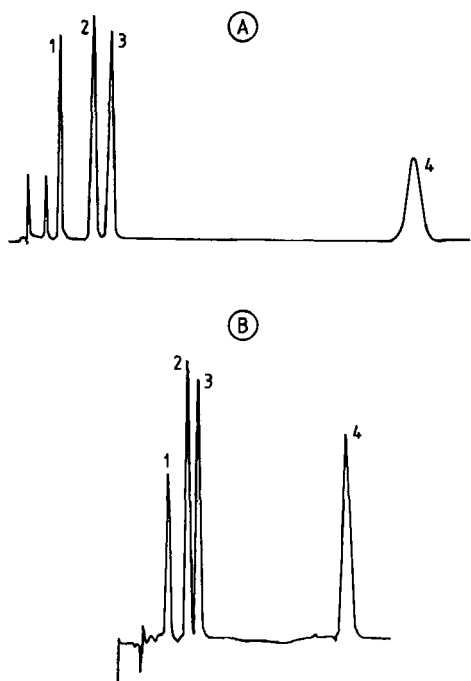
Figure 59. HPLC analysis of steroid esters
A) Column: 10 cm $L\times0.3$ cm $i.d.$, 5 µm ODS, flow 0.5 mL/min methanol – water 85 : 15, UV at 254 nm; B) see A, light scattering detection
Peaks: (1) testosterone propionate, (2) testosterone isocaproate, (3) testosterone phenylpropionate, (4) testosterone decanoate

Figure 60. Capillary SFC analysis of steroid esters
Column: 20 m $L\times0.1$ mm $i.d.$; 0.1 µm d_f MeSi; temperature: 130 °C; pressure 200 bar CO_2; FID
Peaks: (1) testosterone propionate, (2) testosterone isocaproate, (3) testosterone decanoate, (4) testosterone phenylpropionate

Calibration can be avoided by using universal HPLC detection systems, such as refractive index detection or light-scattering detection. Compared to UV, however, both of these detectors are less sensitive. Figure 59 B shows the analysis using light-scattering detection; the sample is five times more concentrated than in the UV trace in Figure 59 A. The chromatographic efficiency was reduced by 20 %, to 3600 plates, because of a detector contribution to band broadening. The quantitative data, calculated using response factors of 1, are listed in Table 5 in the column labeled MD (mass detector). The calculated composition corresponds well to the known sample composition.

Capillary Gas Chromatography. Steroid esters are thermally stable. They can be analyzed easily by CGC. Figure 60 shows the analysis of the sample with cold on-column injection. The plate number for this conventional CGC column operated at 130 °C was 85 000 plates; analysis time was 22 min. The quantitative results (response fac-

tors = 1.0) are given in Table 5. The data for cold on-column injection correspond very well with the known composition of the sample, and % RSD is < 0.8 % ($n = 5$) for all the steroid esters.

Capillary   Supercritical   Fluid Chromatography. In the mid-1980s, considerable progress has been made in CSFC, and the technique is now accessible for routine chromatography. CSFC is claimed to have advantages compared to GC. Because of the high solvating power of CO_2, high molar mass compounds elute at much lower temperature, allowing better analysis of thermally labile compounds. Compared with HPLC, for the same column, the speed of analysis for SFC is roughly four times faster. In addition, because of the lower viscosity of the mobile phase, much longer columns can be used, increasing the effective plate number. Universal GC detectors can be applied, and combination with hyphenated techniques such as mass spectrometry or Fourier transform infrared spectrometry is easier. Analysis of the steroid ester sample on a 20 m×0.1 mm
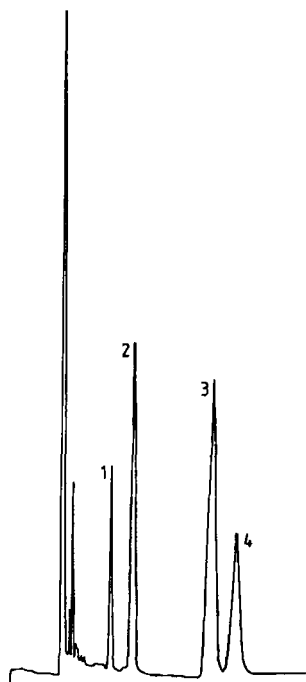
Figure 61. Capillary GC analysis of steroid esters
Column: 25 m L×0.25 mm $i.d.$; 0.15 μm d_f MePhSi; temperature: 60–300 °C at 10 °C/min; cool on-column injection; carrier gas hydrogen 15 psi ($\approx$ 1 bar); FID
Peaks: see Figure 60

Figure 62. High-speed capillary GC analysis of steroid esters
Column: 2 m L×0.25 mm $i.d.$; 0.1 μm d_f MePhSi; temperature: 260 °C; split injection 1/100; carrier gas hydrogen 0.1 psi ($\approx$ 700 Pa); FID

column is shown in Figure 61. In this run at constant density, the plate number is only 5000. This illustrates one of the disadvantages of CSFC. Theoretically the column should offer 150 000 plates, but at a mobile phase velocity of only 0.2–0.3 cm/s. This corresponds to an unretained peak time of 2.5 h, and an analysis time for a compound eluting with a k value of 3.5 (peak 3, testosterone decanoate) of 7.3 h. Such analysis times are, of course, unacceptable. Therefore the column is operated far above the optimal velocity (Figure 61, at 6 cm/s). The slope of the $H - u$ curve is very steep for the 20-m column, and the plate height at 6 cm/s is of the order of 6 mm. Much faster separations can be obtained on short capillary columns with internal diameters of $\approx$ 20 μm, but because of technical problems (especially with sample introduction), quantitation is very difficult. Quantitation of the steroid mixture with CSFC (Table 5) yields higher % RSD values and less accurate results than either CGC or HPLC. The reason for this lies mainly in the injection method (valve split discrimination). Compared to conven-

tional HPLC, the primary advantage of SFC is the use of the universal FID detector. The results presented in Table 5 show that all three techniques can be applied to analysis of the steroid sample. The difference in chromatographic behavior among HPLC, CGC, and CSFC is determined mainly by the mobile-phase properties such as density, diffusivity, viscosity, and polarity. When using dimensionless parameters, equations can be derived that are valid for the three different types of mobile phases and independent of particle size (packed columns) or column diameter (capillary columns) [74]. The reduced parameters will not be discussed in detail but show clearly, as far as efficiency and speed of analysis are concerned, no other chromatographic methods can compete with GC. Baseline separation of the steroid esters can be obtained on a 2-m capillary column offering 7000 plates in 25 s (Figure 62). This is not a plea for gas chromatography. It is just an example to illustrate the possibilities of different chromatographic techniques for the application selected. Often, however, one might be surprised to see the enormous effort spent in trying a separation with a particular technique, while the use of another technique seems straightforward. Each chromatographic technique has its own possibili-
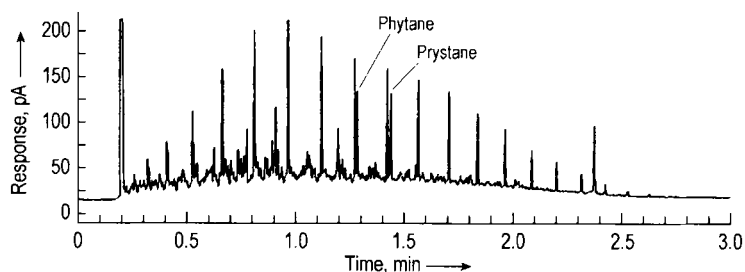
Figure 63. Separation of diesel oil in 2.5 min on standard GC instrumentation

ties, and a good understanding of their performance is a prerequisite before selection for a particular application can be justified.

11.10. Recent and Future Developments

11.10.1. Fast High Resolution Capillary GC

Since the introduction of gas chromatography, there has been an ongoing interest in improving analysis speed. Increasing analysis speed in capillary GC in the first instance is dictated by the problem at hand, the primary objective being the complete separation of the compounds in a mixture. In a large number of cases, the plate number of a capillary column is too high for a given separation problem and the resolution may be impaired. Typical ways to shorten the analysis time in this case are decreasing the column length or increasing the carrier gas flow rate far above the optimum. Recent alternatives include the use of multichannel columns or multicapillaries and flash GC.

Increasing analysis speed for complex profiles without impairing resolution can only be realized by reduction of the internal diameter and the length of the capillary column. A 10 m × 0.1 mm i.d. column offers the same resolution as a 25 m × 0.25 mm i.d. column. Because the column is 2.5 times shorter, the analysis time is reduced drastically. Moreover, since the optimum carrier gas velocity is higher and the H-u curves are flatter for narrow bore columns, higher average carrier gas velocities can be used without loss of resolution. Presently, capillary columns with internal diameters in the order of 100 μm are in the picture for routine operation because state-of-the-art cap

illary GC instrumentation allows installation of such small diameter columns. This is illustrated with the analysis of a diesel sample with baseline separation of the biomarkers pristane and phytane from the preceeding normal hydrocarbons on a standard GC instrument (Fig. 63). The diesel sample was diluted 10 times in cyclohexane and 1 μL was injected in the split mode (1/1000 split ratio). The column was a 10 m × 0.1 mm i.d. × 0.1 μm MeSi. The oven was programmed from 100 °C to 325 °C at 75 °C/min and held for 1 min (4 min program). FID detection with a signal data acquisition rate of 100 Hz was used.

An important obstacle for the use of narrow bore columns, however, is method development and validation. Using narrow bore columns, different operational conditions (inlet pressure, split ratio, temperature program) have to be used. Since little information is yet available on the use of fast capillary GC, the transfer of standard validated operating procedures developed for conventional capillary columns into operating procedures for narrow bore columns might be difficult and will definitely hamper their use in a routine environment. In this respect, the development of method translation software [75], [76] was very helpful for translating a standard operating procedure for a conventional column (whatever its dimensions and stationary phase film thickness) to an operating procedure for a narrow bore column (coated with the same stationary phase). After performing the analyses on the standard column, the optimized conditions are introduced in the method translation program and all operational conditions for the new column are calculated in order to obtain the same resolution. The gain in analysis time is also predicted. The method translation software is available free of charge from the Agilent website. The principle is illustrated with the analysis of an essential oil. Quality control of essential oil samples is routinely done on a

Table 6. Experimental conditions for essential oil analysis on a conventional and on a narrow bore column. The conditions in italic for the narrow bore column have been calculated by the method translation software. The standard operating conditions (SOP) on the conventional column apply broader conditions than needed for the essential oil analysed here

Column	60 m × 0.25 mm i.d. × 1 μm HP-1	20 m × 0.1 mm i.d. × 0.4 μm HP-1
Injection	split, 1 μL, 1/50 split ratio, 250 °C	split, 1 μL, *1/500* split ratio, 250 °C
Carrier type	helium	hydrogen
Carrier pressure	209 kPa, constant pressure	*411 kPa,* constant pressure
Carrier flow	1.74 mL/min at 50 °C	*0.87 mL/min* at 50 °C
Velocity	29.5 cm/s (average at 50 °C)	*58 cm/s* (average at 50 °C)
Hold-up time	3.39 min	*0.57 min*
Oven program	50 °C – 2 °C/min – 275 °C – 40 min	*50 °C – 11.88 °C/min – 275 °C – 7 min*
Detection	FID	FID
Signal data rate	10 Hz	50 Hz
Analysis time	152.5 min	*25.94 min*

50 – 60 m × 0.25 mm i.d. column using a slow (2 °C/min) temperature program, which results in analysis times in the order of 2 – 3 h. The same peak capacity can be obtained on shorter narrow bore capillary columns, but, although analysis times can be reduced drastically, quality control laboratories hesitate to use this approach because changing the column dimensions implies different operational conditions, which results in different selectivities. Details of well-known fingerprints can be lost. The application of the method translation software allows the translation from conventional to narrow bore columns with hardly any change in resolution, selectivity, and thus fingerprint. This is illustrated with the analysis of nutmeg oil.

The analysis was first performed on a "standard" column used for detailed essential oil profiling (60 m × 0.25 mm i.d. × 1 μm MeSi). The operational conditions optimized for routine QC were applied. Secondly, the analyses were repeated on a 20 m × 0.1 mm i.d. × 0.4 μm MeSi. The operational conditions for the narrow bore column were calculated by using the method translation software. The most important operational conditions are summarized in Table 6. From the method translation software program, a speed gain factor of 5.9 is predicted. Note that the carrier gas in the analyses are helium and hydrogen for the 0.25 mm i.d. and the 0.1 mm i.d. columns, respectively.

The chromatograms obtained on the respective columns are compared in Figure 64 top (Nutmeg oil, standard column) and bottom (Nutmeg oil, narrow bore column). From these chromatograms, it is obvious that the resolution is exactly the same on both columns. The analysis on the narrow bore column is much faster. From the last eluting peak, a speed gain factor of 5.7 is measured, which is

close to the predicted speed gain of 5.9. Since for the calculation, the nominal column lengths were used, it might be expected that the correlation could be even better if actual column lengths are measured and used in the method translation software.

11.10.2. The Concept of Retention Time Locking

Recent developments in capillary GC, i.e., electronic pneumatic control (EPC) of the carrier gas, improved oven temperature stability, and excellent reproducibility in column making have led to the concept of retention time locking (RTL) [77], [78]. With retention time locked data bases, absolute retention times instead of retention indices can be used to elucidate the structures of eluting solutes. Moreover, retention time locking can be used in combination with different injectors and detectors. Exact scaling of capillary GC – FID, capillary GC – MS, and capillary GC – AED chromatograms is feasible.

The concept is illustrated with the analysis of some organochlorine pesticides in wine. Figure 65 shows the capillary GC – AED trace in the chloor emission line (837 nm) of a wine extract. RTL conditions were adapted to the analysis and the identities were searched via a pesticide RTL database. A retention time window of 0.5 (± 0.25 min from the measured retention time) and the presence of chloor were introduced in the RTL dialogue box. From the search for the main peak eluting at 21.93 min (Table 7) it is clear that the smallest deviation is obtained for procymidone. The presence of this compound could be confirmed using RTL conditions in capillary GC –

FID1 B, (97058A/004F0401.D)

FID1 B, (97058C/004F0401.D)
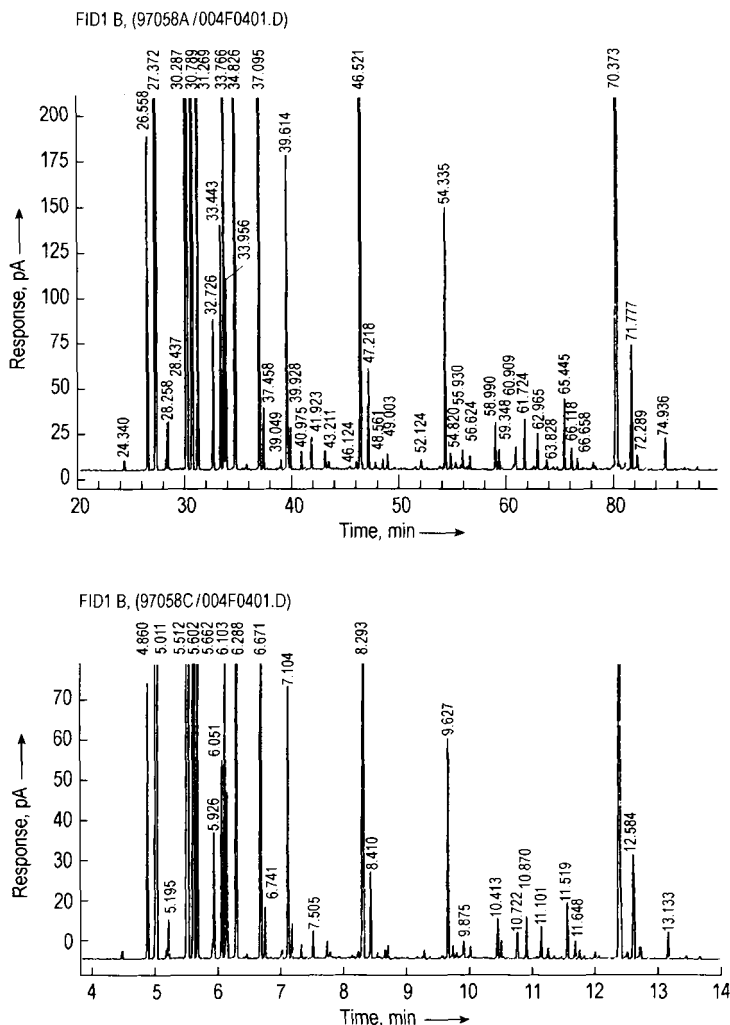
Figure 64. Analyses of Nutmeg oil on a 60 m × 0.25 mm i.d. × 1 μm MeSi column (top trace) and on a 20 m × 0.1 mm × 0.4 μm MeSi column (bottom trace)

MS. The other chlorine compounds, detected by capillary GC – AED analysis, could not be identified directly by capillary GC – MS. Owing to the complexity of the chromatogram no clear spectra were obtained. For these compounds, the result screener was used. Hereby, the TIC is searched for specific ions for all pesticides present in the database (567 pesticides) in their elution window. The peak at 25.64 min in the AED trace was identified as p,p'-DDD. The peak at 28.40 min was identified as iprodione. The peak at 29.76 min could not be identified since no ions corresponding to a pesticide present in the database and eluting in this time window could be detected.

This example clearly demonstrates the complementary nature of capillary GC – AED and capillary GC – MS. The selectivity of the AED results in simple profiles and allows fast screening of pesticides in samples. Capillary GC – MS in combination with the result screener allows confirmation of the detected solutes. Analysis of the samples with capillary GC – MS only results in complex chromatograms and more time consuming data interpretation is needed. Retention time locking also allows the correct matching of the capillary GC – AED chromatograms with the capillary GC – MS chromatograms. The corresponding peaks elute at virtually the same retention times.

PEST0027.D\AED2B
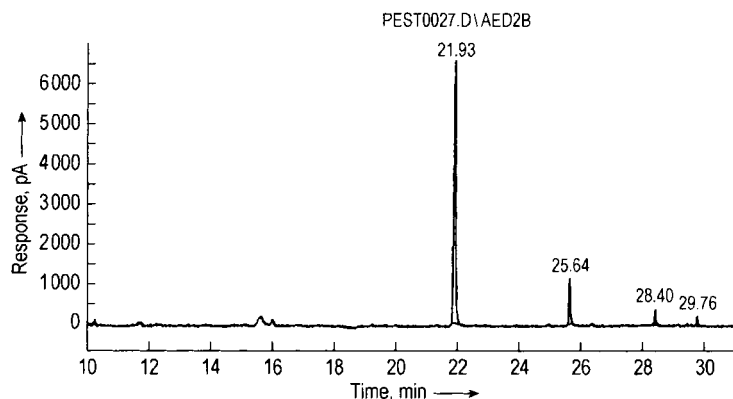
Figure 65. Capillary GC – AED chromatogram at the 837 nm emission line of a white wine extract recorded under RTL conditions

Table 7. Results of retention time locked pesticide data base search

Search results for 21.939 +/– 0.250 min
Contains elements: Cl
Does not contain elements: {no restriction}
RTT file searched: C:\HPCHEM®TL\HPGCPST.RTT

FID_RT	Compound name	MW	Formula	Δ RT (min)
21.825	Chlorbenside	269.19	C:13,H:10,Cl:2,S:1.	–0.114
21.971	Procymidone	284.14	C:13,H:11,Cl:2,N:1,O:2.	+0.032
22.037	trans-Chlordane	409.78	C:10,1,Cl:6,H:8.	+0.098
22.95	Chlorflurecol-Me- ester	274.70	C:15,H:11,Cl:1,O:3,	+0.156

Search results for 25.649 +/– 0.250 min
Contains elements: Cl
Does not contain elements: {no restriction}
RTT file searched: C:\HPCHEM®TL\HPGCPST.RTT

FID_RT	Compound name	MW	Formula	Δ RT (min)
25.409	Chlorobenzilate	325.19	C:16,H:14,Cl:2,O:3,	– 0.240
25.434	Chloropropylate	339.22	C:17,H:16,Cl:2,O:3,	– 0.215
25.581	Diniconazole	326.23	C:15,H:17,Cl:2,N:3,O:1,	– 0.068
25.654	Cyprofuram	279.72	C:14,H:14,Cl:1,N:1,O:3.	+0.005
25.703	p,p′-DDD	320.05	C:14,H:10,Cl:4,	+0.054
25.763	Etaconazole	328.20	C:14,H:15,Cl:2,N:3,O:2.	+0.114
25.786	o,p′-DDT	354.49	C:14,H:9,Cl:5,	+0.137
25.823	Flamprop-isopropyl	363.82	C:19,H:19,Cl:1,F:1,N:1,O:3.	+0.174

Search results for 28.404 +/– 0.250 min
Contains elements: Cl
Does not contain elements: {no restriction}
RTT file searched: C:\HPCHEM®TL\HPGCPST.RTT

FID_RT	Compound name	MW	Formula	Δ RT (min)
28.239	Endrin ketone	380.91	C:12,H:8,Cl:6,O:1,	–0.165
28.362	Benzoylprop ethyl	366.24	C:18,H:17,Cl:2,N:1,O:3.	–0.042
28.424	Iprodione	330.17	C:13,H:13,Cl:2,N:3,O:3.	+0.020
28.436	Dichlorophen	269.13	C:13,H:10,Cl:2,O:2.	+0.032
28.545	Leptophos oxon	396.00	C:13,H:10,Br:1,Cl:2,O:3,P:1,	+0.141
28.572	Chlorthiophos sulfox-ide	377.24	C:11,H:15,Cl:2,O:4,P:1,S:2,	+0.188

Search results for 29.768 +/– 0.250 min
Contains elements: Cl
Does not contain elements: {no restriction}
RTT file searched: C:\HPCHEM®TL\HPGCPST.RTT

FID_RT	Compound name	MW	Formula	Δ RT (min)
29.699	Phosalone	367.80	C:12,H:15,Cl:1,N:1,O:4,P:1,S:2,	–0.069
29.754	Leptophos	412.06	C:13,H:10,Br:1,Cl:2,O:2,P:1,S:1,	–0.014
29.848	Mirex	545.55	C:10.Cl:12.	+0.080

11.10.3. Towards Black Boxes

State-of-the-art GC offers column efficiencies ranging from 5000 to 500 000 plates (packed or short megabore columns to long narrowbore columns), with sample capacities from microgram to subnanogram amounts. Selectivity can be introduced in all parts of the GC system. Column selectivity can be adapted to the specific need by selecting the most suitable *stationary phase*. To enhance resolution of critical pairs, column selectivity can be optimized by *selectivity tuning* between two columns coated with different stationary phases or by *two-dimensional CGC*. *Multidimensional* CGC, HPLC – CGC, SFE – CGC, etc., provide preseparation and enrichment of relevant fractions. In addition to universal inlets a number of *selective inlets* are available in GC. GC enjoys universal detection and sensitive, selective detection. Selective recognition can be performed routinely by CGC – mass spectroscopy, CGC – Fourier transform infrared spectroscopy, and CGC – atomic emission detection, to provide sensitive and selective quantitation of target compounds and structural elucidation or identification of unknowns.

There is much more to come, such as comprehensive capillary GC, hyphenation with time-of-flight MS (capillary GC – TOFMS) and inductively coupled plasma MS (capillary GC – ICP-MS) ... to mention a few recent technological realizations.

Notwithstanding all these capabilities of GC, the data are often disappointing, as evidenced by results from round-robin tests. Deviations can be as high as 100 ppm for BTX measurements in air in the 100 ppm level, as high as 30 ppb for trihalomethanes at the 50 ppb level in drinking water, as high as 60 ppt (ppt = 1 part in 10^{12}) for organochloropesticides in the 70 ppt level in wastewater, etc. How should data such as 3.34 ppt 2,3,7,8 TCDD be interpreted?

What are the reasons for this? Two main aspects can be identified. First of all, there is an educational problem. Technicians in the laboratory often lack experience and expertise, and moreover time and money for training are unavailable. On the other hand, shift of good staff now occurs very quickly without time to transfer the know-how. The second aspect is precolumn errors (e.g., sample introduction and sample preparation), and both parts of the total analytical procedure are interrelated. Future developments in

GC are therefore expected in the direction of fully automated sample preparation and sample introduction systems. Some instrumentation nowadays is already tailor-made for a given application. More can be expected in the future and development may result in specific instruments, such as a pesticide analyzer, a PCB analyzer, or a sugar analyzer.

11.11. References

General textbooks on chromatography and gas chromatography

[1] J. C. Giddings: *Unified Separation Science*, J. Wiley & Sons, New York 1991.
[2] C. F. Poole, S. K. Poole: *Chromatography Today*, Elsevier, Amsterdam 1991.
[3] W. Jennings: *Analytical Gas Chromatography*, Academic Press, San Diego 1987.
[4] G. Schomburg: *Gas Chromatography, A Pratical Course*, VCH Verlagsgesellschaft, Weinheim 1990.
[5] D. Rood: *A Practical Guide to the Care, Maintenance and Troubleshooting of Capillary Gas Chromatographic Systems*, Hüthig Verlag, Heidelberg 1991.
[6] M. L. Lee, F. J. Yang, K. D. Bartle: *Open Tubular Column Gas Chromatography, Theory and Practice*, J. Wiley & Sons, New York 1984.
[7] A. Van Es: *High Speed, Narrow-Bore Capillary Gas Chromatography*, Hüthig Verlag, Heidelberg 1992.

Specific References

[8] A. T. James, A. J. P. Martin, *Biochem. J.* **50** (1952) 679.
[9] L. S. Ettre, *Pure Appl. Chem.* **65** (1993) no. 4, 819.
[10] J. J. Van Deemter, F. J. Zuiderberg, A. Klinkenberg, *Chem. Eng. Sci.* **5** (1956) 271.
[11] M. J. E. Golay in D. Detsy (ed.): *Gas Chromatography*, Butterworths, London 1958, p. 36.
[12] P. Sandra, *J. High Resolut. Chromatogr.* **12** (1989) 82.
[13] P. Sandra, *J. High Resolut. Chromatogr.* **12** (1989) 273.
[14] R. D. Dandeneau, E. H. Zerenner, *J. High Resolut. Chromatogr. Comm.* **2** (1979) 351.
[15] M. Verzele, P. Sandra, *J. Chromatogr.* **158** (1978) 111.
[16] P. Sandra, *LC.GC* **5** (1987) 236.
[17] H. Rotzsche: *Stationary Phases in Gas Chromatography*, Elsevier, Amsterdam 1991.
[18] W. O. McReynolds, *J. Chromatogr. Sci.* **8** (1970) 685.
[19] J. de Zeeuw, R. de Nijs, L. Henrich, *J. Chromatogr. Sci.* **25** (1987) 71.
[20] J. R. Mann, S. T. Preston, *J. Chromatogr. Sci.* **11** (1973) 216.
[21] T. J. Stark, P. A. Larson, R. D. Dandeneau, *J. Chromatogr.* **279** (1983) 31.
[22] P. Sandra et al., *J. High Resolut. Chromatogr. Chromatogr. Commun.* **8** (1985) 782.
[23] P. Sandra, F. David in H. J. Cortes (ed.): *Multidimensional Chromatography, Techniques and Applications*, Marcel Dekker, New York 1990, p. 145.
[24] K. E. Markides et al., *Anal. Chem.* **57** (1985) 1296.
[25] W. A. König: *The Practice of Enantiomer Separation by Capillary Gas Chromatography*, Hüthig Verlag, Heidelberg 1987.

[26] W. A. König: *Gas Chromatographic Enantiomer Separation with Modified Cyclodextrins*, Hüthig Verlag, Heidelberg 1992.

[27] D. E. Bautz, J. W. Dolan, W. D. Raddatz, L. R. Snyder, *Anal. Chem.* **62** (1990) 1560.

[28] G. N. Abbay et al., *LC.GC* Int. 4 (1991) 28.

[29] E. Sippola, F. David, P. Sandra, *J. High Resolut. Chromatogr.* **16** (1993) 95.

[30] B. W. Hermann et al., *J. High Resolut. Chromatogr.* **13** (1990) 361.

[31] S. S. Stafford (ed.): *Electronic Pressure Control in Gas Chromatography*, Hewlett-Packard, Little Falls 1993.

[32] V. Pretorius, W. Bertsch, *J. High Resolut. Chromatogr. Chromatogr. Commun.* **6** (1983) 64.

[33] R. Jenkins, W. Jennings, *J. High Resolut. Chromatogr. Chromatogr. Commun.* **6** (1983) 228.

[34] P. Sandra in P. Sandra (ed.): *Sample Introduction in Capillary Gas Chromatography*, Hüthig Verlag, Heidelberg 1985, p. 1.

[35] M. S. Klee: *GC Inlets, An Introduction*, Hewlett-Packard, Little Falls 1990.

[36] K. Grob: *Classical Split and Splitless Injection in Capillary GC*, Hüthig Verlag, Heidelberg 1986.

[37] K. Grob: *On-Column Injection in Capillary GC*, Hüthig Verlag, Heidelberg 1987.

[38] P. Sandra in K. Hyver (ed.): *High Resolution Gas Chromatography*, 3rd ed., Hewlett-Packard, Little Falls 1987.

[39] D. Desty, A. Goldup, H. Whyman, *J. Inst. Pet.* **45** (1959) 287.

[40] W. D. Snyder, Technical Paper no. 108, Hewlett-Packard, Little Falls 1990.

[41] P. Wyllie, K. J. Klein, M. Q. Thompson, B. W. Hermann, *J. High Resolut. Chromatogr.* **14** (1991) 361.

[42] F. David, P. Sandra, S. S. Stafford, B. Slavica, Hewlett-Packard Application Note 228–223, Little Falls 1993.

[43] A. Zlatkis, H. R. Kaufman, *Nature (London)* **184** (1959) 4010.

[44] P. Sandra in F. Bruner (ed.): *The Science of Chromatography*, Elsevier, Amsterdam 1985, p. 381.

[45] K. Grob, K. Grob Jr., *J. Chromatogr.* **151** (1978) 311.

[46] K. Grob, *J. High Resolut. Chromatogr. Chromatogr. Commun.* **1** (1978) 263.

[47] G. Schomburg et al., *J. Chromatogr.* **142** (1977) 87.

[48] P. Sandra, M. Van Roelenbosch, M. Verzele, C. Bicchi, *J. Chromatogr.* **279** (1983) 287.

[49] F. David, L. Vanderroost, P. Sandra, S. Stafford, *Int. Labmate* **17** (1992) no. 3, 13.

[50] E. Meyer, A. P. De Leenheer, P. Sandra, *J. High Resolut. Chromatogr.* **15** (1992) 637.

[51] W. Vogt, K. Jacob, H. W. Obwexer, *J. Chromatogr.* **174** (1979) 437.

[52] G. Schomburg in R. E. Kaiser (ed.), *Proc. Int. Symp. Capillary Chromatogr. 4th* 1981, 371, 921 A.

[53] F. Poy, S. Visani, F. Terrosi, *J. Chromatogr.* **217** (1981) 81.

[54] F. David, P. Sandra, A. Hoffmann, J. Gerstel, *Chromatographia* **34** (1992) no. 5–8, 259.

[55] R. Buffington, M. K. Wilson: *Detectors for Gas Chromatography, A Practical Primer*, Hewlett-Packard, Little Falls 1987.

[56] H. H. Hill, D. G. McMinn (eds.): *Detectors for Capillary Chromatography*, J. Wiley & Sons, New York 1992.

[57] I. Temmermann, F. David, P. Sandra, R. Soniassy, Hewlett-Packard Co., Application Note 228–135, Little Falls 1991.

[58] F. W. Karasek, R. E. Clement: *Basic Gas Chromatography—Mass Spectroscopy*, Elsevier, Amsterdam 1988.

[59] W. Herres: *Capillary Gas Chromatography—Fourier Transform Infrared Spectroscopy*, Hüthig Verlag, Heidelberg 1987.

[60] R. Buffington: *GC-Atomic Emission Spectroscopy Using Microwave Plasmas*, Hewlett-Packard, Little Falls 1988.

[61] E. Kovats, J. C. Giddings, R. A. Keller in: *Advances in Chromatography*, vol. 1, Marcel Dekker, New York 1965, chap. 7.

[62] H. Van den Dool, P. D. Kratz, *J. Chromatogr.* **11** (1963) 463.

[63] W. Jennings, T. Shibamoto: *Qualitative Analysis of Flavor and Fragrance Volatiles by Gas Chromatography*, Academic Press, New York 1980.

[64] J. F. Sprouse, A. Varano, *Am. Lab. (Fairfield, Conn.)* **16** (1984) 54.

[65] H. J. Cortes (ed.): *Multidimensional Chromatography, Techniques and Applications*, Marcel Dekker, New York 1990, p. 145.

[66] C. Bücchi, A. D'Amato, F. David, P. Sandra, *J. High Resolut. Chromatogr.* **12** (1989) 316.

[67] K. Grob: *On-Line Coupled LC-GC*, Hüthig Verlag, Heidelberg 1991.

[68] F. David, P. Sandra, D. Bremer, R. Bremer, F. Rogies, A. Hoffmann, *Labor Praxis* **21** (1997) 5.

[69] M. Verschuere, P. Sandra, F. David, *J. Chromatogr. Sci.* **30** (1992) 388.

[70] F. David, M. Verschuere, P. Sandra, *Fresenius J. Anal. Chem.* **344** (1992) 479.

[71] K. Blau, G. S. King: *Handbook of Derivatization for Chromatography*, Heyden (Wiley & Sons), New York 1977.

[72] J. Drozd: *Chemical Derivatization in Gas Chromatography*, Elsevier, Amsterdam 1981.

[73] P. Sandra, F. David, *LC.GC* **7** (1989) 746.

[74] L. S. Ettre, J. V. Hinshaw: *Basic Relationships of Gas Chromatography*, Advanstar, Cleveland 1993.

[75] B. D. Quimby, V. Giarrocco, M. S. Klee, Hewlett-Packard Application Note 228-294, February 1995.

[76] F. David, D. R. Gere, F. Scanlan, P. Sandra, *J. Chromatogr. A* **842** (1999) 309.

[77] V. Giarrocco, B. D. Quimby, M. S. Klee, Hewlett-Packard Application Note 228-392, December 1997.

[78] B. D. Quimby, L. M. Blumberg, M. S. Klee, P. L. Wylie, Hewlett-Packard Application Note 228-401, May 1998.

12. Liquid Chromatography

PETER LEMBKE, Belovo-Spain, S.L., Reus, Spain (Chaps. 12, 12.3 – 12.10, Sections 12.2.1 – 12.2.6.7)

GÜNTER HENZE, Institut für Anorganische und Analytische Chemie der Technischen Universität Clausthal (Section 12.2.6.8)

KARIN CABRERA, Merck KGaA, Darmstadt, Federal Republic of Germany (Chap. 12.11)

WOLFGANG BRÜNNER, formerly Merck KGaA, Darmstadt, Federal Republic of Germany (Chap. 12.11)

EGBERT MÜLLER, Merck KGaA, Darmstadt, Federal Republic of Germany (Chap. 12.11)

Abbreviations

2′,5′-ADP	adenosine 2′,5′-diphosphate
5′-AMP	adenosine 5′-monophosphate
A_{ref}	area or height of the reference peak in the standard solution
A_{si}	area or height of component i in the sample
AUFS	bsorption units full scale
C_C	cell capacitance
C_D	double-layer capacitance
CRD	chemical reaction detector
DAD	diode-array detector
D_m	diffusion coefficient
dp	particle size
E	molecular extinction coefficient
ECD	electron capture detector
ELSD	evaporative light-scattering detector
ES	electrospray
F	flow rate
FAB	fast ion bombardment
FD	field desorption
FID	flame ionization detector
FSD	fast-scanning detector
FTIR	Fourier transform IR spectroscopy
GFC	gel filtration chromatography
GPC	gel permeation chromatography
H	plate height, band height
HPIC	high-performance ion chromatography
HPSEC	high-performance size exclusion chromatography
I^*	radiation
I_0/I	transmittance
IEC	ion exclusion chromatography
IEX	ion-exchange chromatography
IMAC	immobilized metal ion affinity chromatography
ISE	ion-selective electrode
K	distribution coefficient
k'	capacity ratio
K_{ij}^{Pot}	selectivity factor
L	length
LC	liquid chromatography
LIF	laser-induced fluorescence
LLC	liquid–liquid chromatography
LSC	liquid–solid chromatography
LSD	light-scattering detector
MSD	mass-selective detector
MWD	multiwavelength detector

N	plate number
NAD	nicotinamide–adenine dinucleotide
NADP	nicotinamide–adenine dinucleotide phosphate
NARP	nonaqueous reversed phase
NP	normal phase
PC	paper chromatography
RI	refractive index
RIG	refractive index gradient
RIU	refractive index units
RP	reversed phase
R_S	electrolyte resistance
Rs	resolution
SAX	strong anion exchanger
SCX	strong cation exchanger
SEC	size exclusion chromatography
SFC	supercritical fluid chromatography
SFE	supercritical fluid extraction
t_0	time analyte spends in the mobile phase
TID	thermoionic detector
TOF	time of flight
t_R	retention time
t_R	time analyte spends in the stationary phase
u	velocity
W	peak width
WAX	weak anion exchanger
WCX	weak cation exchanger
WE	working electrode
z_i	atomic number of ion i
z_j	atomic number of interfering ion j
α	selectivity
Γ	correction factor
λ	correction factor, wavelength
κ	conductivity
ω	angular frequency, correction factor

12.1. General

Liquid chromatography is a separation technique based on a different distribution rate of sample components between a stationary and a liquid

mobile phase. Depending on the stationary phase, a distinction is made between thin layer chromatography (TLC), paper chromatography (PC), and liquid column chromatography (LC). Liquid column chromatography can be divided into open and closed systems. Generally an *open system* consists of a glass column packed with large-sized (≈ 30 µm and larger) particles of the stationary phase onto which the mobile phase is poured. Because of the large particles, the mobile phase can pass the stationary phase only by means of gravity, sometimes assisted by suction. Therefore, no solvent pump is necessary.

However, to increase separation efficiency, high velocities and small-diameter (≤ 10 µm) particles are essential as the stationary phase. Unfortunately, this leads to an increase in back-pressure. To overcome this back-pressure, a pump is necessary for transport of the mobile phase (high-pressure- or high-performance liquid chromatography, HPLC). Separation is performed in a *closed system*.

HPLC is the LC technique with the highest efficiency (ability to separate different sample compounds in a given time with a given resolution). The selectivity of this technique is based not only on the different types of stationary phases but also on the mobile phase. An advantage of gas chromatography (GC) is that selectivity in this technique is based only on the stationary phase. (In GC, interactions occur only between the solute and the stationary phase. In LC, additional interactions between the solute and the liquid mobile phase are found.) Furthermore GC can be used only for samples that can be volatized without decomposition at temperatures up to ca. 400 °C. This limits the use of GC to about 20 % of the samples of interest [1]. The limit of LC is set by the solubility of the sample in the mobile phase. Since nearly every substance is soluble in some kind of a solvent, the application range of LC is very broad. In contrast, capillary GC ($\rightarrow$ Gas Chromatography) has a far higher separation efficiency than HPLC.

Neither technique can be said to be better than the other. Both GC and HPLC have their advantages for particular separation problems.

In LC, two sample introduction techniques can be distinguished the continuous and the discontinuous mode. *Continuous sample introduction*, also known as front analysis or adsorptive filtration, is often used for solvent cleanup. In this technique the sample solution is introduced continuously onto the column; thus only the least-retained compound is obtained in the pure state.

Discontinuous Sample Introduction. Discontinuous sample introduction can divided into elution analysis and displacement analysis. In *elution analysis*, the sample is injected into a continuous mobile phase. If the composition of the mobile phase remains unchanged during the analysis this is called the *isocratic mode*. In *gradient elution* the composition of the mobile phase is changed in such a way that the elution strength is increased (or sometimes also decreased) during analysis. In the discontinuous sample introduction mode, an elution diagram (chromatogram) is obtained in which, ideally, each compound is separated by a pure solvent zone from the next eluting compound.

Displacement analysis is often used in preparative chromatography. After sample injection, the composition of the mobile phase is changed, and the interaction between the mobile and the stationary phases becomes so dominant that the retained sample compounds are displaced by solvent molecules and move with identical velocity (i.e., that of the eluent).

12.1.1. History

In 1905, Tswett, a Russian botanist, published the first paper on liquid chromatography [2], in which he described the separation of plant pigments by means of open-column liquid chromatography. The importance of this new technique was recognized only after Kuhn (Nobel prize, 1938) and Lederer used it for the preparative separation of α- and β-carotin from carrots in 1931 [3], [4]. In 1941 Martin and Synge [5] (Nobel prize, 1951) described the technique of partition chromatography (liquid–liquid chromatography, LLC), and in 1949 Spedding [6] and Tompkins [7] published the first paper on ion-exchange chromatography (IEX). In 1959, Porath and Flodin [8] introduced size exclusion chromatography (SEC). One of the first papers on high-performance liquid chromatography was published by Piel [9] in 1966. Since then, HPLC has become one of the most frequently applied analytical separation methods. Especially because of the rapid development of instrumentation (pumps, detectors, stationary phase, gradient elution, etc.) and column packings during the 1970s and 1980s, very high selectivity and efficiency can be obtained. Today, HPLC is one of the most versatile techniques in the analytical (e.g., trace analysis) and preparative fields.
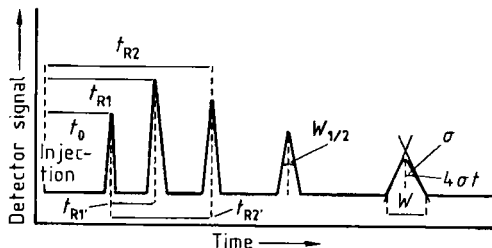
Figure 1. Characteristic information obtainable from a chromatogram

12.1.2. Definition and Theoretical Background (→Basic Principles of Chromatography)

In a chromatographic column, two processes occur simultaneously:

1) The separation process, which is based mainly on the interactions between solute and stationary phase
2) Dilution of the solute in the mobile phase due to diffusion processes; this effect counteracts separation

To understand these effects, several important chromatographic definitions are discussed in this section. Figure 1 shows a typical chromatogram with the most important parameters characterizing a chromatographic separation.

The *retention time* t_R of a particular component is defined by:

$$t_R = t_0 + t_{R'}$$

where t_0 is the time the analyte spends in the mobile phase and $t_{R'}$ is the time the analyte spends in the stationary phase. For given HPLC equipment and a constant flow rate (milliliters per minute), all analytes reside in the mobile phase for the same time t_0. They differ only in the time $t_{R'}$ they spend in the stationary phase. To become independent of column dimension and mobile phase flow the *capacity ratio* k' was introduced:

$$k' = \frac{t_R - t_0}{t_0} = \frac{t_{R'}}{t_0} \quad \frac{\text{time spent in stationary phase}}{\text{time spent in mobile phase}}$$

Another important chromatographic quantity is the relative retention or selectivity of a chromatographic system. The *selectivity* α is defined by equation:

$$\alpha = \frac{t_{R2} - t_0}{t_{R1} - t_0} = \frac{k_2}{k_1} = \frac{t_{R2'}}{t_{R1'}}$$

This represents the ability of a given chromatographic system to separate two components. If $\alpha = 1$, no separation is possible, no matter how high the separation efficiency may be. As in GC, the selectivity can be influenced by the choice of the stationary phase, but in LC the selection of the mobile phase is also a factor. In preparative chromatography the selectivity α is the most important chromatographic parameter to be optimized.

A further quantitative measure for determining the separation of two components in a given sample is the *resolution Rs*.

$$Rs = \frac{2(t_{R2} - t_{R1})}{(W_1 + W_2)}$$

where W is the peak width. A resolution of $Rs = 1.5$ (baseline resolution) is ideal for quantification of particular sample peaks. Yet even with a resolution of $Rs = 1.0$, quantification is possible because only ca. 2 % of the peak areas overlap. Chromatograms with resolutions of less than 1.0 should not be used for quantitative analysis. On the other hand, an $Rs > 1.5$ is not necessary, since no advantage is achieved in quantification, but analysis time is prolonged. The above values for Rs are valid only for Gaussian peaks having the same size. Nevertheless, they can assist and guide the chromatographer.

Other important parameters in HPLC are the plate height H and the plate number N. *Plate height* is calculated from a chromatogram by

$$H = \frac{L}{5.54} \left(\frac{W_{0.5}}{t_R} \right)^2$$

where $W_{0.5}$ is the peak width at half maximum peak height; t_R the retention time; and L the length of the column.

Well-packed HPLC columns should have a plate height of $2 - 4\,dp$ (dp = stationary phase particle size).

The *plate number* is an indicator of column quality. The higher it is, the better is the peak capacity (number of peaks separated in a given time) of the column. It is calculated by:

$$N = \frac{L}{H}$$

The plate number can be increased either by reducing the particle size of the stationary phase
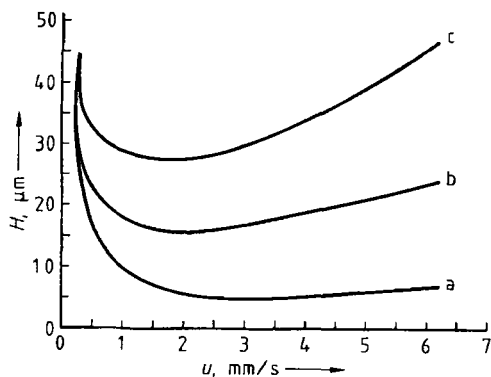
Figure 2. Typical Van Deemter H/u curves
a) 3 μm; b) 5 μm; c) 10 μm

dp and thus the achievable plate height or by lengthening the column. Both have the disadvantage of increasing the pressure drop over the column as well. Unfortunately, the equipment in most HPLC pumps, does not work at back-pressures exceeding 40–50 MPa ("high-pressure shutdown").

The relationship among most of the aforementioned chromatographic parameters is given by the following equation:

$$Rs = \frac{1}{4} \frac{\alpha - 1}{\alpha} \frac{k_2'}{1 + k_2'} \sqrt{N}$$

This is the most important equation in LC because it combines the chromatographic separation parameters selectivity, capacity ratio, and plate number. Since these three terms are more or less independent, one term can be optimized after the other. The selectivity term can be optimized by changing the stationary or the mobile phase; k' can be influenced by the solvent strength of the mobile phase, and N can be adjusted by changing the column length, particle diameter, or solvent linear velocity u.

$$u = \frac{L}{t_0}$$

The different processes contributing to the overall band broadening are described by the Van Deemter equation [10]

$$H = A + B/u + Cu$$

where A is the eddy diffusion term ($\approx 2\,\lambda dp$), B the longitudinal (axial) diffusion term ($\approx 2\,\gamma D_m$), C the mass transfer resistance term ($\approx \omega dp^2/D_m$), and u the linear mobile phase velocity.

$$H = 2\lambda dp + 2\gamma D_m/u + (dp^2 \omega/D_m)u$$

where dp is the mean particle diameter; D_m the diffusion coefficient in the mobile phase, and λ, γ, ω are correction factors.

For liquid chromatography the following conclusions can be drawn from the Van Deemter equation: In contrast to gas chromatography, the C term (which increases tremendously with the linear velocity of the mobile phase) is very important in liquid chromatography because the diffusion coefficients in the fluid phase are much lower than in GC. A good HPLC column should have as low a C as possible (i.e., a small slope in the H versus u curve; see Fig. 2). In LC the B term is not as important as in GC because of the low diffusion coefficients. The B term decreases with increasing linear velocity u. The A term is independent of linear velocity; it depends only on particle size dp. Particle size is very important in LC (the A and C terms). The smaller the particle size of the stationary phase, the smaller is the achievable plate height and the higher is the plate number of a particular column. The development of uniform small particles for stationary phases in LC is one reason for the booming development of HPLC since the 1970s. The smallest particle size currently available is 1.5 μm (standard sizes: 5, 7, and 10 μm). Particles < 3 μm are not so popular because the back-pressure of the column increases dramatically with decreasing particle diameter, while packing the column becomes very difficult. As a rule, reducing the particle diameter by half increases the back-pressure of the column to the second power whereas the resolution Rs improves by only a factor of 1.4. Nevertheless, short columns (2 cm) packed with 1.5 μm material do allow extremely fast analysis combined with low consumption of mobile phase and can be very useful for laboratories that have to deal with a large number of samples every day. Because of their nonporous nature, these small particles have a very low capacity, which may occasionally result in column-overload problems (peak tailing, poor resolution, etc.).
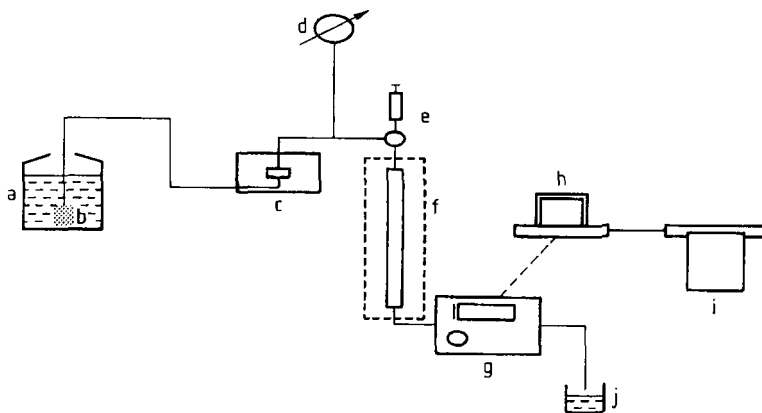
Figure 3. Diagram of a simple HPLC system
a) Solvent reservoir; b) Frit; c) Pump; d) Manometer; e) Injection port; f) Separation column with optional oven; g) Detector;
h) Data processing unit; i) Integrator; j) Waste

12.2. Equipment

To obtain separations sufficient for quantitative analysis in liquid chromatography, good equipment is essential. Only the most important parameters are discussed in this chapter. For further information, see [11]–[15] and other monographs on LC. Figure 3 shows a schematic of a simple HPLC system.

12.2.1. Filters and Connecting Tubing

The connecting tubing between solvent reservoir and pump is usually made of polytetrafluoroethylene (PTFE) with an inner diameter i.d. of 2–4 mm. The end of each pump–inlet tube should be fitted with an inlet-line filter (frit), which keeps out large, inadvertent contaminants and holds the inlet line at the bottom of the reservoir so that no air (bubbles) is drawn into the pump head. These frits, which are normally made of stainless steel (recently also of glass or PTFE), usually have a pore diameter of 5–10 μm and can be cleaned easily with 5 % nitric acid.

The tubing connecting the sample introduction unit, column, and detector consists of small-volume stainless steel capillaries with an inside diameter of 0.25 mm and an outer diameter of 1.25 mm (1/16 inch). For bioseparations that are sensitive to stainless steel, tubing made from polyetheretherketone (PEEK) is recommended. All connecting tubing, especially between sample introduction system and column, should have a small vol-

ume (small inside diameter and shortest length possible), to minimize extra column band broadening and loss in separation efficiency.

12.2.2. Containment of Mobile Phase

Depending on the physicochemical properties of the mobile phase (eluent), different containers are used. The most common are brown laboratory bottles for the protection of light-sensitive eluents. If inorganic ions are to be monitored, polyethylene containers should be used because additional inorganic ions may be dessolved from the glass, thus leading to false results. Safety regulations concerning toxic solvents, fire prevention, etc., must also be taken into account [16].

12.2.3. Pumps

The pump is one of the most important parts of an HPLC system. It is responsible for constant flow of the mobile phase, which is necessary for the quantification and reproducibility of an analysis. The retention time t_R can be correlated with the retention volume V_R and the flow rate F:

$$V_R = F\, t_R$$

Therefore a change in the flow rate or strong pulsation changes the retention time of an analyte, making identification by t_R impossible, because the results are not reproducible. Two types of pump pulsation occur that should be kept as low
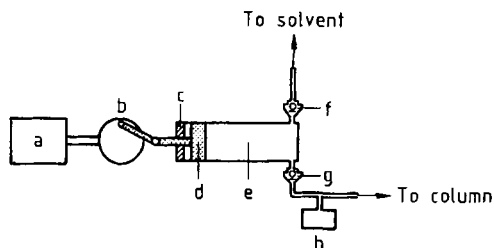
Figure 4. Diagram of a simple reciprocating pump [19]
a) Motor; b) Camshaft; c) Seal; d) Piston; e) Chamber; f) Inlet check valve; g) Outlet check valve; h) Damper

Figure 5. Flow profile of a double-head reciprocating pump (phase shift = 180°)
Courtesy of Laboratory Control/Milton Roy

as possible: The *long-time pulsation* (causes irreproducible retention times) and the *short-time pulsation* (causes irreproducible peak areas). The effects of flow instabilities due to pump pulsation and their measurement are discussed in [17].

Good HPLC pumps should be able to deliver a flow rate of 0.5 – 10 mL/min, with a maximum pressure of ca. 40 – 50 MPa, a constant flow rate with variation of < 0.5 % (see also [18]), and a rest pulsation of < 0.1 %. In addition, they should be equipped with adjustable high- and low-pressure limits to avoid damage to the pump or column; should have a solvent compressibility control; must be easy to maintain; and should be chemically inert. The pump heads should not warm up (evaporation of solvent and bubbles), and the pumps should, preferably, not be too noisy.

Basically, two types of pumps exist: Constant-pressure pumps and constant-flow pumps. Pressure-constant pumps are often used to pack HPLC columns.

Modern HPLC systems work with pumps that deliver a constant flow independent of the back-pressure of the column. These *constant-flow pumps* can be divided into continuous and discontinuous systems. In principle, discontinuous pumps work like a huge syringe and have lost their importance in HPLC. Currently, most HPLC pumps work in the continuous mode to deliver a constant flow. In these reciprocating pumps a plunger piston causes suction of the eluent during the back stroke and delivers the solvent under pressure during the forward stroke. To ensure proper flow direction, the pump head is usually equipped with ball check valves (see Fig. 4 f, g). Single-head, one-and-a-half-head, double-head and triple-head reciprocating pumps are available commercially. To reduce the pulsation caused by reciprocating pumps, many different techniques
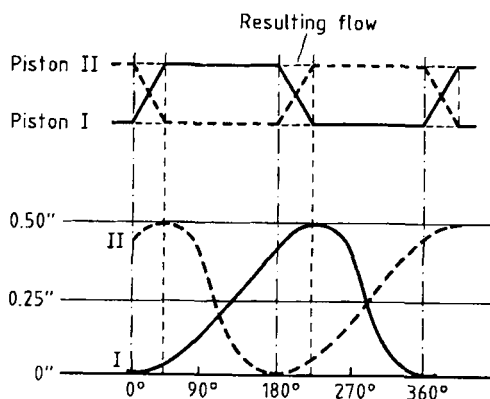
are applied. For double- and triple-head reciprocating pumps, two or three pistons are used at the same time with a 180° or 120° phase shift between each. The resulting mobile phase flow is more or less periodic (see Fig. 5). The rest pulsation remaining can be decreased further if piston movement is not sinusoidal. Because of an eccentric disk, the second piston starts to deliver before the first has reached its end position. At the end position, the second piston has reached its maximum delivery output, while the first piston is refilled during its back stroke.

12.2.4. Sample Introduction Units

Good column performance is correlated directly with proper sample introduction onto the column in form of a narrow plug. Basically, the sample can be introduced into the high-pressure system in two ways: By syringe injectors or by sample valves.

12.2.4.1. Syringe Injectors

Syringe injection is the simplest way to introduce a sample to the system. The technique is very similar to that applied in GC. The sample is transferred with a syringe, through a self-sealing elastomeric septum, into the pressurized column. A specially designed pressure-resistant microsyringe must be used. One disadvantage of this technique is that column inlet plugging through particles of the septum can occur and the reproducibility of
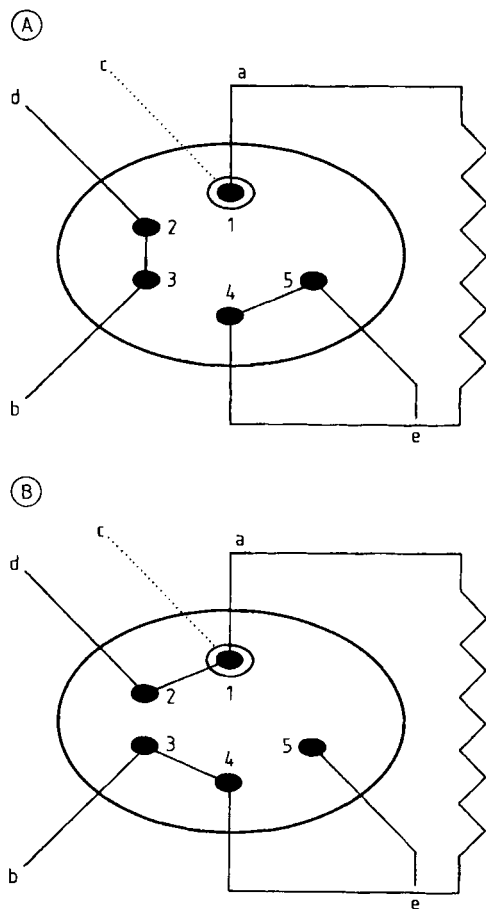
Figure 6. Diagram of a sample loop injection system [20]
A) Load position; B) Inject position
a) Loop; b) Column; c) Injection port; d) Pump; e) Waste

sample injection is often less than 2 % [14]. There-fore, this technique is used rarely, and most LC systems are equipped with sample valves.

12.2.4.2. Sample Valve Injectors

Today, the most common sample valves oper-ate in the loop filling mode and are called *high-pressure sample injectors*. This means that the sample is transferred at atmospheric pressure from a syringe into a sample loop. By means of valving action the sample loop is then connected to the high-pressure mobile phase stream, which carries the sample onto the column (Fig. 6). The valving action can be carried out either manually or auto-matically by electric or pneumatic actuators.

Commercial sample loops are available from 0.2 µL (for microcolumns) to 5 mL or more. On the other hand, a sample loop can be made easily by cutting an appropriate length of capillary (with known i.d.) and connecting it to the sample injec-tor.

Normally the complete loop filling method is preferred to the partial filling method because the former has excellent volumetric precision. When loop volumes exceed 5 – 10 µL, this can be better than 0.1 % RSD. Nevertheless, the partial filling method still has a volumetric precision (dependent on syringe repeatability) better than 1 % relative standard deviation (RSD) [20].

12.2.4.3. Automatic Introduction Systems

Automatic sample introduction units (auto samplers) are often used for routine analysis (e.g., product quality control). A modern, commercially available HPLC auto sampler can handle ca. 100 samples; has a variable injection volume (e.g., from 1 to 1000 µL); can heat or cool the samples (e.g., 4 – 40 °C); and can be programmed while they are waiting to be injected; and can be pro-grammed individually for each sample. The pro-grammed run of the auto sampler should be able to be interrupted for an urgent analysis and continued later without any problems.

The principle of injection is the same as with sample loop injectors (see Section 12.2.4.2).

With an automatic sample introduction system, very efficient work is possible because sample cleanup, pretreatment (e.g., precolumn derivatiza-tion), and sample programming can performed during the day, while the actual analyses are run automatically at night.

12.2.5. Columns

Column Construction. HPLC columns (Fig. 7) are usually manufactured of stainless steel. Cur-rent standard HPLC columns have a 4.0- or 4.6-mm inner diameter and a length of 60 – 250 mm. Columns with smaller inner diame-ter (e.g., 1 or 2 mm — microbore columns) are also available. Although eluent consumption can be reduced with these columns, sensitivity is usually not increased and undesirable "wall effects" and dead volume effects become more noticeable with decreasing column diameter. Microbore columns are preferred only if the sample volume is very limited.
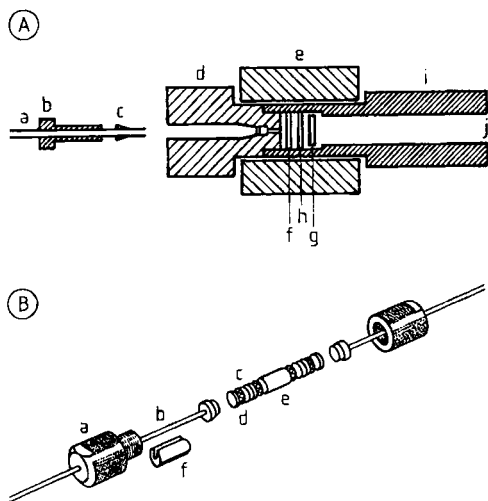
Figure 7. Diagram of the two most common column types for HPLC

A) Conventional HPLC column terminator for 1/16" fittings
a) Capillary; b) Nut; c) Ferrule; d) End fitting with internal thread; e) Coupling body to column; f) Metal frit; g) Filter paper; h) Sealing ring; i) Stainless steel tubing; j) Column packing
B) Cartrige system column
a) Compression nut; b) Column fitting adapter; c) Sieve sandwich; d) PTFE soft seal; e) Column of precolumn; f) Spacer (required when working with precolumns)
Courtesy of Bischoff-Analysentechnik, Leonberg, Germany

For preparative applications, columns with much larger dimensions (i.d. up to 900 mm and more) are available. Generally all parts of the column that do not contain the packing material, should be as small as possible to minimize dead volume. Often, unnecessarily large dead volumes result from false connection of the end fittings.

To ensure good separation, the inner column wall must be well polished and inert to the various mobile phases. Normal glass columns that are inert and have a very smooth inner wall can unfortunately be operated only at pressures less than ca. 5 MPa. To overcome this drawback, glass columns with a surrounding stainless steel jacket were developed that ensure stability at higher operating pressure. Recently, pressure-stable columns made of PEEK which show the same positive properties as glass columns, have come onto the market. These columns are gaining interest for metal-free bioseparations.

12.2.6. Detectors

Another important instrument required in modern LC is a sensitive or selective detector for continuous monitoring of the column effluent. In GC, the differences in physical properties of the mobile phase (carrier gas) and the sample are great enough for universal detectors with good *sensitivity* to be used (e.g., flame ionization detector, thermal conductivity detector, → Gas Chromatography). The problem in LC is that the physical properties of the mobile phase and the sample are often very similar, which makes the use of a universal detector impossible. Nevertheless, presently available LC detectors are very sensitive, are generally selective, and have a relatively wide range of applications (see Table 1).

To compare detectors from different manufacturers, characteristics must be found that determine the performance of the equipment. Besides instrument specifications such as cell geometry, cell volume, dimension of the connecting capillary, or additional special outfitting (e.g., integration marker, auto zero), analytical characteristics such as noise, linearity, sensitivity, and response must be considered. A recently published extensive survey of the different commercially available detectors is given in [21]. A good LC detector should have the following characteristics:

1) High sensitivity and detection limit
2) Good selectivity
3) Fast response
4) Wide range of linearity
5) No contribution to column band broadening (small dwell volume)
6) Reliability and convenience

Detector Sensitivity and Detector Noise. *Detector sensitivity* is the quantitative response of the detector signal to the amount of solute introduced (slope of the plot sample input M versus signal output A: sensitivity $= dA/dM$).

Noise originates from the electronics associated with the detector and from fluctuations of the physical parameters in the detector environment (e.g., flow pulsation or temperature change in refraction detectors, or light intensity changes in UV detectors). For further information, see [11].

Detector Linearity. Detectors differ in their linear dynamic range. The response of an ideal detector should be proportional to either the concentration or the mass flow rate of the solute (see

Table 1. Summary of the most important HPLC detectors

Detector	Type	Gradient possible	Maximum linear range	Maximum sensitivity, g/mL	Sensitive to changes in
UV – VIS	selective	yes	$10^3 - 10^5$	$10^{-9} - 10^{-10}$	pressure (temperature)
Diode-array detector (DAD)	selective	yes	$10^3 - 10^4$	$< 2 \times 10^{-5}$	flow, pressure (temperature)
Refractive index (RI) detector	universal	no	$10^3 - 10^4$	10^{-7}	temperature (10^{-4} RIU $°C$), pressure
Fluorescence – phosphorescence detector	selective	yes	$10^2 - 10^4$	$< 10^{-9}$	flow (temperature)
Fast-scanning detector	selective	yes	10^3	$< 10^{-5}$	flow, temperature
Evaporative light scattering detector (ELSD)	universal*	yes	$10^2 - 10^4$	10^{-8}	flow
Amperometric detector	selective	yes	10^6	10^{-12}	flow, temperature ($1.5 \%/°C$)
FTIR detector	selective	yes	10^4	10^{-4}	flow
Mass selective detector (MS)	universal	yes	$10^2 - 10^4$	$< 10^{-9}$**	flow

* Only universal for nonvolatile solutes.
** Grams per second.
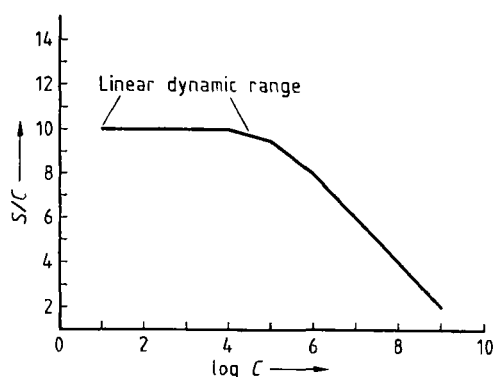
Figure 8. Plot used to determine detector linearity

Section 12.2.6.1). In reality, deviations from linearity usually occur at high concentrations (saturation effect). An easy way to determine the linear dynamic range of a detector is to plot the signal-to-concentration (or sample size) ratio versus the logarithm of concentration (sample size) (see Fig. 8).

12.2.6.1. Concentration-Sensitive and Mass-Flow-Sensitive Detectors

Two kinds of detectors exist, those that respond to solute concentration in the mobile phase and those that respond to changes of the solute mass flow rate into the detector cell. *Concentration-sensitive detectors* are used most frequently in LC. They are independent of the mass flow rate of the solute and are generally nondestructive (e.g., UV and differential detectors). (An easy way to find out whether a detector is concentration sensitive or not is to stop the solvent flow during a run. If the signal remains constant, it is a concentration-sensitive detector.)

The signal of *mass flow-rate-sensitive detectors* is proportional to the product of the solute concentration and the mobile phase flow rate. Consequently, if the flow rate is zero the signal is also zero, whatever the concentration may be. Detectors of this group are usually destructive (e.g., electrochemical detectors and mass spectrometers).

12.2.6.2. Differential Refractive Index (RI) Detector

The RI detector was developed in 1942 by TISELIUS [22] and was one of the first commercially available on-line detectors. It is concentration sensitive and nondestructive, which makes it interesting for preparative chromatography. The detection limit of 10^{-6} g/mL is not as good as that of light absorption detectors (e.g., fluorescence or UV detectors), but the RI detector (RID) is a universal detector. To enhance the detection limit, a mobile phase should be chosen that differs as much as possible in its refraction index from the sample. A great disadvantage of the RID is that no eluent gradient can be run and the detector is very sensitive to even slight temperature variations [1 °C causes a change in the refractive index of
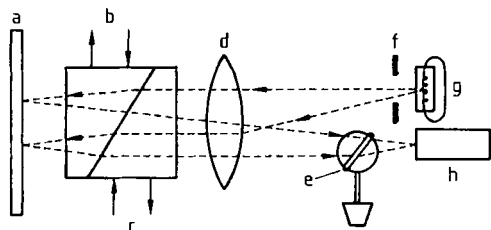
Figure 9. Diagram of a beam deflection (refractive index) detector [24]
a) Mirror; b) Sample cell; c) Reference cell; d) Lens;
e) Optical zero-point adjustment; f) Mask; g) Lamp; h) Detector

Figure 10. Diagram of a simple UV detector Z-pattern flow cell
a) Lamp; b) Flow cell; c) Quartz window; d) Photo detector

ca. 10^{-4} refractive index units (RIU)]. Therefore, very good thermostating must be provided.

The four general methods of measuring refractive index are the Christiansen, the Frensel, the beam deflection, and the interferometric methods. Today, the most commonly applied technique in HPLC is the beam deflection method.

Beam Deflection Refractive Index Detector. A schematic of the beam deflection refractive index detector is shown in Figure 9. Here, the deflection of a light beam is measured in a single compact cell. A light beam from the incandescent lamp (g) passes through the optical mask (f), which confines the beam within the region of the sample (b) and the reference cell (c). The lens (d) collimates the light beam, and the parallel beam passes through the cells to a mirror (a). The mirror reflects the beam back through the sample and reference cell to the lens, which focuses it onto a photo cell (h). The location of the focused beam, rather than its intensity, is determined by the angle of deflection resulting from the difference in refractive index between the two parts of the cell. As the beam changes location on the photo cell, an output signal is generated that is amplified and sent to the recorder.

12.2.6.3. UV – VIS Absorption Detector

The UV – VIS detector is the most frequently used detector in HPLC (Fig. 10). It is simple to handle, concentration sensitive, selective, and nondestructive. The latter also makes it suitable for preparative chromatography.

When light passes through a liquid (mobile phase), the intensity of absorption is proportional to the concentration of the analyte in the mobile

phase C and the optical path length L (Lambert – Beer law):

$$\log I_0/I = E L C$$

where I_0 is the intensity of the light beam passing through the detector cell without the sample; I the intensity of the light beam passing through the detector cell with the sample; I_0/I the transmittance; E the molar extinction coefficient; L the optical path length; and C the concentration of analyte in the mobile phase.

The Lambert – Beer law is theoretically valid only for measurements at the absorption peak maximum and for monochromatic light; it assumes that only one species present, absorbs light of the wavelength used. Only then does the Lambert – Beer law predict a linear relationship between the concentration and the response of the detector. The chromatographer should be aware of the fact that these optimal conditions are usually not fulfilled with real samples and the linearity of the Lambert – Beer law is not always obeyed. This is often the case when working with highly concentrated sample solutions. Generally, a sample peak should not be quantified if the absorption exceeds 1 AUFS (absorption units full scale). In working with UV – VIS detectors, many solvents also absorb — especially in the UV range (see eluotropic order) — which decrease the sensitivity significantly.

The analytes to be monitored must contain a chromophore; otherwise, UV – VIS detection is impossible. If no chromophore is present, one can often be introduced into the molecule by pre- or postcolumn derivatization (see Section 12.8.2).

Besides fixed-wavelength UV detectors, which usually operate at 254 nm, modern UV – VIS de-
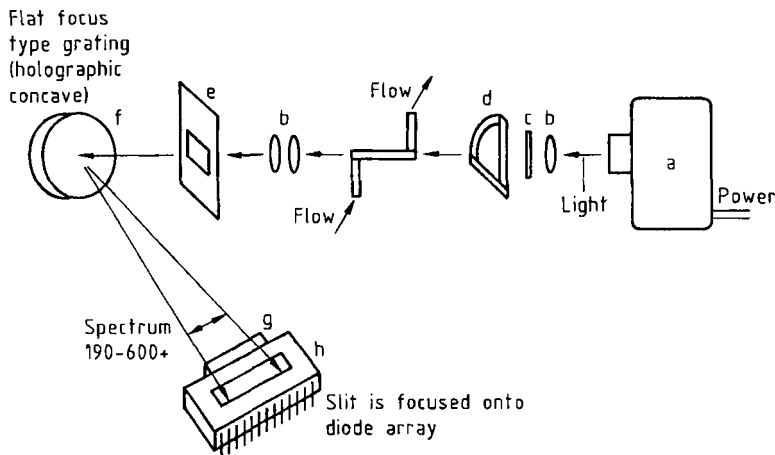
Figure 11. Diagram of a diode-array detector (DAD)
a) D_2 lamp; b) Lens; c) Aperture; d) Shutter; e) Slit; f) Flat focus-type grating (holographic concave); g) Second-order filter; h) Diode array
Courtesy of Waters-Millipore

tectors have an adjustable wavelength from 190 to ca. 600 nm. This enables individual adjustment of the measuring wavelength according to the specific absorption characteristic of the solute. An additional advantage of the adjustable wavelength UV – VIS detector is that a UV – VIS spectrum of the solute can be obtained by simply stopping the mobile phase flow while the solute is inside the detector cell and scanning the sample over a specific range of wavelength. This can be very helpful in determining optimum wavelength at which a particular sample can be detected or identifying an analyte by its characteristic spectrum.

Some UV – VIS detectors, the so-called *multi-wavelength detectors (MWDs)*, are capable of monitoring several different wavelengths simultaneously, thereby enabling the sensitive detection of various analytes that differ in their adsorption maximum.

12.2.6.4. Diode-Array Detector

The diode-array detector (DAD) enables the simultaneous measurement of absorption versus analysis time (chromatogram) and absorption versus wavelength (spectrum), without having to stop the mobile phase flow. Diode-array detectors can record spectra simultaneously every few milliseconds at different points of an elution band (peak) during a chromatographic run. Usually the spectra are taken at the beginning, the maximum, and the

end of an eluting peak. Comparison of the spectra obtained supplies important information concerning the identity and purity of the monitored peak.

Like modern UV – VIS detectors, the DAD can also monitor a sample at more than one wavelength at a time, thus enhancing sensitivity dramatically if the absorption maxima of the test compounds differ. Figure 11 shows a schematic diagram of a typical DAD. Here, an achromatic lens system focuses polychromatic light into the detector flow cell. The beam then disperses on the surface of a diffraction grating and falls onto the photo diode array.

At present, good commercially available DADs have 512 diodes to cover a 600-nm wavelength range (ca. 190 – 800 nm). With decreasing number of diodes (for a set wavelength range) the resolution of the obtained spectra decreases as well, which results in a significant loss of information.

The drawback of a DAD is that because the light is split, up in front of the diodes, the resulting sensitivity is not as good as that obtained with classical UV – VIS detectors. Modern DADs are approximately a factor of 2, and older DADs a factor of 10 or more, less sensitive than normal UV – VIS detectors.

Fast-Scanning Detector (FSD). The fast-scanning detector is a relatively inexpensive alternative to the diode-array detector in LC. Compared
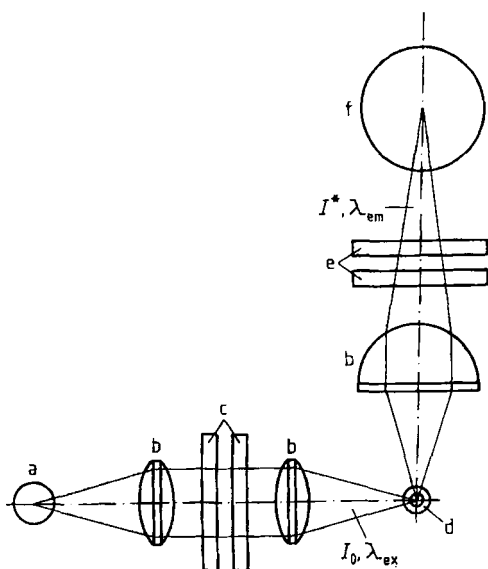
Figure 12. Diagram of a fluorescence detector
a) Lamp; b) Lens; c) Excitation filter; d) Flow cell;
e) Emission filter; f) Photomultiplier

to the DAD it has only one diode (the reason why it is cheaper) and functions in principle as follows: Polychromatic light incident on a rotating grated mirror is split up into single wavelengths, which are then registered one after another by the diode. The required scan time is longer (ca. 1 s) than with a DAD (millisecond range). For liquid chromatography, this does not present a major problem because an average peak width is ca. 20 s, so enough time is available to monitor several spectra per peak for purity control, for example. In capillary electrophoresis (CE) this slowscan time can become a problem because peak width can be extreme short.

Indirect UV Detection. At the end of the 1980s, a new variation of UV detection was developed. This so-called indirect UV detection uses the same equipment as the classical DAD or UV – VIS detector. The difference between these methods is that in the classical mode the analyte contains a chromophore. In indirect UV detection a chromophore is added to the mobile phase so that a continuous, positive baseline signal is generated. If an analyte that has no chromophore passes the detector cell, the adsorption of the mobile phase is decreased and a negative peak is recorded. The advantage of this method is that analytes without chromophores (e.g., alkali and alkaline-earth ions)

can be detected; the present disadvantage is its relatively poor detection limit (between 0.1 and 1 mg/L).

12.2.6.5. Fluorescence Detector

Fluorescence detection provides high selectivity and sensitivity, which is the reason for its use in trace analysis. The high selectivity results from the fact that only a few analytes are able to fluoresce. Typically, two types of spectra are obtained from fluorescing molecules: *excitation spectra* and *emission spectra*. Both spectra are characteristic for each analyte and therefore a great help in identification.

The high selectivity is due to the different detection technique applied (Fig. 12) compared to absorption spectroscopy (e.g., UV detector). The sample that passes the detector cell is irradiated by a light beam with a certain intensity I_0 and excitation wavelength λ_{ex}. Fluorescing solutes are able to emit a radiation I^* with a characteristically higher wavelength λ_{em} which is recorded and converted into an electric signal.

Since the emited radiation is not monochromatic, selectivity can be enhanced by inserting a monochromator into the beam. In this case the increase in selectivity is paid for by a decrease in sensitivity.

With help of pre- or postcolumn derivatization (Section 12.8.2), nonfluorescing analytes can be transformed into detectable ones, with very high selectivity and detection limits.

Since 1990, *laser-induced fluorescence (LIF)* detection has become increasingly important. A good review of the principles and applications of LIF detection is given in [25]. For more detailed information on fluorescence detection, see [26] [27].

12.2.6.6. Evaporative Light-Scattering Detector

The evaporative light-scattering detector (ELSD), sometimes wrongly called a mass detector, is influenced by many variables. SCHULZ and ENGELHARDT [28] showed that its specific response in addition to the analyte's molecular size also depends on the linear flow rate of the mobile phase; consequently, this detector is mass sensitive. Furthermore, they found that the ELSD behaves in certain situations more like a selective and not a universal detector. For example, it can detect only those solutes that are nonvolatile under
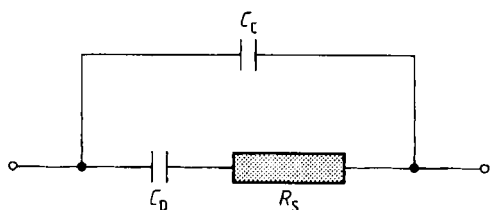
Figure 13. Equivalent circuit of conductivity measuring cell

the applied detection conditions. Nevertheless, ELSD can be a good alternative to the RI detector. Its sensitivity is better by one to two orders of magnitude, and a gradient can be run that would be impossible when measuring the refractive index. Another advantage of the ELSD compared to the RI and UV (< 200 nm) detectors, is the much better signal-to-noise ratio and a reduced baseline drift tendency. Typical applications of the ELSD are in the analysis of sugars and other carbohydrates, polymers, resins, steroids, and bile acids.

12.2.6.7. Infrared Detectors — Fourier Transform Infrared Spectroscopy

Infrared detectors offer the possibility of scanning IR spectra during a chromatographic run. IR detection is limited to a mobile phase that does not absorb in the same range as the analyte.

Modern interfaces joining LC and Fourier transform infrared spectroscopy (FTIR) function in two ways: The first technique is based on a flow cell ("on-line"), in which sensitivity is limited by the dilution of the analyte in the mobile phase. In the second technique the mobile phase is evaporated, leaving the analyte as a small spot that is scanned ("off-line") by FTIR. The advantage of this approach is a significant enhancement in sensitivity. Compared to the refractive index detector the IR detector has a lower sensitivity but theoretically offers the possibility of gradient runs.

12.2.6.8. Electrochemical Detection

In liquid chromatography, electrochemical detection is a superior alternative to optical detection for many analytical problems [29]. Electrochemical detectors monitor currents, electrode potentials, or electrolytic conductivities. In all cases the signal is directly proportional to the concentration or mass flow rate over relatively wide linear ranges. With electronic amplification and simple

signal processing, electrical quantities can be measured with high sensitivity.

Conductometric Detectors. Conductometric detectors are sensitive to ionic solutes in a medium with low conductivity, for instance water or another polar solvent [30]. The response is relatively nonspecific since the conductivity of all ions present in the eluate is determined.

The reciprocal of the resistance is the conductance G (Ω^{-1}); it is expressed in the unit siemens (S). The conductivity κ is the product of the conductance G times the quotient of the cell electrode spacing D and area A (the cell constant):

$$\kappa = G(D/A)$$

The conductivity κ has the dimension microsiemens per centimeter.

The cell constant is important only for absolute measurements; it is determined by calibration with solutions of known conductivity. For relative measurements (i.e., to track eluate conductivity changes on a chromatographic column), the cell constant of the detector need not be known.

Conductivity detectors are important mainly for ion-exchange chromatography. In analytical practice, ion-exchange chromatography with the background conductivity suppressed is used chiefly for the highly sensitive determination of inorganic and organic anions and of alkali and alkaline-earth cations (see Section 12.5.3). The calibration curves (response peak height or area versus concentration) are linear over a wide range, up to three orders of magnitude; the limits of detection are in the nanogram-per-gram range [12].

Conductivity is measured in a cell containing two plates as electrodes (usually made of platinum) through which the column effluent flows. The electrodes have a known surface area and are located in the cell at a fixed distance. The cell is connected in an a.c. circuit and is characterized by the double-layer capacitance C_D, the cell capacitance C_C, and the electrolyte resistance R_S of the sample solution (see the equivalent circuit diagram in Fig. 13).

For the electrolyte resistance R_S to be determined, the effect of C_C must be negligible:

$$\frac{1}{2 \cdot \pi \cdot f \cdot C_C} \gg R_S$$
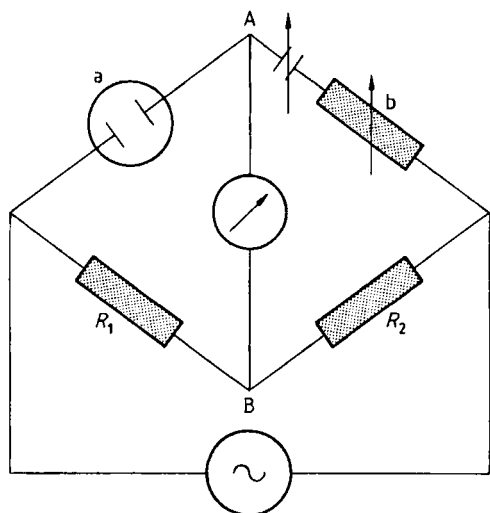
Figure 14. Wheatstone bridge circuit
a) Measuring cell; b) Balancing leg

Figure 15. HPLC conductivity detector [31]
a) Connection to column; b) Platinum capillary; c) Electrical connection; d) Polytetrafluoroethylene capillary; e) Polytetrafluoroethylene; f) Electrical connection; g) Plastic detector housing; h) Stainless steel capillary
Measurement is performed between (b) and (h)

where f is the frequency in Hz of the alternating current.

In practice, this condition is satisfied whenever very dilute solutions or solutions of weak electrolytes are measured with low-frequency alternating currents ($f = 40 - 80$ Hz) or when solutions having conductivities of about 100 µS/cm are measured with frequencies in the kilohertz range. In this way, R_S can be determined fairly accurately with a Wheatstone bridge circuit, shown in simplified form in Figure 14.

Because the impedance of the measuring cell has a capacitive component, the balancing leg in the bridge circuit must include not only the variable resistance but also a variable capacitance. The bridge is balanced or nulled when the a.c. voltage between points A and B is zero. Direct-reading instruments are used to check the balance; for higher measurement accuracy the signals may be amplified first if necessary.

If the conductivity cell is used in combination with a chromatographic column, the measuring electrodes have the form of capillaries. This makes it possible to keep the cell volume very small, as required; the volume is ca. $1-2$ µL. Figure 15 illustrates the principle of an HPLC conductivity detector with two measuring electrodes. The cell must be thermally insulated because the conductivity is temperature dependent.

Four- and six-electrode cells have been designed for precision measurements [13], [14].

Potentiometric Detectors. Essentially, potentiometric detectors consist of ion-selective electrodes (ISEs) to monitor ionic species in chromatographic effluents. Examples of ISEs are glass, solid-state, liquid-membrane, and enzyme electrodes.

The detector response is the potential of the electrode, which depends on the activity of the analyte. This potential is measured with respect to a reference electrode. The response can be described by [32], [33]

$$E = E_0 + \frac{s}{z_i} \log \left(a_i + \sum K_{ij}^{Pot} a_j^{z_i/z_j} \right)$$

where E is the electromotive force of the cell assembly in volts; z_i and a_i are the charge number of the primary ion i and its activity in the sample solution; z_j and a_j are the charge number of the interfering ion j and its activity in the sample solution; K_{ij}^{Pot} is the selectivity factor, a measure of the sensor's preference for interfering ion j rel-
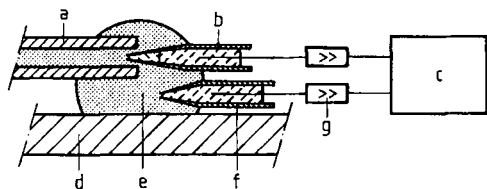
Figure 16. Potentiometric detector with ion-selective microelectrodes [33]
a) Capillary column; b) Ion-selective microelectrode;
c) Voltmeter-recorder; d) Glass plate; e) Eluent drop;
f) Reference electrode; g) Impedance converter

Figure 17. Schematic view of thin-layer amperometric detection (conversion of reactant (R) to product (O) on the surface of the working electrode)
A Detector Cell; B Amplifier; C Chromatogram

ative to measured ion i (an ideally selective membrane electrode would have all $K_{ij}^{Pot} = 0$); E_0 is the potential difference equal to a constant (for a given temperature) plus the liquid junction potential; and $s = 2.303\,RT/z_iF = 59.16$ mV/z_i (at 25 °C for the ideal case).

The activity of an ion is related to its concentration c_i (moles per liter) by $a_i = c_i \cdot \gamma_i$, where γ_i is the activity coefficient of the ion.

Ion-selective electrodes are less important for monitoring chromatographic eluates than for automated flow injection analysis (FIA).

Major characteristics of potentiometric detectors are high specificity, low sensitivity, and slow response.

Only a few examples can be found in the literature in which ISEs have been used as detectors for ion chromatography. They have been employed to detect nitrate in the presence of nitrite [32] and to detect alkali cations [33], [34] and a variety of anions [35], [36] after isolation by ion-exchange liquid chromatography.

Potentiometric detectors vary widely in cell design. One example worth noting is a microtype membrane cell with two compartments separated by an ion-exchange membrane. The column eluate is passed through one compartment of the detector, and the reference solution with the same components of the eluate is passed through the other. This device is useful for monitoring ion-exchange as well as reversed-phase ion-pair chromatography; it offers detection limits in the nanomole range, together with a linear response over about two to three orders of magnitude [34].

The design of a picoliter-volume cell potentiometric detector for open tubular column liquid chromatography is shown in Figure 16. The microelectrode has a tip diameter of 1 μm and is inserted in the open tubular column. A cation-exchange solution can be used as membrane. The
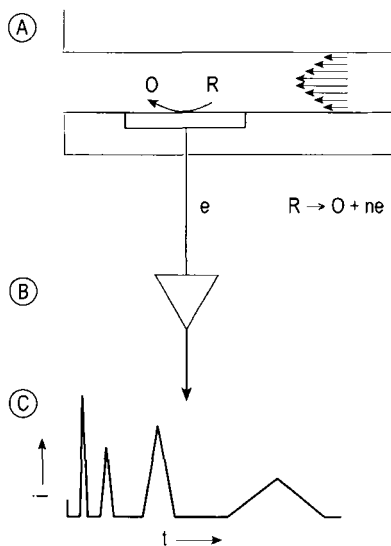
reference electrode, located in a micropipette, is a chlorinated silver wire in 3-mol/L KCl. The response time of the microdetector is less than 10 ms [33].

Amperometric Detectors. Amperometric detectors are the most commonly used electrochemical detectors for highly sensitive and selective determinations in HPLC [30], [31]. The frequently used synonym "electrochemical detector" is not very precise in relation to the detection mode.

The signal current is the result of an electrochemical conversion of the analyte by oxidation or reduction at the working electrode (Fig. 17). The currents are in the pA–μA range and are measured by the three-electrode technique (→ Analytical Voltammetry and Polarography). As a result extremely low detection limits at the 0.1-pmol level can be achieved.

The simplest and most common detection mode is current measurement at constant potential. The reference and counter electrodes should be located downstream of the working electrode so that reaction products at the counter electrode or leakage from the reference electrode do not interfere with the working electrode. The current is amplified and plotted versus time to yield a chromatogram.
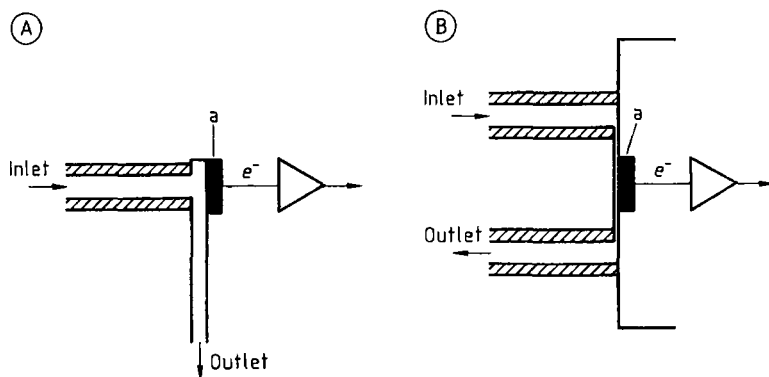
Figure 18. Schematics of the most common amperometric detector cells
A) Wall-jet cell;   B) Thin-layer cell
a) Working electrode

Because the electrode process takes place solely at the surface of the working electrode, amperometric flow cells can determine only local (interfacial) concentrations. Conductometric and potentiometric measuring cells, on the other hand, can track the average volumetric concentrations in the eluate. The same holds for UV detection, which is preferred for routine HPLC studies (see Section 12.2.6.3).

In amperometric detector cells, the analyte is transported to the working electrode surface by diffusion as well as convection; migration is suppressed by the supporting electrolyte. Electrolyte concentrations in the eluent of 0.01 – 0.1 mol/L are sufficient. Several cell geometries employing solid electrodes have been designed and tested. The two types shown in Figure 18 are the most commonly used.

In the *thin-layer cell*, the eluate is forced through a thin, rectangular channel, normally 0.05 – 0.5 mm thick. The working electrode (usually a disk electrode, 2.4 mm in diameter) is embedded in the channel wall. In the "wall-jet" design, the solution is introduced into the cell through a small diameter jet or nozzle and impinges perpendicularly onto the planar electrode. In both cases, the volume is ca. 1 – 2 µL. The advantages of these cells in comparison with other detector designs are good hydrodynamic properties and fast response [37], [38].

The thin-layer cell for HPLC was originally described by KISSINGER and coworkers in 1973 [39]. The wall-jet principle has been known since 1967 and was originally introduced for HPLC in 1974 [40]. The development is based on KEMULA's

work with a mercury drop as detector for column chromatography [41].

When mercury is used as the electrode material, detection is possible over a particularly wide cathodic potential range. The anodic range, however, is limited. It is thus commonly used for the detection of reducible species. Detectors which utilize a static mercury drop are particularly attractive as the drop (working electrode) is easily renewed for each chromatogram; this is quite an advantage for the reproducibility of response [42]. The construction and operation of detectors based on mercury drop electrodes have been reviewed [42], [43].

Solid electrodes, which have a larger anodic range than mercury-based electrodes, include different types of carbon and noble-metal electrodes. The preferred electrode material for commercial detectors is glassy carbon. Its electrochemical properties allow measurements in a wide potential range of about – 0.8 to +1.2 V. The effective range (chiefly the cathodic voltage range) depends on the composition, mainly on the pH value, of the eluate. In contrast to carbon-paste electrodes, which have been proposed for amperometric detection — particularly because of the low background response and the good signal-to-noise ratio — glassy carbon electrodes offer better chemical resistance. They are stable in all solvents employed, such as acetonitrile and methanol as mobile phase.

Noble-metal (Au, Pt) electrodes have been employed for special studies, and gold amalgam electrodes as well as mercury-film electrodes have been used for reductive determinations [44] – [46].
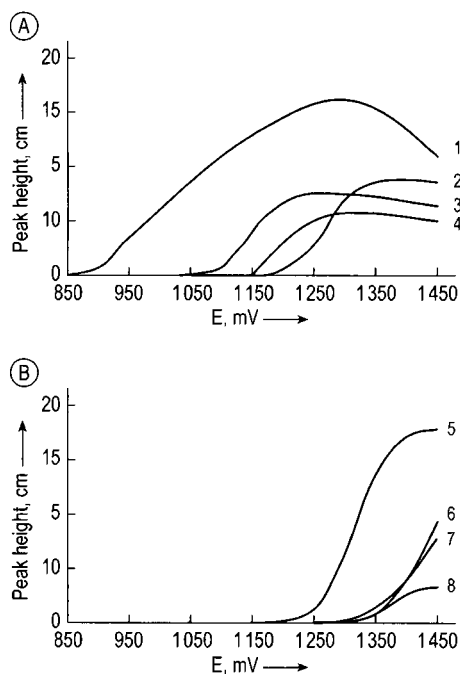
Figure 19. Hydrodynamic voltammograms of polycyclic aromatic hydrocarbons (Eluent: methanol/water 85/15) [47], [67].
1) Benzo[a]pyrene; 2) Benzo[k]fluoranthene; 3) Benzo[ghi]perylene; 4) Indeno[1,2,3-cd]pyrene; 5) Acenaphthene; 6) Fluoranthene; 7) Acenaphthalene; 8) Benzo[b]fluoranthene.

The basis of the good selectivity of amperometric detection is that only compounds with electroactive functional groups (reducible or oxidizable) can undergo reaction at the electrode. The potential at which these electrode reactions occur depends on the redox behavior of the compounds of interest and the working conditions (working and reference electrodes, composition and pH of the mobile phase). The most suitable detection potential is best determined experimentally from the cyclic voltammogram or the hydrodynamic voltammogram of the analyte.

A *hydrodynamic voltammogram* is a current – potential curve which shows the dependence of the chromatographic peak height on the detection potential. The technique used to obtain the necessary information is voltammetric flow injection analysis, in which an aliquot of the analyte is injected into the flowing eluent prior to the detector and the peak current recorded. This is repeated many times, the detector potential being changed after each injection, until the peak current – potential plot reaches a plateau or a maximum, as shown

for some polycyclic aromatic hydrocarbons (PAHs) in Figure 19.

Whereas indeno[1,2,3-cd]pyrene, benzo[ghi]perylene, benzo[k]fluoranthene, and benzo[a]pyrene give a maximum detection signal at potentials between + 1.30 and + 1.35 V (Figure 19 A), the current – potential curves of acenaphthalene, acenaphthene, fluoranthene and benzo[b]fluoranthene do not reach a maximum even at + 1.45 V (Figure 19 B). A measurable signal is first obtained at + 1.35 V for acenaphthalene benzo[b]fluoranthene, and fluoranthene, whereas the other PAHs are oxidized at less positive potentials. Based on these results, it is possible to selectively determine benzo[a]pyrene in a mixture of PAHs by using a detection potential of + 1.0 V. At more positive potentials, other peaks appear in the chromatogram until finally, at a potential of + 1.45 V, peaks for all the PAHs are recorded, as shown in the chromatograms in Figure 20.

When using amperometric detectors, the composition of the eluent should remain as constant as possible (isocratic elution). If gradient elution is used to reduce the analysis time, then base line drift is observed, resulting in a loss of sensitivity. This is particularly so with glassy carbon working electrodes and is due to the slow adjustment of the electrochemical equilibrium at the electrode surface to the changing eluent composition.

Working electrodes made of other materials such as the ultratrace graphite electrode (porous graphite structure impregnated with epoxy resin, produced by Metrohm, Switzerland), have much shorter lead-in times. The relatively rapid response of this kind of electrode arises from its semimicro electrode properties [48].

When used as the working electrode in amperometric detection following gradient elution, e.g., in the analysis of beno[a]pyrene (BaP) and its metabolites, the hydroxybenzo[a]pyrenes, which result from the enzymatic degradation of BaP in animals and humans after uptake in food or from the air, rapid adjustment to steady state conditions is an important requirement and leads to a more stable base line (Figure 21 B). For comparison, the chromatogram using the much slower isocratic elution is shown in Figure 21 A. The detection limits for the analysis of benzo[a]pyrene and its metabolites lie between 0.2 ng and 0.4 ng and the determination limits between 0.3 ng and 0.6 ng when either the glassy carbon electrode or the ultratrace electrode is used.

Amperometric detection after HPLC separation has been used to detect easily oxidizable
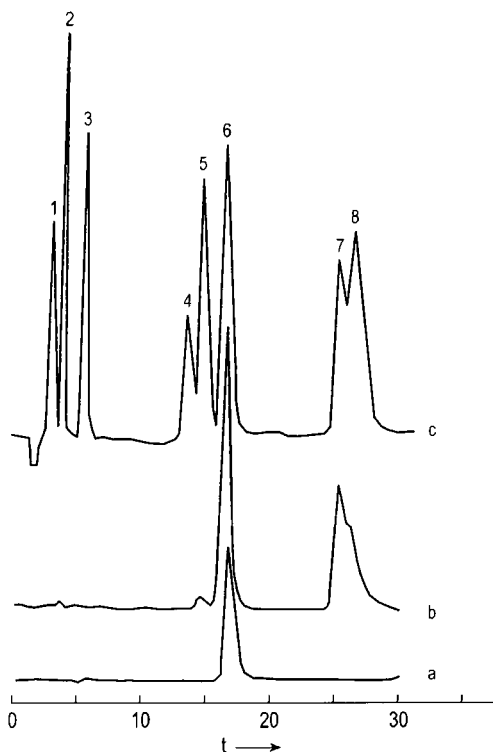
Figure 20. Determination of polycyclic hydrocarbons by HPLC with amperometric detection.
a) Detection potential + 1.0 V : selective determination of 10 ng benzo[a]pyrene (6). b) Detemination of 10 ng benzo[a]pyrene (6) and 10,5 ng benzo[ghi]perylene (7) at + 1.2 V. c) Determination of 8.6 ng acenaphthalene (1), 11.4 ng acenaphthene (2). 51,5 ng fluoranthene (3). 9.5 ng benzo[a]fluoranthene (4), 10.5 ng benzo[k]fluoranthene (5). 10 ng benzo[a]pyrene (6), 10.5 ng benzo[ghi]perylene (7) and 10 ng indeno[1,2,3-cd]pyrene at a detection potential of + 1.45 V.
Eluent : methanol/water 85/15 with 2 g/L trichloracetic acid as supporting electrolyte. Working electrode: glassy carbon; reference electrode: Ag/AgCl in methanolic KCl. The detection and determination limits were in the range 100 – 300 pg per injection [47]. [67].

Figure 21. Chromatograms of BaP and metabolites (each 90 ng). Amperometric detection at + 1.35 V (vs Ag/AgCl, 3 M KCl) using an ultratrace electrode with isocratic elution (A) and a glassy carbon electrode with gradient elution (B) [49].
1) Benzo[a]pyrene-trans-9,10-dihydrodiol; 2) Benzo[a]pyrene-trans-4,5-dihydrodiol; 3) Benzo[a]pyrene-trans-7,8-dihydrodiol; 4) 9-Hydroxybenzo[a]pyrene; 5) 7-Hydroxybenzo[a]pyrene; 6) 1-Hydroxybenzo[a]pyrene; 7) 3-Hydroxybenzo[a]pyrene; 8) Benzo[a]pyrene.

phenols, aromatic amines, indoles, phenothiazines, thiols, and polyaromatic hydrocarbons (Table 2) and easily reducible nitro compounds and quinones (Table 3). These methods have been used in environmental, pharmaceutical, biochemical, and food analysis, as well as for determining hormone residues in soil and water samples and in plant and animal products. The main application is in the analysis of phenols and aromatic amines such as the catecholamines, which can be detected with high sensitivity in the lower picogram region in body fluids and biological samples [51] – [55].

While the current is measured at a constant voltage in amperometric detectors, in voltammetric detectors it is measured while the voltage is changing rapidly. Either linear scan, staircase, square-wave, pulse or a.c. voltammetry can be used for the measurement. When compared to stationary methods, nonstationary methods with superimposed voltage pulses often result in higher selectivity and sensitivity and fewer problems from the adsorption of impurities and electrode reaction products on the electrode surface [53] – [58].

A technique for cleaning the surface of solid working electrodes to improve the reproducibility of the measurements (removal of electrochemical reaction products) is to alternate the potential repeatedly [59].

For the "fast-scan" technique in HPLC, in which the potential is varied rapidly across its full range, the method of square-wave voltammetry (→ Analytical Voltammetry and Polarography) has provided the basis for the development of a "rapid-scan square-wave voltammetric detector"

Table 2. Organic analytes detected amperometrically in the oxidative mode [50]
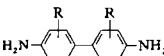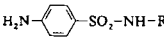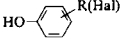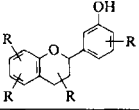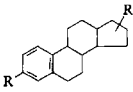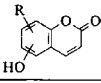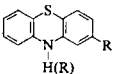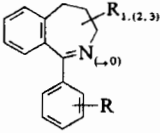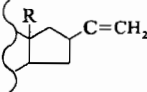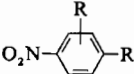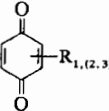
Analyte with electroactive group	Structural formula	Operating potential, mV	Examples
Aromatic amines Anilines	NH₂—⟨⟩—R	+ 1000	phenylenediamine, 3-chloroaniline, aniline, methylaniline, naphthylamine, diphenylamine
Benzidines	H₂N—⟨R⟩—⟨R⟩—NH₂	+ 600	3,3'-dichlorobenzidine
Sulfonamides	H₂N—⟨⟩—SO₂—NH–R	1200	
Aromatic hydroxyls Phenols, halogenated phenols	HO—⟨⟩—R(Hal)	+ 1200	salicylic acid (and derivatives) benzofuranols, chlorophenols (and metabolites), paracetamol, hesperidin, tyrosine
Flavones	flavone structure (OH, O, R substituents)	+ 1100	flavonoles, procyanidins
Catecholamines	HO— / HO—⟨⟩—R	+ 800	DOPA, dopamine nor-epinephrine, epinephrine, and their metabolites
Estrogens	steroid structure with R	+ 1000	estriol, catecholestrogens
Antioxidants	R—⟨R₁,₂⟩—OH	+ 1000 (200)	*tert*-butylhydroxyanisole, *tert*-butylhydroxytoluene
Hydroxycumarins	coumarin structure (R, O, O, HO)	+ 1000	esculin
Aliphatic amines (*sec-* and *tert-*)	R–N⟨R / H(R)	+ 1200	clonidine, diphenhydramine
Indoles	indole structure (H, N, R)	+ 1000 + 450	tryptophan, pindolol, bopindolol metabolit
Thiols	R – SH	+ 800	cysteine, glutathione, penicillamine
Phenothiazones	phenothiazone structure (S, N, H(R), R)	+ 1000	thioridazine, chlorpromazine
Pyrimidines	pyrimidine structure (R, N, (R), (R))	+ 1200	trimethoprim
Purines	purine structure (R, N, N, N, N, H, R)	+ 1000	uric acid
Hydrazines	R – NH – NH₂	+ 600	hydralazine
Miscellaneous Vitamines		+ 800–1000	vitamins A, E, C
Abusive drugs		+ 700–1200	morphine and derivatives, benzodiazepines, cocaine, hallucinogens, tetrahydrocannabinols, tricyclic antidepressants, neuroleptics
Pesticides, herbicides		+ 1100–1300	phenylurea derivatives, carbamates

Table 3. Organic analytes detected amperometrically in the reductive mode [50]

Analyte with electroactive group	Structural formula	Operating potential, mV	Examples
Benzodiazepines (and -4-oxides)		-500 to -1000	chlordiazepoxide, nitrazepam, diazepam
Steroid hormones		-500	ethinylestradiol, levonorgestrel
Aromatic and aliphatic nitro compounds		-400 to -800	explosives (TNT, nitroglycerin), chloramphenicol, nitropoly-aromatics, parathion derivatives, dinitrophenol pesticides
Quinones			vitamin K
Miscellaneous Nitrosamines, thioamides, organometallic compounds, peroxides, azo compounds			prothionamide, benzoyl peroxides, triphenyllead, pyridylazo resorcinol

[60]. In addition, a scanning microvoltammetric detector for open tubular liquid chromatography has been described [61]. This detector consists of a 9-µm (outer diameter) carbon fiber as electrode, inserted into the end of a 15-µm (inner diameter) capillary column. The potential can be scanned at rates up to 1 V/s over a range of 0.0 to +1.5 V. A distinct advantage of this detection method is its three-dimensional resolving power.

Figure 22 shows a conventional chromatogram of five compounds (ascorbic acid, epinephrine, tyrosine, dopamine, and hydroquinone) detected at +1.0 V. The third peak in the chromatogram consists of both tyrosine and dopamine, because
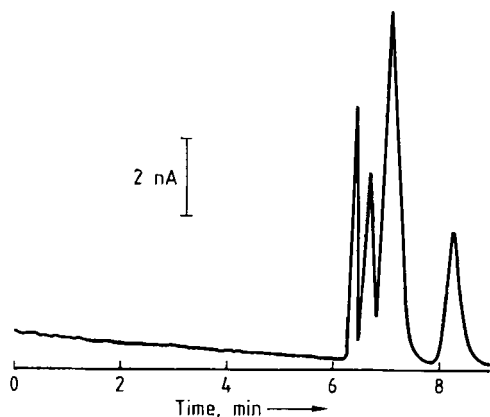
Figure 22. Single-potential chromatogram at +1.0 V versus Ag–AgCl of (in order of elution) 0.1-mmol/L ascorbic acid, 0.1-mmol/L DL-epinephrine, 0.2-mmol/L tyrosine, 0.1-mmol/L dopamine, 0.1-mmol/L hydroquinone
Scan rate was 1 V/s, flow rate 0.59 cm/s; electrode length, 1 mm [61]

Figure 23. Partial threedimensional chromatovoltammogram of mixture shown in Figure 22. [61]

these are not resolved chromatographically. This is revealed by the three-dimensional chromatogram of Figure 23.

No single potential can be used to detect the presence of the two overlapping peaks. However, a potential scan yields a three-dimensional chromatogram of the mixture, in which the two peaks are quite obvious.

Coulometric Detectors. Coulometric detectors are based on potentiostatic coulometry [30]. The signal of the constant-potential measurements, as in amperometric detection, is the current resulting from an electron-transfer process (reduction or oxidation of the analyte arriving with the eluent) while the working electrode is held at constant potential.

In both flow-through detection techniques, the instantaneous measured current i_t is given by

$$i_t = n \cdot F \frac{dN}{dt}$$

where n is the number of electrons per mol exchanged in the reaction, F is the Faraday constant, and dN/dt is the number of molecules per unit time reacting at the working electrode.

The difference between coulometric and amperometric detection lies in the percentage of analyte converted at the working electrode. Of the molecules located in the measuring cell at a given time, only ca. 1–10 % take part in the electrode

reaction in amperometric detection, whereas this conversion should be >95 % in coulometric detection.

To achieve nearly 100 % conversion, the ratio of electrode area to cell volume must be very large in coulometric detectors. The working electrodes employed in coulometric detectors must therefore have a much greater surface area (ca. 5 cm^2) than those used for amperometric detectors. Large electrode areas, however, lower the signal-to-noise ratio and thus increase the detection limits [62]. For this reason, and also because of problems with electrode passivation, coulometric detection has not become an important technique in HPLC, even though the same classes of compounds can be determined as with amperometric detection.

A coulometric detector with two working electrodes and separate reference electrodes, configured one downstream of the other, is called a *dual-electrode cell* [15], [63], [64]. If different potentials are applied to the two working electrodes, two separate chromatograms can be recorded for the same eluate. If the cell is appropriately sized and large-area electrodes (e.g., porous graphite) are employed, current (conversion) efficiencies close to 100 % can be attained. The design of a dual-electrode coulometric detector is illustrated in Figure 24.

Substances with reversible electrode behavior can be determined more sensitively with dual de-
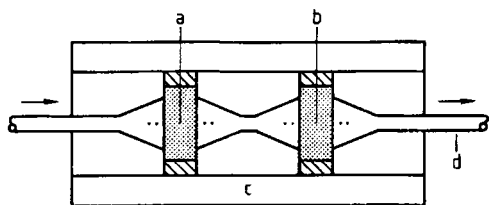
Figure 24. Coulometric detector with dual porous electrodes and total cell volume < 5 μL (Coulochem 5100 A electrochemical detector)
a) Test electrode 1; b) Test electrode 2; c) Stainless steel tubing

tectors than with conventional coulometric detectors. One example is the anodic and cathodic detection (redox mode) of catecholamines [65]. The determination can also be made more sensitive if the two detector signals are fed to a difference circuit so as to eliminate the background that is due chiefly to contaminants in the solvent (differential mode). Dual detectors operated at appropriate electrode potentials can also block out the signals of undesirable compounds (screen-mode detection).

12.3. Solvents (Mobile Phase)

12.3.1. Purification and Pretreatment of LC Solvents

Especially in trace analysis, purification and pretreatment—even of commercially available HPLC-grade solvents—are often necessary (e.g., by filtration or distillation). Occasionally, organic solvents have to be dried as well (normal phase chromatography), for instance, with a molecular sieve. Dissolved atmospheric oxygen in the mobile phase is a major source of trouble in LC. Oxygen increases the UV absorption below 230 nm, thus disturbing sensitive UV detection at lower wavelengths significantly. Dissolved air furthermore, leads to air bubbles inside the chromatographic system (often noticed in "low-pressure" gradient systems), which are responsible for pulsation of the pump or can cause "ghost" peaks when entering the detector cell (e.g., refractive index detector or UV detector). To prevent these problems, the mobile phase has to be degassed. The best method for degassing the eluent is to rinse it with an inert gas (e.g., helium). Often, however, applying vacuum degassing or simply degassing the eluent by

ultrasonic treatment for several minutes before use is sufficient.

12.3.2. Elutropic Order

The more strongly a solvent is adsorbed to the stationary phase, the higher is its elution power. Because the solvent molecules outnumber the analyte molecules with respect to the available adsorption sites of the stationary phase, a relative weak solvent (mobile phase) is still strong enough to elute even quite polar sample compounds. In classical normal phase (NP) adsorption chromatography, the different solvents are listed according to increasing elution power (elution strength), resulting in their elutropic order (see Table 4). The elutropic order was developed for normal phase chromatography; consequently, water is the strongest solvent. Today, most separations are carried out in the reversed-phase (RP) chromatography, so the elutropic order is inversed, turning water into the weakest mobile phase (see Section 12.4.3).

With the help of the elutropic order, appropriate solvents can be chosen, or various solvent mixtures created, that have the same elution strength but different selectivity.

12.3.3. Selection of Mobile Phase

The mobile phase in LC consists of one or more solvents, which additionally may contain modifiers. Generally, the composition of the eluent should be kept as simple as possible. When working with gradients, solvated mobile phase modifiers (e.g., salts) might precipitate in the chromatographic system because of the dramatic change in eluent composition. The following factors must be considered in choosing an appropriate mobile phase for a specific LC problem:

1) Solubility of the sample in the eluent
2) Polarity of the eluent
3) Light transmission of the eluent
4) Refractive index of the eluent
5) Viscosity of the eluent
6) Boiling point of the eluent
7) Purity and stability of the eluent
8) Safety regulations
9) pH

Items 1 and 2 must be seen in the context of the separation system applied (see Chap. 12.5).

Table 4. Important characteristics of the most frequently used mobile phases in LC *

Solvent	Polarity	Solvent strength	Dielectric constant	UV cutoff, nm	Refractive index	Viscosity at 25 °C, mPa · s
Heptane	0.2	0.01	1.92	195	1.385	0.40
Hexane	0.1	0.01	1.88	190	1.372	0.30
Pentane	0.0	0.00	1.84	195	1.355	0.22
Cyclohexane	−0.2	0.04	2.02	200	1.423	0.90
Toluene	2.4	0.29	2.40	285	1.494	0.55
Ethyl ether	2.8	0.38	4.30	218	1.350	0.24
Benzene	2.7	0.52	2.30	280	1.498	0.60
Methylene chloride	3.1	0.42	8.90	233	1.421	0.41
1,2-Dichloroethane	3.5	0.44	10.40	228	1.442	0.78
Butanol	3.9	0.70	17.50	210	1.397	2.6
Propanol	4.0	0.82	20.30	240	1.385	1.9
Tetrahydrofuran	4.0	0.57	7.60	212	1.405	0.46
Ethylacetate	4.4	0.58	6.00	256	1.370	0.43
Isopropanol	3.9	0.82	20.30	205	1.384	1.9
Chloroform	4.1	0.40	4.80	245	1.443	0.53
Dioxane	4.8	0.56	2.20	215	1.420	1.20
Acetone	5.1	0.56		330	1.356	0.30
Ethanol	4.3	0.88	24.60	210	1.359	1.08
Acetic acid	6.0		6.20		1.370	1.10
Acetonitrile	5.8	0.65	37.50	190	1.341	0.34
Methanol	5.1	0.95	32.70	205	1.326	0.54
Nitromethane	6.0	0.64		380	1.380	0.61
Water	10.2		80.00		1.333	0.89

* Data taken from elutropic order in [19].

The *light transmission* of the eluent is of importance in using a UV detector. When monitoring a sample at a wavelength of 230 nm, acetone, for instance, (which absorbs up to 330 nm) cannot be used as a mobile phase. As long as the absorption of the eluent does not exceed 0.5 AUFS, it can be tolerated. Table 4 shows some important characteristics and the UV cutoff from often used eluents. Organic or inorganic trace compounds (impurities) in the mobile phase often shift the UV absorption to higher wavelengths.

When working with a refractive index detector, the difference in the *refractive index* of the sample and the mobile phase should be as great as possible to achieve better sensitivity and lower detection limits.

The *viscosity* of the mobile phase is responsible for the back-pressure of a given LC system. In addition, the higher viscosity of the eluent results in a lower diffusion coefficient for the sample, which leads to reduced mass transfer and peak broadening. Especially during gradient elution with two or more relatively polar organic solvents, unexpected changes in viscosity can occur [66].

The possibilities for influencing the viscosity of the mobile phase are limited. Working at higher temperatures can be an advantage in some cases (e.g., high-temperature liquid chromatography).

The *boiling point* of the solvents should also be considered. For preparative chromatography, a high-purity and low-boiling mobile phase is preferred. On the other hand, the boiling point should not be too low because that could cause gas bubbles in the pump or detector.

The pH of the mobile phase often has a dramatic influence on the selectivity and resolution of investigated samples. For many pharmaceuticals, the optimal pH of the mobile phase is between pH 4 and 5. During method development and optimization, it is recommended to run a sequence of samples between pH 2.5 and 7.5 in pH steps of 0.5. Plotting resolution and selectivity against mobile-phase pH then allows the optimum eluent pH to be identified.

Of course, the eluent must also be sufficiently stable toward thermal or chemical influence. Most organic solvents are easily flammable and have a high vapor pressure at room temperature. Therefore, LC instrumentation should always be placed in a well-ventilated area.

The pH of the mobile phase has a strong influence on the solute retention and the peak shape of eluting ionogenic compounds. It is a critical factor for the reproducibility and ruggedness of a HPLC method (see also [67]).

12.4. Column Packing (Stationary Phase)

A very important factor for successful HPLC separation is the correct choice of column packing. In the past, many different packing materials were tested, for instance, powdered sugar, diatomaceous earth, aluminum oxide, calcium carbonate, calcium hydroxide, magnesium oxide, Fuller's earth, and silica. Today, silica is generally used as the basis for adsorbent chromatography; aluminum oxide is also used to a certain extent. Both materials are pressure stable and are unaffected by pH over a wide range.

Particle Size, Particle-Size Distribution. Standard particle size for analytical HPLC columns is 3, 5 or 10 μm. The smallest commercially available particle size is 1.5 μm. The particle-size distribution should be as narrow as possible so that "good" packings and therefore good separations are obtainable. The smaller the particle-size distribution of the stationary phase, the higher is its quality and therefore its price. However, if the size of stationary phase particles is absolutely identical, this can cause problems in packing the column.

Pore Diameter, Specific Surface Area. In adsorption chromatography a packing material with a very high specific surface area is required. The larger the surface area, the more active sites are available for (adsorptive) interaction. In the case of spherical particles (10-μm i.d.) < 1 % of the geometrical surface contributes to the total surface area. The dominant remainder of the surface area originates from the pores of the particles. A normal silica phase has a specific surface area of ca. 300 m^2/g, a mean pore diameter of 10 nm, and a mean pore volume of ca. 0.5 – 1.0 mL/g.

Most analytes have no access to pores with a diameter smaller than 3 – 4 nm. Therefore the packing material should show a narrow pore-size distribution so that the total number of very small pores (to which the analytes have no access) is as low as possible. Generally, the pore diameter should be about 3 – 4 times larger than the diameter of the analyte in solution, to ensure sufficient accessibility.

Nonporous packings with very small particle diameters have developed into useful tools for fast and economic separations. Due to the reduced resistance to mass transfer of the nonporous material, separation efficiency is increased at high flow rates and consequently analysis times are reduced. To overcome the high back-pressure at increased mobile phase velocities, it is recommended to increase the mobile phase operating temperature. This lowers the mobile phase viscosity and helps to reduce the back-pressure generated over the separation column. However, the boiling point of the mobile phase at the corresponding operating pressure should always be taken into consideration to avoid gas bubbles in the pump or in the detector (see also Chapter 12.3.3). More detailed information regarding nonporous column packing material can be found in [68] – [71].

Short analysis times can be achieved by using short conventional columns (4.5 mm i.d. and in 20 to 100 mm in length) combined with higher mobile phase flow rates (e.g., 3 mL/min) and, if possible, by applying sharp gradient profiles, as described in [72].

12.4.1. Silica

Silica is the traditional and most common material used in normal phase chromatography and serves as the basis for many chemically bonded reverse phases. It is a highly porous, amorphous silicic acid in the form of small, rigid particles. A distinction is made between spherical and irregular particles. Traditionally, *irregular silica* was, and to a certain extent still is, prepared by precipitation of water glass with sulfuric acid. Modern manufacturing techniques for *spherical silica* are, for example, sol – gel processing or spray drying of a silica sol; the latter is apparently the most common and economical technique today [73].

Commercially available silicas differ in their physical (specific surface area, mean pore diameter, pore volume, etc.) and chemical (contamination of different metals such as Al, Mg, Fe) characteristics. These metal contaminations can cause secondary retention effects (e.g., strong tailing of basic solutes). Additionally, commercial silicas differ widely in their Na content. The higher the Na content of the silica, the more alkaline its behavior in contact with water. This often acts as a catalyst for the hydrolysis of the Si – C bonds of chemically modified silica phases (see Section 12.4.3). Good silicas have an Na content of < 30 ppm. Excessively high contents of metal contaminants could easily be removed by the manufacturer by means of a thorough acid-washing
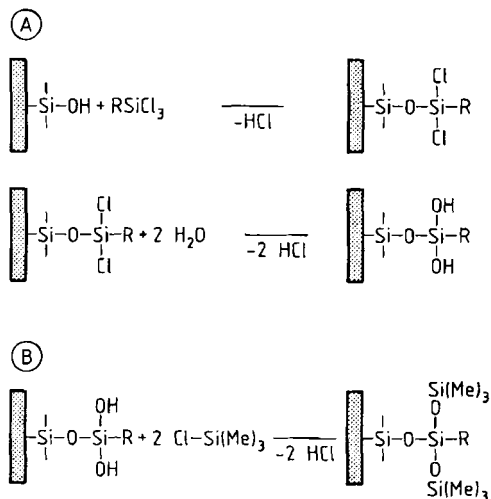
Figure 25. Example (A) derivatization of silica and (B) end capping

step, which would, however, slightly increase the production costs of the material.

The chemical properties of the silica surface are determined by silanol (SiOH) and siloxane (Si–O–Si) groups. There are ca. 7 – 8 μmol SiOH groups per square meter of surface area (i.e., the average distance between silanol groups is 0.4 – 0.5 nm). These silanol groups are responsible for adsorptive interaction with the solute and are the active sites for the derivatization into chemically bonded phases (see Section 12.4.3). Silica is mildly acidic, with a pH of ca. 5. Due to its mildly acidic behavior, problems may occur with very sensitive solutes (acid-catalyzed reaction). At pH > 8, silica goes into solution and the column bed is destroyed.

12.4.2. Alumina

Although alumina is rarely used as the stationary phase in HPLC, in certain applications it could be preferred to silica (e.g., separation of isomeric aromatic hydrocarbons). In contrast to silica, natural alumina is a basic material (basic alumina) with a pH of ca. 12 and a specific surface area of ca. 200 m²/g (depending on the pore-size distribution). Because of the high pH of the packing, acidic analytes (e.g., carboxylic acids) may show strong peak tailing and are sometimes retained irreversibly. Problems with base-sensitive samples

on alumina columns have been reported [74]. Through special treatment, basic alumina can be converted into acidic or neutral alumina; both are commercially available. For adsorption chromatography, neutral alumina is preferred. Acidic alumina can be used as a weak anion exchanger; basic alumina, as a weak cation exchanger.

12.4.3. Chemically Bonded Phases

Chemically bonded phases are usually made by derivatization of the silanol groups of silica (see Fig. 25 A). Remaining unreacted SiOH groups can be removed by end capping with trimethylchlorosilane (TMCS) or hexamethyldisilazane (HMDS) (Fig. 25 B). The positive effect of end capping on the separation of basic solutes on a reversed-phase column is shown in Figure 26. A further advantage of end-capped material is that it is often more stable and therefore has a prolonged service life.

Depending on the size and reactivity of the silica matrix, a maximum of 50 % of all silanol groups present reacts with the silane. This is very important for RP packings because the interactions of basic analytes and the remaining silanol groups can result in severe peak tailing (see Fig. 26 A). Peak tailing can be suppressed by end capping or by the addition of a relatively basic component or a salt to the mobile phase.

A variety of different chemically bonded phases is available (see Table 5). All of them have a stable Si–C bonding. If their organic residue is bonded to a function group (e.g., –OH, –NH₂, –CN, or –NO₂) the phases are often called selective polar phases or chemically bonded phases, and are used in both the normal phase and the reversed-phase chromatographic modes.

To ensure stable bonding, especially Si–C bonds, the polar functional group, which is responsible for the selectivity of the material, is connected to a spacer [–(CH₂)₃ group] that is fixed onto the silica. The result is a nonpolar stationary phase with a relatively polar surface. If the organic residue of a chemically bonded phase consists only of a hydrocarbon chain (C₁ – C₁₈), the stationary phase is very nonpolar or hydrophobic. Here, the elution order of hydrocarbon samples is reversed compared to normal phase chromatography (where retention decreases with increasing number of carbon atoms). Therefore, these phases are called reversed phase and the chromatographic mode *reversed-phase chromatography*. Commercially available packing materials for RP columns
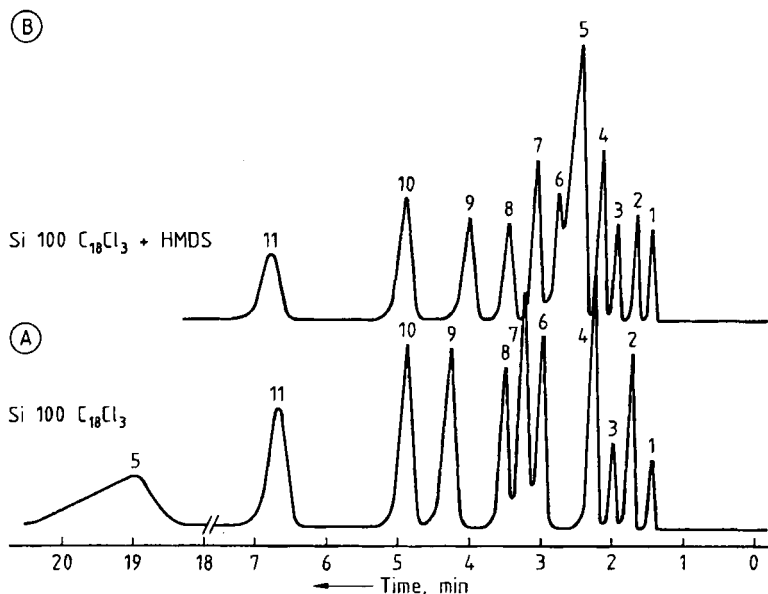
Figure 26. Separation of column performance test on a reversed-phase column
(A) Without end capping; (B) After end capping or basic component or salt to mobile phase
1) H_2O; 2) Benzamide; 3) Phenol; 4) 2-Phenylethanol; 5) Aniline; 6) Nitrobenzene; 7) Methyl benzoate; 8) Benzene; 9) Ethyl benzoate; 10) Toluene; 11) Ethylbenzene
Eluent: H_2O–CH_3OH 34.6 : 65.4 wt %

have a chain length between C_1 and C_{18}. Most common is RP-8 and RP-18 material. RP columns of the same type (e.g., RP-18) from different manufactures, or even from different production batches of the same manufacturer, can differ dramatically in their chromatographic behavior because a different packing material (silica) is often used. For instance, a higher natural alumina contamination of the silica increases the acidity of the Si–OH groups, which again strongly influences the retention of basic analytes (peak tailing).

The development of RP columns by HALASZ and SEBASTIAN [75] was the breakthrough for modern HPLC. Through that development, application of this technique was extended significantly (especially for biological materials), and because of the fast re-equilibration of the column, gradient runs could be performed easily, even in routine analysis.

12.4.4. Column Packing Techniques

Even if excellent equipment is used, but the sample is passed over a "badly packed" column, separation efficiency is poor. Therfore, the goal is to pack the chromatographic bed as densely and

uniformly as possible. Care must be taken that no cracks or channels develop and that no fractionation of particles occurs during the packing process. Because experience and know-how are required to pack columns well, especially with small particles (< 10 µm), purchase of a prepacked column is recommended.

If columns with relatively coarse particles (>30 µm) have to be packed, the *dry-fill packing process* [19] can be applied. Here vertical tapping and simultaneous rapping of the column during the packing process are necessary for well-packed chromatographic beds.

All analytical HPLC columns with particles < 20 µm (usually 5 or 10 µm) are packed by the *slurry technique* [19], which can be further classified into the high-viscosity technique and the balanced-density technique. In the latter, the particles are suspended in a fluid that has a density similar to theirs, so particle segregation by sedimentation decreases. In both techniques, the suspending fluid must be chosen in such a way that particle dispersion is maintained without aggregation and particle agglomeration is avoided by proper selection of the polarity of the slurry [19], [24].

Table 5. Examples of chemically bonded phases and their application range

Chemical modification	Functional group*	Application range
Butyl	$-(CH_2)_3-CH_3$	RP and ion-pairing chromatography
Octyl	$-(CH_2)_7-CH_3$	RP and ion-pairing chromatography
Octadecyl	$-(CH_2)_{17}-CH_3$	RP and ion-pairing chromatography
Phenyl–diphenyl	$-C_6H_5-(C_6H_5)_2$	RP and ion-pairing chromatography, additional selectivity for $\pi-\pi$ interactions
Nitro	$-NO_2$	analytes containing double bonds
Nitrile (cyano)	$-CN$	NP and RP chromatography
Diol	$-CH_2-CH-OH$ $\quad\quad\quad\mid$ $\quad\quad CH_2-OH$	NP and RP chromatography
Amino	$-NH_2$	NP, RP, and weak anion exchanger
Dimethylamino	$-\overset{+}{N}(CH_3)_2$	NP, RP, and weak anion exchanger
Quaternary ammonium	$-\overset{+}{N}(CH_3)_3$	strong anion exchanger
Sulfonic acid	$-SO_3^-{}^+$	strong cation exchanger

* Usually a spacer; for example, a propyl group $[-(CH_2)_3-]$ is placed between the silica surface and the functional group.

The suspension is pressed into the empty column, which is fitted with a frit on its far side, by means of a constant-pressure pump (pressure filtration). The usual packing pressure for a standard analytical column (250×4 mm i.d.) is between 40 and 60 MPa. In this case, 1.7–2.0 g of stationary phase has to be suspended in ca. 30–50 mL of suspension fluid. Detailed discussion of column packing techniques is given in [73], [19], [76].

12.4.5. Column Specifications and Column Diagnosis

The plate number N is often considered the most important column performance specification. A typical plate number for a 250×4 mm i.d. column with 10-μm particles is ca. $N = 10000$. The plate number required for a specific separation can be achieved in two ways: By using either a short column with small particles or a long column with larger particle diameter. As an example, the same separation efficiency can be obtained with 10-μm particles in a 250-mm-long column or with 3-μm particles in 50–100-mm-long columns.

Even though the plate number is a very useful column specification, it is certainly not sufficient to evaluate the suitability of a tested column for a given separation problem. Especially for RP columns, a variety of column performance tests have been developed [77]. These tests give important information about the quality of the column bed (well packed or not), the kinetic and thermodynamic performance, basic and acidic behavior, and

the selectivity and stability (lifetime) of the column (see Table 6)

Especially in routine analysis, an individual test mixture, consisting of characteristic compounds of everyday samples, is often sufficient. This test mixture should be injected onto every new column (to see if selectivity and k' are unchanged) and again from time to time to follow the aging process of the column (peaks get broader, resolution gets worse, k' get lower, peaks become asymmetric, etc.). If one or more of these characteristics are observed, the old column should be replaced by a new one.

12.5. Separation Processes

12.5.1. Adsorption Chromatography

Adsorption chromatography or liquid–solid chromatography (LSC) is the most frequently applied LC mode. Here, specific interactions occur among the solute, the stationary phase, and the mobile phase. Examples are hydrophobic, dipole–dipole, hydrogen bond, charge-transfer, and $\pi-\pi$ interactions.

Every chromatographic separation should give sufficient resolution Rs and good reproducibility, which have great importance for qualitative and quantitative analysis. Good reproducibility in LSC can be achieved only in the range of a linear sorption isotherm. In this case the retention time t_R is independent of sample size and sample vol-

Table 6. Typical reversed-phase column performance tests

Test name	Characterization of	Test compounds	Eluent	pH	Detection
DMD* test	$C_8 - C_{18}$ packing hydrophobic interactions	diphenylhydramine, 5-(p-methylphenyl)-5-phenylhydantoin, diazepam	acetonitrile – potassium dihydrogen-phosphate/phosphoric acid buffer, 156:340 wt%	2.3	UV (220 nm)
Engelhardt – Jungheim test	$C_8 - C_{18}$ packing hydrophobic interaction, behavior of basic and acidic analytes (silanophilic interactions)	aniline, phenol, o-,m-,p-toluidine, N,N-dimethylaniline, toluene, ethylbenzene, methyl benzoate	methanol – water, 50:50 wt%		UV (254 nm)
Aromatics test	$C_8 - C_{18}$ packing hydrophobic interaction, selectivity	benzene, naphthalene, fluorene, anthracene, benzanthracene	acetonitrile – water, 75:25 wt%		UV (254 nm)

* DMD = Diphenylhydramine,5-(p-methylphenyl)-5-phenylhydantoin, diazepam

ume. Generally, a sample size of < 0.1 mg of analyte per gram of stationary phase ensures a linear sorption isotherm and, therefore, reproducible results.

Two groups of stationary phases are used in adsorption chromatography: Polar and nonpolar adsorbents.

12.5.1.1. Polar Adsorbents and Their Properties

The most frequent used polar adsorbents in LSC are silica and alumina (see Sections 12.4.1 and 12.4.2). Polar stationary phases are very selective with respect to the separation of configurational isomers. Figure 27 shows the separation of β-carotin isomers on an alumina column in NP chromatography.

Generally, polar adsorbents show a greater selectivity (higher retention) toward polar solutes, whereas the retention k' of nonpolar analytes decreases with increasing hydrophobicity of the analyte.

Working with polar stationary phases and nonpolar (hydrophobic) eluents is the classical form of LC (which Tswett [2] used in 1903) and is called *normal phase chromatography*. Currently, only ca. 20% of analytical separations are solved by NP chromatography — for instance, chiral analyses — because re-equilibration of the column can be very time consuming (up to 12 h), and in certain cases, exact water-content control is necessary to obtain reproducible results. In NP chromatography, even traces of water (< 100 mg/L) in the solvent cause a dramatic decrease in k'.

12.5.1.2. Polar Chemically Bonded Phases

Diol phases are often used as "polar" column packings. They belong to the group of polar chemically bonded phases, which also includes nitrile, nitro, amino, and similar polar groups. The diol phase is less polar than silica but is therefore easily wettable with water. Because of their rapid equilibration, these packings are much more suitable for gradient elution than unmodified silica. Diol columns are typically used for the separation of peptides, proteins, and generally components that can build up hydrogen bonds.

12.5.1.3. Nonpolar Adsorbents – Reversed Phase

In contrast to NP, in RP chromatography a nonpolar stationary phase and a polar solvent are used. In the reversed-phase mode the elution-strength order of the solvents is the opposite of that of the elutropic row (see Section 12.4.3 and Table 4).

Some separation problems (e.g., triglyceride separation [76]) require a hydrophobic stationary phase and a nonpolar (nonaqueous) mobile phase. This chromatographic mode is called *nonaqueous reversed-phase* (*NARP*) *chromatography*. The most common commercially available reversed-phase materials are C_2, C_4, C_8, C_{18}, and phenyl phases.

As shown in Figure 28 A the retention of a particular hydrophobic analyte increases with growing chain length of the RP material. The selectivity, on the other hand, reaches a plateau near
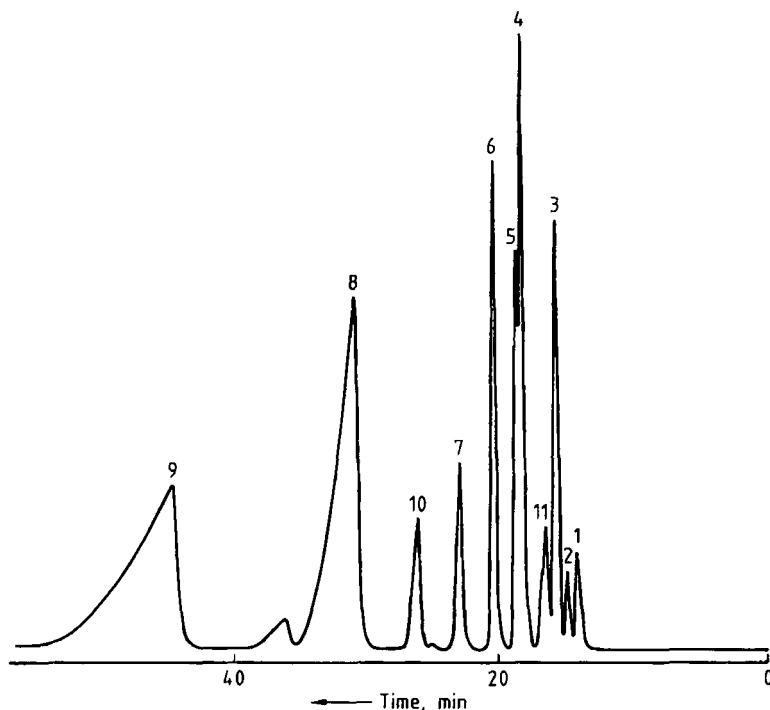
Figure 27. Separation of β-carotin isomers on an alumina column in normal phase chromatography [78]
1) 13,13'-di-*cis*; 2) 9,13,13'-tri-*cis*; 3) 9,13'-di-*cis*; 4) 15-*cis*; 5) 9,13-di-*cis*; 6) 13-*cis*; 7) 9,9'-di-*cis*; 8) all-*trans*; 9) 9-*cis*;
10) α-Carotin; 11) 3 a, b, c

the C_8 column (Fig. 28 B). This means that under identical conditions, the selectivity for the same sample is more or less the same on an RP-8 as on an RP-18 column. The only difference is that the compounds on the RP-18 column are more strongly retained than those on the RP-8 material. Therefore, many separations, which can be performed on an RP-18 column can also be achieved on an RP-8 column by adjusting the elution strength of the mobile phase (in this case by adding more water to the eluent) and vice versa.

Generally, packing materials with shorter alkyl chains (e.g., RP-4 or RP-8) are more suitable for the separation of moderately to highly polar solutes, whereas those with longer alkyl chains (RP-18) are preferred for the separation of moderately polar to nonpolar analytes. Reversed-phase columns are also used for ion-pairing chromatography.

Today, between 50 and 80 % of HPLC applications are run in the RP mode and only 10 to 20 % in the NP mode [79]. This can be explained by the following reasons:

1) The reproducibility of the elution strength of the mobile phase is easier in RP chromatography (more reproducible separations are possible)
2) Re-equilibration between the mobile and stationary phases is much faster in the RP mode (after three to five column volumes) than in NP chromatography, especially if water was added
3) Because of item 2, gradient elution is often not used in NP chromatography
4) RP columns combined with gradient elution are able to separate a wide range of analytes (from hydrophobic to polar components)

Other modes, e.g., IEX (Section 12.5.3) and SEC (Section 12.5.4) are used to a much lesser extent.

The application range of RP chromatography is so large that it is impossible to mention it all. In principle, the technique is applied for nearly all separations in the pharmaceutical, environmental, and biological fields.

A)

~RP-8

~RP-18

log k'

Carbon content of stationary
phase, wt% ———→

B)

~RP-8

~RP-18

α (CH₂)

Carbon content of stationary
phase, wt% ———→
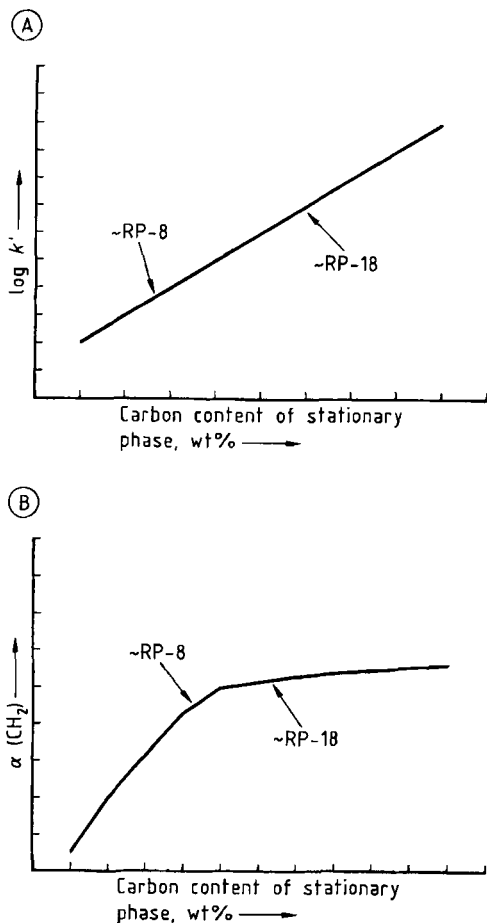
Figure 28. Retention behavior of hydrophobic analytes in
RP chromatography
A) Retention; B) Selectivity

12.5.1.4. Other Adsorbents

Polyamide as a stationary phase is used mainly
for the separation of compounds that can form
hydrogen bonds to the amide group. A typical
example is the separation of phenolic compounds
on polyamide, due to hydrogen bonds between the
phenolic hydroxyl and the amide groups [80]. The
drawback of these polyamide packings is that they
are not as pressure stable as silica and that unde-
sired swelling of the phase can occur with certain
solvents.

Magnesia and magnesia – silica (florisil) are
also sometimes used in adsorption chromatog-

raphy. Magnesia – silica has the disadvantage of
retaining many solutes irreversibly (e.g., aromat-
ics, amines, and ester). Possible applications are in
the separation of lipids or chlorinated pesticides
[80].

Titanium dioxide is also an interesting packing
material because of its good mechanical and
chemical (pH) stability. Its main drawback is the
relatively small surface area (ca. 50 m²/g). How-
ever, this is partially compensated by its high
packing density of 1.5 g/mL, which compares to
only 0.6 g/mL for silica.

12.5.2. Partition Chromatography (LLC)

Partition chromatography, or liquid – liquid
chromatography (LLC), was first mentioned in
1941 in a paper by MARTIN and SYNGE [5]. This
is a well-known technique in gas chromatography,
where the separation is based on distribution of the
analyte between a "liquid" stationary phases and a
gaseous mobile phase. In liquid chromatography,
the distribution takes place between two liquids:
The liquid stationary phase, which is dispersed on
an inert support (e.g., silica), and the liquid mobile
phase. As in the distribution of solutes between
two mobile phases in a separatory funnel, the sta-
tionary and mobile phases in LLC must be immis-
cible. However, liquids are always miscible to a
certain extent, which results in a washing out of
the stationary phase form the column ("column
bleeding").

A major difference between adsorption and
partition chromatography is that in LLC the sep-
aration of a sample is based not on different ad-
sorption energies of the analytes but on different
solubilities of the sample components in the liquid
mobile and stationary phases. This can be a dis-
advantage if binary systems (stationary and only
one solvent as mobile phase) are used. If the solute
is much more soluble in the liquid stationary
phase, the retention will be strong, resulting in
increased dilution and band broadening; thus, the
sensitivity is poor. On the other side, if the solute
is more soluble in the mobile phase, the retention
will be very weak and the solute peak will elute
shortly after the inert peak. To reduce these rel-
atively large differences in polarity (solubility)
between the two liquids, a third liquid ("mod-
ifier") can be introduced to form a ternary system.
In this case the mobile phase is a mixture of two
solvents. Now, the more strongly retained solutes

elute earlier, with better peak symmetry and smaller elution volumes, thereby increasing the sensitivity significantly. A drawback of this ternary system technique is that the stationary phase also becomes more "soluble" in a solvent containing a modifier, which reduces the column lifetime. Furthermore, even a slight temperature increase can cause total miscibility of all three liquids, which destroys the chromatographic system instantly. Such well-thermostated chromatographic equipment (which also includes the solvent reservoirs) is essential (temperature variations 0.1 – 0.2 °C). If a UV detector is used, neither the mobile nor the liquid stationary phase must absorb at the detector wavelength. A further disadvantage of LLC is that gradient elution is impossible. So if a sample contains components with widely different k' ranges, adsorption chromatography is preferred.

The advantages of partition chromatography compared to adsorption chromatography are as follows:

1) Easy exchange of the stationary phase (in situ) and longer column life
2) High loadability (10 times higher than LSC, ca. $10^{-3} - 10^{-2}$ g of analyte per gram of liquid stationary phase)
3) Reproducible results because the supporting material of the column is not, or at least should not be, involved in the separation process
4) Broad range of selectivity due to the large number of different mobile and stationary phase liquids available (with the condition that the liquids are immiscible)

When selecting the stationary phase in LLC, liquids with lower viscosity should be preferred because the use of high-viscosity liquids leads to poorer separation efficiencies due to imparted diffusion and thus sample transport (mass transfer). In addition, the mobile phase must be saturated with stationary phase to reduce column bleeding. Column life can be further extended by the use of a precolumn for proper equilibration and exclusion of oxygen from both phases.

Column Packings and Preparation. Inert materials with relatively large pores (10 – 50 nm) are used (usually silica) as liquid stationary phase supports. Material with larger pores is not suitable because, unlike in narrow pores, the stationary liquid phase is fixed less strongly to the support by capillary forces. Wide-pore supports start to

bleed at low flow rates (< 1 mL/min). Materials with small pores (< 50 nm) are usually stable up to flow rates of 5 mL/min.

Typical stationary phases are, for instance, β,β'-oxydipropionitrile (BOP) or different types of poly(ethylene glycol). Other examples of stationary and mobile phases for LLC are given below:

Stationary phase	Mobile phase
Normal phase LLC	
β,β'-Oxydipropionitrile	pentane, cyclopentane, hexane, heptane,...
Carbowax	modified with up to 10 – 20 % chloroform, THF
Triethylene glycol	dichloromethane, acetonitrile,...
Ethylene glycol	nitromethane
Ethylenediamine	hexane
Water	butanol
Reversed-phase LLC	
Polydimethylsiloxane	acetonitrile – water
Heptane	methanol – water

The stationary phase in LLC can be prepared in different ways, depending on the support particle size chosen. An often-applied preparation technique is the in situ coating method. In this case the stationary phase liquid is pumped through the dry packed (slurry packing) column. Later, the column is flushed with mobile phase to displace excess stationary phase liquid between the particles.

Separation Modes. Like adsorption chromatography, partition chromatography can also be used in the normal and reversed-phase mode.

In *normal phase LLC* the stationary phase is polar and the mobile phase consists of nonpolar solvents. This mode is suited for more polar, water-soluble analytes. The elution order is similar to that in normal phase adsorption chromatography because nonpolar solutes prefer the moving phase and elute with low k' values. Solutes preferring the polar stationary phase elute later with high k'.

In *reversed-phase LLC* the two phases are interchanged and the elution order reversed. This mode is more suitable for hydrophobic analytes that have poor water solubility. For reversed-phase LLC, the support should be hydrophobic to ensure appropriate fixing of the stationary phase liquid. This can be achieved, for instance, through silanization of the silica.

Applications. LLC can be a good alternative to LSC for analytes that are sensitive to catalytic decomposition on the active surface of the column packings. Typical applications of LLC are in the separation of compounds containing metal ions, radioactive metals, and free and underivatized steroids, as well as in preparative chromatography.

Because of the high loadability of the column ($10^{-3} - 10^{-2}$ g of solute per gram of stationary liquid), LLC is often used for preparative chromatography. Since the mobile phase is saturated with stationary phase, high-boiling stationary phases may be difficult to remove from separated fractions and volatile stationary phases are preferred.

To ensure good chromatographic behavior (as in adsorption chromatography), samples should not be injected in solvents with a higher solution power than the mobile phase. Further, the injected sample volume should not exceed one-third the elution volume of th first peak of interest.

12.5.3. Ion-Exchange Chromatography (IEX)

Ion-exchange chromatography (IEX) can be regarded as a special form of adsorption chromatography resulting from the ionic interactions taking place during the separation process. It was first described by SMALL et al. [81] in 1975. Even in the 1980s, IEX was not very well accepted in HPLC because of unstable column packings and detection problems. These drawbacks have been overcome, and today IEX is a very powerful tool in water analysis, amino acid analysis, and the separation cleanup of proteins, for example. Generally all charged solutes (inorganic ions, organic acids, bases, salts, etc.) can be separated by this chromatographic method.

The following abbreviations are often used in ion chromatography (IC):

SCX	strong cation exchanger
WCX	weak cation exchanger
SAX	strong anion exchanger
WAX	weak anion exchanger
IEX	ion-exchange chromatography
IEC	ion exclusion chromatography
IPC	ion-pair chromatography
HPIC	high-performance ion chromatography

The principle of ion-exchange chromatography is as follows: Ions differing in their charges interact with the functional groups of the exchanger with different intensities, and can be selectively eluted from the matrix by changing the elution strength of the mobile phase (variation of ion concentration, i.e., "salt gradient" or pH gradient).

During the ion-exchange process an equilibrium between, for example, the cation in solution and the anion fixed on the matrix is formed. For a particular ion this equilibrium is described by the equilibrium constant K_D. The equilibrium constant depends on the ion size, its charge, and its polarizability; i.e., the affinity for the stationary phase (ion exchanger) increases with

1) Increasing charge
2) Increasing polarizability
3) Decreasing size of the ion investigated

Usually this equilibrium is shifted toward the side of the fixed ions on the exchanger material. Therefore the effluent in IEX always contains ions (buffers, acids, bases) that compete with the analyte for active sites on the matrix so that the analyte can be desorbed from the matrix. The ionic strength of a mobile phase in IEX is a measure of its elution strength. In addition, the pH and dissociation rate of the analyte also have a great influence on retention. Furthermore, the technique shows a much better selectivity for cations differing in their charges than for anions.

Instrumentation. A schematic of an ion-exchange chromatograph is given in Figure 29. In principle, this is the same setup as for a normal HPLC system consisting of a pump (b), sample introduction system (c), separation column (d), and detector (e). IEX differs from normal HPLC in the "suppressor column" that is installed between the separation column and the detector (e.g., conductivity detector). This suppressor column has two tasks: First, it reduces the high background conductivity of the effluent; second, it converts the analytes into species with higher conductivity. The function of the suppressor column can best be explained by a practical example:

The Cl^- content of a water sample is to be determined. The mobile phase consists of an $Na_2CO_3 - NaHCO_3$ buffer solution. The reactions taking place in the suppressor column — which is a strong cation exchanger (SCX) in the protonated form — are the following:

1) Background conductivity is reduced by exchanging the sodium ions of the buffer with protons. The resulting H_2CO_3 is nonconducting
2) At the same time, NaCl is converted into the strong contucting HCl:
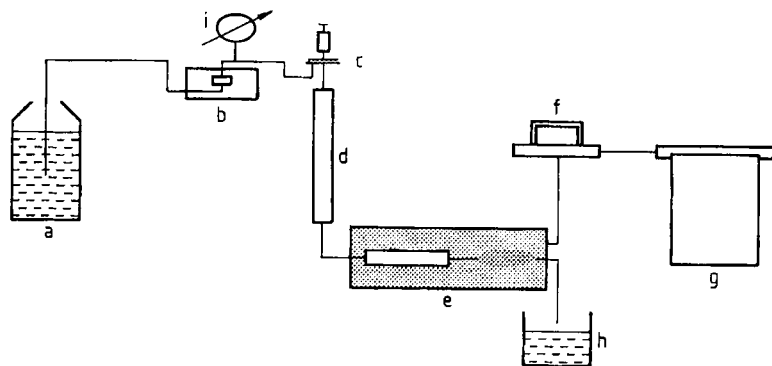
Figure 29. Schematic setup of a simple ion-exchange chromatographic system
a) Solvent; b) Pump; c) Injection port; d) Separation column; e) Detector with suppressor column; f) Control unit; g) Recorder;
h) Waste; i) Manometer

$$\text{Resin-SO}_3\text{H}^+ + \text{NaHCO}_3 \longrightarrow \text{Resin-SO}_3\text{Na}^+ + \text{H}_2\text{CO}_3$$

$$\text{Resin-SO}_3\text{H}^+ + \text{NaCl} \longrightarrow \text{Resin-SO}_3\text{Na}^+ + \text{HCl}$$
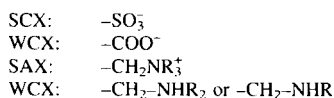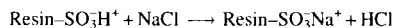
This technique is often referred to as *suppressed conductivity detection*.

Because of the corrosive solvents used in IEX, conventional stainless steel HPLC instrumentation is unsuitable. The pump-head, capillaries, and column should be made of inert material such as PEEK, PTFE, tetrafluoroethylene – ethylene copolymer, polypropylene, ruby and sapphire. Otherwise, contamination problems (especially iron contamination) can occur in the mobile phase, decreasing the separation efficiency and capacity, and increasing the noise of the amperometic detectors.

Detectors. In ion chromatography, suppressed conductivity detection is still the most frequently used technique. Indirect UV detection (Section 12.2.6.3) can also be a powerful tool for the detection of ions, although with respect to sensitivity, conductivity detection (detection power $10-100$ ng/g) cannot be surpassed. In amino acid analysis, UV and fluorescence detection (after post- or precolumn derivatization) is often applied.

Column Packings. Modern column packings for ion-exchange chromatography are made of a polystyrene matrix with different functional groups, depending on whether it is a cation or an anion-exchange resin.

The following functional groups are used:

SCX:	$-\text{SO}_3^-$
WCX:	$-\text{COO}^-$
SAX:	$-\text{CH}_2\text{NR}_3^+$
WCX:	$-\text{CH}_2-\text{NHR}_2$ or $-\text{CH}_2-\text{NHR}$

Polystyrene column packings are also available with small particle sizes (< 10 µm). Compared to inorganic ion exchangers (silicates or zeolites), they have the advantage of being pH resistant over the entire pH range. Many protein separations must be carried out at a pH of ca. 10. A silica-based exchanger would be destroyed at this basic pH, but polystyrene-based exchangers are not.

A disadvantage of the polystyrene matrix for high-performance ion chromatography is the fact that it tends to swell or shrink if certain solvents are used.

An important characteristic of every ion-exchange material is its "total exchange capacity," expressed in equivalents per unit weight (microequivalents per gram). One equivalent corresponds to 1 mol charge. For weak ion exchangers the capacity is a function of pH, whereas the capacity of strong ion exchangers is independent of pH. Figure 30 shows a simultaneous isocratic separation of alkali and alkaline-earth cations obtained by IEX.

12.5.4. Size Exclusion Chromatography

Although separation due to size exclusion has been known for a long time [82], systematic in-
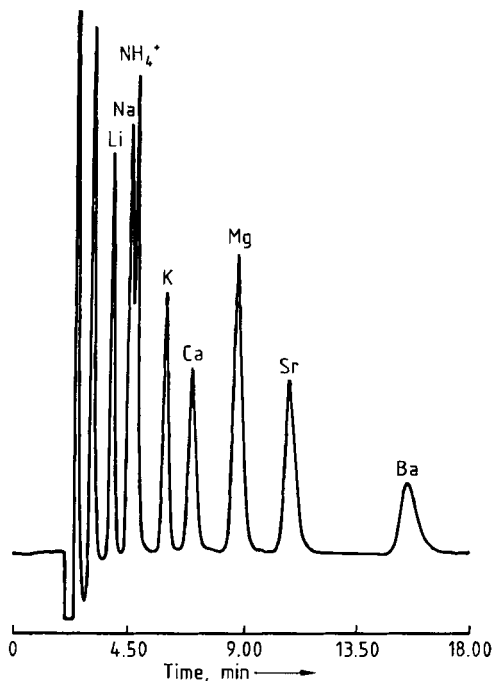
Figure 30. Simultaneous isocratic analysis of alkali and alkaline-earth cations with ion-exchange chromatography Column: Polyencap WCX; eluent: 4 mmol/L citric acid; 0.7 mmol/L dipicolinic acid in water, 1 mL/min; standard solution: 2.5 – 40 μL/L in water; detection: conductivity

vestigation of this method began only after the development of synthetic matrices [83], [84]. If the mobile phase consists of or contains water, size exclusion chromatography (SEC) is often called *gel filtration chromatography* (GFC). If the mobile phase is an organic solvent, the term *gel permeation chromatography* (GPC) is often used. Another frequently used abbreviation is HPSEC, which stands for high-performance size exclusion chromatography.

In SEC an analyte is, by definition, separated only according to its molecular size, not its molecular mass. This is important to remember since the same solute (with its specific molecular mass) can have a different structure, and therefore different sample size, in different solvents. If all other interactions (ionic interactions, adsorption, ect.) with the stationary phase are suppressed, the selectivity of the system depends entirely on the different accessibility of the pores of the matrix for the solutes. Very small molecules (usually solvent molecules) have access to the entire intra particle volume V_i of the stationary phase. In contrast to the moving interparticle volume V_m (which is the effluent volume outsite the particle pores) the intra particle volume V_i is called the stagnant mobile phase. The total solvent volume V_0 inside the column is therefore

$$V_0 = V_m + V_i$$

Since solvent molecules usually have the smallest size, they elute at the end of an SEC chromatogram with a retention time t_0. In other words, they completely permeate the stationary phase ("total permeation"). In all other chromatographic modes the solvent elutes first, if no exclusion has appeared. In SEC, molecules that are so large that they cannot get into the pores stay only in the faster-moving interparticle volume V_m and elute as the first peaks of the chromatogram. For these large analytes the column behaves as if it were filled with nonporous glass beads. The difference between the retention volumes of the largest (excluded) and smallest (solvent peak) molecules is the pore volume or interparticle volume V_i of the stationary phase. Medium-sized molecules permeate the stationary phase only partially, and consequently elute between the total permeation t_0 and the totally excluded solutes (dead or interparticle volume). Generally, the smaller a molecule, the higher is the accessible pore volume and therefore the longer is the time required to pass through the column.

In SEC the retention volume V_R is used instead of the retention time t_R. V_R is defined as the sum of the interparticle volume V_m and the accessible pore volume V_p for the particular solute:

$$V_R = V_m + V_p$$

For very small molecules (e.g., solvent) V_R is

$$V_R = V_m + V_p = V_0$$

For large and therefore excluded particles, V_R is

$$V_R = V_m$$

To become independent of various packings, column volume, and pore volume, and to enable a comparison of data from different packings, a distribution coefficient K_0 was introduced

$$K_0 = \frac{V_R - V_m}{V_i}$$

The above equation shows that the solvent (last eluting peak) has a K_0 of 1. Consequently, the entire sample must elute with $K_0 < 1$. Because all components of a sample elute so early, band broadening is reduced, which results in high and symmetric peak shapes. Thus, even relatively non-sensitive but therefore universal detectors such as the refractive index detector can be used. Since the solvent peak usually elutes at the end of an chromatogram the analysis time is predictable.

Column Packings in SEC. Two types of packings exist: Rigid packings, which are essential for HPSEC because of their stability, and nonrigid gels. The latter are not resistant to higher pressure and are therefore used in open-column SEC.

In *HPSEC* the most common column packings consist of pure or chemically modified silica or, recently, of highly cross-linked macroporous styrene–divinylbenzene copolymer. Due to the differing production processes for these polymer phases (e.g., amount of inserted cross-links), various pore sizes are available with small (5 and 10 µm), pressure-stable (maximum pressure 10–15 MPa), spherical particles [85]. The advantage of these polymer-based columns compared to silica-based columns is that under certain conditions (e.g., pH $> 4 - 5$), silica can act as a weakly acidic ion exchanger. Interactions between the sample and the active silanol groups (e.g., adsorption) are also possible and must be suppressed by proper choice of effluent. Ion exchange, for instance, is suppressed by increasing the ionic strength of the eluent (e.g., inserting a salt or buffer, 0.1–0.5 mol/L). Adsorption effects resulting from silica can be reduced by modifying the silica with chlorotrimethylsilane [86]. An interaction between the analytes and the column packing is indicated either by peak tailing (uncertain) or by the elution of any components after the solvent peak ($t_R > t_0$ or $K_0 > 1$). Nevertheless, silica-based SEC column also have their advantages. Here, pore size and pore volume are independent of the effluent. A silica column can be calibrated and characterized, for instance, with a polystyrene standard in an organic solvent (e.g., CH_2Cl_2); then, the organic solvent is replaced by a water-containing effluent, water-soluble polymers can be investigated without recalibration. In contrast, the pore size and thus the pore volume of polymer-based materials often change with the mobile phase applied. In some cases, a change of solvent can lead to a collapse of the particles, which results in a significant decrease of performance. When poly-styrene gels are used the mobile phase must be degassed very well, because air bubbles can damage the packing. In addition, aromatic sample compounds may also form $\pi - \pi$ interactions with the styrene–divinylbenzene copolymer.

To obtain high separation efficiency, similar to other LC techniques, small particles (≤ 5 µm) and a narrow pore-size distribution are necessary. Common pore sizes in SEC for low molar mass samples are between 6–12 nm, whereas high molecular mass samples (e.g., proteins) require a pore diameter from 30–100 nm or wider (e.g., 400 nm). Since separation capacity is correlated directly with pore size and therefore pore volume, the latter should be as large as possible. To enlarge the pore volume of an SEC system, either the sample can be reintroduced into the same column after the first separation (recycle SEC) or several columns with the same packing can be connected in series. Doubling the column length doubles the elution volume and the analysis time. If the sample contains molecules with a molecular mass that exceed the separation range of the column, several columns with differing sep-aration ranges can be connected for the "first analysis" of an virtually unknown sample, to obtain maximum information about the molar mass distribution. Two or more column packings differing in pore size should not be inserted into the same column because particle sizing can occur, which decreases the separation efficiency. A problem with columns connected in series is that band broadening may occur. This can be reduced by avoiding any unnecessary dwell volumes, by using small particles and solvents with low viscosities (higher diffusion), and by working at higher temperature (high-temperature gel permeation chromatography).

For choice of the appropriate solvent in SEC the following rules apply:

1) The sample must be soluble in the mobile phase
2) The stationary phase must be resistant to the mobile phase
3) Possible interactions between solute and column packing should be suppressed by appropriate additives to the mobile phase or by control of its pH, ionic strength, etc.
4) The mobile phase should be more strongly adsorbed to the column packing than the solute
5) The maximum permitted solute concentration in SEC should not exceed 0.25 wt %; a concentration of 0.1 wt % is sufficient in most cases
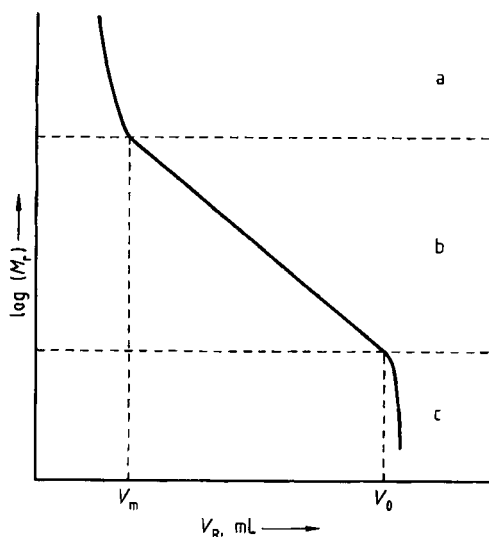
Figure 31. Typical calibration curve in SEC
a) Molecules are too large and elute with interparticle volume; b) Linear range for separation according to the molecular mass of the sample; c) Molecules are too small and penetrate the entire packing

Applications. If all interfering interactions between solute and stationary phase are suppressed, SEC is a very simple, reliable, and fast technique for the separation of many samples, especially oligomers and polymers. The only limitation is that the solutes to be separated must differ by at least 10 % in their molecular masses. Typical applications of SEC:

1) Determination of molar mass distribution of polymers
2) Determination of pore-size distribution of column packings
3) Cleanup technique prior to other high-resolution methods
4) Fingerprint analysis
5) Biopolymer separations (e.g., proteins)

To determine the elution (retention) volume V_R that corresponds to a certain molecular mass, the column must be calibrated with a standard mixture of known molar mass distribution. These standard, high-purity mixtures are available from various column suppliers. Remember that molecule size can change in different solvents and that different solutes (even with comparable molecular masses) can show differing molecular structure and size in the same solvent. In both cases, the retention volume obtained gives an incorrect molar mass.

Therefore, the standard must be chosen in such a way that it corresponds as much as possible to the sample structure. Figure 31 shows a typical schematic calibration curve obtained in SEC.

Because retention volume is the most important factor in SEC (it is related to the logarithm of the molecular mass, even small inaccuracies in the determination of V_R are magnified), the pump must deliver a very constant flow which can be controlled by a continuous flow measurement system, for example. Further information on SEC can be found in [19], [24], [87], [88].

12.6. Gradient Elution Technique

In gradient elution the mobile phase is usually composed of two solvents, A and B. The characteristic of gradient elution is that the composition of the mobile phase is changed during the chromatographic run. Usually the concentration of the stronger solvent B (percent) is increased during the run. In the reversed-phase systems that are generally used, solvent A is normally water.

The three main reasons for using gradient elution are:

1) Separation of sample components showing a wide range of k'
2) Fast evaluation of optimum eluent composition for subsequent isocratic analysis
3) Enhancement of detection limit

In addition, the technique can be applied to remove strongly retained solutes or impurities from the column, thus enhancing sensitivity and column life.

Gradient elution is usually applied in reversed-phase chromatography, although for certain separation problems in normal phase chromatography it can also be a very useful tool. Ideally, the solvent strength of the mobile phase should be adjusted in such a way that the k' range of the sample compounds lies between 2 and 10 (although k' values between 1 and 20 are often acceptable). For small molecules in reversed-phase liquid chromatography a 10 % increase in the percentage of B (organic solvent) decreases the k' value of every solute by a factor of ca. 2.

Compared to the isocratic elution mode, the equipment necessary for gradient elution is extended by a second solvent reservoir, an additional pump (in the case of a high-pressure-mixing gradient system), a mixing chamber, and a propor-
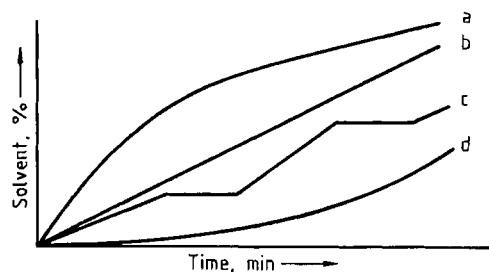
Figure 33. Different possible gradient profiles
a) Convex; b) Linear; c) Step; d) Concave

tioning valve that is regulated by a gradient control system.

Two types of gradient systems are possible (see Fig. 32): The low-pressure-mixing gradient system and the high-pressure-mixing gradient system. In the first case, only one HPLC pump is necessary because the two solvents (A and B) are mixed before the pump (i.e., in the low-pressure area of the system). With the high-pressure gradient system, two pumps are required because solvents A (e.g., water) and B (e.g., organic solvent) are mixed after leaving the pump in the high-pressure area of the system.

Many different gradient profiles are possible (see Fig. 33); the most frequent and useful are the linear and segmented (multilinear) profiles. Besides binary gradients (two solvents), some separations require ternary gradients in which the mobile phase consists of three solvents (A, B, C). With these ternary gradients, special selectivity effects can be achieved [89].

12.7. Quantitative Analysis

Quantitative analysis should deliver reproducible and reliable information about the composition of the sample investigated. Since end of the 1980s, quantitative analysis has become extremely important for product quality control (e.g., food industry, pharmaceutical industry) and environmental analysis. It is of great economic importance, and an erroneous quantitative analysis can result in fatal consequences. Therefore, this chapter introduces different quantification methods for HPLC and describes some often-occurring sources of error.

Correct sampling and sample storage are essential to obtain exact and reliable analytical re-
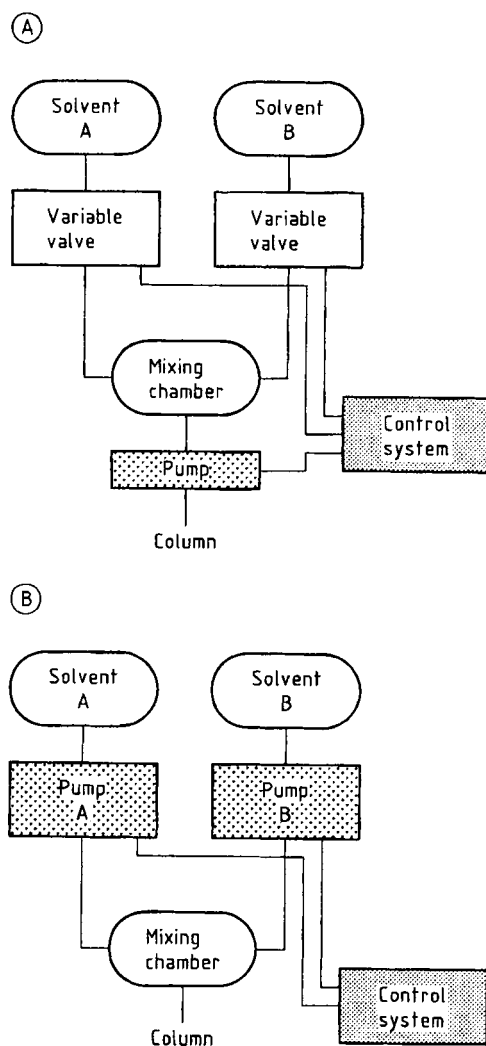
Figure 32. Gradient elution technique
A) Low-pressure system, B) High-pressure system

sults. After a possible cleanup step, the sample is injected into the chromatographic system and a strip chart is obtained from the recorder or electronic integrator, on which the intensity of the detector signal is plotted against analysis time (chromatogram). To be able to make any quantitative statements with the help of this chromatogram, the system must have been previously calibrated. The two most important methods are the external standard method and the internal standard method.

12.7.1. External Standard Method

The external standard method is the technique used most frequently to gather quantitative information from a chromatogram. In this case, a pure reference substance (ideally the same compound as the one to be determined in the sample) is injected in increasing concentrations and the peak areas or peak heights obtained are plotted versus the concentration (calibration curve; →Chemometrics). These calibration curves should show a constant slope (linear curves), and the intercept should be as close to zero as possible. Since the calibration curves usually show nonlinear behavior and flatten off at higher concentrations (see Section 12.2.6), the quantification should be carried out only within the linear part of the curve.

A calibration curve must be recorded for each component of interest. The peak areas/heights of the sample chromatogram are then compared with those in the calibration curve. All chromatographic conditions must remain absolutely constant, for both sample and reference substances. If the volume of injected sample and injected reference solution is constant (which is usually the case if a sample loop is applied), the calculation can be carried out as follows:

$$\frac{A_{si} C_{ref}}{A_{ref}} = C_{si} \tag{1}$$

where A_{si} is the area or height of component i in the sample; A_{ref} the area or height of the reference peak in the standard solution; C_{ref} the concentration of reference substance in the standard solution in milligrams per milliliter; and C_{si} the concentration of component i in the sample solution (milligrams per milliliter).

The amount (percent) of component i in the sample is:

$$\frac{A_{si} C_{ref} V_s}{A_{ref} M_s} \cdot 100 = \% C_{si} \tag{2}$$

where V_s is the volume of the sample solution in milliliter; M_s the mass of sample in volume V_s in milligrams; and $\% C_{si}$ is the percentage component of i in the sample.

12.7.2. Internal Standard Method

The internal standard method is well known from GC (→ Gas Chromatography) and should always be used if exact control of the injection volume is not possible. A defined amount of a chosen compound (internal standard) is added to the sample, so that two peaks (that of the internal standard and that of the compound of interest) must be measured to obtain one result. An internal standard must fulfill several requirements: It must not already be present in the sample; it should elute close to the peak of interest; and the two peak sizes should not differ too much. If possible, peaks should not overlap, and no chemical interaction between the internal standard and other sample components should occur. The internal standard must also have good storage stability and be available in high purity. This method is slightly less accurate than the external standard, but it has several significant advantages:

1) Compensation of varying injection volumes
2) Compensation of recovery of sample during sample cleanup, especially in trace analysis
3) Subjection of the standard to the same chemical reaction (e.g., amino acid analysis) in postcolumn reactions

Calculations for the internal standard method are similar to those mentioned above for the external standard. More detailed information concerning the external and internal standard methods, as well as other methods of evaluation [e.g., area normalization method (100 % method), addition method] is given in [11].

12.7.3. Error Sources in HPLC (→Chemometrics)

Error sources in HPLC are so manifold that they cannot all be discussed completely. Nevertheless, one can distinguish between random errors and systematic errors. Whereas random errors are relatively easy to recognize and avoid, systematic errors often lead to false quantitative interpretation of chromatograms. Two simple ways of recognizing a systematic error in a newly developed (or newly applied) analytical method are either to test this new method with certified materials or—if an "old" (i.e., validated) method already exists—to compare the result from the old and the new methods. If the new method shows good reproducibility, a correction factor can be calculated easily by

$$C_f = \frac{A_{certified}}{A_{new}} \tag{3}$$

where C_f is the quantitative correction factor; $A_{certified}$ the peak area/height or concentration of component i in the certified reference sample or the "old" method; and A_{new} the peak area/height or concentration of component i obtained with the "new" method. Multiplying the calculated amount of component i (see Eq. 1) in the sample by the correction factor C_f compensates for the difference that can be caused by one or more systematic errors. (If possible, however, the systematic error should be identified and eliminated.)

Sample Injection. For every quantitative HPLC analysis, especially in the external standard mode, a constant volume of injected sample solution is essential. This can be achieved only by using sample loops (5, 10, 20 µL, etc.) or automatic sample introduction system (see Section 12.2.4). The constancy in injection volume of these systems has to be checked from time to time. This is done by injecting the same sample several times and comparing the peak areas and peak heights generated. If a constant injected sample volume cannot be ensured, the internal standard method should be applied to compensate for these variations. A variation of the peak area can also result from short-time pulsation of the pump.

Another error in quantification can be caused by overloading the column (rule of thumb: A maximum of 1 mg of analyte per gram of stationary phase). The overloading of a chromatographic column is shown by the appearance of decreasing peak symmetry (fronting or tailing), peak broadening, and decreasing retention time. Obviously, loss of the sample through leakages or strong adsorption to the stationary phase must be eliminated. The latter is often the reason for a baseline rise or for nonlinear behavior of the calibration curve in the lower concentration range.

Pump. A constant flow rate of the mobile phase is another essential condition for quantitative analysis. Especially for peak area quantification, even small variations in flow can have a disastrous effect because the recorder or integrator plots the concentration (signal intensity) versus analysis time (not versus flow rate). Therefore a variation in flow automatically influences the recorded peak area. Quantification over peak height is not so strongly influenced by fluctuations in the flow of the mobile phase.

As a rule of thumb, peak area calculations are sensitive to baseline variations, and therefore many trace analyses are quantified over the peak height. On the other hand, peak height is influenced strongly by retention time (the longer the retention time, the greater is the dilution of the sample and therefore the weaker is the detector output signal). Generally every variation of flow (even very slow variations — often recognizable by a baseline drift) must be avoided for quantitative analysis.

Detector. If possible, the use of highly selective detectors (e.g., fluorescence detector) should be preferred in quantitative analysis because sensitivity is usually enhanced and, due to the high selectivity, peak overlapping is often avoided. However, most HPLC detectors (even the selective ones) are photometric detectors that convert the light absorption of an analyte in the mobile phase into an electrical signal, which is recorded as a function of time. Therefore, their linearity is limited by the Lambert–Beer law (see Section 12.2.6). The influence of temperature on detection (UV, fluorescence, refractive index, etc.) should not be neglected. The solute concentration in the detector cell is indirectly a function of temperature because the density of the mobile phase is a function of temperature [90]. Additionally, the position of the absorption maxima changes with temperature. Generally, for quantitative analysis, temperature variations must be avoided (in hot countries, daily temperature variations of up to 10 °C are quite common) by thermostating the chromatographic system including the detector cell.

Processing of Data. The chromatogram obtained contains qualitative (t_R and k') and quantitative (peak area or height) information. The quantitative information can be evaluated manually or automatically (by an integrator or computer system).

If the baseline shows any drift, *peak heights* are often more correctly determined manually than by an automatic system (important for trace analysis). The *peak area* is very seldom calculated manually, and if it is, a peak area equivalent is used, which is the product of peak height and peak width at baseline. In older publications, the peak area was often calculated by cutting out the peak and weighing it [88]. This method also gave very good results.

12.8. Sample Preparation and Derivatization

12.8.1. Sample Cleanup and Enrichment

Usually, before a sample is injected into the chromatographic system it has to undergo a cleanup procedure. In the simplest case, cleanup consists of only a filtration step. Each sample solution should be filtered routinely before injection onto the column. This prevents sample loop, capillary, or column plugging by small particles in the sample solution and also prolongs column life significantly. Besides filtration, more complex samples require further cleanup and often an additional enrichment step. No general rule for sample cleanup or sample enrichment can be given; each sample matrix is different, and the particular cleanup must be adjusted to the sample in question. Nevertheless, water-soluble (polar) samples can be cleaned up or enriched simply by liquid–solid extractions (e.g., small silica columns) or liquid–liquid extractions in a separatory funnel. The latter is also very important for the isolation of hydrophobic analytes [e.g., polyaromatic hydrocarbons (PAHs), pesticides] from a polar matrix (i.e., water). Isolation of hydrophobic solutes from a polar matrix is done easily by pumping the solution over a lipophilic stationary phase (e.g., RP-18). Here, the hydrophobic analytes are retained while the more polar components do not interact with the stationary phase. After a certain amount of sample solution (depending on the separation problem and the capacity of the solid phase) has been pumped over the stationary phase (either an open-column "solid-phase extraction" or a small HPLC column), the mobile phase is changed to a hydrophobic (organic) solvent, which elutes the retained analytes. After evaporation of the organic solvent, the analytes can be dissolved in a known amount of solvent for analysis.

Another very interesting and promising cleanup technique is supercritical fluid extraction (SFE). This method, which usually employs supercritical carbon dioxide as an extraction medium, shows a high selectivity toward nonpolar solutes. As a rule of thumb, if a solute is soluble in hydrophobic solvents such as heptane or sometimes even in more polar ones such as dichloromethane, it should also be soluble in supercritical carbon dioxide. Through addition of modifiers (e.g., methanol, water, formic acid), even relatively polar molecules are extractable, although the in-

crease in solvent strength is paid for by a loss in selectivity. The selectivity and thus the solvent strength of the supercritical fluid can also be adjusted by variation of the extraction pressure, density of the extractant, or temperature.

SFE with carbon dioxide offers several advantages compared to classical solvent extraction methods: (1) it is very selective; (2) it is fast, because of lower viscosity and approximately 10 times higher diffusion coefficients; and (3) CO_2 is neither flammable nor explosive and is nontoxic. SFE is suitable for oxidizable and temperature-sensitive samples. In addition, no reactions occur between extraction medium and sample compounds and on-line coupling is possible with both supercritical fluid chromatography (SFC) and gas chromatography.

Nevertheless, this technique also has some disadvantages, especially in collecting the extracted analytes. High molecular mass analytes tend to precipitate inside the restrictor (see Section 12.10.4.5) and cause it to plug, during the expansion of the supercritical fluid (expansion of supercritical fluids results in a decrease of density and therefore solvent power). Low molecular mass samples (volatile samples) can be carried away in form of an "analyte fog" to a certain extent during the expansion procedure, resulting in insufficient recovery. Further information on this promising sample cleanup method is given in [91], [92].

12.8.2. Pre- and Postcolumn Derivatization

Sample derivatization is often applied to enhance detection selectivity or sensitivity. This technique is also applied in GC and TLC. Although derivatization is a very helpful tool in chromatography it has the disadvantage that derivatization reactions make the individual sample compounds more similar in adsorption behavior, which can decrease separation selectivity.

In LC the derivatization can be performed either before the chromatographic separation (precolumn derivatization) or after the analytes have passed the column (postcolumn derivatization).

12.8.2.1. Precolumn Derivatization

Precolumn derivatization is the most frequently applied derivatization technique for many reasons. One very important factor is that the chromatographic equipment need not be
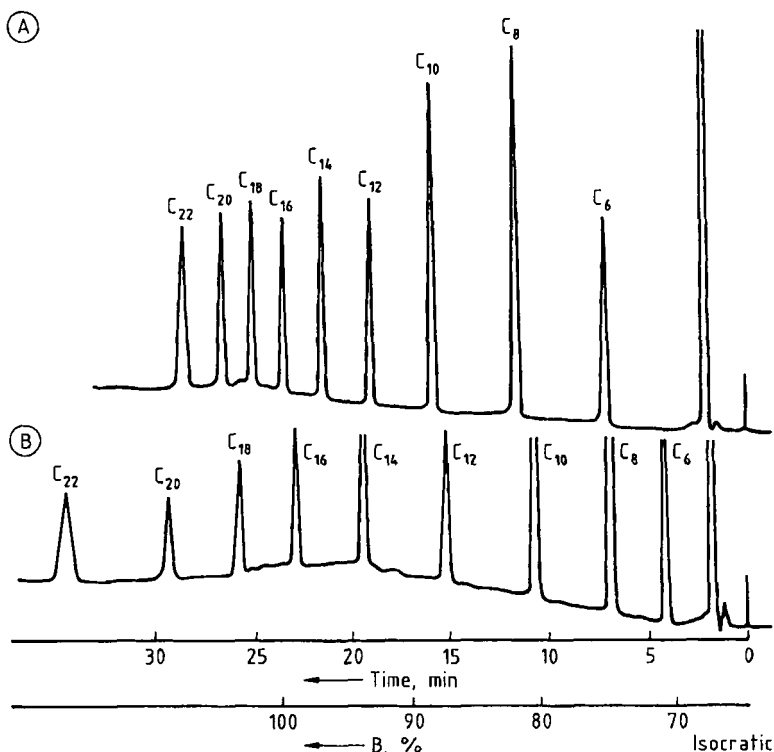
Figure 34. Separation of fatty acid phenacyl esters on an RP-18 column (300×4.2 mm i.d.) with flow = 2 mL/min and t_g= 20 min linear [95]

A) 70 % Methanol in water to 100 % methanol; B) 70 % Acetonitrile in water to 100 % Acetonitrile

changed because precolumn derivatization is performed off-line. Furthermore, the reaction time is of minor importance. Nevertheless, the derivatization should be quantitative, and in certain cases, excess derivatization reagents can lead to poor chromatographic separation.

The most frequent application of precolumn derivatization is in amino acid analysis. Here, the amino acids are derivatized for instance with 9-fluorenyl methoxycarbonyl chloride (FMOC) or *o*-phthalaldehyde (OPA) to form strongly fluorescing compounds, which can be detected easily with a fluorescence detector [93], [94]. Another typical application for precolumn derivatization is in the analysis of fatty acids. Fatty acids have no chromophore, so UV detection is impossible. Even highly unsaturated fatty acids can be detected only in the very far UV range (< 200 nm) with poor sensitivity. Other detectors, such as the refractive index detector, have a disadvantage because they cannot be used in gradient elution, which is essential for fatty acid analysis. Therefore, the fatty

acids are reacted with UV-absorbing reagents (e.g., phenacyl bromide) to form fatty acid phenacyl esters, which are easily detectable at 254 nm [95] (see Fig. 34).

12.8.2.2. Postcolumn Derivatization

In postcolumn derivatization, also known as chemical reaction detector (CRD), the sample compounds are derivatized on-line after separation. Derivatization takes place in a reactor situated between the column outlet and the detector (see Fig. 35). The chromatographic equipment is therefore extended by an additional reagent pump, which delivers the reagent into a zero-void-volume mixing chamber (Fig. 36). Here the reagent is mixed thoroughly with the mobile phase containing the solute. (An ordinary three-way unit is not sufficient as a mixing chamber in postcolumn derivatization.) The mixture is then pumped through a reactor where actual derivatization occurs before the sample reaches the detector. The
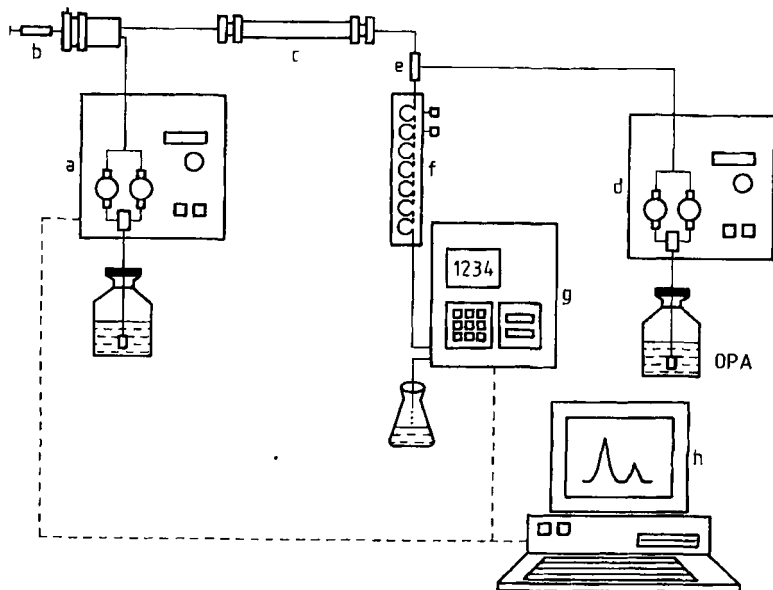
Figure 35. Diagram of postcolumn derivatization system
a) Pump; b) Sample introduction; c) Separation column; d) Reagent pump; e) Mixing chamber; f) Reactor; g) Detector; h) Data control unit

derivatization reagent must be delivered continuously, with as little pulsation as possible.

The reaction time required for a particular derivatization is the limiting factor. It must be short enough so that derivatization can take place during passage of the solutes through the thermostated reactor. Of course, the required reaction time can be prolonged to a certain extent by enlarging the reactor, but this is paid for by increasing backpressure and additional peak dispersion. Even if the required reaction time for a particular derivatization (i.e., for 100% conversion) cannot be achieved with a reactor, this is often no problem in postcolumn derivatization since the derivatization need not to be quantitative, merely reproducible.

The general rule of avoiding any additional void volume between the column outlet and the detector cell cannot be fulfilled in postcolumn derivatization. Therefore, loss of separation efficiency to a certain extent, due to peak dispersion, is unavoidable. To reduce this drawback, several reactor types have been developed. The most important ones are:

1) Packed-bed reactors
2) Open tubes with optimal geometric orientation
3) Open tubes with liquid or gas segmentation
4) Coiled open tubes

Many excellent papers have been published that compare these different types of reactors [94], [96]; therefore, only the most important reactors (open tubes with optimal geographic orientation) are discussed here in more detail.

Coiled Open Tubes. Initially, WALKLING [97] and HALASZ [98] showed that systematic deformation of open tubes can significantly reduce peak dispersion, compared to a simply coiled tube with uniform i.d. A detailed discussion of the different parameters influencing peak dispersion (e.g., relationship of axis, coiling diameter, knot distance, kinematic viscosity of the mobile phase, inner surface of the tube) is given in [99]. The breakthrough in the optimal geometric orientation of open tubes came with the introduction of PTFE capillaries (i.d. $= 0.2 - 0.5$ mm). With this easily deformable material (compared to glass or stainless steel capillaries), not only could coiled and waved reactors be made [99], or figure eights formed around two rods [100], but the capillary could be knitted [100], [101]. Capillaries are knitted in such a way that each right loop is followed by a left one (with small coiling diameter) so that each loop is bent out of the plane of the preceding one (Fig. 37). Because of the small coiling diameter and continuous change in direction,
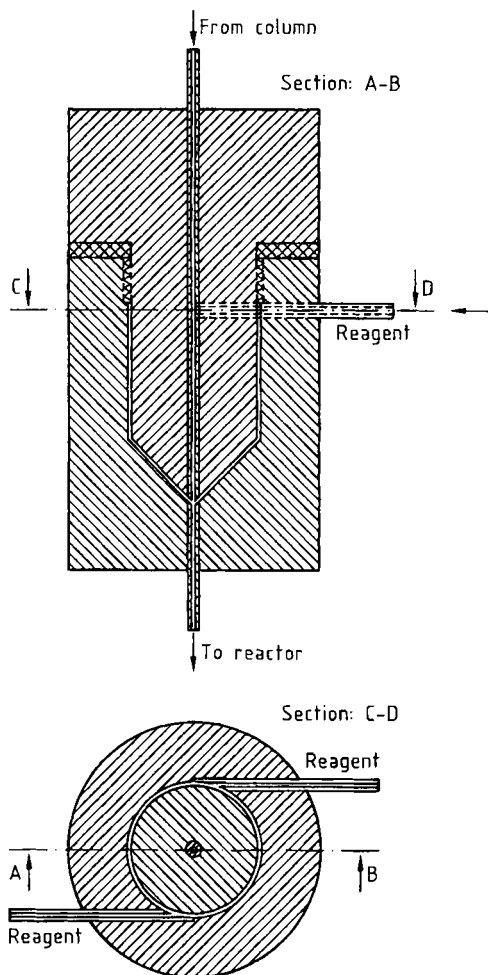
Figure 36. Example of a postcolumn derivatization mixing chamber (cyclon type) [94]
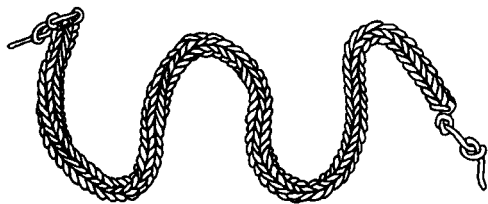
Figure 37. Knitted open tube (KOT) reactor

centrifugal forces are effective even at low flow rates, initiating a secondary flow, that is ultimately responsible for the reduced band broadening. These so-called knitted open tubes (KOTs) have
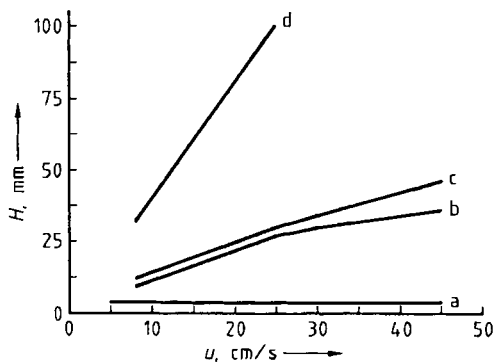
Figure 38. Peak dispersion in open tubes
a) Straight capillary; b) $R = 2.6$ cm; c) $R = 0.35$ cm; d) Knitted open tube (KOT)

the best performance in terms of peak dispersion. With these KOTs, peak dispersion could be made independent of linear velocity over a large velocity range (even at very low flow rates) (Fig. 38).

Packed-Bed Reactors. Packed-bed reactors are packed preferably with small glass beads (20–45 μm). The reactor length and therefore the reaction time are limited by the generated back-pressure of the packed-bed reactor. This is the reason it is generally used only for fast derivatization reactions.

Open Tubes with Liquid or Gas Segmentation. Another often-used reactor type (especially for relatively long-lasting derivatization reactions) is the open tube with liquid or gas segmentation [94]. Segmentation prevents the formation of a parabolic flow profile (axial dispersion), and peak dispersion is due mainly to the "phase separator," which is responsible for removal of the segments prior to detection. The principle of segmentation was first used by MOORE and STEIN [102], who developed an amino acid autoanalyzer (ninhydrin reaction) in 1958.

During the development of a new postcolumn derivatization for a particular separation problem, the following points should be considered:

1) An appropriate derivatization reaction should be sought; these are often described in the literature for classical off-line derivatization
2) The reagent concentration should be optimized: Highly concentrated reagents should be delivered with a low flow rate to prevent any unnecessary sample dilution in the mobile phase

Table 7. Examples of postcolumn derivatization techniques in HPLC

Compound	Reagent,* reaction type	Chromatography	Reference
Amino acids	ninhydrin	IEX	[103]
	fluorescamine	IEX	[104]
	OPA	IEX	[105]
Peptides	fluorescamine	RP	[106]
Mono-, oligosaccharides	3,5-dihydroxytoluene	IEX	[107]
	catalytic hydrolysis/ABAH	IEX	[108]
Carbamate pesticides	alkaline hydrolysis/OPA	RP	[109]
Glyphosate (herbicide)	hypochlorite oxidation/OPA	RP	[110]
Vitamin B_1	thiochrome	RP	[111]
Vitamin B_6	semicarbazide	RP	[112]
Vitamin C	oxidation/o-phenylenediamine	RP	[113]

* OPA = o-phthaladehyde: ABAH = 4-aminobenzoic acid hydrazide.

3) Reaction time should be as short as possible; the longer the required reaction time is, the longer must the reactor be, which results in increased back-pressure and peak dispersion
4) To shorten the required reaction time. the reactor temperature should be as high as possible (an increase of 10 °C doubles the reaction speed)
5) An ca. 80 % conversion of the original solute into the derivatized compound is sufficient to achieve maximum detection sensitivity even for quantitative analysis

Table 7 shows a few derivatization reagents and their typical applications.

12.9. Coupling Techniques

Many different coupling techniques in HPLC were applied since the 1970s but only a few proved to be a significant advantage for the analytical chemist. Even simple HPLC equipment can pose severe handling and reproducibility problems for the beginner. Each additional coupling technique naturally increases these problems, making sophisticated coupling techniques more interesting for academic use than for routine analysis. Nevertheless, coupling techniques such as HPLC – GC and HPLC – MS (mass spectrometry) have become very important tools for analytical chemistry. Another often-applied coupling technique is column switching.

12.9.1. Column Switching

Column switching was described in the late 1960s by DEANS [114] for multidimensional GC

(→ Gas Chromatography). In principle, it is the coupling of different columns (usually with different selectivities) by means of a manual or automatically driven high-pressure valve. The first application of column switching in HPLC was described by HUBER et al. [115] in 1973. They used this technique as a simple and inexpensive alternative to gradient elution. Gradient systems were very expensive in those days. However, in the following years, good and reasonably priced gradient systems came onto the market, making column switching unnecessary as an alternative.

New applications for this technique were found especially in the analysis of very complex samples (e.g., drug analysis in blood serum), where the analyte is selectively "cut out" of the sample matrix. According to the position of the "cut" in a chromatogram, front-cut, heart-cut, and end-cut techniques can be distinguished. Other applications of column switching are on-line sample cleanup, backflush technique, and multidimensional chromatography. Figure 39 shows different examples of switching arrangements.

The high-pressure switching valve is the most important part of the instrumentation in this technique. It must work reliably even after several thousand switches under high pressure, must have a very small dwell volume, and must be chemically inert to both the sample and the mobile phase. Normal switching valves have six ports, but four- and up to ten-port valves are commercially available as well. Their setup is very similar to the usual sample valve injectors, e.g., "rheodyne valves" (Section 12.2.4.2). They can be driven by hand or automatically either by pneumatic force or electronically. Pneumatically or electronically driven valves have the advantage that the entire analysis can be automated.
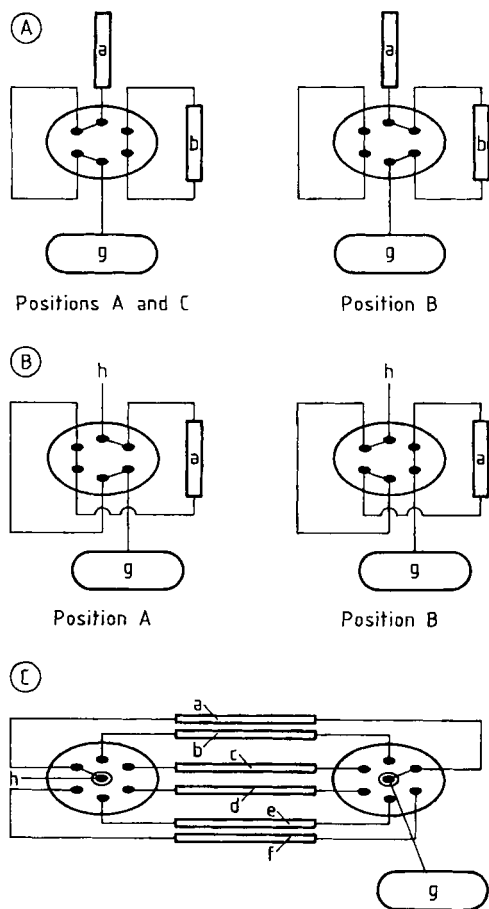
Figure 39. Arrangements for column switching technique
A) Heart-, front-, and end-cut techniques; B) Backflush
technique; C) Multidimensional chromatography
a)–f) Column 1–6; g) Detector; h) Injection valve

12.9.1.1. Front-, Heart-, and End-Cut Techniques

A sample is usually preseparated on a short precolumn and then directed onto an analytical column (with the same or different selectivity) where actual separation occurs. The *heart-cut technique* is most often used when the solute co-elutes with another sample or when it is situated on the flank of another peak. In this case the interesting part of the chromatogram is cut out and "parked" on the analytical column, while the rest of the sample is eluted from the precolumn into the waste. Thereafter, the valve is switched back to the analytical column position, and the parked sample

can now be baseline separated (Fig. 39 A). The *front- and end-cut techniques* are basically the same, differing only in the position of the cut.

12.9.1.2. On-Line Sample Cleanup and Trace Enrichment

Column switching is also very suitable for sample cleanup and can be combined with trace enrichment. A typical example of this technique is the on-line sample cleanup and enrichment of hydrophobic pesticides from aqueous solutions. Here the sample (water, blood, etc.) is pumped over a small RP column, where only the hydrophobic components are retained. Since the amount of sample can be chosen in any size or volume, hydrophobic trace enrichment takes place. After the entire sample volume has passed the precolumn, the high-pressure valve is switched to the analytical column position and an eluent (having a stronger elution strength for hydrophobic analytes e.g., dichloromethane) is passed through the RP precolumn, eluting the hydrophobic components onto the analytical column where separation occurs.

12.9.1.3. Backflush Technique

The switching arrangement for the backflush technique is shown in Figure 39 B. This technique is used mainly if the components of interest in a particular sample are very strongly retained by the stationary phase. Mobile phase flow is maintained in one direction until all early-eluting components have left the column. At this stage, the more strongly retained components are still near the head of the column. Then, the flow direction is reversed (backflush) to allow faster elution of these analytes. The switching arrangement for the backflush technique can be used in the routine analysis of very complex samples that contain components having a high affinity for the stationary phase. In this case, after a certain number of chromatographic runs the flow direction is reversed to flush the analytical column and elute these strongly retained components [116]. This lengthens column life significantly.

12.9.1.4. Multidimensional HPLC

In multidimensional HPLC, several columns with different selectivities are switched parallel to each other with the help of two high-pressure switching valves (see Fig. 39 C). This technique is

used for samples that contain components with a wide range of polarity (from low to high k' values). In this case, one should determine whether gradient elution would be more efficient.

All of the aforementioned column switching techniques can be combined with each other in various ways to provide analytical chemists with important tools for their investigations.

12.9.1.5. Working with Precolumns

In column switching, often at least one column is a precolumn. In the majority of cases the precolumn does not require high separation efficiency. Its purpose is to retain or enrich a certain part of a sample. The actual separation is performed on a second column (the analytical column), which possesses the appropriate separation efficiency. Therefore, the precolumn can be packed with inexpensive, coarse particles. Even a well-packed column bed is unnecessary. If the analyte of interest is retained on the precolumn in a relatively wide zone, this large elution band is reconcentrated as soon as it reaches the head of the analytical column (as long as the analytical column has a better selectivity and higher separation efficiency than the precolumn). Precolumns are usually short stainless steel columns (ca. 20×4 mm i.d.), packed with coarse material (30 – 60 μm), which need not have the small particle-size distribution required for analytical columns. Therefore, these precolumns can easily be homemade, and if purchased, they are inexpensive. If the precolumn is used as a "protection" for the analytical column (its original purpose), it should be well packed and the packing material should be the same as that in the analytical column.

12.9.2. HPLC – Mass Spectrometry

The coupling of high-performance liquid chromatography with mass spectrometry (HPLC – MS) has become a very important tool for the reliable identification and quantification of complex samples. In the pharmaceutical industry, for instance, the metabolic pathways or pharmacological effects of new products are often studied with this system.

The main problem in coupling the LC system with the mass spectrometer is the fact that the mass spectrometer is under a high vacuum and the mobile phase from the HPLC system must be removed before the analyte can enter the MS. This is the reason why HPLC – MS coupling is usually done with microbore or even capillary columns. In using the "normal" 4.0 – 4.6-mm i.d. columns, too much solvent would have to be removed. Initially, solvent removal was performed by the *moving belt* technique, for example. Today, the *direct liquid inlet* technique is often utilized. In this case, solvent eluting from the LC system is sent through an ca. 5-μm hole into the vacuum of the MS. The resulting aerosol comes into a "desolvation chamber," where the solvent is completely removed. With the *thermospray techniqe*, the analyte containing the mobile phase is sent through a heated (60 – 200 °C), narrow pipe (ca. 0.1-mm i.d.) directly into the MS vacuum of a specially designed ion source. This technique has the advantage that no split is necessary after the separation column, so the entire mobile phase can be sent into the MS system, which enhances detection limits significantly. Further, "normal" HPLC columns (4.0 – 4.6-mm i.d.) can be used in this technique, with flow rates between 0.8 and 1.5 mL/min.

For the HPLC – MS systems, many different ionization techniques have been described in the past. Various interface, ionization methods, and operating techniques applicable to LC – MS are discussed in [117] for instance: Thermospray, particle beam, electrospray (ES), field desorption (FD), fast atom bombardment (FAB), time of flight (TOF), etc. The electrospray technique produces a soft ionization for thermally labile compounds, while FAB has the advantage that higher molecular mass samples can be introduced into the mass spectrometer. Table 8 offers a rough guide to the applicability of various LC – MS interfaces. For more detailed information on LC – MS, see [118].

In HPLC – MS coupling, careful sample preparation is recommended. The "dilute-and-shoot" approach is not possible, especially when handling real-world samples such as environmental and biological samples. In these cases sample preparation techniques (see also Chaps. 12.8 and 12.9) such as liquid – liquid extraction, solid-phase extraction [119], or even affinity techniques (see also Chap. 12.11) directly coupled with LC – MS [120] can be useful.

Table 8. Overview of advantages and disadvantages of various LC – MS interfaces * [117]

LC – MS type	LC flow range	Detection limit	Solvent type range	Nonvolatile thermally labile samples	Semivolatile samples	Molecular mass range
Moving belt	+++++	++	+++++	+	+++++	+
Particle beam	+++	++	++++	+	+++++	+
Thermospray	+++++	++	++++	+++	+++++	+++
LC – FAB	+	++	+++	+++++	++++	++++
Electrospray	+	++++	++	+++++	++++	+++++

* Increasing numbers of "+" indicate improved performance.

12.10. Supercritical Fluid Chromatography

12.10.1. History

In 1822, CAGNIARD DE LA TOUR first described the critical state [121]; In 1869, ANDREW [122] published the first paper about a gas – liquid critical point. In 1879, HANNAY and HOGARTH described the solubility of cobalt and iron chlorides in supercritical ethanol [123]. LOVELOCK (1958) first suggested the use of supercritical fluids as a mobile phase in chromatography [124], and in 1962, KLESPER et al. [125] published the first paper on supercritical fluid chromatography. In this paper they described the separation of nickel porphyrins using supercritical chlorofluoromethanes as the mobile phase. This work was done before HPLC was introduced by PIEL in 1966 [9]. In the following years, SIE, RIJNDERS, and GIDDINGS [126] – [134], through their practical and theoretical studies, made significant contributions to understanding of the separation process with supercritical fluids. SIE and RIJNDERS at the Shell Laboratories in Amsterdam were the first to use the term supercritical fluid chromatography in 1967 [135].

In the 1970s, SFC declined in popularity, partly because of the very fast development of HPLC and because, in those days, working with supercritical fluids was not so easy due to several technical problems. Another reason SFC declined in popularity was that expectations had been set far too high in the early use of this technique.

Nevertheless, the development of SFC continued. In 1969 – 1970, JENTOFT and GOUW [136], [137] were the first to work with pressure programming and to use modifiers to increase the elution strength of the mobile phase. SCHNEIDER and coworkers [138] – [140] investigated the physicochemical aspects of supercritical fluids and thereby contributed much to the understanding of the technique.

The introduction of capillary columns in SFC by NOVOTNY and LEE in 1981 [141] led to a revival of this technique, which was followed by a large number of papers on new applications. At the same time (1981) the first commercial packed-column SFC instrument came on the market. Four years later, in 1985, the first capillary column SFC instrument was available commercially.

Since the late 1980s, capillary and packed-column SFC have become more and more popular, as shown by the growing number of published papers, new books, and international meetings devoted exclusively to SFC and SFE. The reason for this recent gain in popularity is that significant progress has been made in improving the instrumentation, which expanded the applications to many problems such as quantitative environmental analysis (although generally the sensitivity is not as good as in GC, for instance). Another reason is that modern chromatographers have begun to realize that the various analytical separation techniques (HPLC, GC, TLC, SFC, CE, etc.) should not compete against each other. Each technique has its advantages, and which technique is preferred should be decided on a case-to-case basis.

12.10.2. Theoretical Background

If a gas is compressed above its critical pressure p_c at temperatures exceeding the critical temperature t_c, then neither a gas nor a liquid is present but a so-called supercritical fluid, which combines properties of both. Table 9 shows the critical parameters of the most commonly used fluids in SFC.

Table 10 compares some important physicochemical properties of a liquid (HPLC), a gas (GC), and a supercritical fluid. The density — and thus the solvation characteristics of the supercritical fluid — are very similar to those of a liquid.

Table 9. Critical parameters of the most frequently used supercritical fluids [142]

Supercritical fluid	p_c, MPa	t_c, °C	ϱ_c, g/cm^3
CO_2	7.39	31.3	0.47
N_2O	7.34	36.5	0.45
SF_6	3.76	45.5	0.74
Xe	5.92	16.6	1.10
NH_3	11.4	132.5	0.24
CCl_2F_2	4.12	111.8	0.56
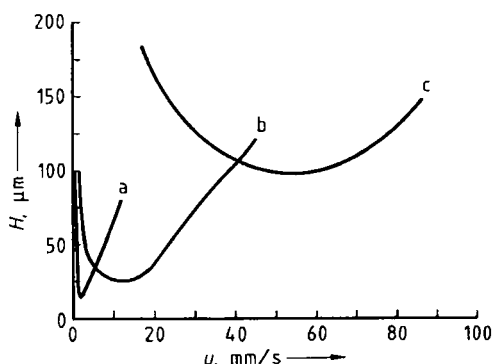
Figure 40. Comparison of Van Deemter curves in GC, HPLC, and SFC
All packed columns (particle size 10 μm); mobile phase in HPLC: methanol, in GC: N_2, in SFC: CO_2; samples in HPLC: benzene, in GC and SFC: methanea) HPLC; b) SFC; c) GC

On the other hand, the viscosity is comparable to a gas, and the diffusion coefficient is about ten times higher than that of a liquid. Therefore in SFC (especially packed-column SFC), analyses can be performed at much higher flow rates than in HPLC without loss of too much efficiency (see Fig. 40), which results in a significantly shorter analysis time. Because of the low viscosity of supercritical fluids, several packed columns can be combined in series, even with small particles (5 μm), without generating too high back-pressure. If several columns packed with the same stationary phase are connected in series, the separation efficiency can be improved dramatically. The combination of different stationary phases (columns) and the variation of their column length result in totally new and interesting selectivity ("selectivity tuning").

In general, a mobile phase for SFC should exhibit the following characteristics:

1) The critical parameters (temperature and pressure) should be achievable with commercial equipment
2) The fluid should be available in high purity and not be too expensive

Table 10. Comparison of physical properties of a liquid, a supercritical fluid, and a gas

	Liquid	Supercritical fluid	Gas
Viscosity η, Pa · s	$< 10^{-1}$	$10^{-4} - 10^{-3}$	10^{-4}
Density ϱ, g/cm^3	0.8 – 1.0	0.2 – 0.9	10^{-3}
Diffusion coefficient D_m, cm^2/s	$< 10^{-5}$	$10^{-3} - 10^{-4}$	10^{-1}

3) The fluid should be chemically compatible with the sample and SFC equipment
4) The fluid should be miscible with many organic solvents (modifiers)

Because of the low critical temperature and pressure of carbon dioxide and its compatibility with much equipment, especially with the flame ionization detector (FID), supercritical carbon dioxide is the most frequently used mobile phase in SFC.

The elution strength of a supercritical fluid is often described by its dielectric constant DC. Figure 41 shows that the influence of pressure on the DC at different temperatures is very similar its influence on the density, which is presented in Figure 42. This diagram is one of the most important for understanding supercritical fluid chromatography and extraction. Generally, the solubility of a solute in a supercritical fluid increases with increasing density of the fluid. Simultaneously, the elution strength of the fluid increases as well, which results in shorter retention times. Figure 42 shows that a small variation of pressure slightly above the critical parameters (p_c and t_c) causes a large change in density and therefore in the elution strength of the fluid. On the other hand, a temperature increase at constant pressure causes a decrease in the density and consequently in the elution strength. In other words, in contrary to gas chromatography, an increase in temperature at constant pressure results in increased retention and therefore analysis time. This explains why negative temperature gradients are often run in SFC.

Further discussions concerning the theory and basics of SFC are given, for example, in [124], [142], [143].

12.10.3. SFC Compared to GC and LC

As mentioned above, SFC should not be regarded as competing with GC or HPLC. SFC is complementary to these two analytical separation
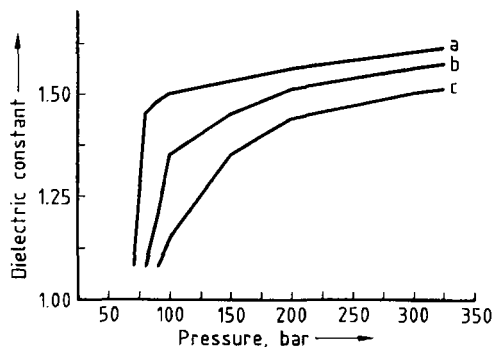
Figure 41. Influence of pressure on the dielectric constant of carbon dioxide at different temperatures
a) $t = 25\,°C$; b) $t = 40\,°C$; c) $t = 60\,°C$

Figure 42. Influence of pressure on the density of carbon dioxide at different temperatures
a) $t = 35\,°C$; b) $t = 40\,°C$; c) $t = 60\,°C$, d) $t = 70\,°C$

techniques, and one should decide from case to case which method is preferred.

Because of its intermediate position, SFC combines several advantages of both GC and HPLC. One significant advantage of SFC compared to HPLC is the fact that besides the normal LC detectors (e.g., UV detector), sensitive universal or selective GC detectors (e.g., FID, electron capture detector) are applicable as long as supercritical carbon dioxide is utilized as mobile phase. Furthermore in packed-column SFC, the same variety of stationary phases (selectivities) can be used as in LC (in contrast to GC), with the additional advantage that the analysis time is significantly shorter.

SFC is superior to GC if thermolabile solutes must be separated. However, this is true only as long as the supercritical fluid has a relatively low critical temperature like carbon dioxide (31 °C), for example.

Nevertheless, SFC also has its drawbacks; for example, only a limited number of samples are soluble in supercritical fluids. In supercritical carbon dioxide, for instance, only relatively nonpolar solutes are soluble. As a rule of thumb, solutes that are soluble in organic solvents having a polarity less than or equal to that of n-heptane are usually also soluble in supercritical carbon dioxide. Samples that are soluble in water are insoluble in CO_2.

If the elution strength of a supercritical fluid is not sufficient, it can be increased by adding an organic modifier (see Section 12.8.1). In this way, especially in combination with capillary columns, the applicability of SFC is extended to more polar analytes. Methanol is the most commonly used modifier because it shows the widest solubility

range for samples that are insoluble in CO_2. If selectivity enhancement is required, a more nonpolar solvent should be chosen. The most common modifiers for SFC are discussed in [144]. Several papers have been published on the mechanism of modifiers in SFC systems [142]. Apparently, the modifier has at least two effects: First, it dramatically alters the solvation properties of the fluid (usually increases the solubility of the sample compound); second, it blocks the most active retention sites of the column packing so that the retention strength of an analyte is decreased as is its analysis time. Even though modifiers have significant advantages, they limit the choice of suitable detectors (e.g., methanol as modifier interferes in FID). Whereas in packed-column SFC, modifier concentrations of 1 – 5 % are usually sufficient, in capillary SFC, modifier concentrations up to 20 % can be necessary. For the separation of acidic or basic samples an additive to the modifier or mobile phase can often be helpful. This improves peak shape and often shortens analysis time considerably. A typical additive for acidic samples is trifluoroacetic acid (TA) whereas for basic samples, primary amines are often used. The function of an additive is not yet clear, although it seems to homogenize the stationary phase (effectively blocking off silanol groups), to improve the solubility of the solute in both phases, and to suppress solute ionization.

A disadvantage of SFC is the fact that even though the same detectors are used as in GC, the detector sensitivity is not as good. In capillary SFC, this results from the very small sample volumes (10 – 250 nL) that must be injected into the column. In the case of packed columns, larger
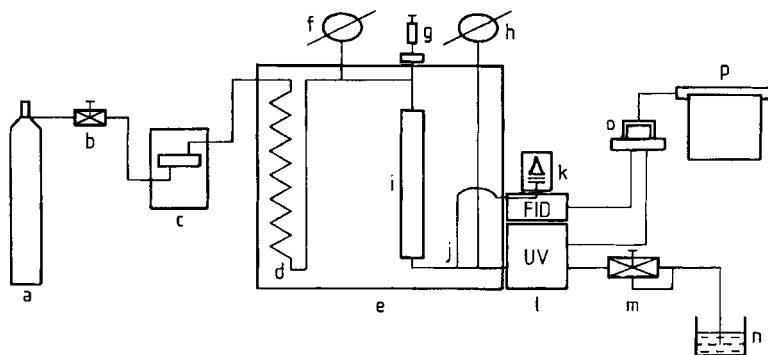
Figure 43. Schematic diagram of an SFC apparatus
a) Gas supply; b) Valve; c) Pump; d) Equilibration coil; e) Oven; f) Manometer; g) Sample introduction unit; h) Manometer; i) Separation column; j) T-unit; k) Flame ionization detector (FID); l) UV detector; m) Back-pressure valve, restrictor, or pressure and flow control unit; n) Fraction collector; o) Data control unit; p) Recorder

sample volumes can be injected but the column outlet going into the combustion detector (e.g., FID, NPD) often has to be split. If not split is applied, the extremely high flow rates, caused by the expanding supercritical fluid, tend to extinguish the flame of the combustion detector.

12.10.4. Equipment

The equipment for packed-column SFC is very similar to that for HPLC, whereas the instrumentation for capillary SFC is more reminiscent of gas chromatography. In Figure 43 a schematic diagram of packed-column SFC instrumentation is shown. It consists of a high-pressure pump, sample introduction system, oven, detector, restrictor, and recorder. In addition, a microcomputer is installed that controls temperature, flow, pressure, and density programming.

12.10.4.1. Pumps

The pump systems used most frequently in open tubular (capillary) and packed-column SFC are the reciprocating piston pumps and the syringe pumps (see Section 12.2.3). The syringe pumps have the advantage of pulseless flow and easy pressure control, while the reciprocating pumps have the advantage of continuous flow. The main difference between the LC pumps described in Section 12.2.3 and SFC pumps is that the latter need a pump head cooling system (temperature of pump head < 5 °C). Reciprocating pumps can deliver only liquids efficiently, not highly compressible supercritical fluids. A further difference is

that in SFC a very effective pulse dampener is required to ensure reproducible separations.

The maximum pressure of a modern SFC pump should be as high as possible, at least 40 – 45 MPa, with a flow range from several microliters per minute for open tubular columns (capillary columns) to ca. 5 – 10 mL/min for packed columns.

12.10.4.2. Oven

The oven of an SFC system should meet the same requirements as a normal GC oven. A constant temperature (variation ±0.1 °C) must prevail in the entire oven at any time of a positive or negative temperature gradient. This is very important for reproducible capillary column SFC analysis. These columns are very sensitive to even slight variations in temperature, which can result in peak shape deformation, peak splitting, or irreproducible retention times.

12.10.4.3. Sample Introduction

For packed columns and capillary columns, different sample introduction systems are used.

In the case of capillary columns, finding appropriate solutions for sample introduction is much more challenging than with packed columns. Since packed columns have dimensions similar to normal HPLC, optimal sample introduction can be achieved with ordinary sample loop injectors (see Section 12.2.4.2).

Onto a normal capillary column (10 m × 50 μm i.d.), < 100 mL should be injected to prevent

Table 11. Detection possibilities with SFC *

Detector	Type	Capillary SFC	Packed-column SFC	Modifier possible	Preferred mobile phase
FID	U/D	++	+**	yes	CO_2, SF_6, N_2O
NPD	S/D	++	+**	yes	CO_2, SF_6
UV	S/ND	+(−)	++	yes	CO_2
ECD	S/D	++	+**	yes	CO_2, Xe
LSD	U/(N)	+	++	yes	CO_2
FTIR	U/N	+	++	no	Xe, CO_2, SF_6
MS	U/D	++	+**	yes	CO_2

* U = universal detection; S = selective detection; D = destructive; N = nondestructive; + = detection possible; ++ = preferred column type for the particular detector; − = column type not recommended.
** Split usually necessary.

more than 1 % loss in resolution [142]. To achieve this, different techniques of sample introduction have been used. In 1988, GREIBROKK et al. [145] described a technique that involves separation of the solvent from the solutes on a precolumn. The solutes are then transferred to the analytical column where they are focused at the column head. Other solvent elimination techniques based on gas purging and solvent backflush were reported by LEE et al. [146]. The most common reproducible techniques to inject such small volumes of sample onto capillary columns are the dynamic split and the time split technique.

In the *dynamic split mode*, which is the most frequently applied technique in open tubular column SFC, the split ratio is determined by the ratio of the flow out of the column to the flow out of the split restrictor. Thus, if the flow out of the split is 500 times faster than the flow out of the column, the split ratio is 1 : 500. The disadvantage of the dynamic split technique is that obtaining reproducible results can be difficult with certain samples.

The *time split injection* technique is based on early work done by J. C. MOORE [147] for LC. In 1986, RICHTER [148] was the first to report on this injection technique in SFC. A standard injection valve, which is pneumatically driven by a low-viscosity gas such as helium, is connected directly to the analytical column (packed column or capillary column) The valve is electronically driven by high-speed pneumatics to ensure very fast switching times (from load position to inject position and back in the millisecond range). The shorter the switching time is, the smaller is the injected sample volume. Many papers have been published comparing the reproducibility of the time split and the dynamic split techniques, some of them differing significantly in their results. Whereas RICHTER et al. [149] showed that time split is superior to dynamic split, SCHOMBURG et

al. [150] saw no significant difference between the precision of the two techniques. Because of problems in reproducibility with these two injection techniques, the use of an internal standard is strongly recommended for quantitative analysis.

12.10.4.4. Detectors

As mentioned before, a unique advantage of SFC is the fact that a wide variety of detection methods can be applied (see Table 11). Besides the traditional LC detection (e.g., UV) the use of GC detectors, especially, enables relatively sensitive, universal (FID), as well as selective [e.g., electron capture detectors (ECD), thermoionic detectors (TID)] methods of detection.

As in LC, both destructive and nondestructive detectors are used in SFC. The detector most commonly used for this technique is the flame ionization detector. It is a destructive, sensitive, and universal detector, which is well known from gas chromatography. Even though RICHTER et al. [151] described, in 1989, a new FID detector with a detection limit (5 pg of carbon per second) and a linear dynamic range (10^6) very similar to that achieved in FID–GC systems, the ordinary FID–SFC system is still a factor of 10–1000 poorer in sensitivity, which consequently calls for relatively concentrated sample solutions. If very sensitive detection is essential, the FID–SFC detection limit can be increased by a factor of ca. 1000 by replacing carbon dioxide with nitrous oxide, which has similar physicochemical characteristics (see Table 9). However, one should be aware of the fact that N_2O can explode under higher pressure and temperature.

A further drawback of the SFC–FID system is the fact that the amount of organic modifier that can be added to the fluid is limited (methanol ca. 3–5 vol %) because the FID responds to organic
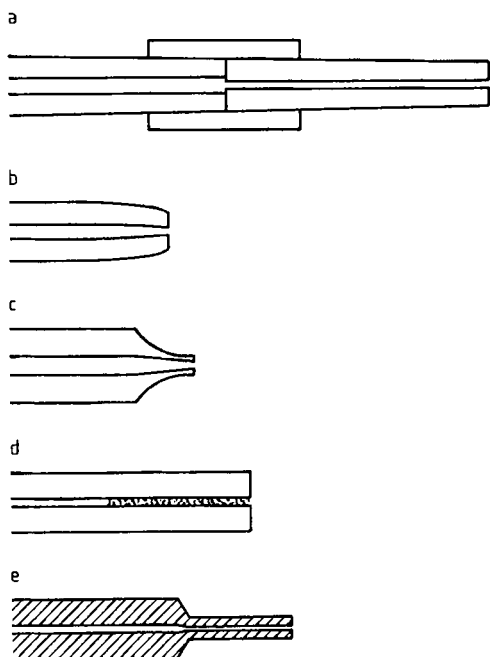
Figure 44. Fixed restrictors
a) Linear capillary restrictor; b) Polished integral restrictor;
c) Fast-drawn capillary restrictor; d) Porous frit restrictor;
e) Pinched restrictor

molecules. Water can be an alternative, but its solubility in supercritical carbon dioxide is low (< 1 %), and the elution strength does not increase too much.

The UV detector is also very frequently used in SFC, mainly in packed column SFC, because of problems in sensitivity with capillary columns. Since pure carbon dioxide shows little absorbance between 200 and 800 nm, UV detection in packed-column SFC is very sensitive. In general, the characteristics of UV detectors discussed in Section 12.2.6.3 are also valid for UV – SFC detectors, except that the latter must be equipped with a pressure-resistant flow cell. A normal UV detector flow cell in LC can resist a maximum pressure of ca. 7 – 10 MPa. In SFC, the UV detector cell must be under the same separation conditions (temperature and pressure) as the column (if the pressure in the detector cell were lower, the fluid would loose its solvation power), so it should resist pressure ≤ 40 MPa.

A nondestructive and universal detector in SFC is the FTIR detector (see Section 12.2.6.7). Although its sensitivity is not too high, this technique is highly informative and often helps to identify analytes by their characteristic functional groups or to solve molecular structure problems.

Compared to LC, SFC – FTIR systems have the significant advantage that carbon dioxide is transparent in a large part of the IR region, allowing FTIR flow cell detection [152]. In LC, the mobile phase (especially organic solvents) must usually be removed prior to FTIR detection, because it absorbs in the required IR region. Consequently, organic modifiers used in SFC – FTIR cause the same problem [153], [154], making flow cell detection impossible. The "ideal" fluid for SFC – FTIR flow cell detection is supercritical xenon. It is transparent over the entire IR range and shows sufficient solvation power [155]. Unfortunately, the high price of xenon limits its application in routine analysis. Many other detection techniques are described for SFC. An excellent discussion of the various detection possibilities in SFC is given in [142].

12.10.4.5. Restrictors

The restrictor is an important part of all SFC equipment. It maintains constant pressure in the system and ensures, or at least should ensure, uniform, pulse-free flow of the supercritical fluid; thus, it is also responsible for reproducible analysis. Two types of restrictors exist: Fixed restrictors (see Fig. 44) and variable restrictors. *Fixed restrictors* have the disadvantage that pressure programming automatically also changes the flow. Therefore, independent pressure and flow programming can be achieved only with computer-driven variable restrictors, which often function in principle like a needle valve. Usually the restrictor must be heated since, due to the expansion of the supercritical fluid to atmospheric pressure, strong cooling and formation of solid carbon dioxide result (Joule – Thomson effect), which causes restrictor plugging and therefore irregularities in the flow and reproducibility of the analysis.

12.10.5. Packed-Column SFC (pc-SFC) and Capillary SFC (c-SFC)

The question of whether packed-column or capillary SFC is preferred, is very dificult to answer. Both techniques have their advantages and disadvantages, depending on the particular separation problem. Lately the trend is more toward the packed-column technique because reproducibility and higher sensitivity are ensured. In addi-

Table 12. Column dimensions and characteristics in SFC

Column type	Internal diameter,* mm	Length,* m	Preferred particle size, μm	Supercritical fluid flow (0.4 g/mL), mL/min	Efficiency per meter,* N/m	Efficiency per time,* N/s
Open tubular capillary	0.025 – 0.1	1 – 35	50 **	0.0005 – 0.002	2000 – 22 000	13 – 33
Packed capillary	0.1 – 0.5	0.05 – 0.5	3, 5, 10			
Packed columns	2.0 – 4.6	0.03 – 0.25	3, 5, 10	0.008 – 2.7	33 000 – 51 000	31 – 83
Microbore packed	0.5 – 2.0	0.03 – 0.25	3, 5, 10			

* Data from [142].
** 50-μm-i.d. open tubular (capillary) column.

tion, independent pressure and flow control is possible only in packed-column SFC.

12.10.5.1. Capillary SFC (c-SFC)

Capillary SFC columns are usually manufactured from fused silica, with an internal diameter of 25 – 100 μm and a total column length between 1 and 35 m. The most commonly used are 3 – 10-m, 50-μm-i.d., fused-silica columns. A column like this has a very low supercritical fluid flow rate, from several microliters per minute to ca. 0.01 mL/min which corresponds to a gas flow rate of about 1 mL/min (see Table 12).

c-SFC has a wider application range than the packed-column technique because the phase ratio is much more advantageous. In other words, in c-SFC the absolute amount of stationary phase interacting with the solute (and therefore also the retention strength) is much lower than in pc-SFC. Thus, the separation in c-SFC (as in GC) is dominated by the volatility of the solutes and their solubility in the mobile phase and not so much by the selectivity gained from the interaction between the analyte and the stationary phase. Nevertheless, because of the very low linear velocity (millimeters per second) and therefore low flow rates (millimeter per minute) in c-SFC, the analysis time is usually much longer than in pc-SFC. Even though the separation efficiency is theoretically higher in capillary SFC than in packed-column SFC, in practice, this difference is often not so pronounced. The coupling of c-SFC with mass spectrometry (c-SFC – MS) is much easier than for LC, and many articles on this hyphenated technique have been published [156], [157].

The disadvantages of capillary SFC are that the number of commercially available, stable stationary phases showing different selectivities is limited. Furthermore, problems with reproducibility can occur because of technical difficulties with quantitative sample introduction. Finally, because of the very low flow rates in c-SFC, analysis time is relatively long and independent pressure programming is impossible; the latter can be done only in combination with a flow program.

12.10.5.2. Packed-Column SFC (pc-SFC)

Typical dimensions of packed columns and flow rates in pc-SFC are given in Table 12. This technique is becoming more and more popular compared to capillary SFC. A wide spectrum of stable stationary phases, showing different selectivities, is available commercially. Reproducible, quantitative sample introduction by means of the HPLC sample loop technique is no problem, nor is independent pressure or flow control.

Very high separation efficiencies (more than 100 000 plates) can be achieved by coupling several packed columns in series (e.g., Fig. 45 shows a separation of triazine pesticides with ten 200×4.6 mm i.d. hypersil silica (5 μm) columns in series—note that the pressure drop over all these columns is only 12.4 MPa. This would not be possible in HPLC since, because of the higher viscosity of the solvents, the back-pressure over the column would become too high (high-pressure shutdown).

Several columns with different lengths and selectivities can also be connected in series. This enables individual adjustment of the total selectivity of the system depending on the separation problem (selectivity tuning). Figure 46 shows the separation of fatty acid ethyl esters from human blood serum. The phytanic acid ester (ethyl 3,7,11,15-tetramethylhexadecanate), an indicator for a rarely occurring inherited disease (Refsums syndrome) can be separated and quantified from
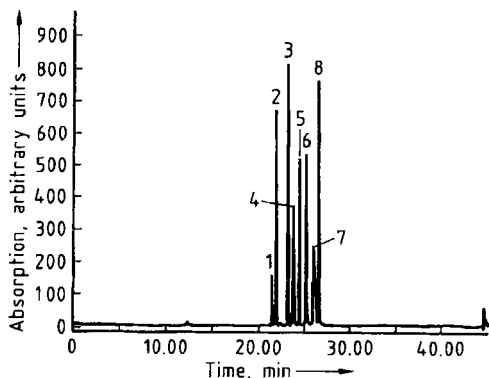
Figure 45. pc-SFC separation of triazine pesticides
Ten 200×4.6 mm i.d. columns hypersil silica columns (particle size 5 µm) in series: program: 2 % methanol, 8 MPa for 5 min, 500 kPa/min + 1 % MeOH/min to 13 MPa and 12 MeOH %; 60 °C, 2.0 mL/min; $\delta p = 12.4$ MPa detector: FID
1) Propazine; 2) Atrazine; 3) Anilazine; 4) Simazine (4-doublet); 5) Prometryne; 6) Ametryne; 7) Bladex; 8) Prometon

Figure 46. pc-SFC separation of lipids in human blood serum
Columns: 250×4.6 mm ICN silica (3–6 µm); 200× 4.6 mm Nucleosil NH_2 (5 µm); 125×4.6 mm Select B (RP-8) column (5 µm) in series; pure CO_2, 40 °C, 14.5 MPa; detector: FID
1) Phytanic acid; 2) 14 : 0; 3) 16 : 3; 4) 18 : 0; 5) 16 : 1; 6) 18 : 1; 7) 18 : 2; 8) 20 : 4 (the figure before the colon denotes the number of carbon atoms in the chain, that after the colon, the number of double bonds in the molecule)

other coeluting fatty acid esters only with the column sequence shown.

The pc-SFC separation technique is rapid (ca. five times faster than HPLC with the same column dimensions). Preparative application and coupling with a mass spectrometer (pc-SFC – MS) are possible. However, capillary SFC, or at least narrowbore columns are preferred for MS coupling because pc-SFC using conventional 100 – 250 mm × 4 – 4.6 mm i.d. columns usually has flow rates that are too high.

The retention mechanisms of capillary and packed-column SFC differ from one another. Whereas packed column SFC with supercritical carbon dioxide shows a similar retention mechanism to HPLC (separation due mainly to specific interactions between the analyte and the stationary phase), capillary SFC behaves more like capillary GC (separation in pc-SFC according to "solubility" and "volatility" of solute; separation in c-SFC according to "volatility" of solute). Investigations concerning the retention mechanism in reversedphase packed-column SFC showed [158] that both a hydrophobic and a polar retention mechanism apply. This can be of significant advantage, as shown in Figure 47. Here, a mixture of salmon oil fatty acid ethyl esters is separated by pc-SFC on an aminopropyl column [159]. A combination of an RP mechanism (retention increases with increasing number of carbon atoms of the esters) and an NP mechanism (retention increases with the growing number of double bonds of the fatty acid

esters) can be observed. This separation is an example of the potential of pc-SFC to obtain very fast separation of complex sample mixtures. If plain silica columns are used, the retention mechanism corresponds to the normal phase mechanism in LC, with the advantage that pc-SFC is significantly faster than the NP – HPLC technique.

12.10.6. Applications

Although the application of SFC is limited to those solutes that are soluble in the particular supercritical fluid (generally carbon dioxide), many interesting separations have been reported in various fields. Polymer analysis (M_r up to ca. 5000 – 8000; even higher with modifier) is certainly one of the most important [160], [161]. Figure 48 shows a typical application in polymer analysis. In the petroleum industry, SFC is becoming more and more important in simulated distillation under mild, nondegradative conditions [142], [162] for analysis of hydrocarbons and of the carcinogenic polycyclic aromatic compounds found in fossil fuels [162] (see Fig. 49). Other
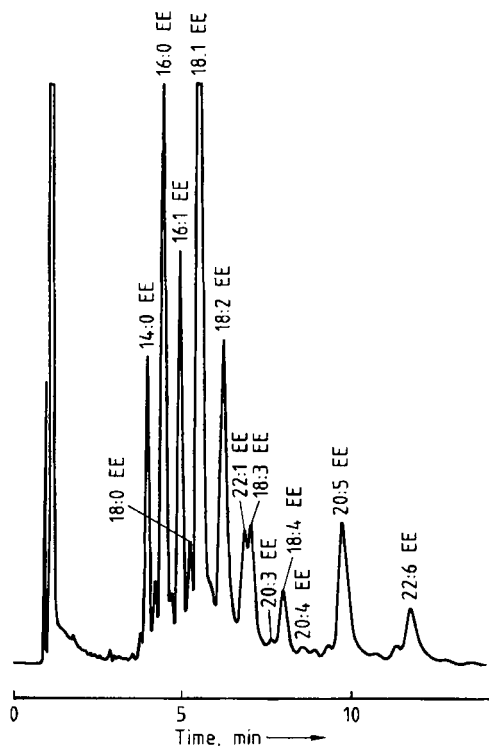
Figure 47. Separation of salmon oil fatty acid ethyl esters with pc-SFC
Column: aminopropyl (particle size 5 μm), 200×4.6 mm i.d.; 34 °C, 1.05 GPa; detector: FID [159]

Figure 48. Capillary column SFC separation of Triton X-100 (surfactant)
Column: 50 μm×10 m SB-methyl-1000 (0.35 μm); mobile phase: pure CO_2; temperature: 120 °C; inlet pressure 10 MPa for 10 min, 1 MPa/min to 35 MPa
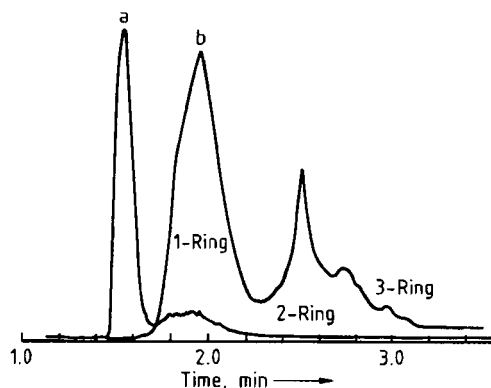
Figure 49. Determination of aromatics in diesel fuel
a) FID detection for saturated compounds and olefins; b) UV detection (at 254 nm) for aromatics
Column: 250×4.6 mm Lichrosorb Si-60; CO_2 isobaric, 15 MPa, 28 °C

typical applications for pc- and c-SFC are the separation of explosives [163], pesticides [164], drugs [165], and pharmaceuticals [166].

Figure 50 shows the potential fo packed-column SFC for an efficient and fast control of the chiral purity of pharmaceuticals. Because of the excellent solubility of lipids in supercritical carbon dioxide, SFC in combination with the universal FID detector is a very helpful tool in lipid analysis.

Although all the applications of pc- and c-SFC would be impossible to mention, their potential in preparative applications should be noted. The latter will certainly become more and more interesting in the future, especially for the food and pharmaceutical industries because very pure extracts are obtainable that are not contaminated with any toxic organic solvents.

In 1997, the worlds first industrial-scale SFC production plant, equipped with separation columns with an internal diameter of over 50 cm and a length of 2 m was designed and started up

[167] for the production of high-purity (> 95 %) polyunsaturated fatty acids from fish oil [167].

12.11. Affinity Chromatography

Affinity chromatography is a special type of adsorption chromatography for the isolation and purification of biologically active macromolecules. It was used successfully for the first time in 1968 for the purification of enzymes [168]. Since then, innumerable proteins (e.g., enzymes,
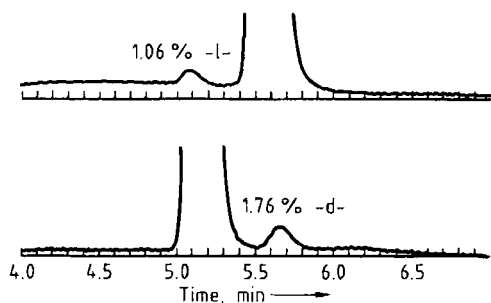
Figure 50. Chiral purity analysis of carbobenzyloxyd,l-alanine at 220 nm with packed-column SFC
Column: 250×4.6 mm Ciralcel-OD; 20 % ethanol (+0.2 % TPA/CO₂) 1 mL/min

Figure 51. Schematic representation of affinity chromatography

receptors, or antibodies) have been isolated with high enrichment (100–10 000 fold) from complex raw extracts in a single chromatographic run. The proteins still possess their full biological activity after separation. The routine application of affinity chromatography has been described in many review articles [169]–[176], [177].

The high selectivity of affinity chromatography is based on biospecific interactions, such as those occurring between two molecules in natural biological processes. One interaction partner (ligand) is covalently attached (immobilization) to an insoluble carrier, while the corresponding partner (frequently a protein) is reversibly adsorbed by the ligand because of its complementary biospecific properties.

In laboratory practice, an affinity matrix (stationary phase) is tailor-made for the protein to be purified and filled into a chromatography column. For process applications many different affinity gels are commercially available. The raw extract is then passed through the column by using a physiological buffer as the mobile phase. In this process, the desired product is adsorbed selectively by the ligand and all unwanted components are washed away.

The chemical composition of the mobile phase is then changed to permit desorption and isolation (elution) of the protein (Fig. 51).

12.11.1. Preparation of an Affinity Matrix

12.11.1.1. Selection of a Carrier [178], [179]

The preparation of an affinity matrix requires a suitable carrier that covalently attaches the se-

lected ligand (immobilization). An ideal carrier should have the following properties:

1) A hydrophilic neutral surface — avoidance of nonspecific adsorption
2) Macroporosity — good permeability
3) Mechanical stability — high flow rates, stability to pressure
4) Chemical and biological stability — insolubility in various solvents, resistance to microbial attack
5) Functional groups on the surface — covalent binding of the ligand

The carriers used can be divided into three groups:

Polysaccharides	agarose
	dextran
	cellulose
Synthetic polymers	polyacrylamides
	polyacrylates
	poly(vinyl alcohols)
Inorganic materials	SiO_2, ZrO_2, TiO_2
	magnetic particles

Agarose, a polysaccharide consisting of galactose units, is so far the carrier of choice because of its favorable chemical and physical properties. However, mechanical instability sometimes results in short column lifetime. In comparison, synthetic polymers and inorganic materials are mechanically more stable.
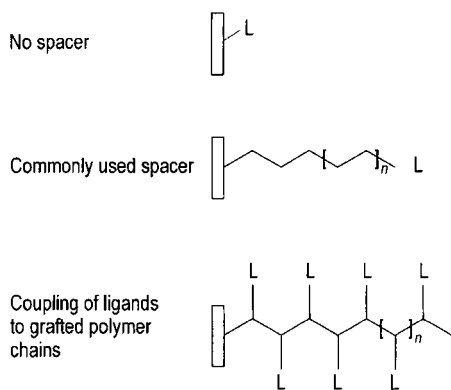
Figure 52. Three general possibilities for the covalent fixation of ligands to a matrix surface: direct coupling, coupling to spacers, and coupling to grafted polymer chains. L = ligand.

The spatial availability of immobilized ligands is often substantially improved by coupling ligands to spacers that keep them at a distinct distance from the matrix [180]. Basically, there are three general possibilities for the covalent fixation of ligands to the surface of appropriate support materials (Fig. 52):

1) Direct coupling to the surface (i.e., with no spacers)
2) Commonly used spacer arms : linear aliphatic hydrocarbons, 6-aminohexanoic acid, hexamethylenediamine, poly(ethylene glycol) [181]. One of the groups (often a primary amine, $-NH_2$) is attached to the matrix, while the group at the other end is selected on the basis of the ligand to be bound. The latter group which also is called the terminal group is usually a carboxyl ($-COOH$) or amino group.
3) Coupling of ligands to grafted polymer chains which freely rotate (i.e., in a "tentacle" arrangement) [182].

12.11.1.2. Immobilization of the Ligand

Covalent binding of a ligand to a carrier requires the presence of functional groups on both binding partners.

Generally, the following groups are available for immobilization of the ligand: $-NH_2$ (amino), $-COOH$ (carboxyl), $-CHO$ (aldehyde), $-SH$ (thiol), or $-OH$ (hydroxyl). High-molecular ligands (e.g., proteins) are usually bound through ε-NH_2 groups of lysine residues or occasionally through

carboxyl groups. The functional group of the ligand must be mainly located in the nonspecific region of the molecule that is of no significance for reversible complex formation with the complementary molecule to be purified.

The functional groups of the carrier are either already present in the natural structure (e.g., OH groups of polysaccharides) or introduced later by the chemical functionalization of the surface:

$$
\begin{array}{ll}
\overset{\displaystyle O}{\overset{\displaystyle /\backslash}{-CH-CH_2}} & \text{epoxide (a)} \\
-CHO & \text{aldehyde (b)} \\
-NCS & \text{isothiocyanate (c)} \\
-SO_2-CH=CH_2 & \text{vinyl sulfone (d)} \\
-OH & \text{hydroxyl (e)} \\
-COOH & \text{carboxyl (f)} \\
-NH_2 & \text{amine (g)}
\end{array}
$$
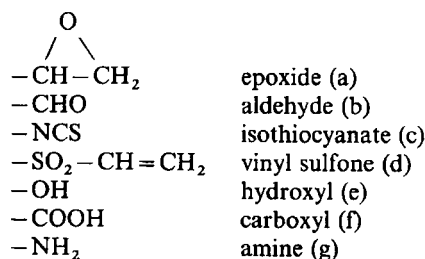
A carrier with reactive functional groups (a)–(d) can be used directly for the immobilization of ligands (e.g., via their NH_2 groups). In the case of epoxides (a), however, the coupling yields obtained are low. In the case of aldehydes (b), chemically instable ligand bonds (Schiff base) are formed that require further reduction (e.g., with boron hydride). Isothiocyanates (c) and vinyl sulfones (d) have the disadvantage that after immobilization, chargeable ligand bonds (thiourea and amine derivatives) are formed. These groups can act as weak ion exchangers and cause additional nonspecific adsorptions.

Functional groups (e)–(g) must be converted to a chemically reactive form (activation) before coupling to the ligand can occur. After immobilization has been carried out, any unreacted reactive groups on the carrier must be deactivated to prevent chemical reactions with the raw extract. This is frequently effected by additional coupling to a low-molecular ligand (e.g., ethanolamine) or by hydrolysis of the remaining activated groups.

Methods of Activation. Functional groups such as hydroxyl, carboxyl, or amino groups (e)–(g) must be converted to an activated form before they can be attached covalently to the ligand.

The classical method for activating *hydroxyl groups* involves reaction with cyanogen bromide [171], [183] to give a reactive cyanate ester (Fig. 53 A). However, the coupling of ligands (e.g., via ε-NH_2 groups of lysine residues) results in chemically unstable isourea linkages that can cause bleeding of the ligand during affinity chro-
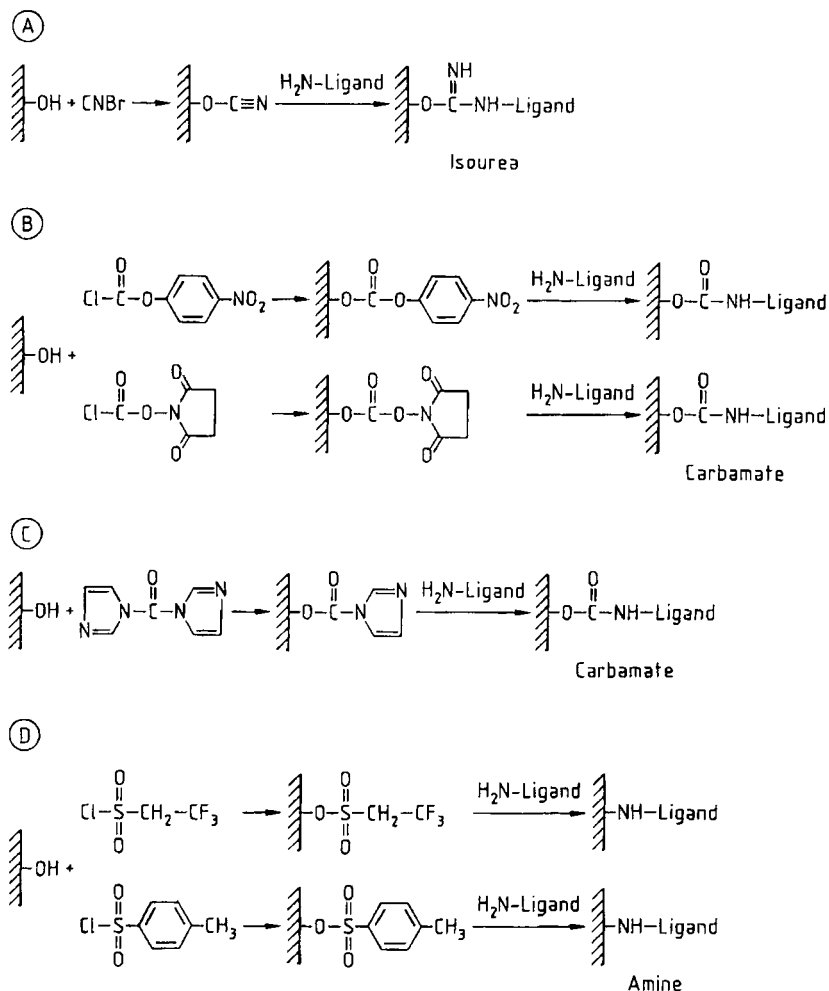
Figure 53. Activation and coupling of ligands to base supports with hydroxyl functions
A) Cyanogen bromide; B) Chloroformates; C) Carbonyldiimidazole; D) Sulfonyl chlorides

matography. This disadvantage can be avoided by activation with other reagents such as 1,1'-carbonyldiimidazole (Fig. 53 C) [184], p-nitrophenyl and N-hydroxysuccinimide chloroformates (Fig. 53 B) [185], [186], or sulfonyl chlorides [187], such as 4-toluenesulfonyl chloride and 2,2,2-trifluoroethanesulfonyl chloride (Fig. 53 D). The resulting imidazoyl, carbonate, and sulfonyl derivatives react, for example, with the amino groups of the ligand, with high coupling yields, and the ligand is attached to the carrier through chemically stable carbamate or amine bonds. The carbamate bond also has the advantage of being

neutral and, consequently, causing no undesirable, nonspecific interactions. Unreacted activated groups can be hydrolyzed easily to reform the original neutral hydroxyl groups.

Carboxyl groups can be converted to active esters (e.g., with p-nitrophenol or N-hydroxysuccinimide), which react with the ligand with high coupling yields. Furthermore, carboxyl groups can be activated with water-soluble carbodiimide.

However, unreacted carboxyl groups act as weak ion exchangers, leading to unwanted nonspecific adsorptions. This disadvantage is also en-

countered with base matrices that are functional-ized with amino groups.

12.11.2. Ligands

A ligand suited to the production of an affinity matrix should have the following properties:

1) Ability to form reversible complexes with the protein to be isolated and purified
2) Suitable specificity for the planned application
3) Large enough complexation constant to obtain stable complexes or sufficient delays during the chromatographic process
4) Complex that can be dissociated easily by sim-ple changes in the mobile phase without irre-versible loss of the biological activity of the protein

A distinction is made between low-molecular and high-molecular ligands as well as between group-specific and monospecific ligands. Group-specific ligands exhibit binding affinity for a range of substances that have similar properties with regard to function or structure. Typical examples are the triazine dyes Cibacron blue F 3 G-A and Procion red HE-3 B, which are used as ligands for many dehydrogenases, kinases, phosphatases, and other proteins. Monospecific ligands (e.g., anti-bodies) are "tailor-made" and can be used for only one application. In general, the binding of mono-specific ligands to their complementary substance is substantially stronger (i.e., unfavorable elution conditions are occasionally required).

12.11.2.1. Group-Specific Ligands

12.11.2.1.1. Low-Molecular Ligands

The low-molecular, group-specific ligands be-long to the largest class of ligands used in affinity chromatography:

Ligand	Target molecule
Cibacron blue F 3 G-A	kinases, phosphatases, dehy-drogenases albumin, clotting factors II and IX, interferon, lipoproteins
Procion red HE-3 B	dehydrogenases, carboxypep-tidase G, interferon, plasmi-nogen, dopamine β-monooxy-genase, inhibin
5'-AMP	NAD$^+$-dependent dehydrogen-ases, ATP (adenosine 5'-tri-phosphate)-dependent kinases
2',5'-ADP	NADP$^+$-dependent dehydro-genases
NAD$^+$ and NADP$^+$	dehydrogenases
Phenylboric acid	nucleosides, nucleotides, car-bohydrates, glycoproteins
Benzamidine	trypsin, urokinase, pronase E, glu- and lys-plasminogen

The coenzymes nicotinamide – adenine dinu-cleotide (NAD) and nicotinamide – adenine dinu-cleotide phosphate (NADP), their reduced forms, and adenosine 5'-monophosphate (5'-AMP) and adenosine 2',5'-diphosphate (2',5'-ADP) represent well-known examples of this class. They can be considered for the purification of ca. 25 % of the known enzymes (especially dehydrogenases).

Moreover, the triazine dyes (e.g., Cibacron blue F 3 G-A and Procion red HE-3 B) [188], [189] mentioned above are also used as ligands for the purification of these proteins. Triazine dyes can replace ligands of the NAD group because they have similar spatial structures. In addition, they bind a series of blood proteins (e.g., albumins, lipoproteins, clotting factors, and interferon).

Thiophilic ligands [190], [191] form a separate class of group-specific ligands that possess high selectivity with respect to immunoglobulins. They are low-molecular compounds, and contain sul-fone and thioether groupings (see below), which show a strong affinity for mono- and polyclonal antibodies in the presence of ammonium sulfate (ca. 1 mol/L).

$$- -CH_2-S-CH_2-CH_2-\overset{O}{\underset{O}{\overset{\uparrow}{\underset{\downarrow}{S}}}}-CH_2-CH_2-S-CH_2-CH_2-OH$$
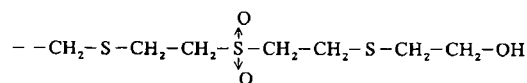
Immobilized metal ion affinity chromatography (IMAC) uses gels with group-specific ligands which display selectivity for certain amino resi-dues (histidine, cysteine, tryptophan) exposed at the outer surface of proteins [192]. The ligands are immobilized chelating agents [e.g., iminodiacetic acid (IDA) or tris (carboxymethyl)ethylenediam-ine (TED)] with bound transition metal ions. The metal ions are mostly by Cu^{2+}, Ni^{2+}, Zn^{2+}, and Co^{2+}, with decreasing complex stability and affinity [193]. The amino acids at the protein surface are bound to a free coordination site of the immobilized complex (Fig. 54) [194].

Adsorbed proteins can be recovered by lower-ing the pH to 3 or 4 or by competitive elution with an increasing concentration of glycine, imidazole, or histidine [195]. Through advances in molecular biology, histidine residues can be inserted into the
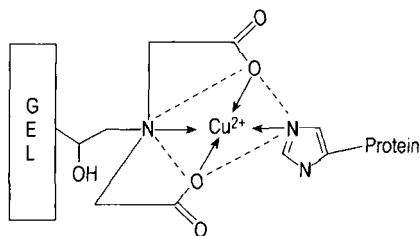
Figure 54. Schematic of binding of a histidine protein residue to an immobilized IDA copper(II) complex

protein sequence (fusion protein), resluting in improved separation in IMAC. The modified protein has a high affinity to the immobilized metal ions due to the addition of the histidine residues in the primary sequence. The fusion part can cleaved chemically or with a specific proteinase after purification.

12.11.2.1.2. Macromolecular Ligands

Macromolecular group-specific ligands include proteins that are used widely in the purification of biologically relevant macromolecules. Important macromolecular ligands are lectins (for the isolation of glycoproteins), protein A and protein G (for the purification of immunoglobulins), calmodulin (for Ca-dependent enzymes), and the polysaccharide heparin (for the purification of clotting factors and other plasma proteins).

Lectins [196]. Lectins exhibit a marked affinity for sugar molecules. They are perfectly suited to the isolation and purification of soluble and membrane-bound glycoproteins and polysaccharides (e.g., enzymes, hormones, plasma proteins, antigens, antibodies, and blood group substances). Among the available lectins, Con A (from *Canavalia ensiformis*) is the most widely used because of its specificity with respect to the frequently occurring α-D-glucose and α-D-mannose structural units. The desorption of the component to be purified from the ligand elution is often achieved by the addition of these low-molecular sugars to the mobile phase (see Section 12.11.3).

Lentil lectin (from *Lens culinaris*) has the same specificity as Con A, but a lower binding constant. For this reason, lentil lectin is suited to the purification of membrane-bound glycoproteins, which often have an exceedingly strong binding affinity for Con A.

Proteins A and G [197], [198]. Apart from the thiophilic ligands mentioned above, proteins A and G from bacterial cell walls are also suited to the purification of immunoglobulins. Protein A binds different subclasses of immunoglobulins to different extents. Accordingly, elution is often achieved with a decreasing stepwise, pH gradient. Mouse immunoglobulins (Igs) are eluted, for example, with 0.1 mol/L citrate buffer at the following pH values:

	IgG 1	IgG 2a	IgG 2b	IgG 3
pH	6.0	5.0	4.0	4.5

Protein G in its natural form can bind not only immunoglobulins, but also albumin. For this reason, it is applied in a genetically engineered form in which the albumin binding domain is eliminated (Fig. 55). Immunoglobulins generally display a higher binding affinity for protein G than for protein A.

12.11.2.2. Monospecific Ligands

Highly specific protein–protein interactions play a large role in many natural biological processes. For instance, the transport and transmission of information proceed via highly specific receptors that can frequently interact with only one molecular species (e.g., hormone or transmitter). In spite of this, the affinity chromatographic isolation of receptors and other highly specific proteins is rather an exception. However, antibodies occupy a special position as monospecific macromolecular ligands.

Antibodies and Antigens (Immunoaffinity Chromatography) [199], [200]. Immunoaffinity chromatography refers to the immobilization of antibodies (monoclonal or polyclonal) or antigens on an insoluble carrier (immunosorbent) for isolation and purification of the corresponding partner.

Antibodies are ligands of special interest because they show high biospecificity and strong binding to the molecules that elicited their synthesis, the antigens. The antibody–antigen interaction takes place between a particular part of the antibody (the antigen-binding site) and its complementary part on the antigen (the antigenic determinant or epitope).

Polyclonal antibodies represent a wide spectrum of antibody molecules that all have specificities with respect to different epitopes and, therefore, different avidities. The association constants

IgG1

pH 2.7

Optical density 280 nm (OD₂₈₀)

95 85 75 65 55 45 35 25 15 0

Elution volume, mL
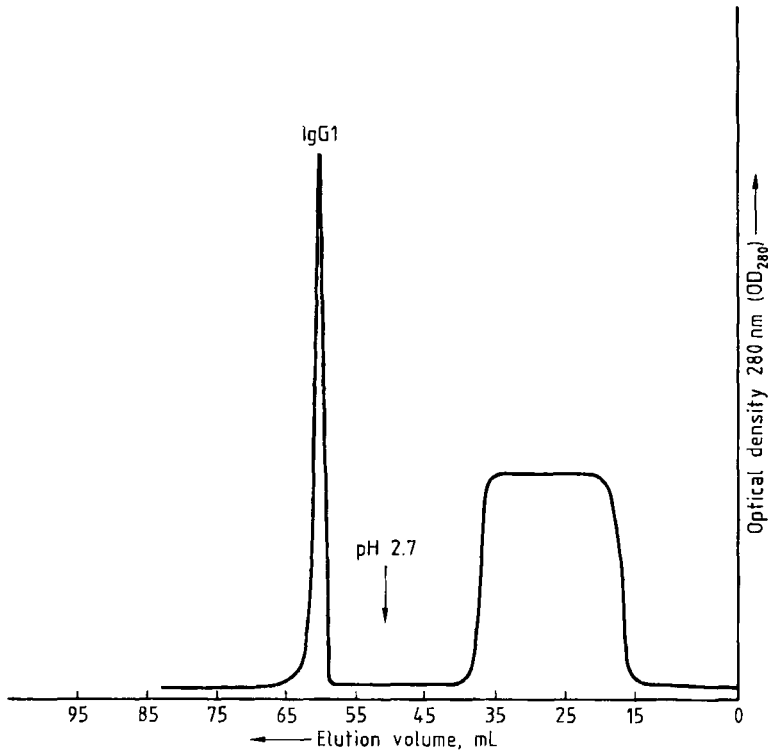
Figure 55. Isolation and purification of IgGl with genetically engineered protein G immobilized onto Sepharose

of an antigen and its corresponding antibody may range from 10^4 to 10^{11} L/mol. As a result, multipoint binding of the antigen to two or more different antibodies can occur on an immunosorbent prepared from polyclonal antibodies, resulting in such tight binding that dissociation and recovery of the antigen may be difficult.

In contrast to polyclonal antibodies, *monoclonal antibodies* show absolute specificity for a single epitope, the smallest immunologically recognized submolecular group of an antigen. Therefore, from a group of monoclonal antibodies the one most suited to antigen purification can be selected. Consequently, immunoaffinity chromatography has gained importance because monoclonal antibodies can be produced in large quantities by a hybridoma technology developed by KÖHLER and MILSTEIN [201]. This allows the use of immunoaffinity chromatography even for industrial-scale applications. Several proteins of potential therapeutic value are now produced by bioengineering techniques and purified with immunoaffinity chromatography (Table 13).

12.11.3. Elution [176]

Above all, successful affinity chromatography requires suitable desorption conditions (i.e., conditions that permit elution of the adsorbed substance with full biological activity). In principle, desorption is achieved by cleaving the binding sites responsible for ligand – protein interaction. These are mainly electrostatic, hydrophobic, and hydrogen bonds that can be specifically or nonspecifically disrupted. A *nonspecific method* of eluting an adsorbed substance is by altering the pH or ionic strength of the mobile phase. A pH decrease to pH 2 – 4 or an increase in ionic strength through a NaCl gradient, for instance, typically causes the dissociation of mainly electrostatic complex binding site. Ligand – protein complexes with high binding constants, such as those encountered in immunoaffinity chromatography, can often be eluted only under drastic conditions that can lead to denaturation. For this purpose, chaotropic salts (e.g., KSCN, KI) are used in concentrations of 1 – 3 mol/L. Urea or guanidin-

Table 13. Proteins of potential therapeutic value purified by immunoaffinity chromatography

Protein	Potential use
Interferons	
α-Interferon	antiviral, antitumor drug
β-Interferon	
γ-Interferon	
Interleukin 2	antitumor drug; also used against immuno-deficiency syndrome
Anticlotting factor	
Urokinase	dissolving blood clots
Urokinase-type plasmino- gen activator	
Zymogen of plasminogen activator	
Plasminogen activator	
Tissue plasminogen activator	
Human protein C	
Clotting factor	
Factor VIII	blood coagulant
Factor IX	
α- and β-Kallikrein	
Insulin-like growth hormone	growth hormone
Vaccines	
Human hepatitis B surface antigen	vaccine
Rubella virus antigen	
Tetanus toxin	
Polio virus	

ium chloride (4 – 6 mol/L) is also employed for the elution of strongly adsorbed substances.

In the case of *specific elution*, low-molecular substances with a certain affinity for the ligand or the adsorbed protein are added to the mobile phase. They displace the adsorbed protein from the ligand by competitive binding. This method of elution is frequently used with group-specific ligands (e.g., lectins). Glycoproteins can be eluted gently from immobilized lectins by the addition of low concentrations (10 – 100 mmol/L) of a monosaccharide to the mobile phase. Another example is the elution of dehydrogenases from coenzyme affinity supports by using NAD and its derivatives. In general, specific elution is a gentle method that usually ensures retention of biological activity.

12.12. References

[1] J. J. Kirkland, *Anal. Chem.* **43** (1971) no. 12, 36 A.
[2] M. Tswett, *Tr. Protok. Varshav. Obshch. Estestvoist- pyt. Otd. Biol.* **14** (1903, Publ. 1905) 20 (Reprinted and translated in G. Hesse, H. Weil: *Michael Tswett's erste chromatographische Schrift*, Woelm, Eschwege 1954, pp. 37.
[3] R. Kuhn, E. Lederer, *Naturwissenschaften* **19** (1931) 306.
[4] R. Kuhn, E. Lederer, *Ber. Dtsch. Chem. Ges.* **64** (1931) 1349 – 1357.
[5] A. Martin, R. Synge, *Biochem. J.* **35** (1941) 1358 – 1368.
[6] F. Spedding, *Discuss. Faraday Soc.* **7** (1949) 214 – 231.
[7] E. Tompkins, *Discuss. Faraday Soc.* **7** (1949) 232 – 237.
[8] J. Porath, P. Flodin, *Nature (London)* **183** (1959) 1657 – 1659.
[9] E. Piel, *Anal. Chem.* **38** (1966) 370 – 372.
[10] J. J. Van Deemter, F. J. Zuiderweg, A. Klinkenberg, *Chem. Eng. Sci.* **5** (1956) 271.
[11] H. Engelhardt (ed.): *Practice of High Performance Analysis*, Springer Verlag, Berlin 1986.
[12] F. C. Smith, R. C. Chang, *CRC Crit. Rev. Anal. Chem.* **9** (1980) 197.
[13] P. L. Bailey, *Anal. Chem.* **50** (1978) 698 A.
[14] K. Stulik, V. Pacakova, *CRC Crit. Rev. Anal. Chem.* **14** (1982) 297.
[15] C. J. Blank, *J. Chromatogr.* **117** (1976) 35.
[16] L. Roth: *Sicherheitsdaten/MAK-Werte*, Verlag moderne Industrie, W. Dummer & Co., München 1978.
[17] H. Schrenker, *Int. Lab.* (1978) 67.
[18] I. Halasz, P. Vogtel, *J. Chromatogr.* **142** (1977) 241.
[19] L. R. Snyder, J. K. Kirkland: *Introduction to Modern Liquid Chromatography*, 2nd ed., John Wiley and Sons, New York 1979.
[20] *Rheodyne. Products for Liquid Chromatography*, Catalog no. 3, Cotati, Calif., USA, 1986.
[21] J. Meister, H. Engelhardt, *Nachr. Chem. Tech. Lab.* **40** (1992) no. 10.
[22] A. Tiselius, D. Claesson, *Ark. Kemi Mineral. Geol.* **15** (1942) no. 22.
[23] Refractive gradient detector
[24] H. Engelhardt: *Hochdruck-Flüssigkeits-Chromatographie*, Springer Verlag, Berlin 1975.
[25] U. A. T. Brinkmann, G. J. De Jong, C. Gooijer, *Methodol. Surv. Biochem. Anal.* **18** (1988) 321 – 338.

[26] T. M. Vickrey (ed.): *Liquid Chromatography Detectors, Chromatographic Science*, vol. 23, Marcel Dekker, New York 1983.

[27] R. P. W. Scott, *J. Chromatogr. Libr. Liquid Chromatography Detectors* **11** (1977).

[28] R. Schulz, H. Engelhardt, *Chromatographia* **29** (1990) nos. 11/12, 517–522.

[29] M. Varadi, J. Balla, E. Pungor, *Pure Appl. Chem.* **51** (1979) 1177.

[30] G. Henze, R. Neeb: *Elektrochemische Analytik*, Springer, Berlin 1986.

[31] G. Henze: *Polarographie und Voltammetrie; Grundlagen und analytische Praxis,* Springer Verlag, Berlin 2000.

[32] F. A. Schultz, D. E. Mathis, *Anal. Chem.* **46** (1974) 2253.

[33] A. Manz, W. Simon, *J. Chromatogr. Sci.* **21** (1983) 326.

[34] R. S. Deelder, H. A. J. Linssen, J. G. Koen, A. J. B. Beeren, *J. Chromatogr.* **203** (1981) 153.

[35] H. Akaiwa, H. Kawamoto, M. Osumi, *Talanta* **29** (1982) 689.

[36] T. Deguchi, T. Kuma, H. Nagai, *J. Chromatogr.* **152** (1978) 379.

[37] T. H. Ryan: *Electrochemical Detectors,* Plenum Press, New York 1984.

[38] K. Brunt: *Electrochemical Detectors in Trace Analysis,* vol. 1, Academic Press, New York 1981.

[39] P. T. Kissinger, C. Refshauge, R. Dreiling, R. N. Adams, *Anal. Lett.* **6** (1973) 465.

[40] B. Fleet, C. J. Little, *J. Chromatogr. Sci.* **12** (1974) 747.

[41] W. Kemula, *Rocz. Chem.* **26** (1952) 281.

[42] P. Just, M. Karakaplan, G. Henze, F. Scholz, *Fresenius J. Anal. Chem.* **32** (1993) 345.

[43] W. W. Kubiak, *Electroanalysis NBS* (1989) 379.

[44] J. Frank, *Chimia* **35** (1981) 24.

[45] P. T. Kissinger, C. S. Brunett, K. Bratin, J. R. Rice, *Spec. Publ. (U.S.)* **519** (1979) 705.

[46] S. Yao, A. Meyer, G. Henze, *Fresenius J. Anal. Chem.* **339** (1991) 207.

[47] H.-P. Nirmaier, E. Fischer, G. Henze, *GIT Labor-Fachzeitschrift* **42** (1998) 12.

[48] J. Schiewe, A. M. Bond, G. Henze, in G. Henze, M. Köhler, J. P. Lay (eds.): *Voltammetrische Spurenanalyse mit Mikroelektroden in Umweltdiagnostik mit Mikrosystemen,* Wiley-VCH Verlag, Weinheim 1999, p. 121.

[49] H.-P. Nirmaier, E. Fischer, G. Henze, *Electroanalysis,* **10** (1998) 187.

[50] R. Wintersteiger, G. Berlitz, *GIT Suppl. 3* (1989) 19.

[51] J. P. Hart, *Electroanalysis of Biologically Important Compounds,* Ellis Horwood, Chichester 1990.

[52] D. M. Radzik, S. M. Lunte: "Application of Liquid Chromatography/Electrochemistry in Pharmaceutical and Biochemical Analysis," *CRC Critical Reviews in Analytical Chemistry* **20** (1989) no. 5, 317.

[53] K. Štulik, V. Pacáková: "Electrochemical Detection in High-Performance Liquid Chromatography," *CRC Critical Reviews in Analytical Chemistry* **14** (1982) no. 4, 297.

[54] K. Štulik, V. Pacáková: "Electrochemical Detection Techniques in High-Performance Liquid Chromatography," *J. Electroanal. Chem.* **129** (1981) 1.

[55] G. Patonay, HPLC Detection, Newer Methods, VCH **1992.**

[56] K. Štulik, V. Pacáková: "Electrochemical Detection Techniques in High-Performance Liquid Chromatography," *J. Electroanal. Chem.* **129** (1991) 1.

[57] R. J. Rucki: "Electrochemical Detectors for flowing Liquid Systems," *Talanta* **27** (1980) 147.

[58] D. C. Johnson, S. G. Weber, A. M. Bond, R. M. Wightman, R. E. Shoup, J. S. Krull: "Electroanalytical Voltammetry in Flowing Solutions," *Anal. Chim. Acta* **180** (1986) 187.

[59] D. S. Austin-Harrison, D. C. Johnson: "Pulsed Amperometric Detection Based on Direct and Indirect Anodic Reactions," *Electroanalysis* **1** (1989) 189.

[60] R. Samuelsson, J. O'Dea, J. Osteryoung, *Anal. Chem.* **52** (1980) 2215.

[61] J. G. White, R. L. St. Claire, J. W. Jorgenson, *Anal. Chem.* **58** (1986) 293.

[62] P. T. Kissinger, *Anal. Chem.* **49** (1977) 447 A.

[63] M. Goto et al., *J. Chromatogr.* **226** (1981) 33.

[64] E. L. Craig, P. T. Kissinger, *Anal. Chem.* **55** (1983) 1458.

[65] J. Dutrieu, Y. A. Delmotte, *Fresenius J. Anal. Chem.* **314** (1983) 416.

[66] C. L. Putzig et al., *Anal. Chem.* **64** (1992) 270 R–302 R.

[67] H.-P. Nirmaier, E. Fischer, A. Meyer, G. Henze, *J. Chromatography A.* 1996, 730, 169.

[68] K. K. Unger, G. Gilge, J. N. Kinkel, M. T. W. Hearn, *J. Chromatogr.* 359, 61 (1986).

[69] H. Chen, C. Horvath, *J. Chromatogr.* 705, 3 (1995).

[70] K. Stulik, V. Pacáková, J. Suchánková, H. Claessens, *Anal. Chim. Acta* 352, 1 (1997).

[71] R. Ohnmacht, B. Boros, I. Kiss, L. Jelinek, *Chromatographia* **50** (1999) no. 1/2.

[72] I. M. Mutton, *Chromatographia* **47** (1998) no. 5/6, 291–298.

[73] K. K. Unger (ed.): "Packings and Stationary Phases in Chromatographic Techniques," *Chromatographic Science Series,* vol. 47, Marcel Dekker, New York 1990.

[74] J. Persek, H. Lin, *Chromatographia* **28** (1989) 565.

[75] I. Halasz, I. Sebastian, *Angew. Chem. Int. Ed. Engl.* **8** (1969) 453.

[76] B. G. Herslöf, *HRC & CC, J. High Resolut. Chromatogr. Chromatogr. Commun.* **4** (1981) 471–473.

[77] A. Sándy et al., *Chromatographia* **45** (1997) 206–214.

[78] M. Vecchi, G. Englert, R. Maurer, V. Meduna, *Helv. Chim. Acta* **64** (1981) 2746.

[79] E. R. Majors, *LC-GC Int.,* **8** (1995) no. 7, 368–374.

[80] K. Aitzetmüller, *J. Chromatogr.* **113** (1975) 231–266.

[81] H. Small, T. S. Stevens, W. C. Baumann, *Anal. Chem.* **47** (1975) 1801.

[82] J. W. BcBain, *Kolloid Z.* **40** (1926) 1.

[83] I. Porath, P. Flodin, *Nature (London)* **183** (1959) 1657.

[84] J. C. Moore, *J. Polymer Sci. Part A* **2** (1964) 835.

[85] Macherey & Nagel & Co.: "HPLC", Düren 1991.

[86] K. K. Unger: "Adsorbents in Column Liquid Chromatography," in [24]

[87] Z. Deyl, K. Macek, J. Janak (eds.): "Liquid Column Chromatography," *J. Chromatogr. Libr. 3* (1975).

[88] V. Meyer: *Praxis der Hochdruck-Flüssigkeits-Chromatographie,* Diesterweg-Salle-Sauerländer, Frankfurt 1979.

[89] K. Aitzetmüller, J. Koch, *J. Chromatogr.* **145** (1987) 195.

[90] J. E. Campbell, M. Hewins, R. J. Lych, D. D. Shrewsbury, *Chromatographia* **16** (1982) 162–165.

[91] S. A. Westwood (ed.): *Supercritical Fluid Extraction and its Use in Sample Preparation*. Blackie Academic & Professional, London 1993.

[92] K. Jinno (ed.): "Hyphenated Techniques in Supercritical Fluid Chromatography and Extraction," *J. Chromatogr. Libr.* **53** (1992).

[93] H. Godel, Th. Graser, P. Földi, *J. Chromatogr.* **297** (1984) 49.

[94] I. S. Krull (ed.): "Reaction Detection in Liquid Chromatography," *Chromatographic Science Series*, vol. 34, Marcel Dekker, New York 1986.

[95] H. Engelhardt, H. Elgass, *Chromatographia* **22** (1986) no. 1–6, June.

[96] A. H. M. T. Scholten, U. A. T. Brinkman, R. W. Frei, *J. Chromatogr.* **205** (1981) 229.

[97] P. Walkling, Ph. D. Thesis, Frankfurt 1968.

[98] I. Halasz, P. Walkling, *Ber. Bunsenges. Phys. Chem.* **74** (1970) 66.

[99] K. Hofmann, I. Halasz, *J. Chromatogr.* **199** (1980) 3.

[100] H. Engelhardt, U. D. Neue, *Chromatographia* **15** (1982) 403.

[101] U. D. Neue, Ph. D. Thesis, Saarbrücken 1976.

[102] D. Moore, W. H. Stein, D. H. Speckman, *Anal. Chem.* **30** (1958) 1190.

[103] D. H. Speckman, W. H. Stein, S. Moore, *Anal. Chem.* **30** (1958) 1190.

[104] M. Weigele, S. L. De Bernardo, J. P. Tengi, W. Leimgruber, *J. Am. Chem. Soc.* **94** (1972) 5927.

[105] M. Roth, *Anal. Chem.* **43** (1971) 880.

[106] S. Stein et al., *Arch. Biochem. Biophys.* **155** (1973) 202.

[107] D. Gottschalk, H. Körner, J. Puls in H. Engelhardt, K.-P. Hupe (eds.): *Kopplungsverfahren in der HPLC*. GIT Verlag, Darmstadt 1985.

[108] P. Vratny, U. A. T. Brinkman, R. W. Frei, *Anal. Chem.* **57** (1985) 224.

[109] L. Nondek, R. W. Frei, U. A. T. Brinkman, *J. Chromatogr.* **282** (1983) 141.

[110] H. A. Moye, C. J. Miles, S. J. Scherer, *J. Agric. Food Chem.* **31** (1983) 69.

[111] J. Schrijver et al., *Ann. Clin. Biochem.* **19** (1982) 52.

[112] J. Schrijver, A. J. Speek, W. H. P. Schreurs, *Int. J. Vitam. Nutr. Res.* **51** (1981) 216.

[113] J. T. Vanderslice, D. J. Higgs, *J. Chromatogr. Sci.* **22** (1984) 485.

[114] Deans-säulenschalten GC.

[115] J. F. K. Huber et al., *J. Chromatogr.* **83** (1973) 267.

[116] S. Wielinski, A. Olszanowski, *Chromatographia* **50** (1999) no. 1/2, 109–112.

[117] A. P. Bruins, *Adv. Mass Spectrom.* **11 A** (1989) 23–31.

[118] A. L. Burlingame, D. S. Millington, D. L. Norwood, D. H. Russel, *Anal. Chem.* **62** (1990) 268 R–303 R.

[119] H. Kataoka, J. Pawliszyn, *Chromatographia* **50** (1999) no. 9/10, 532–538.

[120] M L. Nedved, S. Habbi-Gondarzi, B. Ganem, J.-D. Henion, *Anal. Chem.* **68** (1996) 4228–4236.

[121] Cagnaird de la Tour, *Ann. Chim. Phys.* **21** (1822) 127.

[122] T. Anrews, *Philos. Trans. R. Soc. London* **159** (1869) 575.

[123] J. B. Hannay, J. Hogarth, *Proc. R. Soc. London* **30** (1880) 178.

[124] C. M. White: *Modern Supercritical Fluid Chromatography*. Hüthig, Heidelberg 1988.

[125] E. Klesper, A. H. Corwin, D. A. Turner, *J. Org. Chem.* **27** (1962) 700.

[126] S. T. Sie, W. van Beersum, G. W. A. Rijnders, *Sep. Sci.* **1** (1966) 459.

[127] S. T. Sie, G. W. A. Rijnders, *Sep. Sci.* **2** (1967) 699.

[128] S. T. Sie, G. W. A. Rijnders, *Sep. Sci.* **2** (1967) 729.

[129] S. T. Sie, G. W. A. Rijnders, *Sep. Sci.* **2** (1967) 755.

[130] S. T. Sie, J. P. A. Bleumer, G. W. A. Rijnders in C. L. A. Harbourn (ed.): *Gas Chromatography 1968*, The Institute of Petroleum, London 1969, p. 235.

[131] J. C. Giddings, *Sep. Sci.* **1** (1966) 73.

[132] J. C. Giddings, M. N. Myers, L. McLaren, R. A. Keller; *Science (Washington, D.C. 1883)* **162** (1968) 67.

[133] J. C. Giddings, M. N. Myers, J. W. King, *J. Chromatogr. Sci.* **7** (1969) 276.

[134] J. C. Giddings, L. M. Bowman, N. M. Myers, *Anal. Chem.* **49** (1977) 243.

[135] S. T. Sie, G. W. A. Rijnders, *Anal. Chim. Acta* **38** (1967) 31.

[136] R. E. Jentoft, T. H. Gouw, *J. Chromatogr. Sci.* **8** (1970) 138.

[137] R. E. Jentoft, T. H. Gouw, *J. Polym. Sci. Polym. Lett. Ed.* **7** (1969) 811.

[138] D. Bartmann, G. M. Schneider, *J. Chromatogr.* **83** (1973) 135.

[139] U. van Wasen, G. M. Schneider, *Chromatographia* **8** (1975) 274.

[140] U. van Wasen, G. M. Schneider, *J. Phys. Chem.* **84** (1980) 229.

[141] M. Novotny et al., *Anal. Chem.* **53** (1981) 407 A.

[142] M. L. Lee, K. E. Markides (eds.): *Analytical Supercritical Fluid Chromatography and Extraction*. Chromatography Conferences, Inc., Provo, Utah, 1990.

[143] K. P. Johnston, J. M. L. Penninger (eds.): "Supercritical Fluid Science and Technology," *ACS Symp. Ser.* **406** (1989).

[144] M. L. Lee, *J. Microcol. Sep.* **4** (1992) 91–122.

[145] A. F. Buskhe, B. E. Berg, O. Gyllenhaal, T. Greibrokk, *HRC & CC, J. High Resolut. Chromatogr.* **11** (1988) 16.

[146] M. L. Lee et al., *J. Microcol. Sep.* **1** (1989) 7.

[147] J. C. Moore, *J. Polym. Sci. Part A* **2** (1964) 835.

[148] B. E. Richter: *1986 Pittsburgh Conference and Exposition on Analytical Chemistry and Applied Spectroscopy*, Atlantic City, NJ, March 10–14, 1986, paper 514.

[149] B. E. Richter et al., *HRC & CC, J. High Resolut. Chromatogr.* **11** (1988) 29.

[150] G. Schomburg et al., *HRC & CC, J. High Resolut. Chromatogr.* **12** (1989) 142.

[151] B. E. Richter et al., *J. Chromatogr. Sci.* **27** (1989) 303.

[152] P. Morin, M. Caude, H. Richard, R. Rosset, *Chromatographia* **21** (1986) 523.

[153] J. W. Jordan, L. T. Taylor, *J. Chromatogr. Sci.* **24** (1980) 82.

[154] R. C. Wieboldt, G. E. Adams, D. W. Later, *Anal. Chem.* **60** (1988) 2422.

[155] S. B. French, M. Novotny, *Anal. Chem.* **58** (1986) 164.

[156] J. D. Pinkston, D. J. Bowling in [43]

[157] M. Smith (ed.): "Supercritical Fluid Chromatography," *RSC Chromatography Monographs*, 1989.

[158] J. Zapp, Ph. D. Thesis, Saarbrücken 1990.

[159] K. D. Pharma GmbH: DE P 42 065 39.9, 1992 (P. Lembke, H. Engelhardt, R. Krumbholz).

[160] Y. Hirata, F. Nakata, *J. Chromatogr.* **295** (1984) 315.

[161] H. H. Hill, M. A. Morrissey in G. M. White (ed.): *Modern Supercritical Fluid Chromatography*. Huethig, Heidelberg 1988.

[162] P. Morin, M. Caude, H. Richard, R. Rossel, *Analysis* **15** (1987) 117.

[163] J. M. F. Douse, *J. Chromatogr.* **445** (1988) 244.

[164] B. W. Wright, R. D. Smith, *J. High Resolut. Chromatogr.* **9** (1986) 73.

[165] J. B. Crowther, J. D. Henion, *Anal. Chem.* **57** (1985) 2711.

[166] R. M. Smith, *J. Chromatogr.* **505** (1990) 147.

[167] P. Lembke, in K. Anton, C. Berger (eds.): *Supercritical Fluid Chromatography with Packed Columns,* Marcel Dekker, New York 1998, pp. 429–443.

[168] P. Cuatrecasas, M. Wilchek, C. Anfinsen, *Proc. Natl. Acad. Sci. U.S.A.* **61** (1968) 636.

[169] W. H. Scouten: *Affinity Chromatography.* John Wiley & Sons, New York 1981.

[170] W. Brümmer in R. Bock, W. Fresenius, H. Günzler, W. Huber, G. Tölg (eds.): *Analytiker-Taschenbuch,* **vol. 2,** Springer Verlag, Berlin 1981, pp. 63–96.

[171] M. Wilchek, T. Miron, J. Kohn in W. B. Jacoby (ed.): *Methods of Enzymology,* **vol. 104,** Academic Press, New York 1984, pp. 3–55.

[172] I. R. Birch, C. R. Hill, A. C. Kenney in P. N. Cheremisinoss, R. P. Ovellette (eds.): *Biotechnology Applications and Research,* Techcromic Publishing Company Inc., Lancaster 1985, p. 594.

[173] P. Mohr, K. Pommering: "Affinity Chromatography, Practical and Theoretical Aspects," in J. Cazes (ed.): *Chromatography Science,* vol. 33: Marcel Dekker, New York 1986.

[174] K. Ernst-Cabrera, M. Wilchek in R. Burgess (ed.): *Protein Purification: Micro to Macro, UCLA Symposia on Molecular and Cellular Biology, new series,* **vol. 68,** A.R. Liss Inc., New York 1987, pp. 163–175.

[175] St. Ostrove in J. N. Abelson, M. I. Simon (eds.): *Methods of Enzymology,* **vol. 182,** Academic Press, New York 1990, pp. 357–371.

[176] J. Carlsson, J. C. Janson, M. Sparrmann in J. C. Janson, L. Ryden (eds.): *Protein Purification,* VCH Publishers, New York 1989, pp. 275–329.

[177] G. T. Hermanson, A. K. Mallia, P. K. Smith (eds.): *Immobilized Affinity Ligand Techniques,* Academic Press, New York 1992.

[178] R. F. Taylor, *Anal. Chim. Acta* **172** (1985) 241.

[179] K. Ernst-Cabrera, M. Wilchek, *Makromol. Chem., Makromol. Symp.* **19** (1988) 145–154.

[180] C. R. Lowe, M. J. Harvey, D. B. Craven, P. D. G. Dean, *Biochem. J.* **133** (1973) 499–506.

[181] P. Cuatrecasas, *Nature* **228** (1970) 1327.

[182] E. Müller, "Coupling Reactions" in M. Kastner (eds.): *Protein Liquid Chromatography,* Elsevier, Amsterdam 2000.

[183] J. Kohn, M. Wilchek, *Enzyme Microb. Technol.* **4** (1982) 161–163.

[184] G. S. Bethell, J. S. Ayers, M. T. W. Hearn, W. S. Hancock, *J. Chromatogr.* **219** (1981) 361.

[185] T. Miron, M. Wilchek, *Appl. Biochem. Biotechnol.* **11** (1985) 445–456.

[186] K. Ernst-Cabrera, M. Wilchek, *J. Chromatogr.* **397** (1987) 187.

[187] K. Nilsson, K. Mosbach in W. B. Jacoby (ed.): *Methods of Enzymology,* **vol. 104,** Academic Press, New York 1984.

[188] E. Stellwagen in J. N. Abelson, M. I. Simon (eds.): *Methods of Enzymology,* **vol. 182,** Academic Press, New York 1990, pp. 343–357.

[189] St. J. Burton in A. Kenney, S. Fowell (eds.): *Methods in Molecular Biology,* **vol. 11,** The Humana Press Inc., Totowa, N.Y., 1992, pp. 91–103.

[190] J. Porath, F. Maisano, M. Belew, *FEBS Lett.* **185** (1985) 306–310.

[191] B. Nopper, F. Kohen, M. Wilchek, *Anal. Biochem.* **180** (1989) 66–71.

[192] J. Porath, J. Carlson, I. Olsson et al., *Nature* **258** (1975) 598–599.

[193] Z. Horvath, G. Nagydiosi, *J. Inorg. Nucl. Chem.* **37** (1975) 767–769.

[194] V. A. Davankov, A. V. Semechkin, *J. Chrom.* **141** (1977) 313–353..

[195] E. Sulkowski, *Trends Biotechn.* **3** (1985) 1–7.

[196] P. J. Hogg, D. J. Winzor, *Anal. Biochem.* **163** (1987) 331–338.

[197] J. J. Langone, *Adv. Immunol.* **32** (1982) 157–252.

[198] A. Jungbauer et al., *J. Chromatogr.* **476** (1989) 257–268.

[199] K. Ernst-Cabrera, M. Wilchek, *Med. Sci. Res.* **16** (1988) 305–310.

[200] M. L. Yarmush et al., *Biotechnol. Adv.* **10** (1992) 413–446.

[201] G. Köhler, C. Milstein, *Nature (London)* **256** (1975) 495–497.

13. Thin Layer Chromatography

JOSEPH C. TOUCHSTONE, School of Medicine, University of Pennsylvania, Philadelphia, PA 19104-6080, United States

13.1. Introduction

The following constitutes a general guide to the use of thin layer chromatography (TLC). Pertinent reference sources on the subject include [1]–[3]. The discussion here is focused on precoated plates, since these have largely supplanted homemade layers. For further information on the relationship between TLC and other types of chromatography → Basic Principles of Chromatography.

13.2. Choice of the Sorbent Layer

Precoated plates for TLC have been commercially available since 1961. The sorbents may be coated on glass, plastic, or aluminum supports. Sorbents with and without binder, and with and without UV indicators, are available in a variety of layer thicknesses, ranging from 100 μm in the case of plastic plates and high-performance layers to 2311 μm for preparative layers. The largest selection is presented by glass plates, coated to a thickness of 250 μm in the case of "analytical layers."

Important considerations with respect to all TLC plates are that the layers be uniform in thickness and surface, and that the plates yield reproducible results from package to package. These are difficult characteristics to duplicate with homemade plates, and are of prime concern in analytical work. The availability of high-quality commercial precoated plates is in fact one of the reasons why quantitative thin layer chromatography has been so successful in many applications.

Plates with a plastic support have an additional distinguishing feature relative to glass-backed plates: They can be cut with scissors. This is advantageous, for example, when small plates are desired but only 20×20 cm plates are on hand. With plastic-backed plates it is also easy to elute separated substances from the developed layer by cutting out an appropriate region and immersing the cut piece in a suitable extraction solvent. This is an especially important consideration when working with radioactive substances (→ Radionuclides in Analytical Chemistry). Measuring the radioactivity of such a substance is facilitated by simply immersing the portion of plate containing the radioactive zone in a scintillation vial containing counting fluid. Because the layer need not be scraped, potential contamination and losses caused by flying dust are minimized.

Considering only the cost of the sorbent and plates, precoated plates are obviously more expen-

sive than homemade plates. However, the total expense of homemade plates also includes the cost of the coating apparatus and drying – storage racks, as well as the time involved. After these factors are taken into account it will be seen that precoated plates are highly desirable.

13.2.1. Properties of Presently Available Sorbents

Silica Gel. Silica gel is prepared by the precip- itation of a silicate solution with acid, or by hy- drolysis of silicon derivatives. The surface area and diameter of the silica gel particles depend on the method of precipitation. Variations in pH during precipitation can produce silica gels with surface areas ranging from $200-800$ m^2/g. It has been shown [4] that silica gel provides three types of surface hydroxyl groups: Bound, reactive, and free. Relative reactivity and adsorption follow the order bound > free > reactive. Thus, control of the distribution of surface functions can have a signif- icant effect on the chromatographic properties of a silica.

Special methods have been developed for pre- paring spherical particles with specific pore char- acteristics. Thermal treatment can also affect the pore size. Many of the silicas currently in use have a pore size of 6 nm.

Alumina. Alumina has seen limited use as a sorbent in TLC. Three types of alumina are avail- able: Acidic, basic, and neutral. The amount of water present greatly affects the chromatographic behavior of alumina. For control purposes the plates may be heated to a specific temperature prior to use.

Kieselguhr. Kieselguhr is a purified diatoma- ceous earth usually characterized by wide pores. It is frequently used as a *preadsorbent layer* on other types of TLC plates. Commercial kieselguhr prep- arations include Celite, Filter Cel, and Hyflo Super Cel.

Cellulose. Cellulose has been used as a sorbent since the earliest days of chromatography, origi- nally in the form of paper chromatography. Most methods of preparing celluloses for TLC are based on procedures for preparation of this medium for other uses. The types used for TLC are powdered (fibrous) or microcrystalline celluloses. The fiber length is shorter with TLC cellulose than in paper

chromatography, which leads to less diffusion of the spots relative to paper chromatography.

Polyamides. Polyamides, previously used al- most exclusively in column chromatography, have seen increasing application as sorbents for TLC. Certain separations are accomplished much more readily on polyamide than on conventional layers, since this medium lends itself to the analysis of water-soluble samples.

Ion Exchange. Ion exchange provides greater selectivity than most other forms of TLC, and recent developments have broadened its appli- cability. Ion exchange can be utilized with either organic or inorganic substances. Within these clas- sifications, a further division can be made into anionic or cationic processes according to the na- ture of the ions subject to exchange. Several sor- bents are now available for ion-exchange TLC, and the number will grow as development contin- ues.

Reversed-Phase Sorbents. Reversed-phase (RP) sorbents can be classified into (1) those in which a liquid is impregnated in a solid support, and (2) chemically bonded sorbents, where the liquid phase is bonded chemically instead of phys- ically. Chemical bonding provides layers that are more stable, preventing the stationary phase from being removed by solvent. Three types of reac- tions are used for bonding: Esterification, reaction of a chlorosilane with a selected organic substit- uent, and chlorination of the support followed by reaction with an organolithium compound to produce Si–C bonds. These bonding procedures can lead to various types of structures depending on the mode of attachment to the silica and the nature of substituents present at the other end of the alkyl side chain. The major bonding types have the structures Si–O–C, Si–O–Si–C, and Si–C–O–Si–N. The length of commercially avail- able chains ranges from 2 to 18 carbon atoms.

Gels. Gels provide an entirely different type of medium for separation on thin layers. Sephadex was one of the first gels used as a stationary phase in the resulting *steric exclusion* TLC.

Cyclodextrin. More recently, cyclodextrin- modified layers have been introduced for use in the separation of enantiomers. The increasing im- portance of isomer separation in pharmaceutical applications has led to a demand for these spe-

cially modified layers. Chiral layers can take several different forms. Most are C_{18} layers impregnated with a chiral-selecting protein derivative together with copper ions. Others consist of layers in which a sorbent silica has been chemically bonded to a cyclodextrin derivative.

13.2.2. Binders

Sorbents are prepared both with and without binding substances. A binder changes the properties of a sorbent such that the latter will adhere to its support and not readily flake off during handling, chromatography, visualization, or documentation. Nevertheless, binding agents are not always necessary, nor are they always desirable. Cellulose, for example, does not usually require a binder, since it has good inherent properties of adhesion and strength. Binders are obviously undesirable if they interfere with development or visualization. Calcium sulfate, a commonly used binder, is slightly soluble in water, and it is therefore not recommended in conjunction with aqueous solvent systems. Layers employing starch or polymer binders (so-called organic binders) should not be used when very high-temperature (>150 °C) visualization techniques are contemplated. They are also not suitable with universal visualization reagents such as sulfuric acid spray. Starch is appropriate only when it does not interfere, and when tough, strong layers are essential.

Calcium sulfate binder concentrations range from ca. 5 wt % in some silica gel preparations to 9 wt % in aluminum oxide G sorbent, 13 wt % in the widely used silica gel G, and up to 30 wt % in certain silica gel preparations for preparative (high capacity) TLC, where a "G" in a sorbent name denotes the presence of calcium sulfate (gypsum).

Commercial precoated plates are sometimes made with a small amount of organic polymeric binder, which makes the layer very hard and resistant to abrasion. This binder does not dissolve in relatively strong, polar solvents or in organic solvents, which increases the range of possible applications for separation but precludes certain detection procedures. Such binders, which are of a proprietary nature, absorb less moisture from the atmosphere than calcium sulfate, and for this reason the plates usually do not require activation prior to use.

13.2.3. Suppliers

A number of distributers and manufacturers supply precoated TLC plates, but not all carry a complete selection of sorbent/backing-plate combinations in the four major sizes (5×20, 10×20, 20×20, and 20×40 cm):

Analtech Inc., Newark, Del, United States	full line
J. T. Baker Chemical Co., Phillipsburg, N.J., United States	full line
Brinkmann Instrument Inc., Westbury, N.Y., United States	full line
Desaga GmbH, Heidelberg, Germany	relatively broad line
Eastman Kodak, Rochester, N.Y., United States	mylar backed layers
EM Science, Darmstadt, Germany	full line
Ingenur F. Heidenreith, Oslo, Norway	relatively broad line
Macherey-Nagel GmbH, Düren, Germany	sorbents and layers
Schleicher & Schuell GmbH, Kassel, Germany	specialty layers
Whatman Chemical Separations, Clifton, N.J., United States	full line

Some commercial 20×20 cm plates have pre-scored lines on the back at 5, 10, and 15 cm so they can easily be broken into smaller plates.

There is little literature available comparing plate-layer characteristics. To the benefit of the method, only two or three major manufacturers produce each sorbent type, so diversity with respect to a given sorbent is kept to a minimum. However, different manufacturers of the plates themselves employ different coating techniques, which in turn has an effect on the characteristics of sorbent that has been coated on the plates. Thus, even though two precoated-plate companies buy their sorbent from the same manufacturer, the resulting plates may not display identical separation characteristics.

The recently introduced *hydrophilic modified* precoated plates have broadened the spectrum of selectivity available for TLC. These plates are referred to as "HPTLC-NH$_2$ precoated plates," in which γ-aminopropyl groups are bonded to the surface of a silica-gel skeleton, and "HPTLC-CN precoated plates," with γ-cyanopropyl moieties as the functional groups. In terms of polarity the two hydrophilic modifications can be classified as follows: Normal $SiO_2 > NH_2 > CN > RP$. Since the surface NH_2 and CN groups are of medium polarity, no wetting problems are encountered with the corresponding plates even at the maximum level

of modification. Any mobile phase can be used, without restriction.

The presence of the functional group means that an NH_2-modified support can act as a weakly basic ion exchanger, useful, for example, in separating charged substances. Thus, various nucleotide building-blocks can be distinguished on the basis of differing charge numbers, with substances retained more strongly as the charge numbers increase. Other retention mechanisms can also be involved in sample retention on an NH_2 plate. In the case of oligonucleotides, the observed selectivity arises through differences in hydrophobicity among the various nucleobases.

A CN-modified layer can be used for either normal-phase or reversed-phase chromatography depending on the polarity of the solvent system.

The *hardness* of a layer varies as a function of any binder present, and this in turn affects the speed of development and the resulting R_f for a given substance. Factors such as activation and developing-chamber saturation are also important influences in this regard.

Generalizations regarding cellulose layers are easier to formulate. As noted previously, the two most popular forms of cellulose are fibrous and microcrystalline. The major supplier of fibrous cellulose is Macherey, Nagel and Co., whose main product is MN 300 cellulose. The major supplier of microcrystalline cellulose is FMC Corporation, under the trade name Avicel. Microcrystalline cellulose layers usually develop more rapidly and produce zones with lower R_f values that are more compact in size relative to fibrous cellulose layers.

13.3. Sample Cleanup

Proper preparation of the sample is the most important preliminary step in conducting any analysis (→ Sample Preparation for Trace Analysis), and this is especially true in chromatography. Of particular concern is the amount of extraneous material present in the sample solution. The greater the amount of "junk," the more sample preparation that will be required.

The Sample. Before undertaking the preparation of any sample, an aliquot of material representative of that sample must be obtained. Otherwise the analysis itself may be meaningless. The best way to assay a solid sample is to collect random aliquots from the whole and then mix them together before removing the extraction aliquot. If a package of tablets or capsules is to be analyzed, random portions from the whole should similarly be mixed together prior to selecting an aliquot for analysis.

The Solvent for the Sample. An appropriate solvent must of course be selected for dissolving the sample. When a sample must be derivatized, solubility characteristics may change drastically, or there may not be enough sample remaining for performing general solubility tests. The solubility of a crystalline material can be readily determined with a microscope. A few crystals of the sample are placed on a slide in the field of the microscope and carefully observed as a drop of solvent is added. If the crystals do not dissolve, the solvent can be evaporated and a new one selected for further testing. Testing with a number of solvents greatly expedites selection of the one that is optimal for experimental analysis.

Extraction of the compounds of interest from a complex sample and subsequent dissolution in the proper solvents requires different strategies depending on the compound class under consideration. For example, polar compounds do not dissolve well in nonpolar solvents, and nonpolar compounds will not dissolve in polar solvents. Ionic compounds dissolve best in a polar solvent to which the counter ion has been added. Satisfactory sample application on a TLC layer requires that the solvent be highly volatile and as nonpolar as possible. Nonvolatile solvents spread through the plate-layer matrix during the period of contact, taking sample with them. Polar solvents likewise spread solute through the matrix as a consequence of both high solvent power and low volatility.

Sample Extraction. Several alternatives are available for extracting the sample from its solvent. These include:

1) *Direct application* of the solution, with extraction taking place on the TLC layer
2) Application of a solution of the sample on a *preadsorbent layer*, followed by direct development

   With both of these techniques proper selection of the mobile phase can ensure that extraneous materials are left at the chromatographic origin. In some cases the layer can be predeveloped in such a way that analyte is deposited on a line which can be regarded as an origin for sub-

Table 1. Sorbents for use in microcolumns

Nonpolar	Polar	Ion exchange
C_{18}	cyanopropyl	carboxylic acid
C_8	diol	propylsulfonic acid
C_2	silica gel	benzenesulfonic acid
Phenyl	aminopropyl	diethylaminopropyl
Cyclohexyl	propyl	quaternary or secondary amine

sequent development with the correct mobile phase for separation.

3) *Liquid Extraction of the Sample.* Some samples may be too impure for simple extraction by one of the first two methods, but control of the pH often improves the situation. For example, organic acids are converted into water-soluble salts at high pH. Lowering the pH restores the unionized acids, which are then extractable into organic solvents. At high pH, basic compounds are extracted into organic solvents, returning to the water phase as their salts at low pH.

4) *Solid Phase Extraction.* Columns and microcolumns (also referred to as cartridges) prepacked with sorbents can be used for *solid phase extraction (SPE)* of samples. Such columns are available commercially, but one can easily pack a simple tube with the required sorbent, such as those in Table 1. Some extraction columns are made to fit onto a syringe, whereby the column becomes an extension of the syringe barrel, and the plunger is used to force the sample and eluting solvents through the packing. Table 2 lists a number of United States suppliers of packed SPE columns.

SPE columns require minimal amounts of solvent. Eluants of increasing strength can be used to elute analytes with different polarities in separate fractions. Thus, silica-gel columns release low-polarity substances upon elution with a nonpolar solvent, whereas the analyte is retained as a consequence of its polarity. The analyte is then eluted with a more polar solvent. The least polar solvent capable of eluting the analyte should always be used to ensure that very polar impurities remain behind. The eluant is subsequently evaporated; alternatively, if an analyte concentration is sufficiently high, an aliquot can be taken for application to the TLC plate. This methodology

presents unlimited possibilities for analyte isolation.

5) *Liquid – Liquid Extraction Using Microcolumns.* Special microcolumns buffered at pH 4 – 9 and packed with diatomaceous earth have been developed for liquid – liquid partitioning. These can be utilized for sample volumes ranging from one to several hundred milliliters. The principle of liquid – liquid partitioning is also the basis of a device known as an Extube. This is a disposable plastic syringe packed with an inert fibrous matrix presenting a large surface area. A sample and an extraction solvent are poured into the tube, where liquid – liquid partitioning occurs. Analyte is extracted into the organic phase, which passes through the matrix, while impurities and other polar components, including water, are retained. The eluate is then concentrated by evaporation prior to TLC.

General Guidelines for the Use of SPE Columns
[5]. The SPE cartridge must always be *conditioned* with the particular solvent that will be used. This is necessary both to activate the sorbent and to clean the column. Conditioning procedures vary depending on whether normal or reversed-phase sorbents are employed. Ion-exchange conditioning is a function of the polarity of the sample solvent.

For normal-phase packings the first conditioning wash is often conducted in the solvent in which the analyte is dissolved. With most SPEs 2 mL of solvent is sufficient. An SPE can also be conditioned with 3 – 5 mL of a nonpolar solvent such as hexane. This can be adjusted by adding methylene chloride, but the final mix should always be miscible with the sample solvent.

With reversed-phase sorbents, methanol is used for the first wash. A typical cartridge requires 3 – 5 mL of methanol, which is followed by the solvent used to prepare the sample. A film of water-miscible solvent remains on the sorbent after the aqueous solvent has been added, improving the interaction between aqueous sample and hydrophobic sorbent.

For ion-exchange sorbents, deionized water is usually used first, but if the sample is in a nonpolar solvent, the SPE should be conditioned with 2 – 5 mL of this solvent instead. Aqueous samples require use of 2 mL of methanol followed by 2 mL of water. The sorbent is then adjusted to the organic solvent content and salt concentration of the sample solution. Sorbent must not be allowed to

Table 2. Suppliers of SPE columns in the United States

Company	Column designation	Accessories available
Alltech Inc., Deerfield, Ill.	"Extract-Clean"	manifold
Analtech Inc., Newark, Del.	"Spice" cartridges	manifold
J. T. Baker Chem. Co., Phillipsburg, N.J.	"Baker-Bond"	manifold
Applied Separations, Bethlehem, Pa.	"Spe-ed" cartridges	various accessories
Bodman Chemicals, Aston, Pa.	"Pure-Sep" cartridge	various accessories
Isolab Inc., Akron, Ohio	"Quick-Sep"	manifold
J & W Scientific, Folsom, Ca.	"Accuford" tubes	manifold
Versa-Prep Lab Systems Inc., Nanuet, N.Y.	"Tef-Elutor"	various accessories
Varian Inc., Sunnyvale, Ca.	"Bond-Elute"	various accessories
Waters Associates, Milford, Mass.	"Sep-Pak"	various accessories
Worldwide Monitoring Corp., Hatfield, Pa.	"Cleanscreen" columns	full line of sorbents
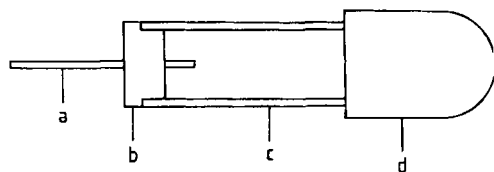
Figure 1. Microcapillary spotting device for TLC
a) Capillary tube; b) Septum; c) Support tube; d) Rubber bulb

dry between conditioning and addition of the sample.

The sample solution is subsequently added to the top of the cartridge and drawn through the sorbent bed by application of a vacuum. Special vacuum manifolds are available for this purpose, some of which handle 12 or more tubes, thus increasing sample throughput.

The next step is *washing* of the column (and sample). For reversed-phase work, washing with a polar solvent generally ensures that the analyte will be retained. Normal-phase applications require a wash with some nonpolar solvent that leaves the analyte on the sorbent. An ion-exchange column is washed with water, again so that the analyte will be retained.

Washing is followed by *elution* of the analyte. In reversed-phase mode a nonpolar solvent is used to remove the analyte; in normal phase, it is a polar solvent that removes the analyte. Adjustment of the ionic strength or modification of the eluting solvent is used to free the solute from an ion-exchange column. In each case, two volumes of $100-500 \mu L$ should suffice for complete elution.

The eluant is then *evaporated*, and the residue (extract) is *reconstituted* in a solvent suitable for introduction into the required chromatographic system. Typically, $5-200 \mu L$ of solvent is used for this purpose, $2-10 \mu L$ of which is taken for

analysis. If the concentration is suitable, an aliquot of the eluant can be analyzed directly without the need for evaporation and reconstitution.

13.4. Sample Application

Numerous devices are available for applying samples to TLC plates. *Microliter syringes*, such as those manufactured by Hamilton, have frequently been used for applying 5 to $10 \mu L$ amounts of samples. Application of small volumes of sample with a syringe results in relatively narrow or small application areas, since little solvent is present to cause diffusion. Sucessful use of these devices takes practice, however, and it necessitates washing of the delivery device between sample applications. *Microcapillaries* for this purpose are available from many suppliers as an alternative, and these offer the best volume-delivery precision. Capillaries are also disposable, ruling out contamination between samples. These capillaries with holders, as illustrated in Figure 1, are available in a wide range of fixed volumes, and they have proven to be ideal for TLC applications.

Automated or *semiautomated sample applicators* are available as well. These devices apply consistent and reproducible sample spots, but it is a misconception that they are necessary for quantitative work. With proper technique, manual methods of sample application can provide results entirely comparable to those from automatic devices. Commercial automated units employ syringes or rows of microcapillaries to apply a spot or band of sample, and some actually "spray" the sample onto the layer. Many are designed for preparative separations, applying large amounts of sample as streaks across the sorbent layer.

The Sample. In the case of silica-gel layers the sample should be dissolved in the least polar solvent in which it is soluble. With a solution of the proper concentration, 1–5 µL should suffice to permit detection of the analyte of interest after chromatography. Concentrating may prove to be the best procedure with a dilute solution. However, dilute solutions can also be applied in the form of several successive aliquots. If the sample is too dilute, spreading of the application area can be prevented by accelerating evaporation of solvent with a stream of air blown over the area in question. A hair dryer is often used for this purpose. Correct application ensures that the separated zones will be both symmetrical and compact.

Sample Amounts. The amounts of the sample and the analyte are critical factors in achieving good results, since applied load affects the shapes and positions of developed zones. Application of a sample that is too concentrated can result in overloading. Overloaded zones produce comet-like vertical streaks after development, which can cause overlap of previously separated components.

TLC is a sensitive technique, and it is important to "think small" with respect to sample size. The highest resolution will always be obtained with the smallest sample commensurate with the ability to visualize the analyte. This is also one of the prerequisites for reproducibility, especially if densitometry is to be used for quantitation (see Chap. 13.8). Zones that are too concentrated result in integrated curve areas that behave in a nonlinear way on regression plots. Depending on the nature of the analyte, sample sizes as small as picogram amounts are not uncommon, especially in high-performance thin layer chromatography (HPTLC). This is especially true in the analysis of fluorescent substances. Thus, for ultimate sensitivity it is sometimes desirable that a fluorescent derivative of the analyte be prepared.

With regard to preadsorbent plates, the applied sample volume has an effect on preadsorbent efficiency. Up to certain levels, increasing the sample volume appears to increase plate efficiency, since a larger volume prevents compacted zones, and uniformity of the starting line after predevelopment is maximized.

Using Micropipettes for Manual Application of Analytical Samples. Care should be taken to ensure that separate spots are placed on a straight line horizontally across the plate. With hard-layer plates a guideline can be drawn on the plate sur-

face itself with a soft pencil to act as a spotting guide. A white sheet of paper with a dark line drawn on it can also serve as a guide, since such a line will be visible through a layer positioned over it. Use of a guideline makes it easier to keep the sample origins evenly spaced across the plate and parallel to the edge. In the case of plates with a preadsorbent zone no guide or guideline is required, because the sample can be applied anywhere within the preadsorbent zone. This zone is inert, so it will not retain the analyte, permitting samples to move with the solvent front until they reach the adsorbent zone. A preadsorbent provides a convenient means of applying a dilute sample without the need for evaporation. No matter how the sample is applied, the result after predevelopment will be a fine line at the juncture between the adsorbent and the preadsorbent material.

Samples are applied to a 20×20 cm TLC plate 2 cm from the bottom. The two outermost samples should be 1.5 to 2 cm from the edges. Normally, individual "lanes" are kept about 1 cm apart. The origin should be located 2 cm from the bottom to ensure that sample will not be eluted inadvertently into the mobile-phase reservoir. Standards are interspersed at every third or fourth position. A single plate can provide space for 20 or more samples.

Usually 2–10 µL of sample solution is applied to the TLC plate. A fixed-volume micropipette is completely filled with the solution, and the end of the capillary is then touched to the surface of the TLC adsorbent, causing solution to be "wicked" into the adsorbent, and producing a round "spot." The capillary is most conveniently held by thrusting it through a hole in a septum, which is in turn inserted into a larger piece of glass tubing. Some manufacturers provide a rubber bulb (similar to the top of a medicine dropper) that fits over this larger tube. The bulb has a hole in it to equilibrate pressure while the capillary is draining (see Fig. 1).

Since it is known that small samples lead to improved separation, a method known as *overspotting* is sometimes used for sample application. Thus, a 10 µL sample might be applied to the plate as five spots of 2 µL each, all delivered at the same place. Successive spots must not be applied until solvent from the previous application has completely evaporated. Failure to observe this precaution can result in distorted zones. Sample application in this way can be used to form a single spot, or a series of spots can be joined to form a band. Application of a sample as a band gives greater

resolution [6]. Treatment between applications with a hair dryer or a stream of filtered air facilitates solvent evaporation. The zone of application should be kept as narrow as possible compatible with the degree of concentration required. Spots or zones that are too concentrated result in poor separation, since the mobile phase tends to travel around the concentrated center, which leads to distorted zones or tailing.

Applying Preparative Samples. The preferred method for introducing a sample onto a preparative layer is to apply it as a narrow streak across the entire width of the plate. Both manual and automatic means are available for applying such a streak. The streak should be made as straight (horizontally) and as narrow as possible. Most commercial streakers are capable of producing a sample zone only 1 – 3 mm wide. A layer 1000 μm (1 mm) thick can readily handle 100 mg of sample.

Some successful preparative work has been reported based on the same manual techniques developed for applying analytical spots. In these cases a small (i.e., 2 – 4 μL) pipette is used to apply successive round spots in direct contact with each other across a plate, but the method is time-consuming and tedious, and it requires great care. Any spotting irregularities are greatly magnified in the separation process, and this often makes separation of closely spaced components very difficult. However, if the separated components are widely spaced, then satisfactory results are achievable in this way.

The use of preparative plates with a preadsorbent zone reduces the amount of care necessary to produce a narrow band. Sample in this case is applied to the relatively inert preadsorbent layer as a crude "streak" across the plate. Even so, it is best to exercise some care and apply the sample evenly across the plate, because a high local sample concentration can cause an overloading effect in the corresponding region of the plate.

The Importance of Sample Drying. Thorough drying of applied samples is a necessary step that is frequently ignored. Errors can result if a sample carrier solvent is not allowed to evaporate completely from the TLC plate. Many chromatographers erroneously assume that the sample spot is sufficiently dry when solvent is no longer visible on the plate. The preferred practice is to allow a drying time of five to ten minutes, perhaps assisted by the air stream from a hair dryer.

13.5. The Mobile Phase

Proper selection of the mobile phase is essential to achieving optimal TLC results. Indeed, the great strength of TLC lies in the wide range of applicable mobile-phase solvents, extending from the extremely nonpolar to the extremely polar. A highly polar mobile phase is generally used when the separation involves very polar substances, whereas a less polar mobile phase is best for less polar components. A very polar substance is strongly attracted to silica gel, so a polar solvent is required to displace it. Chromatographic conditions cited in references may be useful as starting points, but some knowledge of solvent – solute characteristics is required for optimal separation.

Selecting a Mobile Phase. It is quite rare for nothing whatsoever to be known about the chemical nature and solubility characteristics of a sample. If one knows something about what constitutes a "good" solvent for the sample, this may imply what will be a "poor" solvent. When a "good" solvent is used as the carrier, adsorbed materials will move in the solvent front, while a "poor" solvent will probably not move anything. What is required for effective TLC is a "mediocre" solvent, one that will cause partial movement and some measure of selective discrimination. Usually the first step is to find a solvent or mobile phase that will move the analyte with an R_f of 0.5. In difficult separations, a mobile phase with a changing composition may be necessary.

Interaction forces involved in TLC include: Intramolecular forces, inductive forces, hydrogen bonding, charge transfer, and covalent bonding [7]. Understanding selectivity thus requires that something be known about the nature of these various types of interaction.

Selecting a Mobile Phase Based on Solvent – Solute Characteristics. The mobile phase in TLC is the source of mobility, but more importantly it produces selectivity. Selectivity — the degree of interaction between components of the mobile phase and analytes in the solute — determines the extent of any resolution that will be achieved. Solvent strength is the factor that determines mobility. Knowing the characteristics of a solute makes it easier to decide which solvents to try during the search for a mobile phase that will produce the desired separation.

Important general characteristics of mobile-phase solvents important in TLC include:

1) *Water content;* this can vary widely, and it is a common source of reproducibility problems, since traces of water may lead to large changes in solvent strength and selectivity
2) *Autoxidation* sometimes occurs during solvent storage, especially over prolonged periods
3) Solvents kept in metal containers may be subject to *decomposition* and unknown chemical reactions, which may change the solvent characteristics
4) Solvent *impurities* may differ from batch to batch, and such impurities may interact with the solute, the solvent, or the sorbent
5) *Chloroform* often contains ethanol as a preservative, and since ethanol has high solvent strength, small amounts may have a dramatic effect

The optimal composition for a mobile phase depends on the nature and number of functional groups in the material to be analyzed. A molecule containing many oxygen functions will display high interactivity, whereas a hydrocarbon with no functional groups will show little interaction. On the other hand, a hydrocarbon solvent will dissolve other hydrocarbons, but not molecules characterized by numerous polar functional groups. Functional groups affect the solvent–solute interaction in the order RH < $ROCH_3$ < $RN(CH_3)$ < ROR < RNH_2 < ROH < $RCONH_2$ < RCO_2H. This sequence also applies to components of the mobile phase [8].

The level of interaction between the solute and the mobile phase as opposed to the sorbent determines the distribution of the solute between these two phases. The mobile phase is generally less reactive than the sorbent. If the analyte interacts more strongly with the mobile phase it will prefer to remain in the mobile phase, and the separation will be poor. If the solute has a more powerful interaction with the sorbent the separation potential will be greater. Chromatography thus consists of an interplay between solute–sorbent and solute–mobile-phase interactions (→ Basic Principles of Chromatography).

The *solvent strength parameter* (see below) provides a quantitative measure of the relationship between solvent strength and the type of sorbent used, although solvent behavior differs greatly in normal- vs. reversed-phase chromatography. The interaction of a solvent with the sorbent is thus a function of the type of sorbent used. Methanol is more polar than chloroform. As polarity increases, so does elution strength with respect to a normal-phase sorbent such as silica gel. Depending on the analyte, methanol will lead to more rapid analyte migration up the layer relative to a less polar solvent. The opposite is the case with a reversed-phase system. Here, the more polar the mobile phase the greater the tendency toward retention by the sorbent. Thus, water will not readily move analytes in a reversed-phase system, but will do so easily under normal-phase conditions. Even so, if the analyte does not dissolve, nothing will happen. Hexane, a nonpolar solvent, is at the opposite end of the scale from methanol.

Perhaps the most widely recognized characteristic solvent property is the *solvent strength parameter $E°$* [7]. This was defined by SNYDER as the adsorption energy for a solvent per unit of standard sorbent. The higher the solvent strength, the greater will be the R_f value of an analyte in normal-phase TLC. The opposite is true in the case of reversed-phase sorbents. Solvent strength parameters are listed below for a variety of common solvents [7]. The values cited are based on heats of adsorption with alumina, but a similar sequence and similar values apply to silica gel. The same principles also hold true for reversed phases, but the corresponding solvent-strength values are reversed.

Solvent	$E°$, Al_2O_3
Fluoroalkanes	−0.25
n-Pentane	0.00
Hexane	0.00
Isooctane	0.01
Petroleum ether, Skellysolve B, etc.	0.01
n-Decane	0.04
Cyclohexane	0.04
Cyclopentane	0.05
Diisobutene	0.06
1-Pentene	0.08
Carbon disulfide	0.15
Carbon tetrachloride	0.18
Amyl chloride	0.26
Butyl chloride	0.26
Xylene	0.26
Diisopropyl ether	0.28
2-Chloropropane	0.29
Toluene	0.29
1-Chloropropane	0.30
Chlorobenzene	0.30
Benzene	0.32
Ethyl bromide	0.37
Ethyl ether	0.38
Ethyl sulfide	0.38
Chloroform	0.40
Methylene chloride	0.42

Solvent	$E°$, Al_2O_3
Methyl isobutyl ketone	0.43
Tetrahydrofuran	0.45
Ethylene dichloride	0.49
Methyl ethyl ketone	0.51
1-Nitropropane	0.53
Acetone	0.56
Dioxane	0.56
Ethyl acetate	0.58
Methyl acetate	0.60
Amyl alcohol	0.61
Dimethyl sulfoxide	0.62
Aniline	0.62
Diethyl amine	0.63
Nitromethane	0.64
Acetonitrile	0.65
Pyridine	0.71
Butyl cellosolve	0.74
Isopropyl alcohol, n-propyl alcohol	0.82
Ethanol	0.88
Methanol	0.95
Ethylene glycol	1.11
Acetic acid	large
Water	larger
Salts and buffers	very large

Solvent *selectivity* refers to the ability of a solvent to establish interaction differences for two different analytes. Resolution is dependent on three factors:

1) The nature of the sorbent
2) The migration rate of the two solutes, which in turn depends on the solvent strength
3) Separation factors, which are determined by interactions between the solvent and the mobile phase

If two solutes cannot be separated by a particular mobile phase of predetermined strength, it may still be possible to achieve separation with a different mobile phase of the same strength. The above list of solvent strengths can be used to illustrate application of the solvent selectivity concept. Selectivity arises from the fact that a given solvent will interact differently with two solutes of different functional makeup. The solvent-strength list includes several examples of solvents with identical strength values. Ethyl acetate and diethylamine for example, have approximately the same solvent strength, but entirely different chemical structures, and thus different reactivities with respect to other molecules. In particular, two solutes that cannot be separated with a given mobile-phase system containing ethyl acetate may separate when diethylamine is substituted for the ethyl acetate despite a minimal change in overall mobility.

The mechanisms of retention on chemically bonded reversed phases are not clearly understood,

but observed R_f values for a series of solutes usually turn out to be reversed compared to the corresponding sequence on silica gel provided water constitutes a large fraction of the mobile phase. It is also possible to achieve separations on RP plates using an entirely organic mixture as the mobile phase.

Most RP chromatography is conducted with two-solvent mixtures composed of water together with an alcohol, acetonitrile, acetone, or dioxane. Methanol – water (8:2) is a convenient starting point for work with RP layers. The solvent strength is then varied by changing the water – organic modifier ratio, and selectivity is improved by using a different modifier. More polar samples require relatively more water. Small changes in mobile-phase strength have less effect on R_f values in RPTLC than in normal-phase TLC, making mobile-phase selection easier in the former case.

Mobile phases for gel TLC must be capable of dissolving the sample and at the same time causing the gel network to swell. Organic solvents such as tetrahydrofuran (THF) are common for gel permeation TLC on organic polymer layers (e.g., styrene – divinylbenzene), while aqueous buffers are employed for gel filtration chromatography on layers such as dextran.

Useful general rules related to selectivity and solvent strength include the following:

1) Increasing the solvent strength will decrease separation and increase R_f.
2) Decreasing the solvent strength will increase separation, but decrease R_f.
3) Selectivity is greatest when the concentration of the stronger component of the mobile phase is greater than 5 vol % and less than 50 vol %. In other words, more favorable solute – solvent interactions result when the solvent of greater strength is not dominant.
4) Introduction of a solvent with greater hydrogen-bonding potential may increase resolution.
5) With respect to (3), it is generally best to examine the effect of small changes in the amount of the stronger modifying solvent.
6) If the observed R_f value is too high, a primary solvent of lower solvent strength should be tried rather than drastically changing the proportions of a mixture, and vice versa.

13.6. Development

13.6.1. Apparatus

Most TLC development is accomplished by allowing the mobile phase to ascend a plate standing in a suitable-sized chamber or tank containing the appropriate mobile phase. Glass chambers are by far the most common. Simple chambers can be made from wide-mouthed screw-cap, battery, or museum jars. Chambers specifically made for TLC are available in a variety of sizes, both rectangular and circular, accommodating all sizes and formats of TLC plates. Aluminum racks are available for supporting and developing several conventional plates at one time in a single tank. One company has introduced a special tank with a raised glass hump running lengthwise on the inside bottom to act as a tank divider. The latter requires only 20 mL of solution per side, and makes it possible for two large plates to be developed simultaneously. The divided bottom also allows the tank to be saturated with two mobile-phase systems simultaneously, placed on opposite sides of the divider, so that a plate can be introduced on either side for development. Small chambers have been designed specifically for use with 10×10 cm or HPTLC plates. These small chambers occupy much less bench space, and require considerably less solvent.

A *sandwich chamber* for use with 10×10 cm plates offers several advantages, including small solvent volume (6 mL), rapid chamber saturation, and rapid development. A glass-backed TLC plate forms one wall of the chamber, with the adsorbent facing inward toward a second glass plate. The sides are clamped together with metal clips and the open end of the chamber is inserted in a solvent trough containing the developing solvent. Other commercially available sandwich chambers accommodate larger plates (20×20 and 10×20 cm). Such a chamber can be purchased commercially from a number of suppliers for about the same price as a normal glass TLC tank.

Ascending or Linear Development. The separation process in TLC occurs in a three-phase system consisting of stationary, mobile, and gas phases, all of which interact with each other until equilibrium is reached. If a plate is developed in a chamber whose atmosphere is not saturated with the mobile phase, flow up the plate may occur in an unsymmetrical way because some mobile phase

evaporates from the plate into the chamber atmosphere in an attempt to establish equilibrium. The entire chromatographic process then depends on the volatility of the mobile phase, the temperature, and the rate of development. It is for this reason that the development chamber should be made as small as conveniently possible based on the plate size and the number of plates to be developed at one time.

Reproducible results are achieved only with a constant environment. Laboratory benches exposed to the sun during part of the day should be avoided because of temperature variation. Most development chambers are of such a size that they may conveniently be placed in an environmentally controlled space, room, or bath as a way of maintaining consistency. It is advisable to weigh down the lid of the chamber so that it cannot be dislodged by vapor pressure or heat produced by the mobile phase, which would disturb the equilibrium within the system. Some newer chambers include latches that actually "lock down" the lids.

Chamber Saturation [9]. Saturation of the developing chamber atmosphere is facilitated by lining the inner walls with blotting paper or thick chromatography paper (e.g., Whatman 3 MM), or by using saturation pads specially designed for TLC, although some separations can be accomplished in unlined tanks. The individual components of the mobile phase should be mixed thoroughly before they are introduced into the developing chamber. Reported solvent proportions are assumed to be by volume unless stated otherwise. It is advantageous to pour the mixed mobile phase over the lining paper as it is added to the developing chamber, since this hastens the process of saturating the paper and the chamber. Enough time should be allowed for complete equilibration of the chamber.

Plate development may be described in terms of either time or distance. With timed development a plate is allowed to develop for a specific period of time, after which it is carefully lifted out of the chamber. The location of the solvent front is then marked with a pencil or a sharp object such as a syringe needle. If mobile phase is instead expected to travel a fixed distance (e.g., 10 cm), this distance should be marked on the plate at the time of sample application. It then becomes easy to observe when the mobile phase has reached the desired point so the plate can be removed promptly from the chamber. A fixed development distance is used to standardize R_f values; because
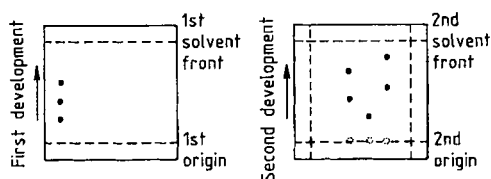
Figure 2. An example of two-dimensional development

R_f values by definition fall between 0.0 and 1.0, 10 cm is a convenient standard distance.

13.6.2. Development Techniques

Standard Development. The standard method of development has already been described in the preceding paragraphs. If the TLC characteristics of the analyte are unknown, a simple single-component mobile phase is preferable for preliminary experiments aimed at establishing separation behavior. If no single-component mobile phase provides the separation desired, however, multicomponent systems should be examined.

Multiple Development. If a single development of a plate suggests that a partial separation has been achieved, the plate should be allowed to dry, after which it can be reinserted into the developing chamber for redevelopment. This procedure may be repeated as many times as necessary to achieve a satisfactory separation. Repeated development increases the effective length of the layer, thereby increasing the likelihood of separation, because the various molecules will have an opportunity to interact over a greater distance, perhaps culminating in complete resolution.

With a mixture containing compounds that differ considerably in their polarities, development with a single mobile phase may not provide the desired separation. However, separation may result if such a plate is developed successively with various mobile phases that have different selectivities or strengths. Multiple mobile phases can be used in a number of ways. For example the least polar phase might be introduced first, followed by a more polar phase, or vice versa.More than two developments can also be tried, and the development distance on the plate might be changed with each development. For example, a very polar mobile phase, such as methanol, could be used first with a given sample mixture in order first to separate the polar substances. After the meth-

anol had developed part way up the layer the plate would be removed from the chromatography tank and dried. Then a second mobile phase, a nonpolar solvent such as cyclohexane, would be allowed to develop the entire length of the plate in an attempt to resolve in the upper portion those nonpolar substances that were not resolved in the previous (polar-phase) development. Depending on the circumstances, the same procedure might also be carried out with the reverse sequence.

Continuous Development. Continuous development involves a continuous flow of mobile phase along the length of the plate, with the mobile phase being permitted to evaporate as it reaches the end. The chromatography tank in this case is not as tall as a normal tank, which means that a normal-sized TLC plate extends beyond the top. Mobile phase then evaporates from the exposed end of the plate, causing a continuous flow to be maintained.

Continuous development offers the advantage of greater separation without the need for changing the mobile phase. Solvent selection is simplified because one can exploit the high separation power of a low-polarity solvent without suffering the usual consequence of reduced R_f. Sensitivity may be increased as well, because zone diffusion is reduced as a result of the shorter distance traveled by the analyte. The technique is essentially equivalent to development on a very long TLC plate without the requirement for special equipment.

Two-Dimensional Development. This developing technique involves the application of normal, multiple, or stepwise development in two successive dimensions. A very versatile method, it has been widely used in paper chromatography as well as in certain TLC separations. A single sample is applied at one corner of a plate and developed in the usual way. The plate is then removed from the developing chamber and dried, after which it is rotated by 90° and placed in a chamber containing a different mobile phase so that it will develop again, this time in a direction perpendicular to that of the first development. The line of partially resolved sample components from the first development becomes the origin for the second development (see Fig. 2). It is important that initial sample application not be too close to the edge of the plate, since otherwise the resulting line of developed spots might fall below the level of solvent in the second developing chamber. A
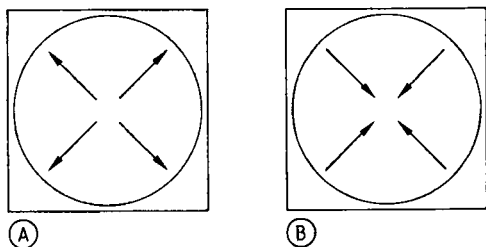
Figure 3. Circular and anticircular TLC development
A) Circular development; a wick is used to apply mobile phase to the center, causing radial flow outward; B) Anticircular development, in which solvent is applied with a wick to the outer ring, causing solvent migration inward toward the center

distance of 2.5 – 3 cm from the edge is usually sufficient. The versatility of the technique arises not only from the second development with a different mobile phase, but also from the opportunity provided for treating or modifying the layer and the sample prior to the second development. Standards can be applied to the second origin before the second development to permit direct comparisons and calculation of R_f values after development and visualization. The technique is also applicable to plates containing two different sorbents, such as silica gel on one half and RP_{18} on the other.

Circular and Anticircular Development. Circular and anticircular development have been accomplished in a number of ways. Because of the need for solvent-delivery control in such systems, chromatography is best carried out in equipment specially designed for the purpose. It is also important that the chamber be kept level. Camag, Analtech, and Anspec all offer chambers for circular development. Samples are applied at the center, and mobile phase is wicked to the surface. Development causes the analyte mixture to separate into a series of concentric rings. Figure 3 illustrates these devices.

Reversed-Phase Partition Development. Reversed-phase development employs a relatively nonpolar stationary phase, which means that the layer is hydrophobic rather than hydrophilic as in normal partition TLC. The corresponding mobile phase is relatively polar, and may be water or an alcohol. Most of the common TLC sorbents have been used in conjunction with partition development, including cellulose, silica gel, and kiesel-

guhr. This technique is generally applicable to the separation of polar compounds, which often remain at the origin during conventional development. With reversed-phase development, nonpolar compounds are the ones that usually remain near the origin. The need for polar eluents results in development times longer than those normally encountered with ordinary silica gel and less polar eluents.

Horizontal Development. Linear development can also be conducted from two directions simultaneously, with mobile phase entering from two opposing sides of a single plate. Such development requires the use of a horizontal developing chamber. The main advantage of the procedure, which was originally designed for use with 10×10 HPTLC layers, is that the number of samples per plate can be doubled. In closed systems and with small samples it is possible to separate in this way as many as 72 samples at the same time.

13.7. Visualization

Once a sample has been transferred to a TLC plate, all the components of that sample will be located somewhere on the layer. If the analyte is colored, localization is simple. Most organic compounds are not colored, however, so special techniques are required to change the materials in such a way that some optical contrast is established in the visible or the ultraviolet region. Reagents useful in color formation have been listed in [1].

A satisfactory visualization technique must take account of the following considerations:

1) Significant contrast is essential between the separated zone and the background.
2) A high degree of sensitivity is important to ensure detection of minute amounts of analyte.
3) Many compounds are not stable once separated on a sorbent; fluorescent compounds in particular are often unstable. This may not be a serious concern however, if only qualitative information is required. Nonetheless, the stability characteristics of the solutes should be ascertained.
4) A universal reagent capable of detecting all the solutes present is most desirable.
5) Satisfactory detection methods always result in well-defined zones of separation. Spraying or

dipping procedures can be problematic as a result of dispersion of the separated solutes.

6) An ideal detection method would permit quantitative analysis based on proportional absorbance intensities (see Chap. 13.8).

Methods for inducing detectability may be either *nondestructive* or *destructive*, where nondestructive methods are ones that do not destroy the analyte.

13.7.1. Nondestructive Visualization Methods

Ultraviolet absorbance can be utilized for detection purposes in several ways. For example, ultraviolet light often makes it possible to locate analytes separated on a fluorescent TLC plate. Thus, a UV-absorbing component may quench the background fluorescence, appearing as a dark spot against a luminescent field. Zinc silicate is a common phosphor in plates marked "F" for fluorescent, which fluoresce under the influence of light with a wavelength of 254 nm. Plastic fluorescent plates sometimes contain the fluorescent dye Rhodamine 6 G.

However, this fluorescence quenching is not a very sensitive detection method. As a rule it is recommended that visualization of the thin layer chromatogram take place in a dark room under the influence of lamps with emission maxima at 254 and 366 nm.

Inherent fluorescence can also be used for detecting separated zones, although the commonly available 254- and 366-nm lamps are suitable only for detecting compounds that fluoresce within the output range of the lamps themselves. Inherent fluorescence is the most sensitive aid to detection, occasionally revealing as little as 1 pg of an analyte.

Nondestructive detection is sometimes possible simply by spraying a plate with water. Free polar adsorbents (silica gel and alumina) are able to adsorb large amounts of water, whereas regions on the plate occupied by spots of organic material adsorb much less moisture, and appear as fatty zones against an almost translucent white background.

Detection with iodine vapor is one of the most widely applicable visualization techniques. If iodine crystals are placed in the bottom of a covered glass development chamber, within a few hours the atmosphere in the chamber will be saturated with the vapors. Analyte components on a thin layer chromatogram placed in such an iodine chamber tend to adsorb iodine, subsequently appearing as dark brown spots against a lighter tan background [10]. This method is characterized by either destructive or nondestructive detection depending on the analyte. After the developed layer in an iodine tank has been examined for evidence of reactive zones, the plate can be removed and areas of interest can be marked. Once the iodine has evaporated, analyte can sometimes be recovered for use in another procedure — provided it has not reacted with the iodine. Some analytes do undergo reaction with iodine, however, in which case the detection process is nonreversible. For example, phenolic compounds are known to react irreversibly under these conditions.

The iodine reaction can be surprisingly sensitive in some cases. Thus, morphine and codeine have been detected in this way at the 0.1-μg level. Iodine reagent can also be applied by spraying with a 1 % solution of iodine in an inert, volatile solvent such as hexane.

13.7.2. Destructive Visualization Methods

Charring is a destructive method that can be regarded as essentially universal. When an organic material on a TLC plate is heated in the presence of acid, an easily distinguishable black or brownish-black spot is formed. In many cases, quantitation of such a spot is possible by densitometry (see Chap. 13.8).

Acid solutions frequently used for TLC charring include:

sulfuric acid : acetic anhydride (1 : 3)
sulfuric acid : sodium dichromate (the well known "chromic acid cleaning agent")
sulfuric acid : nitric acid (1 : 1)
sulfuric acid : copper acetate or copper sulfate
phosphoric acid : copper sulfate or copper acetate

The plate under investigation is first treated with the acid solution and then heated. Charring is accomplished at 120–280 °C over a period of 1–40 min. Initial experiments should be conducted at 120 °C for 1 min, since this may produce dehydration products that fluoresce with different colors. For example, steroids generally produce fluorescent products when sprayed with 1 % sulfuric acid in ethanol followed by heating to as little as 100 °C. As noted previously, this type of rea-

gent has the inherent advantage of offering universal detection: most analytes produce some type of reaction, and if no reaction is initially apparent the sample on the layer can be heated to increasingly higher temperatures until charring occurs. The obvious disadvantage is destruction of the analyte (to say nothing of surrounding structures and clothing if care is not exercised!).

13.7.3. Reactive Visualization Methods

This category includes the preparation of derivatives, either before or after development of the chromatogram. It is sometimes feasible to improve the detectability of an analyte by subjecting it to a chemical reaction prior to chromatography. The separation that follows such a reaction is obviously not of the original materials, but rather of the reaction products. Confirmatory identification should be based on R_f values for both the unchanged and the reacted product. Reactive derivatization often has the additional advantage of improving the resolution of compounds that are otherwise difficult to separate.

13.7.4. Application of Visualization Reagents

The most common approach to application of a visualization reagent is probably spraying, using a spray gun powered either by compressed air from a laboratory line or compressed gas in a canister. The usual device incorporates a propellant container mounted on a screw cap that fits the top of a glass reagent jar. Reagent is drawn out of the jar by the venturi effect, and is dispersed as fine droplets in the emerging gas "cone." Spraying should be conducted in a fume hood to eliminate or minimize the exposure of personnel to reactive and toxic materials. Many distributers offer a wide variety of specially designed plate supports, spraying racks, and disposable or washable spray booths.

A TLC plate should be sprayed as evenly as possible, combining side-to-side and up-and-down motions of the spray equipment. Careful work is required to ensure uniform distribution of the sprayed reactant. Quantitation becomes particularly uncertain if spot intensity is dependent upon both the concentration of the spot component and the amount of reagent locally applied.

Application of reactive solutions is also possible by immersion of the TLC plate in some type of "dipping tank." The chief disadvantage of dipping techniques, especially with highly reactive agents, is that the back of the plate also becomes wet, and must be wiped in order to remove excess reagent. Care must also be taken when attempting to dip "soft" plates, since a soft coating is easily damaged by the forces associated with moving the plate in a surrounding liquid. One alternative is allowing the reactive reagent to ascend the plate by capillary action, just as in development.

If the need arises to spray a large number of plates it may be advantageous to consider setting up a *vapor charring* system, in which a plate is heated in a chamber through which SO_3 gas is passed, which in turn induces charring. The gaseous approach has the advantage that it distributes charring reagent evenly over the entire plate. The procedure utilizes a square Pyroceram dish approximately 10×10×25 cm in size and equipped with a curved glass cover. These are readily available as accessories for microwave cooking. The TLC plate is placed in the bottom of the vessel, adsorbent side up. A small amount of fuming sulfuric acid is then swabbed around the glass cover about 2.5–5 cm from the edge. The prepared cover is placed over the base, and the entire assembly is transferred to a hot plate or an oven. Appropriate conditions of time and temperature must be precisely established for each required analysis. It must also be emphasized that extreme precautions are necessary withrespect to the fuming acid, particularly when opening the vessel after charring. The advantages, of course, are very even charring and highly reproducible results.

Both Camag and Analtech also distribute sealed chambers commercially, utilizing ammonium hydrogencarbonate for visualization purposes. Vapors formed as the salt is heated tend to char or otherwise cause visual changes in a plate which is heated simultaneously. Many compounds form fluorescent derivatives under these conditions.

13.8. Quantitation

Modern approaches to *quantitative* TLC require the use of densitometry. Alternative visual quantitation techniques will not be discussed here, nor will elution techniques and the associated spectrophotometric methodology. For a discussion

in this area one should consult the literature, for instance the "Practice of Thin Layer Chromatography" by TOUCHSTONE [1] now in its third edition. The focus here will be densitometric methodology, which is clearly becoming the method of choice. For a detailed discussion of theoretical aspects of densitometry see POLLAK [11].

In situ methods based on densitometers and spectrodensitometers are now available as well, and these can be very precise, with relative standard deviations of less than 5%. Instruments currently on the market (1994) for in situ scanning of TLC layers include the following:

Ambias Inc., San Diego, Ca., USA	scanner for combined densitometry and isotope detection
Analtech Inc., Newark, Del., USA	one of the only video scanners designed for TLC
Camag, Muttenz, Switzerland	linear scanner used in many laboratories; equipped with a monochromator as a standard feature
Desaga GmbH, Heidelberg, FRG	linear scanner with monochromator
Helena Laboratories, Beaumont, Tex., USA	densitometer originally designed for electrophoresis scanning
Shimadzu Scientific Instruments, Columbia, Md. USA	linear scanner with monochromator

Video Scanners. Video scanners (such as the Uniscan device from Analtech) permit very rapid data acquisition. An entire TLC plate can beanalyzed at once with this device. Three detector modes are available: transmittance, reflectance, and fluorescence. The scanner incorporates its own computer with a video digitizer, light sources (visible and UV), and an imaging detector. Bandpass filters in front of the lens are used to select specific emission wavelengths, which can extend the usefulness of the instrument. In the case of fluorescent compounds separated by HPTLC, quantities in the low picogram range are readily assayed. For rapid quantitation in many regions, multichannel detector instruments offer a great advantage over conventional detection systems. Inexpensive instruments now permit digital storage of large amounts of data, opening the way to mathematical methods of data analysis. The most satisfactory instruments take advantage of a charge-coupled discharge video or standard camera, which can capture a complete image of a TLC layer in fractions of a second. The Analtech Uniscan comes with an enclosed view box featuring visible and short- or long-wave UV illumination. The image is projected on a monitor, and it is processed with the aid of a cursor for setting scan limits. Signals are processed in such a way as to provide both a chromatogram and tabulated quantitative data in the form of a printout. The quantitative information is calibrated on the basis of data from serial applications of standards that have been separated on the same chromatogram.

Linear Scanners. Linear scanners usually rely on a fixed light source under which the TLC layer must be moved. Changes in density are recorded by a computer programmed to convert the signal into a chromatogram, which is than evaluated on the basis of parameters so adjusted as to provide optimal scanning. Both Shimadzu and Camag market linear scanners of this type, which are also capable of determining absorption spectra of separated materials suitably positioned unter the light beam from a monochromator. The same instruments can be used to scan fluorescent compounds as well. Judicious use of filters can enhance the specificity of an analysis. For example, a 400-nm cut-off filter provides useful screening in the case of fluorescence analyses.

In *double-beam* operation the primary signals are amplified and fed directly into logarithmic converters. At the output, the reference signal is substracted from the sample signal, and the result is fed into an integrator. From there the signal passes to a conventional chart recorder that records the ratio of the two beam signals.

Important properties to consider in evaluating an instrument include the nature of the light source, the lines of highest emission energy, and the response curve of the detecting element, which may be either a photomultiplier or a cadmium sulfide diode. For example, the following spectral characteristics were observed with one instrument based on a xenon–mercury lamp (Hanovia 901.B-11) and photomultiplier detectors: Mercury bands from the xenon–mercury lamp were apparent at 312–313 nm and 365–366 nm, with the band at 365–366 nm by far the most intense. The lamp had an intensity curve rising gradually from 200 nm to a hump at ca. 500 nm, after which it leveled off. The photomultipliers showed a maximum response to light at 380 nm. This wavelength was preceded by a shoulder at the end of a steady rise. Above 380 nm the response declined gradually, becoming almost zero at 570 nm. The peaks of highest emission of the lamp coincided with the bands of greatest activation for the fluorescence of quinine sulfate separated on a silica-gel layer. This is significant, because it is usually possible to obtain greater sensitivity by using light

with a wavelength close to the lines of maximum emission of the lamp.

It is important to scan at the wavelength of maximum absorption of the solute in the sorbent, where the absorption maximum for a compound adsorbed on silica gel particles is typically not the same as that for the same compound in an alcohol solution. Precise and sensitive determination requires that the absorption maximum be determined in situ on the plate. This is possible with an instrument equipped with sources of variable wavelength or monochromators. Narrow-bandpass filters are available for instruments based instead on filters.

Single-beam recordings show the expected background variations, together with an increased noise level in both transmission and reflectance measurements. (The background is much smoother with a double-beam system.) Over short distances near the center of the plate the background variations and noise experienced in the course of reflectance measurements in single-beam operation may appear acceptable, but in reflectance mode the peak response for a given spot is substantially lower than in transmission mode, so the S/N value (i.e., precision) should be expected to decrease. Peak signals obtained in transmission are about five times as large as the corresponding signals in reflectance, whereas the noise level is only twice as great.

Errors in sample application are among the most significant sources of inaccuracy. The technique of spotting can have a marked effect on delivery volumes between 1 and 5 µL. A constant *starting zone size* should be maintained so that a given amount of substance applied in the form of different volumes will yield spots with identical areas, whereas different amounts spotted at constant volume will exhibit a linear relationship in a plot of peak area (from recordings of densitometer scans) versus mass.

Errors introduced during movement of the solute through the sorbent layer include lateral diffusion during development, and problems can also arise from structural variations for different plates or variations between tanks. Lateral diffusion is a function of the relative rates of diffusion of solute and mobile phase ($\rightarrow$ Basic Principles of Chromatography). Errors due to lateral diffusion when substantially different amounts of solute are applied next to each other can be prevented by scoring the plate into strips ca. 1 cm wide prior to spotting. In this way two dissimilar samples can safely be developed side by side. Variation in

structure from one layer to another leads to incorrect extrapolation of data obtained when a known quantity of standard is compared to the quantity of an unknown. The ratio between the quantity in a final spot and the quantity spotted varies for different layers, so it is recommended that a standard be applied to each test plate. Factors such as the spot size before and after development and the R_f value also become important in this context.

There are several possible sources of error in quantifying the results obtained from a densitometer or radiochromatogram scan. Important considerations include sampling methods, separation variables, variations in detector response, and calibration and measurement techniques.

Many of the errors in quantitation originate in the way a sample is applied to the layer, since this affects the shape of the spot after chromatogram development. The chromatographic process itself determines whether there will be tailing or overlap of the spots. High R_f values result in broad peaks that are not compatible with reproducible results from scanning in situ.

Another source of error is sample decomposition during chromatography. Evidence of decomposition can be obtained by developing chromatograms with standards of known serial dilutions. If the peak area does not increase linearly as a function of concentration, sample decomposition should be suspected. Errors due to tailing and overlap are quite common. These can be eliminated by changing the mobile phase to one that produces better separation.

The peak-area calibration procedure is usually less subject to influence from changes in instrumental parameters than calibration based on peak height. In addition, it offers improved results with peaks whose shapes are not Gaussian. However, results obtained with the peak-area quantitative method are adversely affected by neighboring peaks. Peak areas should always be used when precision is important.

The electronic integrators used in GC analysis are generally satisfactory for measuring peak areas in TLC scans as well. An electronic digital integrator measures peak areas automatically and converts them into numerical data, which are then printed. Sophisticated versions of such devices also correct for baseline drift.

Thin layer chromatography peak areas derived from scans can also be measured conveniently with computer systems developed for GC applications, often without any change. Most devices of this type are programmed to print out a complete

report, including compound names, retention times, peak areas, area correction factors, and weight-percent composition with respect to the various sample components.

13.9. References

[1] J. C. Touchstone: *Practice of Thin Layer Chromatography*, 3rd ed., Wiley, New York 1992.
[2] J. Sherma, B. Fried: *Handbook of Thin-Layer Chromatography*, Marcel Dekker, New York 1991.
[3] F. Geiss: *Fundamentals of Thin Layer Chromatography*, Hüthig-Verlag, New York 1987.
[4] T. E. Beesley in J. C. Touchstone, D. Rogers (eds.): *Thin Layer Chromatography, Quantitative Environ-*
mental and Clinical Applications, J. Wiley, New York 1980, p. 10.
[5] M. Zief, R. Riser, *Am. Lab. (Fairfield, Conn.)* **6** (1990) 70.
[6] D. E. Jaenchen, H. Issaq, *J. Liq. Chromatogr.* **11** (1988) 1941.
[7] L. R. Snyder: *Principles of Adsorption Chromatography*, Marcel Dekker, New York 1968.
[8] L. R. Snyder, J. J. Kirkland: *Introduction to Modern Liquid Chromatography*, Wiley, New York 1979.
[9] R. A. de Zeeuw, *J. Chromatogr.* **33** (1968) 222.
[10] G. C. Barrett in J. C. Giddings, R. A. Keller (eds): *Advances in Chromatography*, **vol. 11**, Marcel Dekker, New York 1974, p. 151.
[11] V. Pollak in J. C. Giddings et al. (eds.): *Advances in Chromatography*, Marcel Dekker, New York 1979, pp. 1–50.

14. Electrophoresis

PIER GIORGIO RIGHETTI, Department of Agricultural and Industrial Biotechnologies, University of Verona, I-37134 Verona, Italy

CECILIA GELFI, ITBA, CNR, L.I.T.A., Segrate 20090 (Milano), Italy

14.1. Introduction

The theory of electrophoresis has been adequately covered in the excellent textbooks of GIDDINGS [1] and ANDREWS [2] as well as in specific manuals [3], [4]. For discussion on electrophoresis in free liquid media, e.g., curtain, free-flow, endless belt, field-flow-fractionation, particle, and cell electrophoresis the reader is referred to a comprehensive review by VAN OSS [5] and to a book largely devoted to continuous-flow electrophoresis [6]. Here the focus is mostly on electrophoresis in a capillary support, i.e., in gel-stabilized media, and discussion is limited to protein applications.

Electrophoresis is based on the differential migration of electrically charged particles in an elec-

tric field. As such, the method is applicable only to ionic or ionogenic materials, i.e., substances convertible to ionic species (a classical example: neutral sugars, which form negatively charged complexes with borate). In fact, with the advent of capillary zone electrophoresis (CZE) it has been found that a host of neutral substances can be induced to migrate in an electric field by inclusion in charged micelles, e.g., of anionic (sodium dodecyl sulfate, SDS) or cationic (cetyltrimethylammonium bromide, CTAB) surfactants [7]. Even compounds that are not ionic, ionogenic, or complexable can often be analyzed by CZE as they are transported past the detector by the strong electrosmotic flow on the capillary walls [8].

Electrophoretic techniques can be divided into four main types: zone electrophoresis (ZE), mov-

ing-boundary electrophoresis (MBE), isotacho-phoresis (ITP), and isoelectric focusing (IEF). This classification is based on two criteria: (1) initial component distribution (uniform vs. sharply discontinuous); and (2) boundary permeability (to all ions vs. only to H^+ and OH^-). Alternatively, electrophoretic techniques may be enumerated in chronological order, as follows: moving-boundary electrophoresis, zone electrophoresis, disc electrophoresis, isoelectric focusing, sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE), two-dimensional (2-D) maps, isotachophoresis (ITP), staining techniques, immobilized pH gradients (IPG), and capillary zone electrophoresis.

Zone electrophoresis became practicable when hydrophilic gels (acting as an anticonvective support) were discovered. GRABAR and WILLIAMS in 1953 first proposed the use of an agar matrix (later abandoned in favor of a highly purified agar fraction, agarose) [9]. They also combined, for the first time, electrophoresis on a hydrophilic support with biospecific detection (immunoelectrophoresis). Barely two years after that, SMITHIES [10] applied another gel, potato starch. The starch blocks were highly concentrated matrices (12–14% solids) and introduced a new parameter in electrophoretic separation: molecular sieving. Human sera, which in cellulose acetate or paper electrophoresis were resolved into barely five bands, now produced a spectrum of 15 zones.

14.2. Basic Principles

In a free buffer solution, in the absence of molecular sieving, the velocity v of a particle is proportional to the field strength E, multiplied by its electrophoretic mobility μ:

$$v = \mu E$$

The velocity and E are both vectors, while μ is a scalar, being positive for cations and negative for anions. A distinction is made between absolute mobility (the μ value of species at infinite dilution and complete ionization) and effective mobility (the μ value of species under the actual experimental conditions, e.g., as a result of sieving in gel matrices, or in the presence of a strong electrosmotic flow). In deriving an expression for μ which takes into account parameters such as charge, size, and solvent effects, one can consider the different

forces acting on a species. At infinite dilution, the electric field is exerting a force on the charge Q of the particle and this force is opposed by the frictional force, given by the friction factor f_c multiplied by the velocity v. At steady state:

$$Q E = f_c v$$

For a spherical particle obeying Stokes' law, f_c is given by:

$$f_c = 6\pi\eta r$$

where η is the solvent viscosity and r the particle radius. By substituting and rearranging one can derive the expression for the absolute mobility:

$$\mu = v/E = Q/(6\pi\eta r)$$

According to the Debye–Hückel theory, the total charge Q of a particle is correlated to the dielectric constant ε and to the zeta potential by the expression:

$$Q = \varepsilon r \zeta$$

whereby the expression for μ follows:

$$\mu = \varepsilon\zeta/(6\pi\eta)$$

The mobility is directly proportional to the dielectric constant and zeta potential and inversely related to the solvent viscosity. Thus, it follows that μ is strongly reduced in organic solvents, is zero at the isoelectric point of an amphoteric species (where the net charge is zero), and is greatly reduced in any form of gel media, which can be viewed as highly viscous solutions.

14.3. Electrophoretic Matrices

Only three matrices are in use today: cellulose acetate, agarose, and polyacrylamide.

14.3.1. Cellulose Acetate

Cellulose acetate is a cellulose derivative in which each hexose ring of the polysaccharide chain has two hydroxyl groups esterified to acetate (in general, in the C-3 and C-6 positions). This medium was developed for electrophoresis largely through the work of KOHN [11], who also designed the original tank, produced by Shandon Southern

Instruments as Kohn's Electrophoresis Apparatus Model U77 and imitated throughout the world.

Cellulose acetate is still very popular in clinical chemistry, as it offers a ready-made support that can be equilibrated with buffer in a few seconds and produces good separations of proteins from biological fluids in ca. 20 min. Completely automatic systems (for sample loading, staining, destaining, and densitometry) have been built around it, so that this system has become the first example of combined electrophoresis and robotics, an integration which had its testing ground and underwent development in clinical chemistry and is now spreading to basic research laboratories.

Cellulose acetate allows migration of even large serum proteins, such as pre-β-lipoprotein ($M_r > 5 \times 10^6$) and is amenable to immunofixation in situ [12], [13] without any need for prior transfer by blotting (as required, e.g., after SDS-PAGE). In addition, the immunoprecipitate, obtained directly on the cellulose acetate, can be enhanced by use of a second antibody, conjugated with, e.g., alkaline phosphatase [14]. Moreover, owing to their large porosity and their availability as thin films, cellulose acetate matrices can be stained with gold micelles, increasing the detection sensitivity to the subnanogram range [15], whereas agarose and polyacrylamide are incompatible with this gilding process [16]. However, the use of cellulose acetate is limited to clinical electrophoresis and is not much in vogue in basic research for high-resolution runs, even though focusing and two-dimensional techniques on cellulose acetate membranes have been described [17].

14.3.2. Agarose Gels

Agarose is a purified linear galactan hydrocolloid, isolated from agar or recovered directly from agar-bearing marine algae, such as the Rhodophyta. Genera from which agarose is extracted include *Gelidium, Gracilaria, Acanthopeltis, Ceramium, Pterocladia,* and *Campylaephora.* Agar is extracted from red algae by boiling in water and is then isolated by filtering off the particulates, gelling and freeze-thawing the colloid to remove water-soluble impurities, and precipitation with ethanol [18]. The resultant product is a mixture of polysaccharides, which are alternating copolymers of 1,4-linked 3,6-anhydro-α-L-galactose and 1,3-linked β-D-galactose. The repeating disaccharide is referred to as agarobiose [19]. Of course, this is an idealized structure, since even purified

agarose contains appreciable amounts of the following substituents: sulfate, pyruvate, and methoxyl groups. Sulfate is generally esterified to the hydroxyl at C-4 of β-D-galactopyranose, and C-2 and C-6 of 3,6-anhydro-α-L-galactose. Pyruvate is linked to both C-4 and C-6 of some residues of β-D-galactopyranose. The resulting compound is referred to as 4,6-O-carboxyethylidene [20]. In addition, the hydroxyl group on C-6 of some residues of β-D-galactopyranose can also be methylated.

Agarose was first used to form gels for electrophoresis by HJERTÉN [21] and also for immunodiffusion [22] and chromatography [23]. Agarose can be separated from agaropectin by several procedures, including preferential precipitation of agaropectin by a quaternary ammonium salt, followed by fractionation of agarose with polyethylene glycol. By using a combination of polyethylene glycol precipitation and anion-exchange chromatography, an agarose with no detectable pyruvate and 0.02 % sulfate has been prepared [24]. A further treatment consists of alkaline desulfation in the presence of $NaBH_4$ followed by reduction with $LiAlH_4$ in dioxane [25]. When an agarose solution is cooled, during gelation, single strands of agarose (average molecular mass 120 kDa or ca. 392 agarobiose units) [26] dimerize to form a double helix. The double helix has a pitch of 1.9 nm, and each strand has threefold helical symmetry [27]. However, formation of the double helix is not sufficient to produce a gel; it is assumed that these helices aggregate laterally during gelation to form suprafibers (Fig. 1). In fact, the width of an agarose double helix is no more than 1.4 nm, whereas in gelled agarose, fibers one order of magnitude wider than this have been observed by electron microscopy and light scattering. The greater the number of aggregated double helices (i.e., the thicker the diameter of the "pillar"), the larger is the "pore" defined by the supporting beam [28]. Thus, it is reasonable to assume that it is the presence of these suprafibers that causes agarose gels to have strengths and pore sizes greater than those of gels that do not form suprafibers, such as starch and 2.5–5 % cross-linked polyacrylamide gels. In electrophoresis, agarose gels are used for fractionation of several types of particles, including serum lipoproteins [29], nucleic acids [30], virus and related particles [31], and subcellular organelles, such as microsomes. However, a suspension of 0.14 % agarose particles is used for the latter, rather than a coherent gel mass [32]. More recently, by ex-
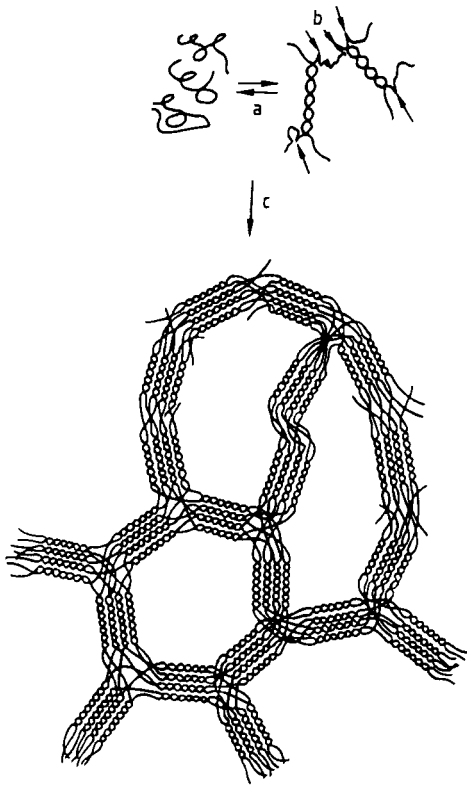
Figure 1. Formation of an agarose gel
Strands of agarose join to form double helices (a). At kinks (b) the strands exchange partners. Higher-order aggregation results in suprafibers (c). In hydroxyethylated agarose, it appears that the pillar looses its structure and more of the double helices are dispersed into the surrounding space: this results in a decrease in average pore size, i.e., higher sieving for the same % matrix as compared with underivatized agarose [31].

ploiting ultradilute gels (barely 0.03 % agarose, a special high-strength gel, Seakem LE, from Marine Colloids, Rockwille, ME, United States) SERWER et al. [33] have even been able to fractionate *E. coli* cells. Previously, this had been thought to be impossible, and in fact RIGHETTI et al. [34] reported a maximum pore diameter, for even the most dilute agaroses (0.16 %) of 500 – 600 nm. According to SERWER et al. [33] the effective pore radius P_E of a 0.05 % agarose gel is 1.2 μm; hence, the average pore of a 0.03 % gel should be several micrometers wide.

FMC BioProducts (formerly Marine Colloids, a division of FMC Corporation) of Rockland, ME, United States, has produced a number of highly sophisticated agaroses to meet the diverse require-

ments of electrophoresis users; the reader is referred to an extensive review on this topic [35]. Mixed-bed gels, containing 0.8 % agarose and variable amounts (2.5 – 9 %T) of acrylamide (%T indicates the total amount of monomers) have also been produced [36], [37]. Such gels were quite popular in the 1970s for nucleic acid fractionation, as they allowed fine resolution of lower M_r DNAs and RNAs, but they are seldom used in modern electrokinetic methods. However, these gels (under the trade name Ultrogels) in a spherical form are still popular for chromatographic purposes. Five types of bead exist, with a fractionation range of 1 – 1200 kDa [38].

14.3.3. Polyacrylamide Gels

The most versatile of all matrices are the polyacrylamide gels. Their popularity stems from several fundamental properties: (1) optical clarity, including ultraviolet (280 nm) transparency; (2) electrical neutrality, due to the absence of charged groups; and (3) availability in a wide range of pore sizes. Their chemical formula, as commonly polymerized from acrylamide and N,N'-methylenebisacrylamide (Bis), is shown in Figure 2, together with that of the two most widely employed catalysts, persulfate (ammonium or potassium) and $N,N,N'N'$-tetramethylethylenediamine (TEMED) [39]. Four groups simultaneously and independently developed polyacrylamide gel electrophoresis: RAYMOND and WEINTRAUB [40], DAVIS [41], ORNSTEIN [42], and HJERTÉN [43]. In addition, HJERTÉN [44] introduced polyacrylamide beads as a matrix for chromatography. Normal polyacrylamide gels (i.e., cross-linked with standard amounts of 3 – 5 % Bis) have porosities that decrease linearly with the monomer content (%T). Their porosities are substantially smaller than those of corresponding agarose gels. However, highly cross-linked gels (> 30 %C) can have porosities comparable to or even greater than those of agarose matrices (%C is the mass in grams of cross-linker per 100 g total monomers) [45], [46]. However, high-C polyacrylamide matrices are brittle and opaque, and tend to collapse and exude water above 30 %C [34]. Thus, at present their use is limited.

Hydrophilic gels are considered to be a network of flexible polymer chains with interstices into which macromolecules are forced to migrate by an applied potential difference, according to a partition governed by steric factors. Large mole-

Tetramethylethylenediamine   Acrylamide   Bisacrylamide
(TEMED)
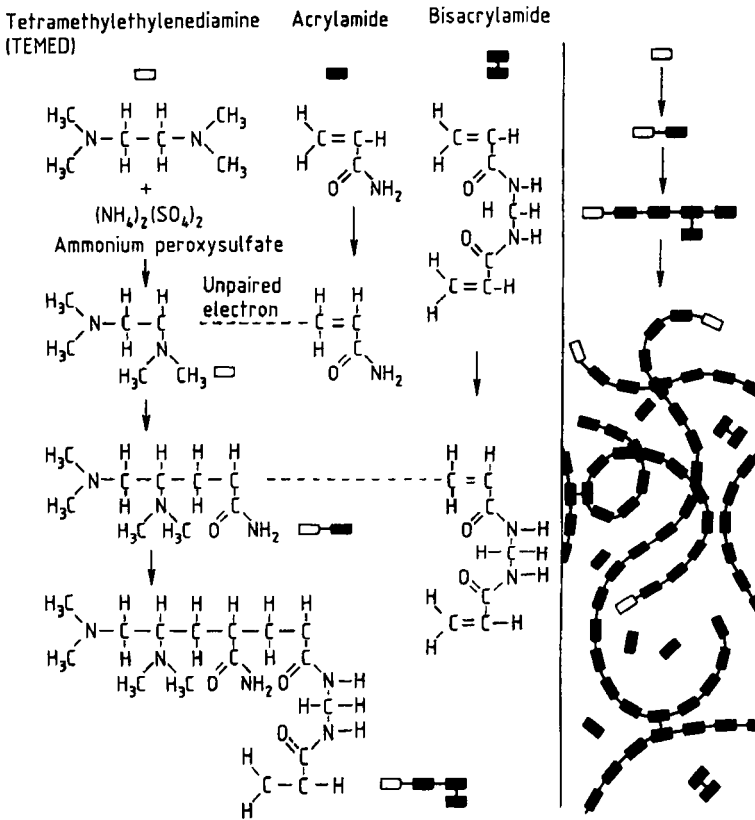
Figure 2. The polymerization reaction of acrylamide
The structure of acrylamide, *N,N'*-methylenebisacrylamide (Bis), and of a representative segment of cross-linked polyacrylamide are shown. Initiators shown are peroxysulfate and *N,N,N',N'*-tetramethylethylendiamine (TEMED).

cules can only penetrate into regions where the meshes in the net are large, while small molecules find their way into tightly knit regions of the network. Perhaps the best way to envision a hydrophilic matrix is to consider it as being composed of two interlaced fluid compartments, with different frictional coefficients [47]. In general, agaroses are more porous than polyacrylamides and these two matrices are used to complement each other.

The versatility of polyacrylamide gels is also shown by the large number of cross-linkers, besides Bis, that can be used to cast gels with particular properties for different fractionation purposes. *N,N'*-(1,2-dihydroxyethylene)bisacrylamide (DHEBA) can be used for casting reversible gels (i.e., gels that can be liquefied), since the 1,2-diol structure of DHEBA renders them susceptible to cleavage by oxidation with periodic acid [48]. The same principle should also apply to *N,N'*-dial-

lyltartardiamide (DATD) gels [49]. Alternatively, ethylenediacrylate (EDA) gels [50] may be used, since this cross-linker contains ester bonds that undergo base-catalyzed hydrolytic cleavage. The poly(ethylene glycol) diacrylate cross-link belongs to the same series of ester derivatives [34]. As a further addition to the series, *N,N'*-bisacrylylcystamine (BAC) gels, which contain a disulfide bridge cleavable by thiols, have been described [51]. Practically any cross-linker can be used, but DATD and *N,N',N''*-triallylcitrictriamide (TACT) should be avoided, since, being allyl derivatives, they are inhibitors of gel polymerization when mixed with compounds containing acrylic double bonds. Their use at high %C is simply disastrous [52], [53].

Other monomers, in addition to acrylamide, have been described. One of them is Trisacryl, *N*-acryloyl-2-amino-2-hydroxymethyl-1,3-pro-

panediol. The Trisacryl monomer creates a micro-environment that favors the approach of hydrophilic solutes (proteins) to the gel polymer surface, since the polyethylene backbone is buried underneath a layer of hydroxymethyl groups. In addition, owing to the much larger M_r of the monomer, as compared with acrylamide, gels made with the same nominal matrix content (%T) should be more porous, since they would have on average thicker fibers than the corresponding polyacrylamide matrix. This type of gel is extensively used by IBF Reactifs (VilleneuveLa-Garenne, France) for preparing gel filtration media or ion exchangers and also for electrophoresis [54], [55]. However, this monomer degrades with zero-order kinetics and hence seems to be of limited use in electrophoresis [56].

In 1984, ARTONI et al. [57] described two novel monomers that impart peculiar properties to polyacrylamide-type gels: acryloylmorpholine (ACM) and bisacrylylpiperazine (BAP). These gels exhibited some interesting features: owing to the presence of the morpholino ring on the nitrogen atom involved in the amido bond, the latter was rendered stable to alkaline hydrolysis, which bedevils conventional polyacrylamide matrices. In addition, such matrices are fully compatible with a host of hydro-organic solvents, thus allowing electrophoresis in mixed phases. The use of the BAP cross-linker has been reported to have beneficial effects when silver staining polyacrylamide matrices.
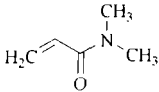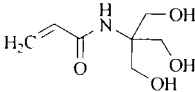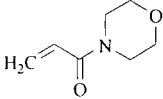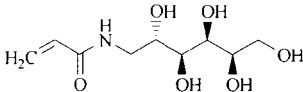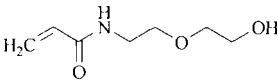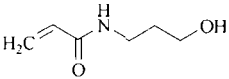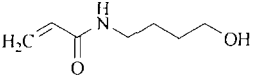
More recently, novel monomers combining high hydrophilicity with extreme hydrolytic stability have been described. One of them is N-acryloylaminoethoxyethanol (AAEE). As a free monomer, it has ten times higher resistance to hydrolysis than plain acrylamide, but as a polymer its stability (in 0.1 N NaOH, 70 °C) is 500 times higher [58]. The strategy: the ω-OH group in the N-substituent was sufficiently remote from the amido bond so as to impede formation of hydrogen bonds with the amido carbonyl group; if at all, the oxygen in the ethoxy moiety of the N-substituent would act as a preferential partner for hydrogen-bond formation with the ω-OH group. The second is acryloylaminopropanol (AAP), in which the ether group on the N-substituent was removed and the chain shortened [59]. As a result, this monomer was even more hydrophilic than AAEE and highly resistant to hydrolysis as compared to free acrylamide (hydrolysis constant of AAP 0.008 L mol^{-1} min^{-1} versus 0.05 L mol^{-1} min^{-1} for acrylamide, in an alkaline

milieu). The AAP monomer was found to have a unique performance in DNA separations when used as a sieving liquid polymer in capillary zone electrophoresis (CZE) [60]. In addition, poly(AAP) has excellent performance in CZE at high temperatures (60 °C and higher), which are necessary for screening for DNA point mutations in temperature-programmed CZE [61]. It is known that acrylamide is a neurotoxin and possibly even (at high doses) a carcinogen. Now the situation could have dramatically changed: if we assess toxicity in terms of the ability of these monomers to alkylate free SH groups in proteins (thus forming a cysteinyl-S-β-propionamide adduct) and evaluate this reaction by high-resolution, delayed extraction MALDI-TOF MS (matrix-assisted laser desorption ionization time of flight mass spectrometry) both in the linear and reflectron mode, this picture will change substantially: at one end of the scale the major offenders remain indeed acrylamide and DMA (strongly alkylating, hence highly toxic agents), whereas at the opposed end (minimally alkylating) one can locate only AAP [62]. Tables 1 and 2 list a number of monomers and cross-linkers, respectively, in use today.

14.4. Discontinuous Electrophoresis

In 1959 RAYMOND and WEINTRAUB [40] described the use of polyacrylamide gels (PAG) in ZE, offering UV and visible transparency (starch gels are opalescent) and the ability to sieve macromolecules over a wide range of sizes. In 1964, ORNSTEIN [42] and DAVIS [41] created discontinuous (disc) electrophoresis by applying a series of discontinuities to PAG (leading and terminating ions, pH, conductivity, and porosity), thus further increasing the resolving power of the technique. In disc electrophoresis (Fig. 3) the proteins are separated on the basis of two parameters: surface charge and molecular mass. FERGUSON [63] showed that one can distinguish between the two by plotting the results of a series of experiments with polyacrylamide gels of varying porosity. For each protein under analysis, the slope of the curve log m_T (electrophoretic mobility) versus gel density (%T) is proportional to molecular mass, while the y-intercept is a measure of surface charge [64]. Nonlinear Ferguson plots have been reported [65], related to the reptation mode of DNA in sieving media.

Table 1. Monomers utilized for preparing polyacrylamide gels

Name	Chemical formula	M_r
N,N-Dimethylacrylamide (DMA)		99
N-Acryloyl-2-amino-2-hydroxy-methyl-1,3-propane (Trisacryl)		175
N-Acryloylmorpholine		141
N-Acryloyl-1-amino-1-deoxy-D-glucitol		235
N-Acryloylaminoethoxyethanol		159
N-Acryloylaminopropanol		129
N-Acryloylaminobutanol		143

14.5. Isoelectric Focusing

New routes were explored toward a simpler assessment of charge and mass of proteins. As early as 1961 SVENSSON [66] *experimented with* isoelectric focusing (IEF), an electrophoretic technique by which amphoteric compounds are fractionated according to their isoelectric point (pI) along a continuous pH gradient. In contrast to ZE, where the constant pH of the separation medium establishes a constant charge density at the surface of the molecule and causes it to migrate with constant mobility (in the absence of sieving), the surface charge of an amphoteric compound in IEF keeps changing and decreasing according to its titration curve, as it moves along the pH gradient toward its steady-state position, i.e., the region where the pH matches its pI. There its mobility is zero and the molecule comes to a stop.

The gradient is created and maintained by the passage of current through a solution of amphoteric compounds with closely spaced pI values, encompassing a given pH range. The electrophoretic transport causes these buffers (carrier ampho-

Table 2. Cross-linkers described for gelling polyacrylamide matrices
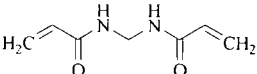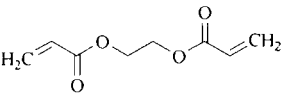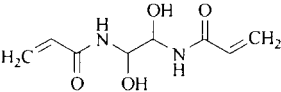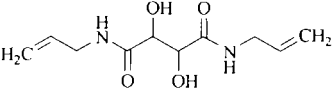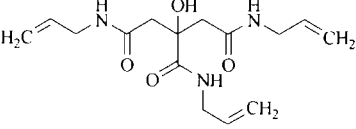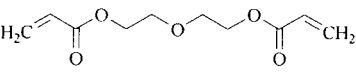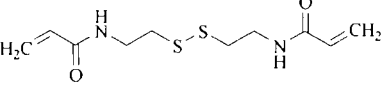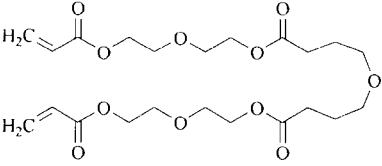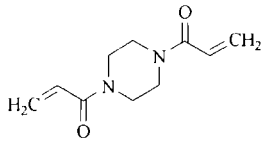
Name	Chemical formula	M_r	n^*
N,N'-Methylene bisacrylamide (Bis)		154	9
Ethylene diacrylate (EDA)		170	10
N,N'-(1.2-Dihydroxyethylene) bisacrylamide (DHEBA)		200	10
N,N'-Diallyltartardiamide (DATD)		228	12
N,N',N''-Triallyl citric triamide (TACT)		309	12 – 13
Polyethylene glycol diacrylate 200 (PEGDA$_{200}$)		214	13
N,N'-Bisacrylylcystamine (BAC)		260	14
Polyethylene glycol diacrylate 400 (PEGDA$_{400}$)		400	25

Table 2. (continued)

Name	Chemical formula	M_r	$n*$
N,N'-Bis-acrylylpipera-zine (BAP)		194	10

* Chain length.
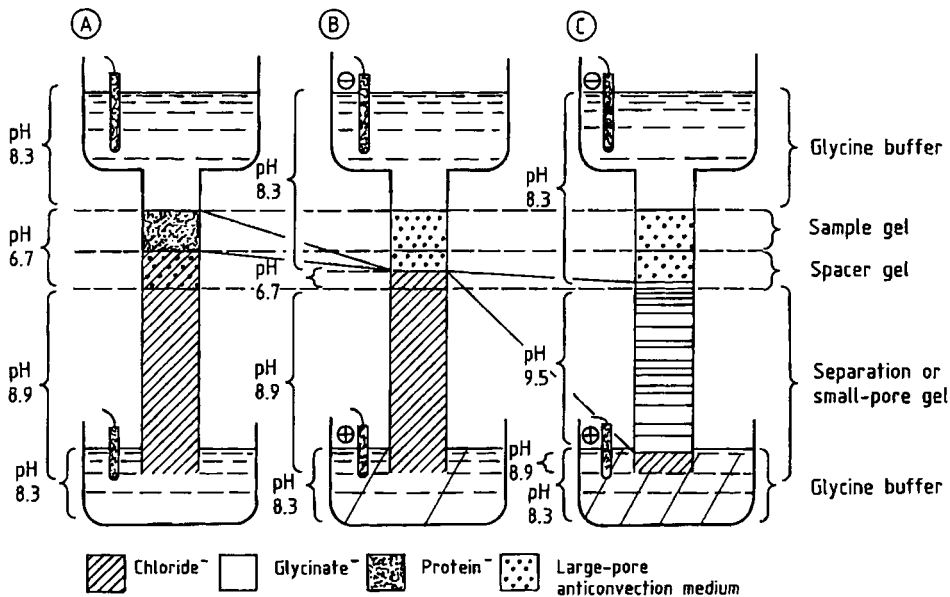
Figure 3. Principle of discontinuous electrophoresis
A) Sample in sample gel; B) Sample concentration in stacking gel; C) Sample separation in small-pore, separation gel [42]

lytes, CA) to stack according to their pI values, and a pH gradient, increasing from anode to cathode, is established. This process is shown in Figure 4, which gives the formulae of the CA buffers and portions of their titration curves in the proximity of the pI value. When the system has achieved a steady state, maintained by the electric field, no further mass transport is expected except from symmetric to-and-fro micromovements of each species about its pI, generated by the action of two opposite forces, diffusion and voltage gradient, acting on each focused component. Thus, this diffusion–electrophoresis process is the primary cause of the residual current under isoelectric steady-state conditions.

The technique applies only to amphoteric compounds and, more precisely, to good ampholytes with small $|pI-pK_1|$ values, i.e., with a steep titration curve around their pI (Fig. 4): the essential condition for any compound to focus in a narrow band. In practice, notwithstanding the availability of CAs covering the pH range 2.5 – 11, the limit of CA-IEF is in the pH 3.5 – 10 range, due to solvent conductivity outside this interval. When a restrictive support, such as polyacrylamide, is used, a limit also exists for the size of the largest molecules exhibiting an acceptable mobility through the gel. A conservative evaluation sets an upper limit around 750 kDa for standard techniques. The molecular form in which the proteins are separated
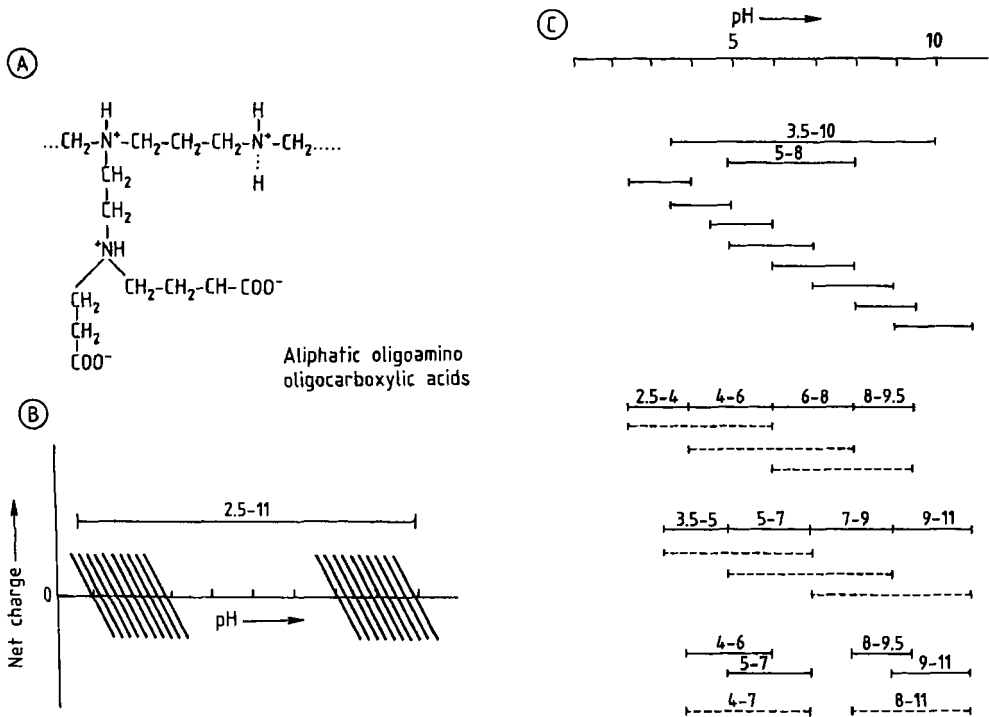
Figure 4. Composition of carrier ampholytes for isoelectric focusing
A) Representative chemical formula (oligoamino oligocarboxylic acids); B) Portions of titration curves of carrier ampholytes near pI; C) Narrow pH intervals, obtained by subfractionation of the wide pH 3.5–10 gradient components

depends strongly on the presence of additives, such as urea and/or detergents. Moreover, supra-molecular aggregates or complexes with charged ligands can be focused only if their K_d is lower than 1 μmol/L and if the complex is stable at pH = pI. Aggregates with higher K_d are easily split by the pulling force of current, whereas most chromatographic procedures are unable to modify the native molecular form.

The general properties of the carrier ampholytes, i.e., of the amphoteric buffers used to generate and stabilize the pH gradient in IEF, can be summarized as follows:

1) Fundamental "classical" properties: (a) zero buffering ion mobility at pI; (b) good conductance; (c) good buffering capacity
2) Performance properties: (a) good solubility; (b) no influence on detection system; (c) no influence on sample; (d) separable from sample
3) "Phenomena" properties: (a) plateau effect, drift of the pH gradient; (b) chemical change in the sample; (c) complex formation

In chemical terms, CAs are oligoamino, oligocarboxylic acids, available from different suppliers under different trade names (Ampholine from LKB Produkter, Pharmalyte from Pharmacia Fine Chemicals, Biolyte from Bio-Rad, Servalyte from Serva, Resolyte from BDH). There are two basic synthetic approaches: (1) the Vesterberg approach [67], allowing different oligoamines (tetra-, penta-, and hexa-amines) to react with acrylic acid; and (2) the Pharmacia synthetic process [68], which involves the copolymerization of amines, amino acids, and dipeptides with epichlorohydrin. The wide-range synthetic mixture (pH 3–10) seems to contain hundreds, possibly thousands of different amphoteric chemicals, having pI values evenly distributed along the pH scale. CAs, from any source, should have an average molecular mass around 750 Da (range 600–900 Da, the higher values referring to the more acidic CA species). Thus, they should be easily separable from macromolecules by gel filtration, unless they are hydrophobically complexed to proteins. A further complication arises from the chelating effect

of acidic CAs, especially toward Cu^{2+} ions, which could inactivate metalloenzymes. In addition, focused CAs represent a medium of very low ionic strength (less than 1 mequiv./L at steady state) [3], [69]. Since the isoelectric state involves a minimum of solvation and thus of solubility for the protein macroion, there could be a tendency for some proteins (e.g., globulins) to precipitate during IEF close to the pI. This is a severe problem in preparative IEF, but in analytical procedures it can be lessened by reducing the total amount of sample applied.

14.6. Sodium Dodecyl Sulfate Electrophoresis

Sodium dodecyl sulfate (SDS) electrophoresis fractionates polypeptide chains essentially on the basis of their size. It is therefore a simple, yet powerful and reliable method for molecular mass determination. In 1967 SHAPIRO et al. [70] first reported that electrophoretic migration in SDS is proportional to the effective molecular radius and, thus, to the M_r of the polypeptide chain. This means that SDS must bind to proteins and cancel out differences in molecular charge, so that all components migrate solely according to size. Surprisingly large amounts of SDS appear to be bound (an average of 1.4 g/g protein). This means that the number of SDS molecules bound is approximately half the number of amino acid residues in a polypeptide chain. This amount of highly charged surfactant molecules is sufficient to overwhelm effectively the intrinsic charges of the polymer coil, so that their net charge per unit mass becomes approximately constant. If migration in SDS (and disulfide reducing agents, such as 2-mercaptoethanol, in the denaturing step, for proper unfolding of the proteins) is proportional only to M_r, then, in addition to canceling out charge differences, SDS also equalizes molecular shape differences (e.g., globular vs. rod-shaped molecules). This seems to be the case for protein – SDS mixed micelles. These complexes can be assumed to behave as ellipsoids of constant minor axis (ca. 1.8 nm) and a major axis proportional to the length of the amino acid chain (i.e., molecular mass) of the protein. The rod length for the 1.4 g/g protein complex is of the order of 0.074 nm per amino acid residue. For further information on detergent properties, see HELENIUS and SIMONS [71].

In SDS electrophoresis, the proteins can be prelabeled with dyes that covalently bind to their NH_2 residues. The dyes can be conventional (e.g., the blue dye Remazol) or fluorescent, such as dansyl chloride, fluorescamine, O-phthaldialdehyde, and MDPF (2-methoxy-2,4-diphenyl-3[2H]-furanone). Prelabeling is compatible with SDS electrophoresis, as the size increase is minimal, but would be disastrous in disc electrophoresis or IEF, as it would generate a series of bands of slightly altered mobility or pI from an otherwise homogeneous protein.

For data treatment the sample and M_r standards are electrophoresed side-by-side in a gel slab. After detection of the polypeptide zones, the migration distance (or R_F) is plotted against log M_r to produce a calibration curve [72] from which the M_r of the sample can be calculated. Note that, in a gel of constant %T, linearity is obtained only in a certain range of molecular size. Outside this range a new gel matrix of appropriate porosity should be used. Two classes of protein show anomalous behavior in SDS electrophoresis: glycoproteins (because their hydrophilic oligosaccharide units prevent hydrophobic binding of SDS micelles) and strongly basic proteins, e.g., histones (because of electrostatic binding of SDS micelles through their sulfate groups). The first can be partially alleviated by using alkaline Tris/borate buffers [73], which will increase the net negative charge on the glycoprotein and thus produce migration rates well correlated with molecular size. The migration of histones can be improved by using pore gradient gels and allowing the polypeptide chains to approach the pore limit [74].

14.7. Porosity Gradient Gels

When macromolecules are electrophoresed in a continuously varying matrix concentration (which results in a porosity gradient), rather than in a gel of uniform concentration, the protein zones are compacted along their track, because the band front is, at any given time, at a gel concentration somewhat higher than at the rear of the band, so that the former is decelerated continuously. A progressive band sharpening thus results. There are other reasons for resorting to gels of graded porosity. Disc electrophoresis separates macromolecules on the basis of both size and charge differences. If the influence of molecular charge could be eliminated, then clearly the

method could be used, with a suitable calibration, for measuring molecular size. This has been accomplished by overcoming charge effects in two main ways. In one, a relatively large amount of charged ligand, such as SDS, is bound to the protein, effectively swamping the initial charges present on the protein molecules and giving a quasi-constant charge/mass ratio. However, in SDS electrophoresis proteins are generally dissociated into their constituent polypeptide subunits, and the concomitant loss of functional integrity and antigenic properties cannot be prevented. Therefore, the size of the original, native molecule must be evaluated in the absence of denaturing substances.

The second method for M_r measurements relies on a mathematical canceling of charge effects, following measurements of the mobility of native proteins in gels of different concentrations. This is the so-called Ferguson plot [63], discussed in Chapter 14.4. A third method for molecular size measurements uses gels of graded porosity. This method is characterized by high resolving power and relative insensitivity to variations in experimental conditions. Under appropriate conditions (at least $10 \, \text{kV} \cdot \text{h}$), the mobility of most proteins becomes constant and eventually ceases as each constituent reaches a gel density region in which the average pore size approaches the diameter of the protein (pore limit) [75]. Thus, the ratio of the migration distance of a given protein to that of any other becomes constant after the proteins have all entered a gel region in which they are subjected to drastic sieving conditions. This causes the electrophoretic pattern to become constant after prolonged migration in a gel gradient. The gel concentration at which the migration rate for a given protein becomes constant is called the "pore limit": If this porosity is properly mapped with the aid of a suitable set of marker proteins, it is possible to correlate the migration distance to the molecular mass of any constituent in the mixture.

After electrophoresis is over and the proper experimental data are gathered, they can be handled by two-step or one-step methods. Among the former, the most promising approach appears to be that of LAMBIN and FINE [76], who observed that there is a linear relationship between the migration distance of proteins and the square root of electrophoresis time, provided that time is kept between 1 and 8 h. The slopes of the regression lines of each protein are an indication of molecular size. When the slopes of the various regression lines thus obtained are plotted against the respective

molecular masses, a good linear fit is obtained, which allows molecular mass measurements of proteins between 2×10^5 and 10^6 Da. The shape of the proteins (globular or fibrillar), their carbohydrate content (up to 45 %), and their free electrophoretic mobilities (between 2.1 and $5.9 \times 10^{-5} \, \text{cm}^2 \, \text{V}^{-1} \, \text{s}^{-1}$) do not seem to be critical for proper M_r measurements by this procedure. One-step methods have been described by ROTHE and PURKHANBABA [77] who found that, when $\log M_r$ is plotted against either D (distance migrated) or %T (acrylamide + Bis), a nonlinear correlation is always obtained. However, when $\log M_r$ is plotted against $\sqrt{\%T}$ or $\sqrt{D}$, a linear regression line is obtained, which allows the accurate determination of M_r values of proteins (standard deviation $\pm 3.7 \%$; Fig. 5). The correlations $\log M_r - \sqrt{\%T}$ or $\log M_r - \sqrt{D}$ are not significantly altered by the duration of electrophoresis. Therefore, a constant M_r value should be obtained for a stable protein, no matter how long electrophoresis has been going on. ROTHE and MAURER [78] have demonstrated that the relationship $\log M_r - \sqrt{D}$ is also applicable to SDS electrophoresis in linear polyacrylamide gel gradients.

14.8. Two-Dimensional Maps (Proteome Analysis)

By sequentially coupling pure charge (IEF) to pure size fractionation (SDS-PAGE; the latter orthogonal to the first), one can distribute the polypeptide chains two-dimensionally, with charge and mass as coordinates (IEF-SDS or ISO-DALT, according to ANDERSON's nomenclature: ISOelectric for charge and DALT for mass separation) [79]. When the first dimension is performed in immobilized pH gradients, the technique is called IPG-DALT [4]. The technique was first reported by BARRETT and GOULD [80] and described in more detail by O'FARRELL [81], KLOSE [82], and SCHEELE [83]. Large gels (e.g., 30×40 cm) [84] and prolonged exposure to radiolabeled material (up to two months) have allowed resolution of as many as 12 000 labeled peptides in a total mammalian cell lysate. Thus, there is a good chance that, in a properly prepared 2-D map, a spot will represent an individual polypeptide chain, uncontaminated by other material. On this assumption, and provided that enough material is present in an individual spot (about 1 µg), it is possible to blot it
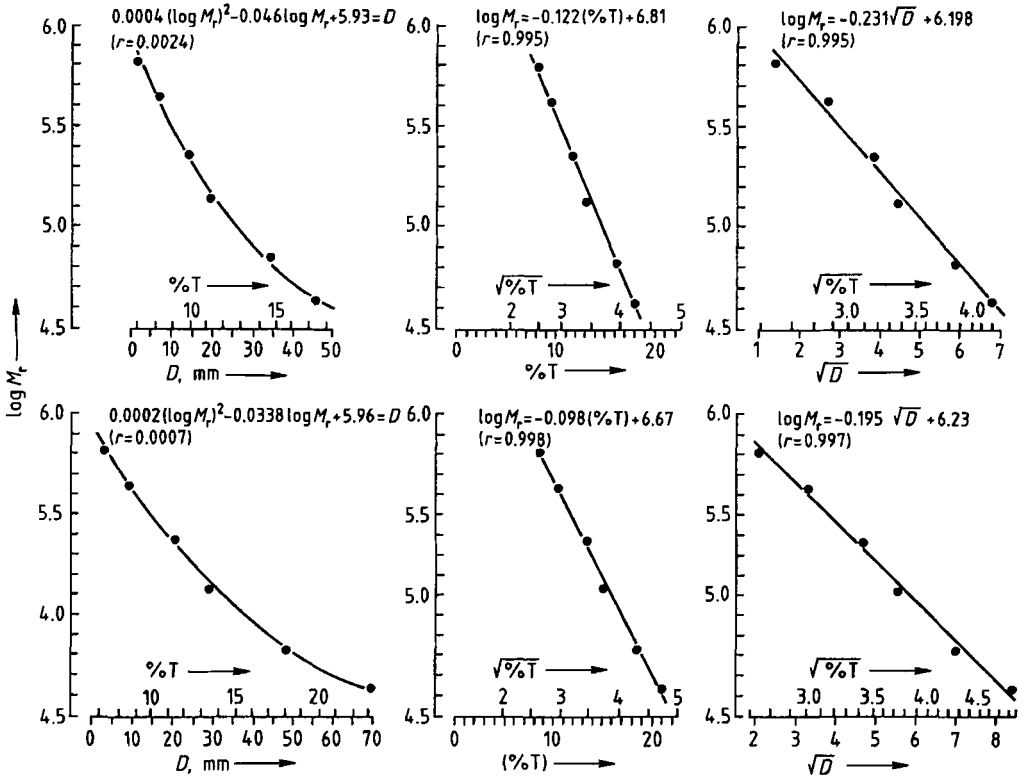
Figure 5. Typical plots of log M_r against migration distance D or gel composition %T after pore gradient electrophoresis Note that these plots are nonlinear, whereas when log M_r is plotted against $\sqrt{D}$ or $\sqrt{\%T}$ a linear relationship is obtained [77].

onto a glass fiber filter and perform sequencing on it [85].

Two-dimensional electrophoresis (2-DE) excited grandiose projects, like the Human Protein Index System of Anderson & Anderson [79], who initiated the far-reaching goal of mapping all possible phenotypes expressed by all the different cells of the human organism. An Herculean task, if one considers that there may be close to 75 000 such phenotypes (assuming a total of ca. 250 differentiated cells, each expressing > 3000 polypeptides, of which 300 are specific to a given cell). This started a series of meetings of the 2-DE group. particularly strong in the field of clinical chemistry [86], [87]. A number of books were devoted to this 2-DE issue [88], [89], and a series of meetings was started [90]. Today, 2-DE has found a proper forum in the journal *Electrophoresis*, which began hosting individual papers dealing with variagate topics in 2-D maps. Starting in 1988, *Electrophoresis* launched special issues devoted to 2-D maps, not only in clinical chemistry

and human molecular anatomy, but in fact in every possible living organism and tissue, the first being dedicated to plant proteins [91]. Soon. a host of such "Paper Symposia" appeared, collecting databases on any new spots of which a sequence and a function could be elucidated, under the following editorships: Celis [92]–[101]; Dunn [102]–[105]; Lottspeich [106]; Tümmler [107]; Williams [108]; Humphery-Smith [109]; Appel et al. [110], [111] and Cash [112]. This collection of "Paper Symposia" is a gold mine of new information and novel evolutionary steps on the IEF technique (such as new solubilizers. new staining procedures, sample pretreatment before IEF and the like). The world "proteome" is a recent neologism that refers to PROTEins expressed by the genOME or tissue. Just as genome has become a generic term for "big-science" molecular biology, so proteome is becoming a synonym of bioinformatics, since this project requires the building and continuous updating of a vast body of information, such as the amino acid sequence and biological

function of the newly discovered proteins. An idea of the complexity of this field is given in [113], [114]. A partial list of the many sites available on the WWW for consulting databases and for visualizing codified 2-D maps is given in the following:

1) A 2D PAGE protein database of *Drosophila melanogaster*
   http://tyr.cmb.ki.se/

2) ExPASy Server (2D liver, plasma, etc. SWISS-(PROT, 2DPAGE, 3DIMAGE), BIOSCI, Melanie software)
   http://expasy.ch/

3) CSH QUEST Protein Database Center (2D REF52 rat, mouse embryo, yeast, Quest software)
   http://siva.cshl.org/

4) NCI/FCRDC LMMB Image Processing Section (GELLAB software)
   http://www-lmmb.ncifcrf.gov/lemkin/gel-lab.html

5) 2-D Images Meta-Database
   http://www-lmmb.ncifcrf.gov/2dwgDB/

6) *E. coli* Gene-Protein Database Project-ECO2DBASE (in NCBI repository)
   ftp://ncbi.nlm.nih.gov/repository/ECO2DBASE/

7) Argonne Protein Mapping Group Server (mouse liver, human breast cell, etc.)
   http://www.anl.gov/CMB/PMG/

8) Cambridge 2D PAGE (a rat neuronal database)
   http://sunspot.bioc.cam.uk/Neuron.html

9) Heart Science Centre, Harefield Hospital (Human Heart 2D Gel Protein DB)
   http://www.harefield.mthames.nhs.uk/nhli/protein/

10) Berlin Human Myocardial 2D Electrophoresis Protein Database
    http://www.chemie.fu-berlin.de/user/pleiss/

11) Max Delbruck Ctr. for Molecular Medicine-Myocardial 2D Gel DB
    http://www.mdc-berlin.de/~emu/heart/

12) The World of Electrophoresis (EP literature, ElphoFit)
    http://www.uni-giessen.de/~gh43/electrophoresis.html

13) Human Colon Carcinoma Protein Database (Joint Protein Structure Lab)
    http://www.ludwig.edu.au/www/jpsl/jpslhome.html

14) Large Scale Biology Corp (2D maps: rat, mouse, and human liver)
    http://www.lsbc.com/

15) Yeast 2D PAGE
    http://yeast-2dpage.gmm.gu.se/

16) Proteome Inc. YPD Yeast Protein Database
    http://www.proteome.com/YPDhome.html

17) Keratinocyte, cDNA Database (Danish Centre for Human Genome Research)
    http://biobase.dk/cgi-bin/celis/

18) Institut de Biochemie et Genetique-Yeast 2D gel DB
    http://www.ibgc.u-bordeaux2.fr/YPM/

19) Swiss-Flash newsletter
    http://www.expasy.ch/swiss-flash/

20) Phosphoprotein Database
    http://www-lmmb.ncifcrf.gov/phosphoDB/

Figure 6 shows the complexity of one such 2-D map, downloaded from the ExPASy site, representing a partial panorama (limited to the pH gradient 5.5 – 8.5) of the thousands of polypeptide spots present in the human liver.

14.9. Isotachophoresis

The theoretical foundations of isotachophoresis (ITP) were laid in 1897 by KOHLRAUSCH [115], who showed that, at a migrating boundary between two salt solutions, the concentrations of ions were related to their effective mobilities (Kohlrausch autoregulating function) [116]. The term isotachophoresis underlines the most important aspect of this technique, namely the identical velocities of the sample zones at equilibrium [117]. ITP will take place when an electric field is applied to a system of electrolytes, consisting of:

1) A leading electrolyte, which must contain only one ionic species, the leading ion L^-, having the same sign as the sample ions to be separated, and an effective mobility higher than that of any sample ions

2) A second, terminating electrolyte, which contains one ion species, the terminating ion T^-, having the same sign as the sample ions to be separated, and an effective mobility lower than that of any sample ions

3) An intermediate zone of sample ions, having the same sign as the leading and terminating ions and intermediate mobilities

The three zones are juxtaposed, with the proviso that sharp boundaries must be created at the start of the experiment. The polarity of the electric field must be such that the leading ion migrates in front
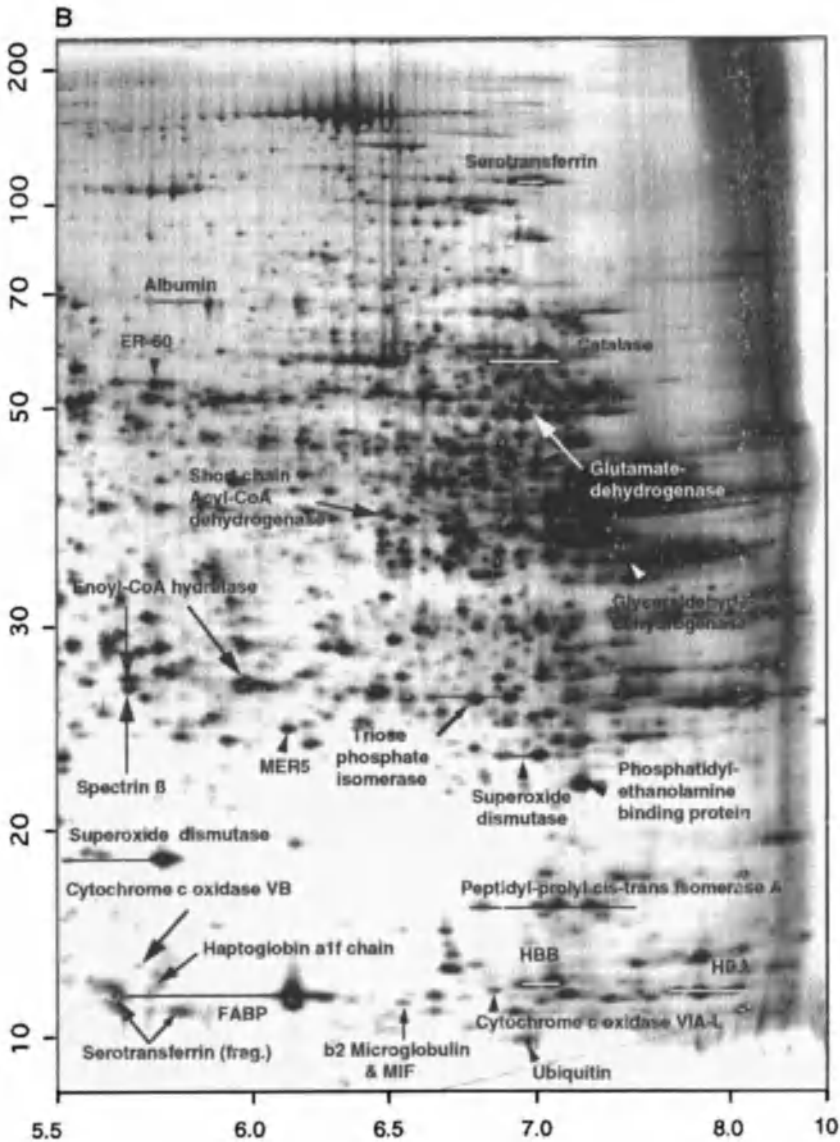
Figure 6. Partial 2-D map of human liver proteins, as resolved by IEF in the first dimension (x-axis, pH 5.5 – 8.5) and SDS-PAGE in the second dimension (y-axis, molecular mass).

of the ITP train at all times. When the system has reached the steady state, all ions move with the same speed, individually separated into a number of consecutive zones, in immediate contact with each other, and arranged in order of effective mobilities. Once the ITP train is formed, the ionic concentration in a separated sample zone adapts itself to the concentration of the preceding zone.

The Kohlrausch function, which is given at the leading/terminating ion boundary, in fact gives the conditions at any boundary between two adjacent ions A^-, B^-, with one common counterion R^+ when the boundary migrates in the electric field.

There are two fundamental properties of ITP built into the autoregulating function: the concentrating effect and the zone-sharpening effect
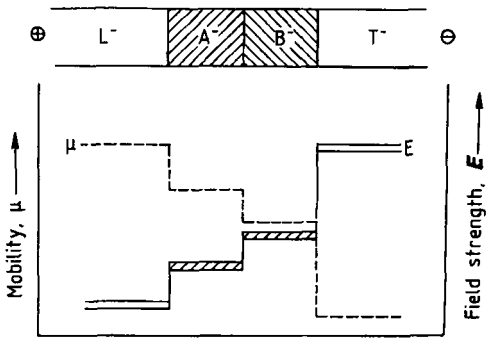
Figure 7. Scheme of an isotachophoretic process, with adjacent zones of leading ion L⁻, samples A⁻ and B⁻ and terminating ion T⁻
The respective mobilities (dotted line) and voltage gradients (solid line) in each zone are represented.

(Fig. 7). Suppose component A⁻ is introduced at very low concentration, even lower than that of the terminating ion T⁻. Since the mobility of A⁻ is intermediate between that of L⁻ and T⁻, its concentration must also be intermediate between those of L⁻ and T⁻. This will result in A⁻ being concentrated (decrease in zone length) until it reaches the theoretically defined concentration. Conversely, if the concentration of A⁻ is too high (even higher than that of L⁻), the A⁻ zone will be diluted (increase in zone length) until the correct equilibrium concentration is reached. This is an example of the zone-sharpening effect: if A⁻ diffuses into the B⁻ zone, it is accelerated, since it enters a zone of higher voltage gradient, therefore, it automatically migrates back into its zone. Conversely, if it enters the L⁻ zone, it finds a region of lower voltage, is decelerated, and thus falls back into its zone (Fig. 7). This applies to all ions in the system. Note that the first few minutes of disc electrophoresis, the period of migration in the sample and stacking gels, represent in fact an ITP migration. This is what produces the spectacular effects of disc electrophoresis: stacking into extremely thin starting zones (barely a few micrometers thick, from a zone up to 1 cm in thickness, a concentration factor of 1000–10000) and sharpening of the zone boundaries. A characteristic of ITP is that the peaks, unlike those in other separation techniques (except for displacement chromatography), are square-shaped as a result of the Kohlrausch autoregulating function, i.e., the concentration of the substance within a homogeneous zone is constant from front to rear boundary [118]. In fact, under ideal conditions,

the diffusive forces, which cause an eluted peak to spread into a Gaussian shape, are effectively counteracted. Although ITP is not much in vogue today, it is now used for stacking purposes as a transient step in CZE, for concentrating and sharpening dilute sample zones [119].

14.10. Immunoelectrophoresis

The basic technique of immunoelectrophoresis was first described by GRABAR and WILLIAMS in 1953 [9]. The antigen sample is applied to a hole in the middle of a glass plate, coated with a 1-mm layer of agarose gel. Electrophoresis separates the antigen mixture into various zones. A longitudinal trough, parallel to the long edge of the plate, is then made in the gel and filled with the antiserum to the antigen mixture. A passive, double-diffusion takes place: the antiserum diffuses into the gel (advancing with a linear front) while the antigens diffuse radially in all directions from the electrophoresis zones. The antigen–antibody complexes are formed at equivalence points, the number of precipitates formed corresponding to the number of independent antigens present. The precipitates form a system of arcs, resulting from the combination of linear and circular fronts. A number of variants have been described over the years, but in this chapter only some of the most popular immunoelectrophoretic techniques are reviewed. For a more extensive treatise, see [126]–[128].

14.10.1. Rocket Immunoelectrophoresis

In rocket immunoelectrophoresis, the antigen–antibody reaction occurs during the electrophoresis of an antigen in an antibody-containing medium. Both antigen and antibody move according to their electrophoretic mobilities, and they also react with each other, resulting in flame-shaped precipitation zones of antigen–antibody complexes (LAURELL's rocket technique) [120]. Under the influence of the electric field, unbound antigen within the peak of the flame-shaped precipitate migrates into the precipitate, which redissolves in the excess antigen. Thus, the leading edge of the flame is gradually displaced in the direction in which the antigen is migrating. The amount of antigen within the leading boundary edge is successively diminished because of the formation of soluble antigen–antibody complex-

es. When the antigen is diminished to equivalence with the antibody, the complexes can no longer be dissolved and a stable precipitate forms at the leading edge, which is thereafter stationary. The distance finally traveled by the peak depends on the relative excess of antigen over antibody and can be used as a measure of the amount of antigen present. Every precipitation band featured as a flame represents an individual antigen [120].

14.10.2. Crossed Immunoelectrophoresis

In the case of polydisperse samples, crossed immunoelectrophoresis (CIE) as described by CLARKE and FREEMAN [121] is employed. First, a mixture of protein antigens is subjected to conventional electrophoresis in agarose. Then the agarose gel is cut into strips, which are transferred to a second glass plate. Melted agarose (1 % w/v) containing the antiserum is then cast to form a gel of the same thickness as that of the first-dimension gel, and a secure junction between the two layers is achieved by melting to fuse their edges. During the second-dimension electrophoresis, performed perpendicular to the first, each antigen migrates independently, and precipitation zones are formed which resemble the Laurell rockets but have a wider base. Remarkably, two antigens present in a single zone, i.e., possessing equal electrophoretic mobilities, can be distinguished by the second-dimension electroimmunoassay as the shape and height of the respective rockets generally do not coincide exactly. In a typical product of the Clarke and Freeman method, ca. 50 rockets can be counted in human serum, and their relative abundance can be assessed by measuring the areas of the respective peaks [121]. Peak assessment and evaluation have been greatly facilitated by the development of a computer system for the specific analysis of CIE patterns [122].

14.10.3. Tandem Crossed Immunoelectrophoresis

It would be quite difficult to identify a single antigen in the complex patterns of CIE. One of the proposed methods for complex mixtures is the tandem CIE technique. Before the electrophoresis in the first dimension, two wells rather than one are cut into the gel strip, positioned one after the other in the direction of electrophoretic migration. One well is loaded with a mixture of the antigens to be analyzed, while the pure antigen whose peak is to be identified in the mixture is introduced into the other. The remaining manipulations are performed according to classical CIE. In the final pattern there will be two peaks which fuse smoothly and are separated by exactly the distance between the two wells. These two peaks may be of different heights (owing to difference in antigen concentration in the sample and reference wells), whereas fusion between them is indicative of the fact that they are caused by the same protein antigen. It is this canceling of the inner flanks of the two rockets and the fusion process that allow the unknown antigen to be located. By repeating this process with different purified antigens, it is possible to map, one by one, the components of a heterogeneous mixture [123].

14.10.4. Intermediate Gel Crossed Immunoelectrophoresis

Tandem CIE aimed at the identification of a single antigen in a mixture requires the availability of a pure marker antigen (see Section 14.10.3). In an alternative approach, identification can also be made when monospecific antiserum to the antigen is available. This approach is called intermediate gel CIE. After the first-dimension electrophoresis, the slab is prepared as for conventional CIE. Then a 1- to 2-cm strip of polyspecific antiserum-containing agarose gel nearest to the first-dimension gel is excised and a monospecific antiserum-containing agarose is cast instead. In the course of the second-dimension electrophoresis, the antigen to be detected will precipitate with the specific antiserum just as it enters the intermediate gel, while other antigens will pass through this gel without being retarded. The bases of their precipitation peaks, formed in the polyspecific antiserum-containing agarose, will be positioned on the borderline between the two latter gels. The antigen under study is thus distinguishable from the others [124].

14.10.5. Fused-Rocket Crossed Immunoelectrophoresis

In contrast to the Laurell rocket technique, this method is not used for quantitative determinations. Rather, it is applied to detect heterogeneity in seemingly homogeneous protein fractions, obtained by gel chromatography or ion exchange chromatography. A set of wells is punched in a

checkerboard pattern in a strip of antibody-free agarose gel. Aliquots of each fraction eluted from the column are placed in the wells and the gel is left in a humid chamber for 30–45 min to allow the proteins in the wells to migrate and fuse, thus reproducing the continuous elution profile on an extended scale (hence the term "fused rockets"). On the remainder of the plate a thin layer of agarose is cast containing a polyvalent antiserum, and electrophoresis is performed as usual. If the collected chromatographic peak is indeed homogeneous, a single fused rocket will appear, while inhomogeneity will be revealed by an envelope of subpeaks within the main eluate fraction [125].

14.11. Staining Techniques and Blotting

MERRIL et al. [129] described a silver-staining procedure in which the sensitivity — which in Coomassie Blue is merely on the order of a few micrograms per zone — is increased to a few nanograms of protein per zone, thus approaching the sensitivity of radioisotope labeling. In the gilding technique of MOEREMANS et al. [130] polypeptide chains are coated with 20-nm particles of colloidal gold and detected with a sensitivity of < 1 ng/ mm^2. Proteins can also be stained with micelles of Fe^{3+}, although with a sensitivity about one order of magnitude lower than with gold micelles [131]. These last two staining techniques became a reality only after the discovery of yet another electrophoretic method, the so-called Southern [132] and Western [133] blots, in which nucleic acids or proteins are transferred from hydrophilic gels to nitrocellulose or any of a number of other membranes, where they are immobilized by hydrophobic adsorption or covalent bonding. The very large porosity of these membranes makes them accessible to colloidal dyes. In addition, transfer of proteins to thin membranes greatly facilitates detection by immunological methods. This has resulted in new, high-sensitivity methods called "immunoblotting." After saturation of potential binding sites, the antigens transferred to the membrane are first made to react with a primary antibody, and then the precipitate is detected with a secondary antibody, tagged with horseradish peroxidase, alkaline phosphatase, gold particles, or biotin, which is then allowed to react with enzyme-linked avidin [134]. In all cases the sensitivity is greatly augmented. It appears that immunoblotting will

pose a serious threat to quite a few of the standard immunoelectrophoretic techniques [135]. In terms of colloidal staining, a direct staining method for polyacrylamide gels with colloidal particles of Coomassie Blue G-250 is said to be as sensitive as silver staining [136]. For an extensive review covering all aspects of staining techniques, together with a vast bibliography on polypeptide detection methods, see [37].

14.12. Immobilized pH Gradients

In 1982, immobilized pH gradients (IPG) were introduced, resulting in an increase in resolution by one order of magnitude compared with conventional IEF [138]. Conventional IEF was besieged by several problems: (1) very low and unknown ionic strength; (2) uneven buffering capacity; (3) uneven conductivity; (4) unknown chemical environment; (5) nonamenable to pH gradient engineering; (6) cathodic drift (pH gradient instability); and (7) low sample loadability. In particular, a most vexing phenomenon was the near-isoelectric precipitation of samples of low solubility at the isoelectric point or of components present in large amounts in heterogeneous samples. The inability to reach stable steady-state conditions (resulting in a slow pH gradient loss at the cathodic gel end) and to obtain narrow and ultranarrow pH gradients, aggravated matters. Perhaps, most annoying was the lack of reproducibility and linearity of pH gradients produced by the so-called carrier ampholyte buffers [3]. IPGs proved able to solve all these problems.

IPGs are based on the principle that the pH gradient, which exists prior to IEF itself, is copolymerized, and thus rendered insoluble, within the fibers of a polyacrylamide matrix (Fig. 8). This is achieved by using, as buffers, a set of eight commercial chemicals (known as Immobilines, by analogy with Ampholine, produced by Pharmacia-LKB Biotechnologies, Uppsala, Sweden) having pK values distributed over the pH range 3.1–10.3. They are acrylamido derivatives with the general formula $CH_2 = CH–CO–NH–R$, where R denotes one of three weak carboxyl groups, with pK values of 3.1, 3.6, and 4.6, for the acidic compounds, or one of five tertiary amino groups, with pK values of 6.2, 7.0, 8.5, 9.3, and 10.3, for the basic buffers [139]. Two additional compounds are needed: a strongly acidic (pK 1.0) and a strongly basic (pK > 12) titrant for producing linear pH gradients
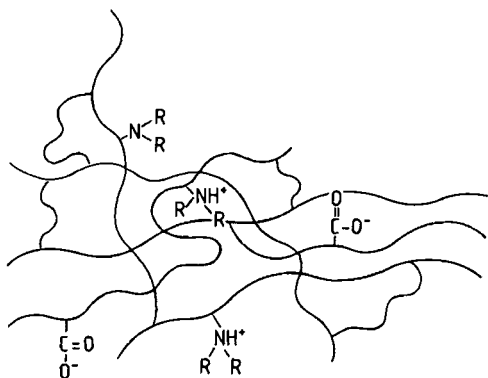
Figure 8. Scheme of a polyacrylamide matrix with a grafted pH gradient, depicted as bound negative and positive charges The ratio of such charges in each gel layer between anode and cathode defines a unique pH value.

covering the entire pH range 3–10 [140]. Computer simulations had shown that, in the absence of these two titrants, extended pH intervals would exhibit strong deviations from linearity at the two extremes, as the most acidic and most basic of the commercial Immobilines would act simultaneously as buffers and titrants [141]. The 2-acrylamidoglycolic acid (pK 3.1) species [142] is useful for separating strongly acidic proteins, since it extends the pH gradient to as low as pH 2.5. N,N-Diethylaminopropylacrylamide   (pK 10.3) has been used for analysis of strongly alkaline proteins [143]. Given the fairly evenly spacing of the pK values along the pH scale, it is clear that the set of ten chemicals proposed here (eight buffers and two titrants) is quite adequate to ensure linear pH gradients in the range pH 2.5–11 (the ideal ΔpK for linearity would be 1 pH unit between two adjacent buffers). The rule ΔpK = 1 is fairly well obeyed, except for two "holes" between pK 4.6 and 6.2 and between pK 7.0 and 8.5. For a more detailed treatise on how to use an IPG gel, and IPG recipes, the reader is referred to an extensive manual [4] and to reviews [144]–[148]. Owing to the much increased resolution of IPG, quite a number of so-called electrophoretically silent mutations (bearing amino acid replacements with no ionizable groups in the side chains) have now been fully resolved.

14.13. Capillary Zone Electrophoresis

Capillary zone electrophoresis (CZE) appears to be a most powerful technique, perhaps equaling the resolving power of IPG. If one assumes that longitudinal diffusion is the only significant source of zone broadening, then the number of theoretical plates N in CZE is given by [149]:

$$N = \mu V / 2D$$

where μ and D are the electrophoretic mobility and diffusion coefficient, respectively, of the analyte, and V is the applied voltage. This equation shows that high voltage gradients are the most direct way to high separation efficiencies. For nucleic acids, it has been calculated that N could be as high as 10^6 theoretical plates. Figure 9 is a schematic drawing of a CZE system, illustrating one of the possible injection systems, namely gravity feed [150]. The fused-silica capillary has a diameter of 50–100 µm and a length up to 1 m. It is suspended between two reservoirs, connected to a power supply that is able to deliver up to 30 kV (typical operating currents 10–100 µA). One of the simplest ways to introduce the sample into the capillary is by electromigration, i.e., by dipping the capillary extremity into the sample reservoir, under voltage, for a few seconds. Detection is usually accomplished by on-column fluorescence and/or UV absorption. Conductivity and thermal detectors, as usually employed in ITP, exhibit too low a sensitivity in CZE. The reason for this stems from the fact that the flow cell where sample monitoring occurs has a volume of barely 0.5 nL, allowing sensitivities down to the femtomole level. In fact, with laser-induced fluorescence detection, a sensitivity on the order of 10^{-21} mol is claimed [151]. By forming a chiral complex with a component of the background electrolyte (copper aspartame) it is possible to resolve racemates of amino acids [8]. Even neutral organic molecules can be made to migrate in CZE by complexing them with charged ligands, such as SDS. This introduces a new parameter, a hydrophobicity scale, in electrokinetic migrations. For more on CZE, readers are referred to the Proceedings of the International Symposia on High-Performance Capillary Electrophoresis [152]–[162], which give an account of the evolution of and new developments in the CZE technique, spanning a ten-year period. Other special issues dedicated to CZE can be found in *J. Chromatogr. B* [163] and *J. Biochem. Biophys. Methods* [164]. In addition, the
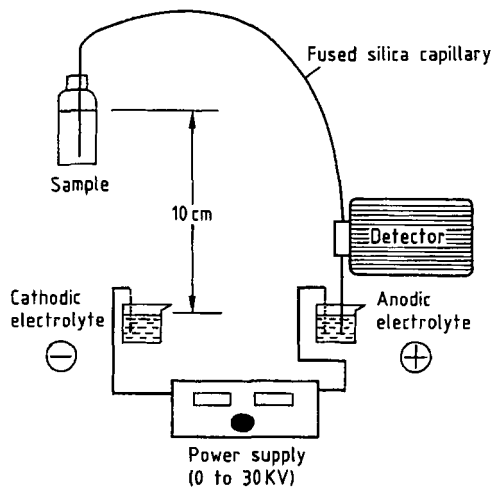
Figure 9. Scheme of a capillary zone electrophoresis apparatus

The detector includes a high-pressure mercury – xenon arc lamp, oriented perpendicular to the migration path, at the end of the capillary. The sample signal is measured with a photomultiplier and a photometer connected to the analog/digital converter of a multifunction interface board, connected to a microcomputer.

journal Electrophoresis has devoted a number of Paper Symposia to various aspects of CZE [165] – [171]. Two books which offer a broad coverage of the variegated aspects of CZE are [172] and [173].

It is of interest to note here some aspects of the CZE technique, which offers a unique performance, unrivalled by other electrokinetic methods, except for 2-D maps. These aspects regard the use of sieving liquid polymers instead of true gels for separation of macromolecules, notably proteins and nucleic acids. Any polymer above the entanglement threshold can exhibit sieving and separate macro-analytes almost as efficiently as true gels. Such polymers embrace the classical polyacrylamides, as well as a number of others, such as celluloses, poly(vinyl alcohol), and poly(vinyl pyrrolidone). The advantages: after each run, the polymers can be expelled from the lumen of the capillary and thus each new run can be performed under highly controlled and reproducible starting conditions, in the absence of carryover from previous runs. In addition, all celluloses are transparent to low wavelengths, down to 200 nm, thus permitting on-line detection of proteins and peptides via the adsorption of the peptide bond, with much increased sensitivity. Other interesting, recent developments regard the use of isoelectric
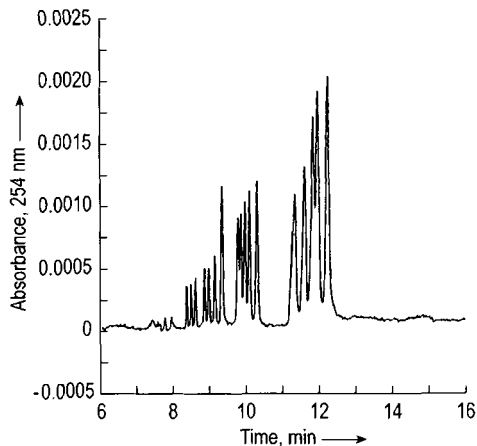
Figure 10. CZE of DNA marker V in isoelectric buffers. Conditions: 150 mM His buffer, pH=pI=7.6 in presence of 1.5 % poly(DMA); poly(AAP)-coated capillary, i.d. 100 μm, total length 30 cm (25.4 cm to the detector). UV absorbance detection at 254 nm. Run at 25 °C and 100 V/cm [174].

buffers, i.e., amphoteric species as the sole buffering ion at pH = pI. Such buffers, due to their very low conductivity, allow delivery of high voltage gradients, with much increased resolution and very short analysis times. Examples on separations of nucleic acids in isoelectric His [174] and of peptides in isoelectric Asp [175] are given in Figures 10 and 11, respectively.

14.14. Preparative Electrophoresis

Several variants of electrophoresis have been described for preparative runs. They utilize either completely fluid beds or gel phases. This chapter restricts itself to those techniques exploiting isoelectric focusing principles, both in carrier ampholytes or in immobilized, nonamphoteric buffers, since they couple, in general, high loading capacities to a high resolving power. A classical example involves vertical columns filled with a sucrose density gradient, which is then emptied and collected in fractions at the end of the IEF run. This was Rilbe's original idea for performing IEF [176]; however, even though LKB has sold more than 3000 columns over the years, they are seldom used today. The following sections describe a few methods which appear to be more accessible; a complete survey of all other variants is available [3].
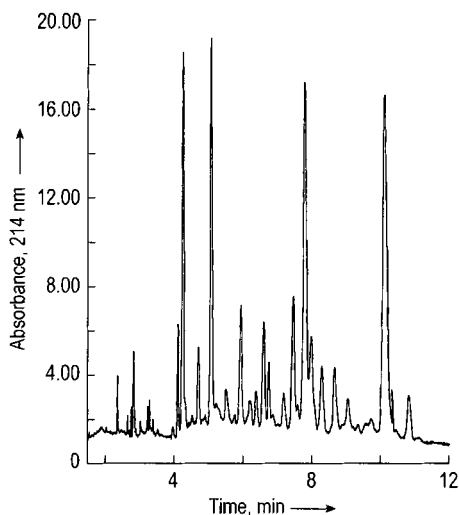
Figure 11. CZE of a tryptic digest of β-casein in a 37 × 100 μm i.d. capillary, bathed in 50 mM isoelectric aspartic acid (pH approximating the pI value of 2.77) added with 0.5 % hydroxyethyl cellulose. Run: 600 V (current 58 μA). The three major peptides (pI 6.1, 6.93, and 3.95) were eluted and analyzed by mass spectrometry [175].

14.14.1. Preparative Isoelectric Focusing in Granulated Gel Layers

RADOLA [177], [178] described a method for large-scale preparative IEF in troughs coated with a suspension of granular gels, such as Sephadex G-25 superfine (7.5 g/100 mL), Sephadex G-200 superfine (4 g/100 mL), or Bio Gel P-60 (4 g/100 mL). The trough consists of a glass or quartz plate at the bottom of a Lucite frame (Fig. 12). Various sizes of trough may be used; typical dimensions are 40 × 20 cm or 20 × 20 cm, with a gel layer thickness of up to 10 mm. The total gel volume in the trough varies from 300 to 800 mL. A slurry of Sephadex (containing 1 % Ampholine) is poured into the trough and allowed to dry in air to the correct consistency. Achieving the correct consistency of the slurry before focusing seems to be the key to successful results. Best results seem to be obtained when 25 % of the water in the gel has evaporated and the gel does not move when the plate is inclined at 45°. The plate is run on a cooling block maintained at 2 – 10°C. The electric field is applied via flat electrodes or platinum bands which make contact with the gel through absorbent paper pads soaked in 1 mol/L sulfuric acid at the anode, and 2 mol/L ethylenediamine at the cathode. In most prepara-
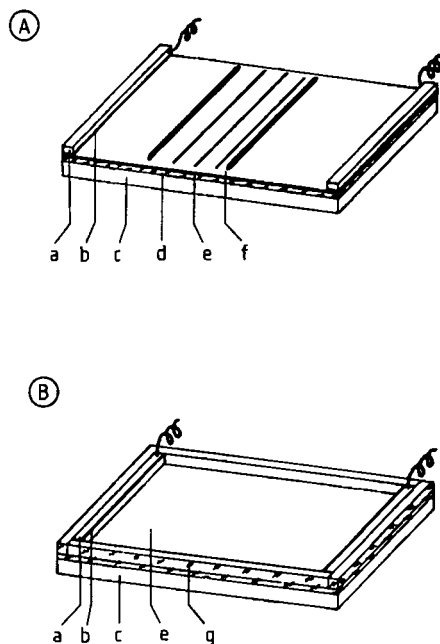
Figure 12. Preparative IEF in granulated Sephadex layers A) Small-scale separation; B) Large-scale separation a) Electrode; b) Filter paper pad soaked in electrode solution; c) Cooling block; d) Glass plate; e) Gel layer; f) Focused proteins; g) Trough [178].

tive experiments, initial voltages of 10 – 15 V/cm and terminal voltages of 20 – 40 V/cm are used. Power gradients as high as 0.05 W/cm are well tolerated in 1-cm thick layers. Samples may be mixed with the gel suspension or added to the surface of preformed gels, either as a streak or from the edge of a glass slide. Larger sample volumes may be mixed with the dry gel (ca. 60 mg Sephadex per mL) and poured into a slot in the gel slab. Samples may be applied at any position between the electrodes. After focusing, proteins are located by the paper print technique. Focused proteins in the surface gel layer are absorbed onto a strip of filter paper which is then dried at 110°C. The proteins may then be stained directly with dyes of low sensitivity, such as Light Green SF or Coomassie Violet R-150, after removing ampholytes by washing in acid. The pH gradient in the gel can be measured in situ with a combination microelectrode sliding on a calibrated ruler. Radola's technique offers the advantage of combining high resolution, high sample load, and easy recovery of focused components. As much as 5 – 10 mg/mL protein per milliliter gel suspension can be fractionated over wide pH ranges. Purifi-
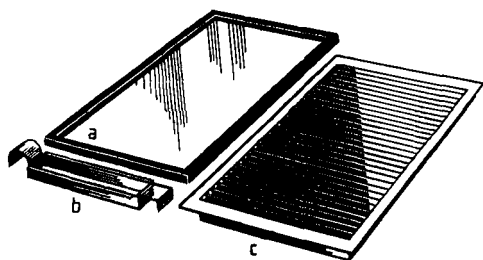
Figure 13. Trough (a), sample applicator (b), and fractionation grid (c) for preparative IEF in granulated gel layers (LKB Multiphor 2117 apparatus)

cation of 10 g pronase E in 800 mL gel suspension has been reported. At these high protein loads, even colorless samples can be easily detected, since they appear in the gel as translucent zones. As there is no sieving effect for macromolecules above the exclusion limits of the Sephadex, high molecular mass substances, such as virus particles, can be focused without steric hindrance. The system has a high flexibility, since it allows analytical and small- and large-scale preparative runs in the same trough, merely by varying the gel thickness.

For recovery of protein bands at the end of the IEF run, a fractionation grid pressed onto the gel layer allows recovery of 30 fractions, i.e., one fraction each 8 mm along a separation axis of 25 cm (Fig. 13). After scraping off each fraction with a spatula, the protein is eluted by placing the gel into a syringe equipped with glass wool as bottom filter, adding a buffer volume, and ejecting the eluate with the syringe piston.

14.14.2. Continuous-Flow, Recycling Isoelectric Focusing

Another interesting approach, called recycling IEF, has been described by BIER's group [179], [180]. It is well known that continuous-flow techniques, which appear essential for large-scale preparative work, are disturbed by parabolic and electrosmotic flows, as well as by convective flows due to thermal gradients. BIER et al. have improved this system by separating the actual flow-through focusing cell, which is miniaturized, from the sample and heat-exchange reservoir, which can be built up to any size. Minimization of parabolic flow, electrosmosis, and convective liquid flow is achieved by flowing the sample to be separated through a thin focusing cell (the actual dis-

tance from anode to cathode is only 3 cm) built of an array of closely spaced filter elements oriented parallel to the electrodes and parallel to the direction of flow. Increased sample load is achieved by recirculating the process fluid through external heat-exchange reservoirs, where the Joule heat is dissipated. During each passage through the focusing cell, only small sample migrations toward their eventual pI are obtained, but through recycling a final steady state is achieved. The IEF cell has ten input and output ports for sample flow-through, monitored by an array of ten miniaturized pH electrodes and ten UV sensors. The entire system is controlled and operated by a computer. A scheme of the entire apparatus can be seen in Figure 14. By activating pumps at two extreme channels, the computer can alter the slope of the pH gradient, counteracting any effect of cathodic drift, which results in a net migration of the sample zones towards the cathode.

14.14.3. The Rotofor

The Rotofor cell is a preparative-scale, free-solution, isoelectric focusing apparatus that provides a useful alternative to the recycling system described above. This cell [181] consists of a cylindrical focusing chamber that rotates around its axis, achieving liquid stabilization, not by recycling, but by gravity stabilization, as reported in 1967 by HJERTÉN [182]. The inside of the cylindrical focusing apparatus is divided by 19 parallel, monofilament, polyester membrane screens into 20 discrete compartments. Figure 15 gives a schematic view of the focusing chamber with the assembled Nylon screens supported by a cooling tube. There are several advantages of such a horizontal, rotating system. First of all, in the classical vertical density gradient IEF column of VESTERBERG et al. [176] isoelectric protein precipitates (not at all uncommon at the high protein loads of preparative runs) used to sediment along the supporting density gradient, thus contaminating other focused zones along the separation path. The Rotofor cell overcomes this problem, since, by focusing in a horizontal axis, potential protein precipitates stay focused in their compartment and do not disturb the remainder of the gradient. Secondly, the vertical density gradient column for IEF had a weak point in the elution funnel: all the liquid in the column was emptied via a bottom outlet, and this resulted in severe remixing of the separated protein zones and in contamination
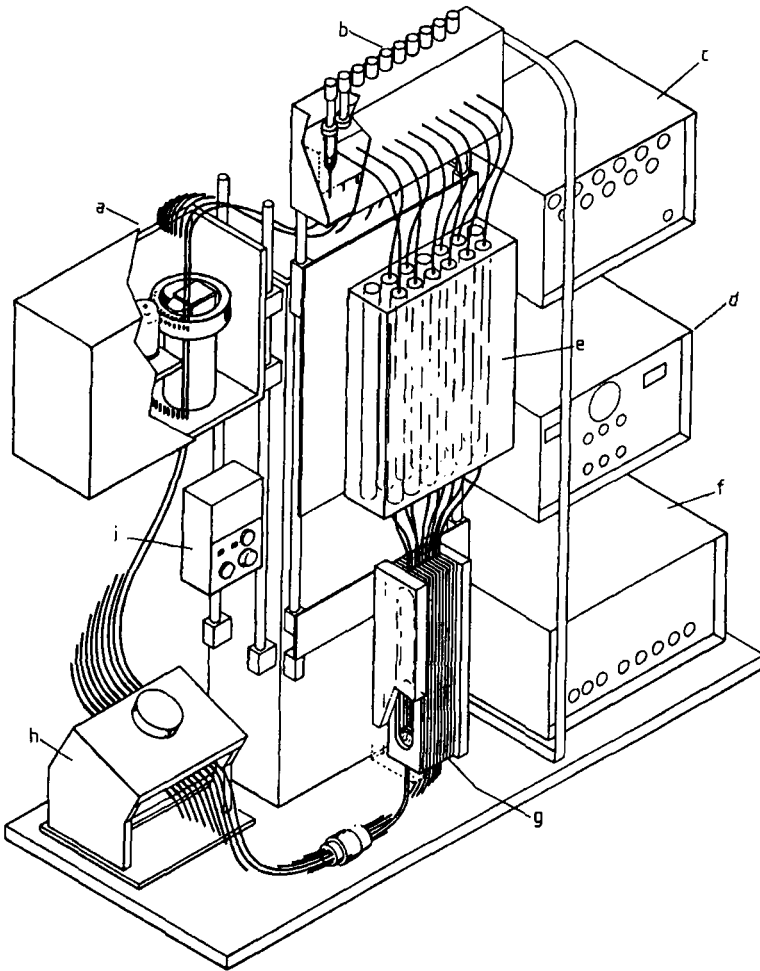
Figure 14. Scheme of the recycling IEF apparatus [180]
a) UV monitor; b) pH meter; c) Data interface; d) Power supply; e) Heat-exchange reservoir; f) UV control; g) Focusing cell; h) Pump; i) Pump control

by diffusion, owing to the long times required for harvesting the column content (usually 30–60 min, depending on column size). These problems are eliminated in the Rotofor apparatus, since the 20 chambers inside the column are emptied simultaneously and instantaneously by an array of 20 collection needles. The cylindrical focusing cell holds a total of 55 ml of solution for preparative protein purification; during operation, the chamber is rotated at 10 rpm around its axis to stabilize against convective disturbances. The run is usually performed at 12 W constant power (up to 2000 V maximum voltage); focusing is in general accomplished in as little as 4 h. Once the system is at steady state, collection is quickly per-

formed by using an array of 20 collection needles that lead to 20 test tubes nested in a specially designed vacuum collection box. The needles simultaneously pierce a tape sealing 20 small holes, each at the bottom of an individual chamber in the Rotofor. By applying vacuum to the collection box, the 20 compartments are simultaneously emptied. In common with other preparative IEF techniques, the Rotofor can be used in a cascade set-up, that is, the contents of a single cell, containing the protein of interest, can be spread over the entire 20-cell assembly, and focused a second time over a shallower pH gradient, upon dilution with an appropriate range of carrier ampholytes.
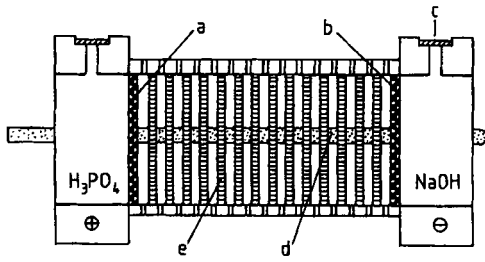
Figure 15. Scheme of the Rotofor cell
a) Cation-exchange membrane: b) Anion-exchange membrane: c) Vent caps; d) Ceramic cold finger; e) Screening material (6-μm polyester)

14.14.4. Recycling Free-Flow Focusing (RF3)

In the recycling IEF apparatus, stabilization of fluid flow is performed by screen elements; in the Rotofor, stabilization is achieved by screen elements and by rotation. In the RF3 apparatus shown schematically in Figure 16, fluid stabilization is achieved by means of rapid flow through narrow gaps, with continuous recycling. The key to stability of laminar flow through the cavity is its shallow depth (0.75 mm), combined with the rapid flow of process fluid (the latter necessary for avoiding electrohydrodynamic distortion). The lateral electrodes are separated from the focusing cavity by ion-permselective membranes. This instrument can also be used in the electrodialysis mode, by having the cation-selective membrane facing the cathodic compartment and the anion-selective membrane in front of the anodic chamber [183].

14.14.5. Multicompartment Electrolyzers with Isoelectric Membranes

Perhaps the final evolution of preparative IEF is the concept of a membrane apparatus, in which the entire IPG gel is reduced to isoelectric membranes delimiting a series of flow chambers. Figure 17 gives an exploded view of this novel apparatus [184], [185]: it consists of a stack of chambers sandwiched between an anodic and a cathodic reservoir. The apparatus is modular and in the present version can accommodate up to eight flow chambers. Figure 17 shows a stack of three chambers already assembled to the left, a central compartment, and a thinner chamber to
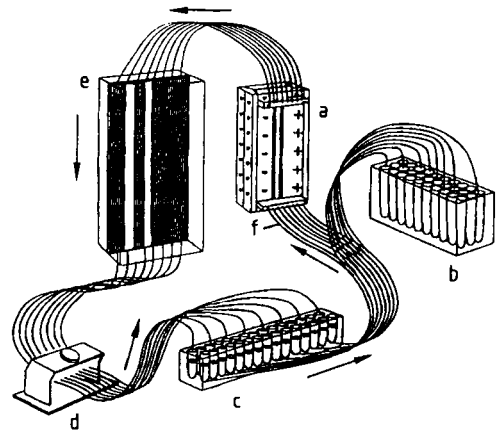
Figure 16. Scheme of the recycling free-flow focusing (RF3) apparatus
The arrows indicate the fluid flow pattern. At the end of the process, the flow pattern is switched to the harvesting position toward the fraction collector [183]
a) Focusing cell: b) Fraction collector; c) Bubble trap;
d) Peristaltic pump; e) Heat exchanger; f) 30-Channel manifold

the right for connection to the other electrode. All flow chambers are provided with inlet and outlet for sample or electrolyte recycling, an O-ring for ensuring flow-tight connections, and four holes for threading four long metal rods which can be tightened by hand-driven butterfly nuts for assembling the apparatus. Several versions of these cells have been built, capable of housing Immobiline membranes from 4.7 cm (the present apparatus) up to 9 cm diameter. The pH-controlling membranes are housed in the central depression between two 1-cm-wide caoutchouc rings. After assembling and tightening the apparatus, each compartment is flow-tight, so that no net liquid bulk flow ensues (except, when applicable, that generated by electrosmosis). The Pt electrodes are housed in two rectangular Perspex mountings, which also act as legs on which the electrolyzer stands. The distance between adjacent cells is only 10 mm, so that each chamber holds ca. 5 ml of liquid. The reason why this system works is shown in Figure 18: two isoelectric membranes facing each flow chamber act by continuously titrating the protein of interest to its isoelectric point. They can be envisaged as highly selective membranes, which retain any protein having pIs in between their limiting values, and which allow transmigration of any nonamphoteric, nonisoelectric species. The only condition required is that $pI_{cm} > pI_p > pI_{am}$, where the subscripts cm and am denote cathodic and anodic membranes, and p is the protein
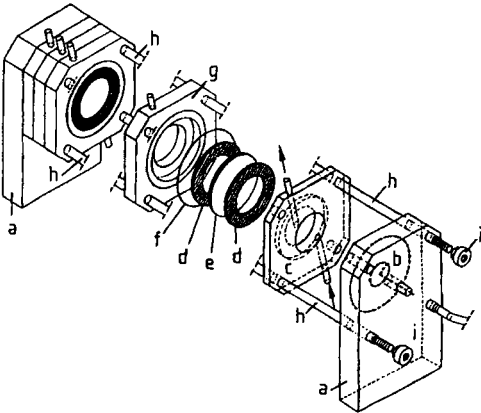
Figure 17. Exploded view of the multicompartment electrolyzer
a) Rectangular supporting legs; b) Pt electrode; c) Thin terminal flow chamber; d) Rubber rings for supporting the membrane; e) Isoelectric Immobiline membrane cast onto the glass-fiber filter; f) O-ring; g) One of the sample flow chambers; h) Threaded metal rods for assembling the apparatus; i) Nuts for fastening the metal bolts [128]

Figure 18. Mechanism of the purification process in the multicompartment electrolyzer
In the case shown, a protein with pI 6.5 is trapped between the pI 6 anodic and pI 7 cathodic membranes, whereas proteins with higher pI move toward the cathode and a species with lower pI (5.5) crosses the two membranes and collects at the anode [128].

having a given isoelectric point between the two membranes [184]. For this mechanism to be operative, the two isoelectric membranes must possess good conductivity and good buffering capacity, so as to effectively titrate the protein present in the flow chamber to its pI, while ensuring good current flow through the system. WENGER et al. [186] had in fact synthesized amphoteric, isoelectric Immobiline membranes and demonstrated that they are good conductors and good buffers at their pI.

14.15. References

[1] J. C. Giddings: *Unified Separation Science,* Wiley-Interscience, New York 1991, chap. 8.

[2] A. T. Andrews: *Electrophoresis: Theory, Techniques and Biochemical and Clinical Applications,* Clarendon Press, Oxford 1986.

[3] P. G. Righetti: *Isoelectric Focusing: Theory, Methodology and Applications,* Elsevier, Amsterdam 1983.

[4] P. G. Righetti: *Immobilized pH Gradients: Theory and Methodology,* Elsevier, Amsterdam 1990.

[5] C. J. Van Oss, *Sep. Purif. Methods* **8** (1979) 119–198.

[6] P. G. Righetti, C. J. Van Oss, J. W. Vanderhoff (eds.): *Electrokinetic Separation Methods,* Elsevier, Amsterdam 1979.

[7] B. L. Karger, *Nature (London)* **339** (1989) 641–642.

[8] P. Gozel, E. Gassmann, H. Michelsen, R. N. Zare, *Anal. Chem.* **59** (1987) 44–49.

[9] P. Grabar, C. A. Williams, *Biochim. Biophys. Acta* **10** (1953) 193–201.

[10] O. Smithies, *Biochem. J.* **61** (1955) 629–636.

[11] J. Kohn, *Nature (London)* **180** (1957) 986–987.

[12] L. P. Cawley, B J. Minard, W. W. Tourtellotte, P. J. Ma, C. Clelle, *Clin. Chem. (Winston Salem N.C.)* **22** (1976) 1262–1268.

[13] A. A. Keshgegian, P. Peiffer, *Clin. Chim. Acta* **108** (1981) 337–340.

[14] J. Kohn, P. Priches, J. C. Raymond, *J. Immunol. Methods* **76** (1985) 11–16.

[15] P. G. Righetti, P. Casero, G. B. Del Campo, *Clin. Chim. Acta* **157** (1986) 167–174.

[16] P. Casero, G. B. Del Campo, P. G. Righetti, *Electrophoresis (Weinheim Fed. Repub. Ger.)* **6** (1985) 373–376.

[17] T. Toda, T. Fujita, M. Ohashi in R. C. Allen, P. Arnaud (eds.): *Electrophoresis '81,* De Gruyter, Berlin 1981, pp. 271–280.

[18] P. Grabar, *Methods Biochem. Anal.* **7** (1957) 1–38.

[19] C. Araki, *Proceedings 5th International Seaweed Symposium,* Halifax, 1965, pp. 3–17.

[20] M. Duckworth, W. Yaphe, *Carbohydr. Res.* **16** (1971) 189–197.

[21] S. Hjertén, *Biochim. Biophys. Acta* **53** (1961) 514–517.

[22] S. Brishammar, S. Hjertén, B. Van Hofsten, *Biochim. Biophys. Acta* **53** (1961) 518–521.

[23] S. Hjertén, *Arch. Biochem. Biophys.* **99** (1962) 466–475.

[24] M. Duckworth, W. Yaphe, *Anal. Biochem.* **44** (1971) 636–641.

[25] T. J. Låås, *J. Chromatogr.* **66** (1972) 347–355.

[26] T. G. L. Hickson, A. Polson, *Biochim. Biophys. Acta* **168** (1965) 43–58.

[27] S. Arnott, A. Fulmer, W. E. Scott, I. C. M. Dea, R. Moorhourse, D. A. Rees, *J. Mol. Biol.* **90** (1974) 269–284.

[28] P. G. Righetti in R. C. Allen, R. Arnaud (eds.): *Electrophoresis '81,* De Gruyter, Berlin 1981, pp. 3–16.

[29] J. J. Opplt in L. A. Lewis, J. J. Opplt (eds.): *Handbook of Electrophoresis*, CRC Press, Boca Raton 1980, pp. 151–180.

[30] N. C. Stellwagen, *Adv. Electrophor.* **1** (1987) 177–228.

[31] P. Serwer, *Electrophoresis (Weinheim Fed. Repub. Ger.)* **4** (1983) 375–382.

[32] S. Hjertén, *J. Chromatogr.* **12** (1963) 510–526.

[33] P. Serwer, E. T. Moreno, G. A. Griess in C. Schafer-Nielsen (ed.): *Electrophoresis '88*, VCH, Weinheim 1988, pp. 216–222.

[34] P. G. Righetti, B. C. W. Brost, R. S. Snyder, *J. Biochem. Biophys. Methods* **4** (1981) 347–363.

[35] P. G. Righetti, *J. Biochem. Biophys. Methods* **19** (1989) 1–20.

[36] J. Uriel, J. Berges, *C. R. Hebd. Seances Acad. Sci. Ser. 3* **262** (1966) 164–170.

[37] A. C. Peacock, C. W. Dingman, *Biochemistry* **7** (1968) 668–673.

[38] E. Boschetti in P. G. D. Dean, W. S. Johnson, F. A. Middle (eds): *Affinity Chromatography*, IRL Press, Oxford 1985, pp. 11–15.

[39] A. Chrambach, D. Rodbard, *Science (Washington D.C.)* **172** (1971) 440–451.

[40] S. Raymond, L. Weintraub, *Science (Washington D.C.)* **130** (1959) 711–712.

[41] B. J. Davis, *Ann. N.Y. Acad. Sci.* **121** (1964) 404–427.

[42] L. Ornstein, *Ann. N.Y. Acad. Sci.* **121** (1964) 321–349.

[43] S. Hjertén, *J. Chromatogr.* **11** (1963) 66–70.

[44] S. Hjertén, *Anal. Biochem.* **3** (1962) 109–118.

[45] S. Hjertén, *Arch. Biochem. Biophys. Suppl. 1* (1962) 276–282.

[46] S. Hjertén, S. Jersted, A. Tiselius, *Anal. Biochem.* **27** (1969) 108–129.

[47] J. H. Bode in B. J. Radola (ed.): *Electrophoresis '79*, De Gruyter, Berlin 1980, pp. 39–52.

[48] P. B. H. O'Connell, C. J. Brady, *Anal. Biochem.* **76** (1976) 63–76.

[49] H. S. Anker, *FEBS Lett.* **7** (1970) 293–296.

[50] P. N. Paus, *Anal. Biochem.* **42** (1971) 327–376.

[51] J. N. Hansen, *Anal. Biochem.* **76** (1976) 37–44.

[52] C. Gelfi, P. G. Righetti, *Electrophoresis (Weinheim Fed. Repub. Ger.)* **2** (1981) 213–219.

[53] A. Bianchi-Bosisio, C. Loherlein, R. S. Snyder, P. G. Righetti, *J. Chromatogr.* **189** (1980) 317–330.

[54] P. G. Righetti, C. Gelfi, M. L. Bossi, E. Boschetti, *Electrophoresis (Weinheim Fed. Repub. Ger.)* **8** (1987) 62–70.

[55] B. Kozulic, K. Mosbach, M. Pietrzak, *Anal. Biochem.* **170** (1988) 478–484.

[56] C. Gelfi, P. De Besi, A. Alloni, P. G. Righetti, *J. Chromatogr.* **608** (1992) 333–341.

[57] G. Artoni, E. Gianazza, M. Zanoni, C. Gelfi, M. C. Tanzi, C. Barozzi, P. Ferruti, P. G. Righetti, *Anal. Biochem.* **137** (1984) 420–428.

[58] M. Chiari, C. Micheletti, M. Nesi, M. Fazio, P. G. Righetti, *Electrophoresis* **15** (1994) 177–186.

[59] E. Simò-Alfonso, C. Gelfi, R. Sebastiano, A. Cittero, P. G. Righetti, *Electrophoresis* **17** (1996) 723–731; **17** (1996) 732–737; **17** (1996) 738–743.

[60] C. Gelfi, M. Perego, F. Libbra, P. G. Righetti, *Electrophoresis* **17** (1996) 1342–1347.

[61] E. Simò-Alfonso, C. Gelfi, M. Lucisano, P. G. Righetti, *J. Chromatogr. A* **756** (1996) 255–262.

[62] E. Bordini, M. Hamdan, P. G. Righetti, *Rapid Commun. Mass Spectrom.* **13** (1999) 2209–2215.

[63] K. A. Ferguson, *Metab. Clin. Exp.* **13** (1964) 985–995.

[64] H. R. Maurer: *Disc Electrophoresis*, De Gruyter, Berlin 1972.

[65] A. Chrambach in C. Schafer-Nielsen (ed.): *Electrophoresis '88*, VCH, Weinheim 1988, pp. 28–40.

[66] H. Svensson, *Acta Chem. Scand.* **15** (1961) 325–341;**16** (1962) 456–466.

[67] O. Vesterberg, *Acta Chem. Scand.* **23** (1969) 2653–2666.

[68] L. Söderberg, D. Buckley, G. Hagström, *Protides Biol. Fluids* **27** (1980) 687–691.

[69] C. Schafer-Nielsen in C. Schafer-Nielsen (ed.): *Electrophoresis '88*, VCH, Weinheim 1988, pp. 41–48.

[70] A. L. Shapiro, E. Vinuela, J. V. Maizel, Jr., *Biochem. Biophys. Res. Commun.* **28** (1967) 815–822.

[71] A. Helenius, K. Simons, *Biochim. Biophys. Acta* **415** (1975) 29–79.

[72] J. L. Neff, N. Munez, J. L. Colburn, A. F. de Castro in R. C. Allen, P. Arnauds (eds.): *Electrophoresis '81*, De Gruyter, Berlin 1981, pp. 49–63.

[73] J. F. Poduslo, *Anal. Biochem.* **114** (1981) 131–140.

[74] P. Lambin, *Anal. Biochem.* **85** (1978) 114–124.

[75] J. Margolis, K. G. Kenrick, *Anal. Biochem.* **25** (1968) 347–358.

[76] P. Lambin, J. M. Fine, *Anal. Biochem.* **98** (1979) 160–168.

[77] G. M. Rothe, M. Purkhanbaba, *Electrophoresis (Weinheim Fed. Repub. Ger.)* **3** (1982) 33–42.

[78] G. M. Rothe, W. D. Maurer in M. J. Dunn (ed.): *Gel Electrophoresis of Proteins*, Wright, Bristol 1986, pp. 37–140.

[79] N. G. Anderson, N. L. Anderson, *Clin. Chem. (Winston Salem N.C.)* **28** (1982) 739–748.

[80] T. Barrett, H. J. Gould, *Biochim. Biophys. Acta* **294** (1973) 165–170.

[81] P. O'Farrell, *J. Biol. Chem.* **250** (1975) 4007–4021.

[82] J. Klose, *Humangenetik* **26** (1975) 231–243.

[83] G. A. Scheele, *J. Biol. Chem.* **250** (1975) 5375–5385.

[84] R. A. Colbert, J. M. Amatruda, D. S. Young, *Clin. Chem. (Winston Salem N.C.)* **30** (1984) 2053–2058.

[85] R. H. Aebersold, J. Leavitt, L. E. Hood, S. B. H. Kent in K. Walsh (ed.): *Methods in Protein Sequence Analysis*, Humana Press, Clifton 1987, pp. 277–294.

[86] D. S. Young, N. G. Anderson (eds.): Special Issue on Two Dimensional Electrophoresis, *Clin. Chem.* **28** (1982) 737–1092.

[87] J. S. King (ed.): Special Issue on Two Dimensional Electrophoresis, *Clin. Chem.* **30** (1984) 1897–2108.

[88] J. E. Celis, R. Bravo (eds.): *Two Dimensional Gel Electrophoresis of Proteins*, Academic Press, Orlando 1984, pp. 1–487.

[89] B. S. Dunbar: *Two-Dimensional Electrophoresis and Immunological Techniques*, Plenum Press, New York 1987, pp. 1–372.

[90] M. J. Dunn (ed.): *2-D PAGE '91*, Zebra Printing, Perivale 1991, pp. 1–325.

[91] C. Damerval, D. de Vienne (eds.): Paper Symposium: Two Dimensional Electrophoresis of Plant Proteins, *Electrophoresis* **9** (1988) 679–796.

[92] J. E. Celis (ed.): Paper Symposium: Protein Databases in Two Dimensional Electrophoresis, *Electrophoresis* **10** (1989) 71–164.

[93] J. E. Celis (ed.): Paper Symposium: Cell Biology, *Electrophoresis* **11** (1990) 189–280.

[94] J. E. Celis (ed.): Paper Symposium: Two Dimensional Gel Protein Databases, *Electrophoresis* **11** (1990) 987–1168.

[95] J. E. Celis (ed.): Paper Symposium: Two Dimensional Gel Protein Databases, *Electrophoresis* **12** (1991) 763–996.

[96] J. E. Celis (ed.): Paper Symposium: Two Dimensional Gel Protein Databases, *Electrophoresis* **13** (1992) 891–1062.

[97] J. E. Celis (ed.): Paper Symposium: Electrophoresis in Cancer Research, *Electrophoresis* **15** (1994) 307–556.

[98] J. E. Celis (ed.): Paper Symposium: Two Dimensional Gel Protein Databases, *Electrophoresis* **15** (1994) 1347–1492.

[99] J. E. Celis (ed.): Paper Symposium: Two Dimensional Gel Protein Databases, *Electrophoresis* **16** (1995) 2175–2264.

[100] J. E. Celis (ed.): Paper Symposium: Two Dimensional Gel Protein Databases, *Electrophoresis* **17** (1996) 1653–1798.

[101] J. E. Celis (ed.): Genomics and Proteomics of Cancer, *Electrophoresis* **20** (1999) 223–243.

[102] M. J. Dunn (ed.): Paper Symposium: Biomedical Applications of Two-Dimensional Gel Electrophoresis, *Electrophoresis* **12** (1991) 459–606.

[103] M. J. Dunn (ed.): 2D Electrophoresis: from Protein Maps to Genomes, *Electrophoresis* **16** (1995) 1077–1326.

[104] M. J. Dunn (ed.): From Protein Maps to Genomes, Proceedings of the Second Siena Two-Dimensional Electrophoresis Meeting, *Electrophoresis* **18** (1997) 305–662.

[105] M. J. Dunn (ed.): From Genome to Proteome: Proceedings of the Third Siena Two-Dimensional Electrophoresis Meeting, *Electrophoresis* **20** (1999) 643–845.

[106] F. Lottspeich (ed.): Paper Symposium: Electrophoresis and Amino Acid Sequencing, *Electrophoresis* **17** (1996) 811–966.

[107] B. Tümmler (ed.): Microbial Genomes: Biology and Technology, *Electrophoresis* **19** (1998) 467–624.

[108] K. L. Williams (ed.): Strategies in Proteome Research, *Electrophoresis* **19** (1998) 1853–2050.

[109] I. Humphrey-Smith (ed.): Paper Symposium: Microbial Proteomes, *Electrophoresis* **18** (1997) 1207–1497.

[110] R. D. Appel, M. J. Dunn, D. F. Hochstrasser (eds): Paper Symposium: Biomedicin and Bioinformatics, *Electrophoresis* **18** (1997) 2703–2842.

[111] R. D. Appel, M. J. Dunn, D. F. Hochstrasser (eds): Paper Symposium: Biomedicin and Bioinformatics, *Electrophoresis* **20** (1999) 3481–3686.

[112] P. Cash (ed.): Paper Symposium: Microbial Proteomes, *Electrophoresis* **20** (1999) 2149–2285.

[113] M. R. Wilkins, K. L. Williams, R. D. Appel, D. F. Hochstrasser (eds): *Proteome Research: New Frontiers in Functional Genomics*, Springer, Berlin 1997.

[114] S. M. Hanash, in B. D. Hames (ed.): *Gel Electrophoresis of Proteins: a Practical Approach*, Oxford University Press, Oxford 1998, pp. 189–211.

[115] F. Kohlrausch, *Ann. Phys. Chem.* **62** (1897) 209–234.

[116] F. M. Everaerts, J. L. Becker, T. P. E. M. Verheggen: *Isotachophoresis: Theory, Instrumentation and Applications*, Elsevier, Amsterdam 1976.

[117] P. Gebauer, V. Dolnik, M. Deml, P. Bocek, *Adv. Electrophor.* **1** (1987) 281–359.

[118] S. G. Hjalmarsson, A. Baldesten, *CRC Crit. Rev. Anal. Chem.* **18** (1981) 261–352.

[119] L. Krivankovà, P. Bocek, in M. G. Khaledi (ed.): *High Performance Capillary Electrophoresis*, John Wiley & Sons, New York 1998, pp. 251–275.

[120] C. B. Laurell, *Anal. Biochem.* **15** (1966) 45–52.

[121] H. G. M. Clarke, T. Freeman, *Protides Biol. Fluids* **14** (1967) 503–509.

[122] I. Sondergaard, L. K. Poulsen, M. Hagerup, K. Conradsen, *Anal. Biochem.* **165** (1987) 384–391.

[123] J. Kroll in N. H. Axelsen, J. Kroll, B. Weeke (eds.): *A Manual of Quantitative Immunoelectrophoresis*, Universitetsforlaget, Oslo 1973, pp. 61–67.

[124] P. J. Svendsen, N. H. Axelsen, *J. Immunol. Methods* **2** (1972) 169–176.

[125] P. J. Svendsen, C. Rose, *Sci. Tools* **17** (1970) 13–17.

[126] N. H. Axelsen, J. Kroll, B. Weeke (eds.): "A Manual of Quantitative Immunoelectrophoresis," *Scand. J. Immunol. Suppl.* **1** (1973) no. 2, pp. 1–169.

[127] N. H. Axelsen (ed.): "Quantitative Immunoelectrophoresis," *Scand. J. Immunol. Suppl.* **2** (1975) pp. 1–230.

[128] N. H. Axelsen (ed.), *Scand. J. Immunol. Suppl.* **10** (1983) no. 17, pp. 1–280.

[129] C. R. Merril, R. C. Switzer, M. L. Van Keuren, *Proc. Natl. Acad. Sci. U.S.A.* **76** (1979) 4335–4339.

[130] M. Moeremans, G. Daneels, J. De Mey, *Anal. Biochem.* **145** (1985) 315–321.

[131] M. Moeremans, D. De Raeymaeker, G. Daneels, J. De Mey, *Anal. Biochem.* **153** (1986) 18–22.

[132] E. M. Southern, *J. Mol. Biol.* **98** (1975) 503–510.

[133] H. Towbin, T. Staehelin, J. Gordon. *Proc. Natl. Acad. Sci. U.S.A.* **76** (1979) 4350–4355.

[134] H. Towbin, J. Gordon, *J. Immunol. Methods* **72** (1984) 313–340.

[135] O. J. Bjerrum, N. H. H. Heegaard, *J. Chromatogr.* **470** (1989) 351–367.

[136] V. Neuhoff, R. Stamm, H. Eibl, *Electrophoresis* **6** (1985) 427–448.

[137] C. R. Merril, K. M. Washart, in: *Gel Electrophoresis of Proteins, a Practical Approach*. Oxford University Press, Oxford 1998, pp. 53–92, 319–343.

[138] B. Bjellqvist, K. Ek, P. G. Righetti, E. Gianazza, A. Görg, W. Postel, R. Westermeier, *J. Biochem. Biophys. Methods* **6** (1982) 317–339.

[139] M. Chiari, E. Casale, E. Santaniello, P. G. Righetti, *Appl. Theor. Electrophor.* **1** (1989) 99–102; *ibid.* **1** (1989) 103–107.

[140] E. Gianazza, F. Celentano, G. Dossi, B. Bjellqvist, P. G. Righetti, *Electrophoresis* **5** (1984) 88–97.

[141] G. Dossi, F. Celentano, E. Gianazza, P. G. Righetti, *J. Biochem. Biophys. Methods* **7** (1983) 123–142.

[142] P. G. Righetti, M. Chiari, P. K. Sinha, E. Santaniello, *J. Biochem. Biophys. Methods* **16** (1988) 185–192.

[143] C. Gelfi, M. L. Bossi, B. Bjellqvist, P. G. Righetti, *J. Biochem. Biophys. Methods* **15** (1987) 41–48; P. K. Sinha, P. G. Righetti, *ibid.* **15** (1987) 199–206.

[144] P. G. Righetti, E. Gianazza, C. Gelfi, M. Chiari, P. K. Sinha, *Anal. Chem.* **61** (1989) 1602–1612.

[145] P. G. Righetti, C. Gelfi, M. Chiari, in B. L. Karger, W. S. Hancock (eds.): *Methods in Enzymology: High Resolution Separation and Analysis of Biological Macromolecules, Part A: Fundamentals*, **vol. 270**, Academic Press, San Diego 1996, pp. 235–255.

[146] P. G. Righetti, A. Bossi, *Anal. Biochem.* **247** (1997) 1–10.

[147] P. G. Righetti, A. Bossi, *J. Chromatogr. B* **699** (1997) 77–89.

[148] P. G. Righetti, A. Bossi, C. Gelfi, in B. D. Hames (ed.): *Gel Electrophoresis of Proteins*, 3rd ed., Oxford University Press, Oxford 1998, pp. 127–187.

[149] J. W. Jorgenson in J. W. Jorgenson, M. Phillips (eds.): "New Directions in Electrophoretic Methods," *ACS Symp. Ser. 335*, Am. Chem. Soc., Washington, 1987, pp. 70–93.

[150] D. J. Rose, Jr., J. W. Jorgenson, *J. Chromatogr.* **447** (1988) 117–131.

[151] F. Foret, P. Bocek, *Advances Electrophor.* **3** (1989) 271–347.

[152] B. L. Karger (ed.): "Proceedings of the 1st International Symposium on High-Performance Capillary Electrophoresis," *J. Chromatogr.* **480** (1989) 1–435.

[153] B. L. Karger (ed.): "Proceedings of the 2nd International Symposium on High-Performance Capillary Electrophoresis," *J. Chromatogr.* **516** (1990) 1–298.

[154] J. W. Jorgenson (ed.): "Proceedings of the 3rd International Symposium on High-Performance Capillary Electrophoresis," *J. Chromatogr.* **559** (1991) 1–561.

[155] F. M. Everaerts, T. P. E. M. Verheggen (eds.): "Proceedings of the 4th International Symposium on High-Performance Capillary Electrophoresis," *J. Chromatogr.* **608** (1992) 1–429.

[156] B. L. Karger (ed.): "Proceedings of the 5th International Symposium on High-Performance Capillary Electrophoresis," *J. Chromatogr. A* **652** (1993) 1–574.

[157] B. L. Karger, S. Terabe (eds.): "Proceedings of the 6th International Symposium on High-Performance Capillary Electrophoresis," *J. Chromatogr. A* **680** (1994) 1–689.

[158] H. Engelhardt (ed.): "Proceedings of the 7th International Symposium on High-Performance Capillary Electrophoresis," *J. Chromatogr. A* **716** (1995) 1–412; **717** (1995) 1–431.

[159] B. L. Karger (ed.): "Proceedings of the 8th International Symposium on High-Performance Capillary Electrophoresis," *J. Chromatogr. A* **744** (1996) 1–354; **745** (1996) 1–303.

[160] W. S. Hancock (ed.): "Proceedings of the 9th International Symposium on High-Performance Capillary Electrophoresis," *J. Chromatogr. A* **781** (1997) 1–568.

[161] S. Fanali, B. L. Karger (eds.): "Proceedings of the 10th International Symposium on High-Performance Capillary Electrophoresis," *J. Chromatogr. A* **817** (1998) 1–382.

[162] E. S. Yeung (ed.): "Proceedings of the 11th International Symposium on High-Performance Capillary Electrophoresis," *J. Chromatogr. A* **853** (1999) 1–576.

[163] A. M. Krstulovic (ed.): "Capillary Electrophoresis in the Life Sciences," *J. Chromatogr. B* **697** (1997) 1–290.

[164] Y. Baba (ed.): "Analysis of DNA by Capillary Electrophoresis," *J. Biochem. Biophys. Methods* **41** (1999) 75–165.

[165] B. L. Karger (ed.): Paper Symposium: Capillary Electrophoresis, *Electrophoresis* **14** (1993) 373–558.

[166] N. J. Dovichi (ed.): Paper Symposium: Nucleic Acid Electrophoresis, *Electrophoresis* **17** (1996) 1407–1517.

[167] F. Foret, P. Bocek (eds.): Paper Symposium: Capillary Electrophoresis: Instrumentation and Methodology, *Electrophoresis* **17** (1996) 1801–1963.

[168] J. P. Landers (ed.): Paper Symposium: Capillary Electrophoresis in the Clinical Sciences, *Electrophoresis* **18** (1997) 1709–1905.

[169] Z. El Rassi (ed.): "Capillary Electrophoresis and Electrochromatography," *Electrophoresis* **18** (1997) 2123–2501.

[170] B. R. McCord (ed.): Paper Symposium: Capillary Electrophoresis in Forensic Science, *Electrophoresis* **19** (1998) 3–124.

[171] Z. El Rassi (ed.): Paper Symposium: Capillary Electrophoresis and Electrochromatography Reviews, *Electrophoresis* **20** (1999) 2989–3328.

[172] P. G. Righetti (ed.): *Capillary Electrophoresis: an Analytical Tool in Biotechnology*, CRC Press, Boca Raton, FL 1996.

[173] M. G. Khaledi (ed.): *High Performance Capillary Electrophoresis: Theory, Techniques and Applications*, John Wiley & Sons, New York 1998.

[174] P. G. Righetti, C. Gelfi, *J. Biochem. Biophys. Methods* **41** (1999) 75–90.

[175] P. G. Righetti, A. Bossi, E. Olivieri, C. Gelfi, *J. Biochem. Biophys. Methods* **40** (1999) 1–16.

[176] O. Vesterberg, T. Waldström, K. Vesterberg, H. Svensson, B. Malmgren, *Biochim. Biophys. Acta* **133** (1967) 435–445.

[177] B. J. Radola, *Ann. N.Y. Acad. Sci.* **209** (1973) 127–143.

[178] B. J. Radola, *Biochim. Biophys. Acta* **295** (1973) 412–428.

[179] M. Bier, N. Egen in H. Haglund, J. C. Westerfeld, J. T. Ball Jr. (eds.): *Electrofocus '78*, Elsevier, Amsterdam 1979, pp. 35–48.

[180] M. Bier, N. Egen, T. T. Allgyer, G. E. Twitty, R. A. Mosher in E. Gross, J. Meienhofer (eds.): *Peptides: Structure and Biological Function*, Pierce Chem. Co., Rockford 1979, pp. 79–89.

[181] N. Egen, W. Thormann, G. E. Twitty, M. Bier in H. Hirai (ed.): *Electrophoresis '83*, De Gruyter, Berlin 1984, pp. 547–550.

[182] M. Bier, T. Long, *J. Chromatogr.* **604** (1992) 73–83.

[183] P. G. Righetti, E. Wenisch, M. Faupel, *J. Chromatogr.* **475** (1989) 293–309.

[184] P. G. Righetti, E. Wenisch, A. Jungbauer, H. Katinger, M. Faupel, *J. Chromatogr.* **500** (1990) 681–696.

[185] P. G. Righetti, E. Wenisch, M. Faupel, *Advances Electrophor.* **5** (1992) 159–200.

[186] P. Wenger, M. de Zuanni, P. Javet, P. G. Righetti, *J. Biochem. Biophys. Methods* **14** (1987) 29–43.

15. Structure Analysis by Diffraction

ERICH F. PAULUS, Institute of Mineralogy and Crystallography of the University of Frankfurt, Frankfurt/Main, Federal Republic of Germany

ALFRED GIEREN, formerly Institute for Analytical Chemistry and Radiochemistry, University of Innsbruck, Innsbruck, Austria

15.1. General Principles

When electromagnetic or corpuscular radiation is diffracted by matter, i.e., when it is involved in an elastic interaction with matter, a scattering diagram or diffraction pattern is produced; this pattern is the Fourier transform of the diffracting object. If the inverse transform is applied, going from "reciprocal" to "direct" space, an image of the object is obtained. This inverse transformation is called Fourier synthesis; the transformation from direct to reciprocal space is called Fourier analysis. The principle of Fourier analysis (diffraction) followed by Fourier synthesis (imaging) controls the classical process of magnification in an optical microscope. Here, an object is irradiated, and elastic scattering takes place. The objective lens causes the scattered rays to interfere, generating a magnified image. To apply this principle at atomic resolution (ca. 10^{-10} m), the first thing needed is radiation with wavelength of the same order of magnitude (maximum resolution is $\lambda/2$)

[55]. A further condition is that the destruction caused by the interaction of the radiation with the object should be as slight as possible, so that the image is not substantially degraded by radiation damage.

The chief kinds of radiation used are X rays, neutrons, and electrons. X rays and neutrons cause relatively little damage, but electrons have a scattering power several orders of magnitude greater, because of their charge, which leads to more radiation damage. Electrons have the advantage that they can be imaged with lenses in a microscope configuration. The use of a lens for imaging (Fourier synthesis) is not, however, essential. Rapid advances in electronic computers, make it possible to replace the lens by a computer. However, this is possible only if all the relevant parameters of the scattering diagram, including the phases of the scattered waves, are known. As a rule, the phases of X ray and neutron radiation are not accessible to direct measurement, so that computer-intensive techniques are required to obtain them.

X-ray and neutron structure analysis generally deal with crystalline matter, either single crystals or crystalline powders. Nonperiodic objects, such as glasses, play only a secondary role.

15.2. Structure Analysis of Solids

15.2.1. Diffraction by Crystal Lattices

With a few exceptions, this article deals with the crystalline solid state; other aggregations such as liquid crystals are not covered.

Naturally occurring crystals differ from ideal crystals in two main aspects:

1) They are finite with surfaces
2) They contain structural defects and vacancies, e.g., the mosaic structure of real crystals

The crystals available in practice thus lie somewhere between amorphous and ideally crystalline, but nearer the ideal end of the scale ("ideal imperfect crystals"). A crystalline specimen may be a single crystal (smallest dimension ca. $0.05-0.5$ mm) or a polycrystalline material with crystallite size ca. 0.04 mm.

High-polymer plastics occupy an intermediate position. According to the two-phase model {56] the amorphous phase is present along with the crystalline phase. The unit cells in the crystalline portion may be oriented in some preferred direction (as in fibers), or there may be no preferred direction.

15.2.1.1. Materials and Methods

"Single crystals" are the classical objects investigated by crystallography. The diffraction of X rays by crystals was discovered in 1912 by VON LAUE, FRIEDRICH, and KNIPPING in a study of copper sulfate pentahydrate crystals. It was later found that diffraction also takes place from crystalline powders. The methodological principles are explained here for single crystals and to a certain extent also for crystalline powder.

Single-crystal structure analysis, the determination of the structure of the asymmetric unit of a crystal, is so far the only analytical method that can yield direct and unambiguous information about the three-dimensional atomic structure of a substance. Crystalline powder structure analysis needs in the average also a model.

Simple Inorganic Compounds. The class of simple inorganic compounds includes substances such as metals and salts, that contain very few atoms in the asymmetric unit. In such cases, the complete crystal structure can often be derived as soon as the crystal symmetry (space group) has been determined. If several possibilities exist, the correct structure can be arrived at by trial and error. When this approach also fails, a Patterson synthesis (see Eq. 27) can be used to solve the structure.

Complex Inorganic and Organic Compounds. Two methods can be used for the solution of a more complex molecular and crystal structure problem.

1) Patterson synthesis followed by successive Fourier syntheses of the electron density
2) "Direct" determination of the phases of the structure factors, possibly followed by successive Fourier syntheses of the electron density (direct methods)

These methods complement each other in many cases. When there are several possible solutions for the Patterson function, some may be eliminated if the direct methods yield a possible solution by which the Patterson synthesis can be interpreted. The contrary can also occur, when a Patterson synthesis is used to decide which phase set is correct.
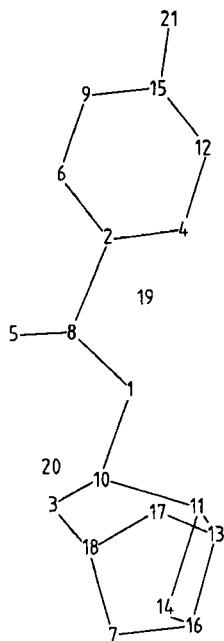
Figure 1. Computer-generated molecular structure diagram after solution of a crystal structure by direct methods

A prerequisite for the use of Patterson synthesis is that concrete assumptions about the complete molecular structure exist, or at least that parts of the structure are known. The position and orientation of the known molecular part can then be determined, for example with the aid of the convolution molecule method [57] or minimum search programs. In the case of heavy-atom derivatives, the known molecular part consists of one or several heavy atoms [58].

Direct methods are most commonly used to solve crystal structures. A number of advanced program systems exist and, in many cases, make solution of the phase problem a routine matter for the expert. The proposed structure, which often proves to be correct and complete, can be output directly in graphical form (Fig. 1). The numbers represent atomic positions; a small numerical value indicates a larger electron density maximum and a large value indicates a smaller one. The program determines which electron density maxima should be connected to each other to obtain the structure; tables are generated that show the positions of the atoms in the cluster (see also Table 3). However, if this method does not succeed immediately, generally much experience, including an understanding of the principles of direct methods, is required to solve the phase problem. Given

sufficient time, virtually all structure problems (ca. 95 %) satisfying certain minima criteria are solvable; these include structures with up to 100 and many more nonhydrogen atoms in the asymmetric unit (the atomic cluster from which the entire unit cell can be generated by symmetry operations), and correspond to those structures of greatest interest to the synthetic chemist. If the data set is improved by cooling the crystal (below– 100 °C) to reduce the thermal vibrations of the atoms, this limit can be extended in special cases to more than 300.

Proteins. In these much larger molecules, it is often possible to determine the phases by the method of isomorphous replacement. Most commonly, this means allowing ions of heavy atoms to diffuse into the solvent-filled vacancies in protein crystals, where they adopt well-defined positions. Thus, this represents the incorporation of atoms in completely new positions, rather than an atom-replacement process. One exception is insulin, in which zinc ions are replaced by lead or mercury ions. Newer methods allow the structure to be solved by direct methods when excellent protein crystals are available.

A number of techniques have been devised for determining the positions of heavy atoms in the unit cells of proteins [6, pp. 337 – 371], but these are not discussed here.

In the noncentrosymmetric case, which always holds for proteins, phase determination of the structure factors requires at least two heavy-atom derivatives to estimate the phases for the native protein from the heavy atom positions [6, pp. 337 – 371]. As a rule, three or four derivatives are investigated in order to reduce phase errors.

All reflections whose phases are to be determined have to be measured on the native protein and the isomorphous crystals. A computer-controlled diffractometer is used. Alternatively, photographic data and automatic photometers can be employed; this has the advantage that the crystals do not have to be exposed to X rays for such a long time, reducing radiation damage. Recently, films have been replaced by area detectors which produce the intensities in digital form. If a synchrotron is used as radiation source, and the pattern is recorded by the Laue method (see Section 15.2.1.5), data acquisition takes only a few minutes.

The extent to which the Fourier synthesis of the electron density can be interpreted depends on the resolution with which the data could be meas-

ured. Resolution is said to be low when the smallest lattice plane spacings to which measured reflections can be assigned are $> 5 \times 10^{-10}$ m; aside from the dimensions and overall shape of the molecule, no detail can be extracted from the Fourier synthesis. At a resolution of 3×10^{-10} m, the shape of polypeptide chains can be identified; at 1.5×10^{-10} m, single atoms can be distinguished and side chains can be identified without knowledge of the amino acid sequence. The atomic positions cannot be refined by conventional least-squares methods, because the computational problem becomes immense and the number of observed data seldom exceeds the number of parameters. For special methods, see [6, pp. 420 – 442].

Supplementary Structure Investigations by Neutron Diffraction. Single-crystal X-ray structure analysis cannot determine the positions of hydrogen atoms in many cases. Even with error-free measurements, there are essentially two reasons:

1) Because hydrogen atoms generally occupy peripheral locations in the molecule, they are smeared out by static and/or dynamic disorder (small conformational differences), and their scattering contribution at higher scattering angles is not significantly higher than the background.

2) The compound under study may contain not only light atoms, but also heavy atoms, such as bromine and iodine. The intensity measurements are not accurate enough to detect the small relative intensity differences due to hydrogen atoms.

In most such cases, well-known stereochemical rules make it possible to calculate the positions of most hydrogen atoms from those of the other atoms. But if one wishes to investigate, say, hydrogen bonds more closely, the use of neutron diffraction is indicated, because H or, better, D atoms (because of the smaller incoherent scattering) are strong neutron scatterers.

Neutron diffraction analysis may also be necessary when elements near one another in the periodic table need to be distinguished. A well-known example is differentiation between the isoelectronic ions Si^{4+} and Al^{3+}.

By virtue of their magnetic moment, neutrons can also be used to investigate magnetic structures [5, pp. 395 – 511], [59].

Chemical bonds cause a displacement of the bonding electrons. Because X rays are also scattered by these electrons, a slight distortion of the apparent atomic positions is unavoidable. For example, X-ray diffraction always gives carbon –hydrogen bond lengths that are too short, because the hydrogen electron is pulled closer to its bonding partner. Neutron diffraction analysis, in contrast, determines the position of the atomic nucleus, not the maximum of the electron density distribution. The deformation density resulting from chemical bonding may be directly revealed by difference Fourier syntheses of electron densities from X-ray diffraction data, and the structure factors calculated from the atomic positions obtained by neutron diffraction analysis (X – N method).

The measurements must be performed with the utmost care, because the effects are comparable to the standard deviations. X-ray and neutron diffraction patterns should be recorded at the lowest possible temperatures, to minimize thermal motion.

For reasons of cost, neutron diffraction is virtually always preceded by single-crystal X-ray structure analysis. In many cases, the deformation densities can be determined by X-ray measurements alone, in view of the fact that the scattering contribution of the bonding electrons vanishes at large scattering angles (X – X method).

15.2.1.2. Introduction to X-Ray Crystallography; History

DEMOCRITOS of Abdera (born ca. 465 B.C.) wrote, "It is customary to say that there is color, sweetness, bitterness; in actuality, however, there are atoms and void" [60]. In this sentence, the atom, as the fundamental building block of matter, was introduced into the discussion of the structure of matter. In 1807, DALTON had only to invoke this hypothesis in order to impose order on the experimental facts known at the time. For example, a crystalline structure consisting of identical particles was also in agreement with these principles.

NIKOLAUS STENO, as early as 1669, formulated the law of constant angles, which states that two like edges of crystals of the same substance always make equal angles with each other. In 1823, FRANZ ERNST NEUMANN discovered the law of simple rational indices. Planes drawn through faces of a crystal intersect three coordinate axes, constructed parallel to crystal edges, at minima lengths a, b, c that are in a constant ratio to one another. The other faces of the crystal can be
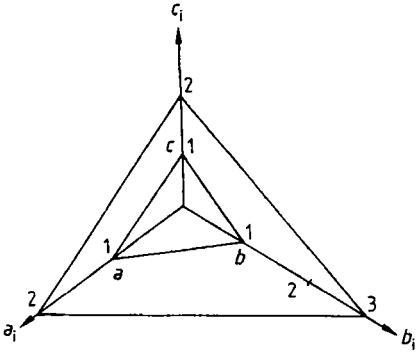
Figure 2. The law of simple rational indices
Planes shown are: $u = v = w = 1$; and $u = 2$, $v = 3$, $w = 2$
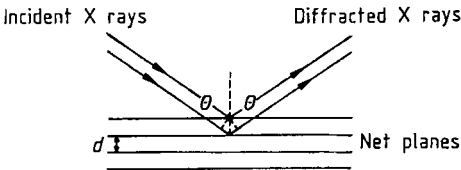
Figure 3. Derivation of the Bragg equation

Figure 4. Ewald construction
a) Incident beam; b) Reflection sphere; c) Limiting sphere; d) Crystal; e) Diffracted beam; f) Lattice points on a plane of the reciprocal lattice perpendicular to the plane of the diagram; g) Reflecting lattice plane

characterized by planes drawn through them that cut the axes at lengths a_i, b_i, c_i which are in whole-number (possibly negative) ratios to the intercepts a, b, c $(a_i : b_j : c_i = u \cdot a : v \cdot b : w \cdot c)$ (Fig. 2).

About 1890, FEDOROV, SCHOENFLIES, and BARLOW more or less simultaneously showed that for crystals built up from discrete particles in a three-dimensionally ordered manner, there can be no more than 230 different combinations of elements, the 230 space groups. In 1912, on the basis of their key diffraction experiment, which yielded discrete X-ray reflections from a crystal, LAUE, FRIEDRICH, and KNIPPING demonstrated both the periodic construction of crystals from atoms and the wave nature of X rays.

W. H. and W. L. BRAGG were the first to show that the models proposed by BARLOW for simple compounds, such as NaCl, CsCl, and ZnS are in agreement with the scattered X-ray intensities. PAUL NIGGLI found that space groups could be determined by X-ray methods via the systematic extinction laws.

The Bragg equation:

$$\sin\theta = n\frac{\lambda}{2d} \tag{1}$$

$$n\lambda = 2d\sin\theta \tag{2}$$

describes the diffraction of X rays by a crystal as reflection from the net planes. Constructive interference occurs only if the difference in path length between two rays is an integer multiple of the wavelength λ (Fig. 3). Here θ is the glancing angle (2θ is the diffraction angle), and d is the net plane spacing.

A construction due to EWALD illustrates the importance of the reciprocal lattice in X-ray crystallography. As Figure 4 shows, the Bragg equation is satisfied where the reflection sphere is cut by a lattice point of the reciprocal lattice constructed around the center of the crystal. Rotating the crystal together with the reciprocal lattice around a few different directions in the crystal fulfills the reflection condition for all points of the reciprocal space within the limiting sphere. The reciprocal lattice vector S is perpendicular to the set of net planes, and has the absolute magnitude $1/d$. In vector notation:

$$S = s - s_0 \tag{3}$$

Vectors s and s_0 each have modulus $1/\lambda$.

The reciprocal lattice provides a simple way to illustrate the positions of diffraction points in space. Figure 4 also shows that the Ewald limiting sphere determines how many reflections can be measured. All reflections inside the limiting sphere with radius $2/\lambda$ are obtained only if the crystal axis AB can take up any orientation. This is not so in many data acquisition methods, e.g., the rotating crystal method. If the crystal is rotated
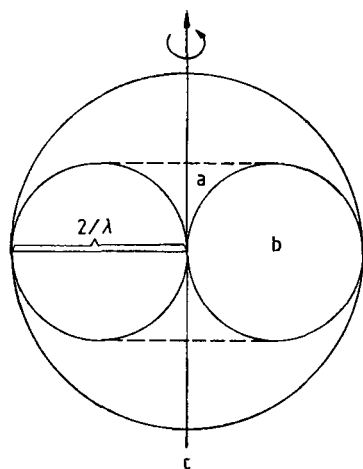
Figure 5. The region of reciprocal space that can be measured for a single setting of a rotating crystal
a) Blind spot; b) Toroidal region of reciprocal space covered for a single setting; c) Crystal axis AB

about AB, reciprocal lattice points lying on reciprocal lattice planes perpendicular to AB follow circular paths, cutting the reflection sphere. Every such cut corresponds to a reflection.

The portion of reciprocal space inside the Ewald limiting sphere that can be measured with a single crystal setting is shown in Figure 5. Almost all lattice points of the limiting sphere are registered when measurements are done with two crystal settings about axes forming an angle ca. 90°.

It is generally true that a vector of the reciprocal lattice is perpendicular to a plane in the direct lattice and, conversely, a vector of the direct lattice is perpendicular to a plane of the reciprocal lattice.

The deeper connection between the real and reciprocal lattices becomes clear when it is shown (see Eqs. 13, 18) that the absolute magnitudes of the functional values of the Fourier transform of the entire crystal are equal to the square roots of the measured intensities. The position coordinates of these functional values are coordinates in reciprocal space (i.e., the indices of the reflections) if nonzero functional values are allowed only at reciprocal lattice points. A reflection can then be defined completely generally as a reciprocal lattice point. Further details are presented in the textbooks cited at the end of this article.

15.2.1.3. Experimental Principles, Applications

If suitable single crystals of a compound are available, it is a routine matter to determine the cell constants. The crystalline specimen should not be much smaller than 0.3–0.5 mm. A crystal volume of 0.005 mm^3 is adequate for measurements on organic compounds; much smaller crystals can be used in the case of inorganic compounds with strongly scattering atoms. Air-sensitive substances can be sealed in borosilicate capillaries.

The following method has rapidly gained acceptance where it is possible or desirable to measure the crystals at low temperature (ca. –100 °C). The crystalline specimen is embedded in an oil that is liquid at the embedding temperature, but solidifies at low temperature [61]. Measurements can be performed on extremely air-sensitive substances, and on those that are liquid at room temperature. The method has the drawback that the crystal employed for data collection is normally not recoverable at the end of the measurement.

Before the introduction of laboratory computers, virtually the only way of determining cell constants was the photographic method. With the Weissenberg and precession cameras, as well as the Space Explorer, a "mixture" of cell constants in both reciprocal and direct space is obtained. The relationships between these cell constants are expressed by the following equations:

$$a = \frac{1}{V^*} b^* c^* \sin\alpha^* \tag{4}$$

$$b = \frac{1}{V^*} c^* a^* \sin\beta^* \tag{5}$$

$$c = \frac{1}{V^*} a^* b^* \sin\gamma^* \tag{6}$$

$$\cos\alpha = (\cos\beta^* \cos\gamma^* - \cos\alpha^*)/(\sin\beta^* \sin\gamma^*) \tag{7}$$

$$\cos\beta = (\cos\alpha^* \cos\gamma^* - \cos\beta^*)/(\sin\alpha^* \sin\gamma^*) \tag{8}$$

$$\cos\gamma = (\cos\alpha^* \cos\beta^* - \cos\gamma^*)/(\sin\alpha^* \sin\beta^*) \tag{9}$$

$$V^* = a^* b^* c^* \sqrt{1 + 2\cos\alpha^* \cos\beta^* \cos\gamma^* - \cos^2\alpha^* - \cos^2\beta^* - \cos^2\gamma^*} \tag{10}$$

Symbols with asterisks relate to the reciprocal lattice; a, b, c are lengths, α, β, γ are angles, and V is the cell volume. If the left-hand sides of these equations are replaced by the cell constants of the reciprocal lattice, the right-hand sides contain the corresponding lattice values in direct space.
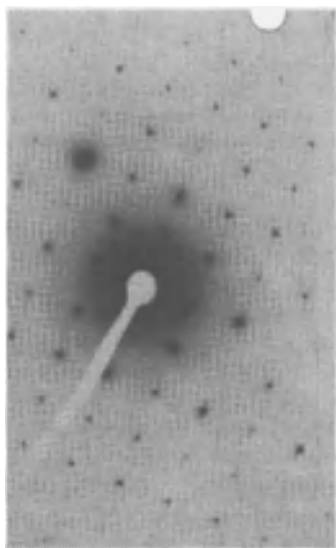
Figure 6. Part of the electron diffraction pattern of a single crystal (MoO_3)

The determination of cell constants is particularly convenient with the precession camera, in which the crystal has to be set on a goniometer head; one reciprocal lattice axis must lie parallel to the goniometer head axis. Two settings of the goniometer spindle yield two principal planes in the reciprocal lattice. The third angle of the reciprocal lattice, obtained in terms of the corresponding angle in direct space, is the difference between the two spindle settings. A photographic cone record (see also Fig. 25) represents a complete projection of the reciprocal lattice; this avoids mistakes in the choice of the cell dimensions due to missing planes of the reciprocal lattice. The first two layer lines supply the remainder of the data needed for space group determination.

Equations 4 – 10 make it possible to calculate the cell constants in direct and reciprocal space. From the symmetries present and the systematic extinction of reflections, the crystal system, the lattice type (Bravais lattice), and the space group can be derived [4, pp. 124 – 159]. Because the diffraction pattern always has a center of symmetry (Friedel's law, which holds when anomalous scattering is neglected), ambiguities in the space group determination are still possible (11 Laue groups instead of 32 crystal classes). In principle, the space group is not definite until the structure has been completely solved, although in most cases there is a high prior probability for a particular space group.

If the slightest doubt exists whether the crystal system is right, no time should be lost in carrying out a cell reduction by the Azaroff–Buerger method [4]; many computer programs are on the market for this operation, but it is also reasonably easy to perform "manually".

If a computer-controlled single-crystal diffractometer is available, the cell constants, and space group can be determined by the computer. Three methods exist, differing only in the means of determining the reflection coordinates in reciprocal space. In "peak-hunting" programs, the region of reciprocal space within specified boundaries is systematically searched for reflections [62]; the Acedeva program makes a "systematically unsystematic" search for reflections [63], with a random number generator determining the directions in which to drive the diffractometer. Normally, either of these two methods yields a number of usable reflections in reciprocal space, from which the three shortest noncoplanar vectors can be extracted and used to index all the reflections. Finally, the reduced cell and the Niggli matrix are calculated, and thus the crystal system and Bravais lattice can be derived [4]. The third method employs an oscillation photograph, made on the diffractometer by using a polaroid film, from which two-dimensional coordinates of selected reflections are obtained; the search for about ten reflections is based on these coordinates [64].

A diffractometer which performs only point measurements, does not give the cell constants unambiguously, since it deals only with discrete points in the reciprocal space. "Continuous" measurement in the reciprocal space is necessary to make certain that no essential points for unit cell determination have been missed. This fact, together with pedagogical considerations, is the reason that film techniques are still used in the X-ray diffraction laboratory. A modern four-circle diffractometer is equipped to perform the necessary photographic measurements which a three-circle diffractometer is unable to do.

Electron diffraction studies on tiny crystals ($6 \leq 10^{-8}$ m) are possible in the electron microscope. The resulting single-crystal diffraction patterns can be indexed similarly to precession photographs (Fig. 6), provided multiple orientations of the same specimen relative to the electron beam can be set up. The results should always be combined with those of X-ray powder patterns. This technique has the advantage that it does not require the laborious growth of single crystals, but its greatest disadvantage is that the intensities are
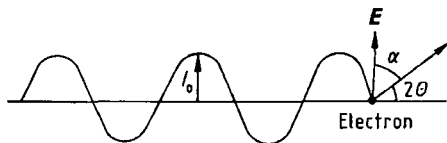
Figure 7. Scattering of X rays by an electron

often totally falsified (this is, however, of no importance for unit cell determination).

Up to this point, only the geometry of diffraction has been considered. However, the intensity ratios of the reflections are the basis for determining the structure of the unit cell contents.

For the intensity ratios to be quantified, scattering must be discussed in more detail. Consider the scattering of X rays by electrons in the atomic electron shells. The scattering power I_e of an electron at point A is given by Maxwell's theory as:

$$I_e = \frac{I_0 e^4}{r^2 m^2 c^4} \sin^2 \alpha \tag{11}$$

where I_0 is the incident X-ray intensity, e is the electron charge, m is the electron mass, r is the distance from the scattering electron to the observer, c is the speed of light, and α is the angle between the electric field vector E and the direction to the observer (Fig. 7).

If the incident X rays are unpolarized, $\sin^2 \alpha$ must be replaced by the factor $(\frac{1}{2} + \frac{1}{2} \cos^2 2\theta)$, the polarization factor. Here, 2θ is the deflection angle.

X rays employed in structure analysis normally have wavelengths $0.5 - 2.3 \times 10^{-10}$ m, the same order of magnitude as the size of atoms and their electron shells. Only at a scattering angle of 0° can e and m in Equation (11) be replaced by Ze and Zm, for an atom with atomic number Z. At other angles, the X rays scattered by the electrons interfere, and the total intensity is reduced as a function of scattering angle. The loss of intensity is greater, the higher the glancing angle θ and the larger the electron shell are.

The best-known example demonstrating the effect of atomic radius is the case of O^{2-} and Si^{4+} (important in the analysis of silicates). As θ increases, the scattered intensity falls off much faster for O^{2-} than it does for the smaller Si^{4+}. In this way, it is possible to distinguish between these two atomic species (Fig. 8). Atomic form factors f for atoms and ions [29, vol. III, pp. 201 – 246] are a function of the scattering
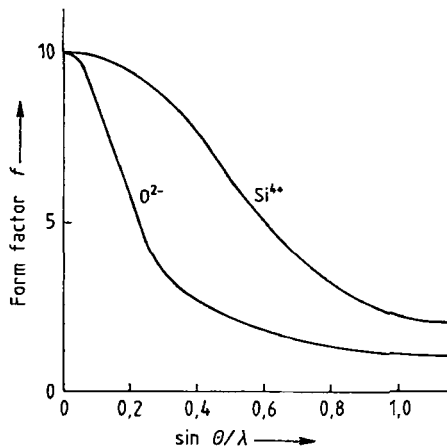
Figure 8. Atomic form factors f for O^{2-} and Si^{4+}

power of the corresponding atom, and are expressed in units of a free electron.

If the frequency of the incident X rays is close to the frequency of the K absorption edge of the irradiated atom, scattering can no longer be regarded as elastic; some of the incident energy is used to excite the electrons in the atom. The atomic form factors have to be corrected for anomalous dispersion; the tables in [29] include this correction. The significance of these corrections for the determination of the absolute configuration is discussed in Section 15.2.4.

Among the key properties of electron beams relating to their scattering by atoms, the electron charge is particularly important. Electrons are scattered by inhomogeneities in the electrostatic potential distribution, chiefly at atomic nuclei. The process is described by the equation:

$$f^c = (Z - f) \frac{e^2 m}{2h^2} \cdot \frac{\lambda^2}{\sin^2 \theta} \tag{12}$$

where Z is the atomic number, f is the atomic form factor for X rays, f^c is the atomic form factor for electron beams, and e and m are the electron charge and mass, respectively.

Neutrons are scattered only by atomic nuclei (provided there is no magnetic scattering due to unpaired $3d$ or $4f$ electrons). The size of the atomic nucleus is orders of magnitude smaller than the wavelength of thermal neutrons, and so no dependence of scattered amplitude on glancing angle can be observed.

The X-ray intensities I of reflections produced by lattice planes having Miller indices (hkl) are

measured. These are related in the following way to the absolute magnitudes of the structure factors $|F|$, i.e., the structure amplitudes:

$$\|F(hkl)\| = k\sqrt{I(hkl)} \tag{13}$$

Similarly to Equation (11), the following can be derived: Given a crystal of volume V, z unit cells per unit volume, X rays with wavelength λ, and rotation of the crystal through the reflection position at rotational frequency ω; the intensity I of a reflection having structure factor F is:

$$I = \frac{e^4}{m^2 c^4} I_0 \|F\|^2 \frac{\lambda^3 z^2 V}{\omega} \frac{1 + \cos^2 2\theta}{2} \frac{1}{\sin 2\theta} \tag{14}$$

where c is the speed of light. The last two factors depend on the recording technique, and are commonly combined in the Lorentz-polarization (LP) factor. Strictly speaking, Equation (14) holds only if the coherent interaction between incident and diffracted X rays is neglected (primary extinction), and if one ignores the fact that the incident X-ray beam is attenuated by diffraction as it propagates through the crystal (secondary extinction), so that not all portions of the crystal contribute equally to diffraction.

All these falsifications become negligibly small if the crystal has an "ideal mosaic" structure, and the "mosaic particles" are smaller than 1 µm. Generally:

$$I \sim \|F\|^n \tag{15}$$

where n is a number between 1 and 2. If $n = 1$, Equation (14) gives:

$$I = \frac{8}{3\pi} \frac{e^2}{mc^2} I_0 \|F\| \frac{\lambda^2 z}{\omega} \frac{1 + \|\cos^2 2\theta\|}{2 \sin 2\theta} \tag{16}$$

Diffraction is controlled by Equation (14) if the kinematic theory is viewed as completely valid ($n = 2$), and by Equation (16) if the dynamic theory ($n = 1$) is applied. Normally, neither of these extremes is exactly fulfilled, and there is a combination of both. The kinematic theory is usually adopted, with "dynamic" contributions being handled by an extinction correction, although this (like the absorption correction) often proves unnecessary for measurements of standard level.

The structure factor F is needed to determine the electron density ϱ at point x, y, z (in fractional units of the cell constants) and thus to determine the crystal structure:

$$\varrho(xyz) = \frac{1}{V} \sum_{h,k,l=-\infty}^{+\infty} F(hkl) \exp[-2\pi i(hx + ky + lz)] \tag{17}$$

$$F(hkl) = \sum_{n=1}^{N} f_n \exp[2\pi i(hx_n + ky_n + lz_n)] \tag{18}$$

The summation in Equation (17) is over all measurable reflections (hkl), and that in Equation (18) is over all atoms N in the unit cell; V is the cell volume; x_n, y_n, z_n are the coordinates of atom n; and f_n is its atomic form factor.

The significance of systematic extinctions for space-group determination can be seen from Equation (18). In space group $P2_1$, for example, for an atom at x, y, z there is another at $-x, y+\frac{1}{2}, -z$. The effect on the ($0k0$) reflections is that the contributions of these two atoms to the structure factor differ in phase by 180°. If k is odd they therefore cancel each other, but are additive if k is even. The following equation represents the situation for space group $P2_1$:

$$F(0k0) = \sum_{n=1}^{N/2} f_n \left[\exp(2\pi i k y_n) + \exp\left(2\pi i k \left(y_n + \frac{1}{2}\right)\right) \right] \tag{19}$$

Corresponding relations have been tabulated for all space groups [29], [30].

Equations (17) and (18) further show a relationship between real and reciprocal space. The function $F(hkl)$ is the Fourier transform of the unit cell contents, expressed in the reciprocal space coordinates h, k, and l. Because the symmetry operation of translation holds for all three spatial directions in crystals, the Fourier transform of the entire crystal is zero, except at reciprocal lattice points.

15.2.1.4. Crystal Growth for X-Ray Structure Analysis

While the growing of single crystals from the melt or the gas phase may sometimes be important for industrial purposes (as in the crucible pulling method used for semiconductors such as Si and Ge), in X-ray structure analysis such methods are secondary to crystal growth from solution, and they are employed only when other techniques are not possible.

In principle, suitable crystals can be grown in either of two ways:

Cooling water
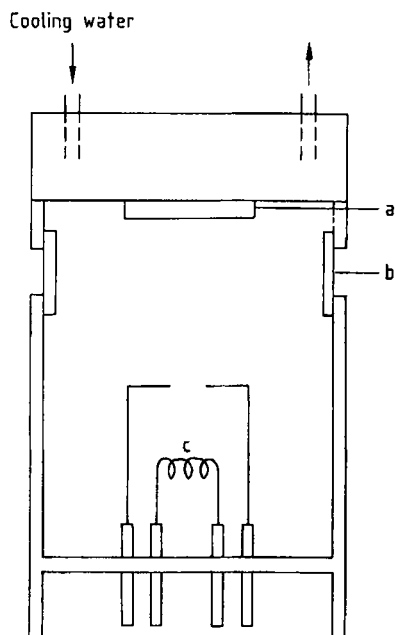
Figure 9. X-ray tube
a) Anode; b) Be window; c) Hot cathode

1) The first approach uses the temperature dependence of solubility. In accordance with the maxim "like dissolves like", the solvent employed is a compound with the same or similar parent structure, and/or having the same or similar functional groups. Normally, compounds are more soluble at higher temperature. Slow cooling, which can be effected under programmed control, yields crystals suitable for structure analysis in most instances (maximum dimension 0.3 – 0.8 mm). A simple way to bring about slow cooling is to place the solution in a Dewar flask filled with a liquid that can be heated to the boiling point of the solvent without itself vaporizing to any great extent.

2) The second approach depends on raising the concentration. The simplest method is to evaporate the solvent; generally, the more slowly this is done, the better are the crystals. A straightforward way of doing this is to pierce one or more small holes in the stopper of the crystallization flask.

Evaporation is not the only way of raising the concentration. Another possibility is to use the diffusion of two liquids into each other; the liquids should be miscible, and the substance to be crys-

tallized should have greatly different solubilities in them. The diffusion can take place in the liquid or the gas phase. In the first case, a layer of the "weaker" solvent (lower solubility) is formed on top of the solution, and the first crystal nuclei will form at the interface between the two solvents. In the simplest example of diffusion via the gas phase, an open vessel containing the solution with the "stronger" solvent (higher solubility) is placed inside a vessel containing the "weaker", and the whole assembly is hermetically sealed.

In principle, the handling and growing of crystals at low temperature can be done by the same methods, with appropriate modifications [61]. It is even possible to grow a crystal in the measuring instrument itself [65].

15.2.1.5. X Rays, Neutrons, and Electrons

X rays of sufficient intensity for diffraction studies are obtained from an X-ray generator, which essentially consists of two portions, the high-voltage generator and the tube housing. The generator transforms 220 V alternating current to a d.c. voltage of ≤ 60 kV. The total power is generally in the range 3 – 6 kW, but the output of the sealed X-ray tubes (Fig. 9) used for diffraction purposes is preferably 2.5 kW. In general, there is no improvement in luminance with a 3 kW tube; the only difference is that the focus is longer.

In "normal-focus" tubes, the focal spot measures 1×10 mm. If this spot is viewed at a 6° angle in the direction of its longest dimension, the effective spot size is 1×1 mm (point focus); in the perpendicular direction, the effective size is 0.1×10 mm (line focus). A typical X-ray tube has four windows, two with point focus and two with line focus. Specimens extended in two dimensions are irradiated with the line focus; point, sphere, or needle-shaped specimens, with the point focus. Fine-focus tubes (0.4×8 mm) are commonly used for single-crystal studies; these offer markedly higher luminance at the same energy output.

The residual gas pressure in the envelope of an X-ray tube (where the hot cathode and the anode are located) is $< 10^{-1}$ Pa. The appropriate voltage for the anode material is applied between the cathode and the anode (Table 1). Electrons emitted by the hot cathode are accelerated by the applied voltage, then braked when they strike the anode. If the electrons have sufficient kinetic energy to cause internal ionization of the atoms in the anode, a small part of the energy released (ca. 1 %) is consumed in ejecting electrons from the K shell

Table 1. Data on X-ray generation

Anode material	Excitation voltage, kV	Operating voltage, kV	Kα_1, nm	Kα_2, nm	Kα, nm	Kβ_1, nm	Kβ filter
Cr	6.0	20–25	0.22896	0.22935	0.22909	0.20848	V
Fe	7.5	25–30	0.19360	0.19399	0.19373	0.17565	Mn
Co	7.8	30–35	0.17889	0.17928	0.17902	0.16207	Fe
Ni	8.5	30–35	0.16578	0.16617	0.16591	0.15001	Co
Cu	9	35–40	0.15405	0.15443	0.15418	0.13922	Ni
Mo	20	50–60	0.07093	0.07135	0.07107	0.06323	Zr
Ag	25.5	60	0.05594	0.05638	0.05609	0.04970	Pd
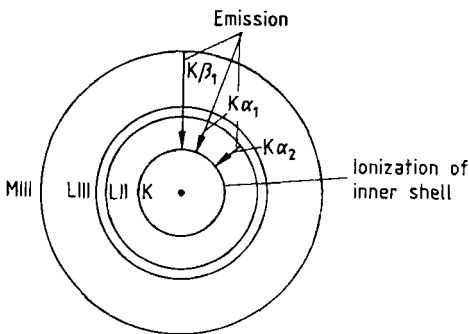
Figure 10. Generation of X rays

of the atoms. The vacant energy levels are filled, chiefly by electrons from the L shell (L III and L II levels) and also from the M shell (M III level). The energy thus gained is emitted as X rays (Fig. 10). The quantum transitions differ in probability, which can be varied to some extent through the operating voltage (see Table 1).

Given the proper tube voltage, the intensities of the principal components are in the ratio:

$$K\alpha_1 : K\alpha_2 : K\beta_1 = 10 : 5 : 2 \qquad (20)$$

A mixture of Kα_1 and Kα_2 radiation can be used for most diffraction studies; the Kβ_1 is removed by a filter or a crystal monochromator. The main constituent of the filter material is an element with an atomic number one less than the anode material; the X-ray spectrum of the filter material then has an absorption edge at the wavelength of the Kβ radiation to be removed. When a monochromator is used, the various wavelengths are separated by diffraction through a crystal, usually a graphite or α-quartz single crystal. The Bragg equation then gives different glancing an-

gles θ for the several wavelengths, so that the unwanted radiation components can be eliminated.

The most common anode materials for X-ray tubes are copper and molybdenum; silver, iron, and chromium are also used. Compounds containing only light atoms (C, N, and O) show only small absorption for molybdenum radiation, while heavier elements have correspondingly higher absorption, but a much more reduced one compared to Cu K$_\alpha$ radiation. Because the scattered intensity is directly proportional to λ^3 (Eq. 14), a copper anode has to be used for extremely small and weakly scattering specimens. The opposite is the case for larger specimens: The low absorption of Mo radiation shifts the intensity ratios in favor of molybdenum. X-ray tubes with chromium anodes generally have low output, air scattering is very high, and absorption problems are difficult or impossible to cope with, at least in single-crystal studies.

Contrasted with these sealed X-ray tubes are rotating anode tubes. The tube is not sealed and the vacuum must be maintained by a pump. The rotation of the anode and the more intensive cooling of the anode material mean that the specific energy loading of the anode can be made far higher; e.g., a generator of power 1.2 kW can yield a specific output of 12 kW/mm^2 with focal spot 0.1×1 mm. For comparison, a fine-focus tube at 90 % of maximum loading has a specific output of only 0.56 kW/mm^2. Rotating anode tubes have now become so reliable that sealed X-ray tubes may disappear from practical use in the near future.

Another kind of X-ray generators are the microfocus X-ray sources. The focus has an area of only 30 μm^2, and with an operating power of 24 W it attains the same brightness as a 5 kW rotating anode generator. The future will show whether its promise holds for very small crystals. A two orders of magnitude higher X-ray intensity is ob-

tained by using synchrotron radiation. This highly monochromatic, parallel, and brilliant X-ray radiation should only be used when the experiments demand this very expensive X-ray source (small crystals, protein crystals, high scattering angle resolution, very special experiments).

Neutron Beams. The neutron was discovered by CHADWICK in 1932. The equation:

$$m v = \frac{h}{\lambda} \tag{21}$$

where h is Planck's constant, m is the mass of the particle, λ is the wavelength, and v is the particle velocity, makes it possible to regard any moving particle as a wave. The neutron has not only mass but also magnetic moment; accordingly, it can interact with atomic nuclei and with unpaired electrons. Practical neutron sources are high-flux nuclear reactors such as those at Grenoble, Oak Ridge, and Brookhaven, which offer neutron fluxes of roughly 1.5×10^{15} cm^{-2} s^{-1}. Reactors available in Germany have fluxes about an order of magnitude lower (BERII in HMI Berlin: 1.2×10^{14} cm^{-2} s^{-1}).

The wavelength of neutron radiation is controlled by the absolute ambient temperature T:

$$\frac{1}{2} m v^2 = \frac{2}{3} k T \tag{22}$$

where k is Boltzmann's constant, and Equation (21) gives the value of λ. Neutrons in equilibrium at room temperature (thermal neutrons) have a wavelength of ca. 10^{-10} m. Monochromators for neutrons are generally diffracting crystals. Time-of-flight spectrometers make use of the fact that neutrons with a certain velocity correspond to neutron radiation with a certain wavelength.

Electron Beams. Electrons, like neutrons, are a form of particle radiation. If matter absorbs neutrons far less than X rays, exactly the opposite holds for electrons. They are so strongly absorbed by air that electron diffraction by the solid phase must generally be carried out in the high vacuum of the electron microscope. The wavelength, and thus the energy of electron radiation are controlled by the voltage drop V through which the electrons pass:

$$\frac{1}{2} m v^2 = e V \tag{23}$$

where e is the elementary charge, m is the electron mass, and v is the electron velocity. This, in combination with Equation (21), gives the wavelength:

$$\lambda = 12.236 V^{-1/2} (10^{-10} \text{ m}) \tag{24}$$

For example, if the voltage drop is 100 kV, the electron beam has a wavelength of 0.039×10^{-10} m. Because electrons do not penetrate to a great depth, they are particularly suited to the study of thin films and surface structures. The scattering length is four orders of magnitude larger than for X rays and neutrons. This leads not only to a correspondingly higher absorption, but also to an extremely high scattering power of the atoms.

15.2.1.6. Cameras and Diffractometers

Three types of detector are available for measuring the X rays diffracted by a crystal: Point detectors, linear and area detectors. The film method involves a two-dimensional detector, which can provide simultaneous information about every point in the region of reciprocal space investigated. The measurements are not immediately available in digital form, but powerful computer-controlled photometers can digitize them.

This technology is sometimes still used in protein crystallography, but the electronically controlled area counter is being introduced in this field. This type of position-sensitive counter directly supplies digital values as a function of position in reciprocal space; it also has a dynamic range several orders of magnitude wider than the film. In conjunction with a rotating anode X-ray source, it forms a data acquisition system with which not just proteins, but also other large molecules with poorly scattering atoms or very small crystals can be measured in an optimal way [66] [67]. One-dimensional position-sensitive detectors are available for powder diffractometry [68].

Point measurements with scintillation or proportional counters yield equally precise values. However, important parts of the information can be lost if it is not borne in mind that only those points can be measured that are moved into reflection position with complete prior knowledge of that position.

Recording methods also differ with which completeness a region of reciprocal space can be measured. In the analysis of crystalline powders, where the orientations of the lattice planes are expected to show a statistical distribution, the largest measurable glancing angle controls the ex-
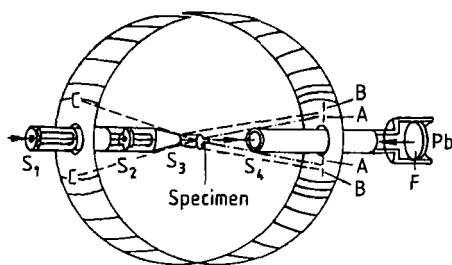
Figure 11. Debye – Scherrer camera

Figure 12. Diffraction cones in the crystalline powder method

Figure 13. Debye – Scherrer diffraction pattern of LiF

tent of reciprocal space that can be covered. The situation is similar with the four-circle single-crystal diffractometer; every net plane can be brought to reflection position, provided a maximum glancing angle (characteristic of the instrument) is not exceeded. In all other single-crystal goniometers, there are "blind regions" in the Ewald sphere that cannot be measured.

Film Methods. The prerequisite for using the film method is a suitable film material. The Committee for Crystallographic Apparatus of the International Union of Crystallography has tested a variety of X-ray films [69]. The properties of a usable X-ray film can be characterized as: Fine grain and homogeneous emulsion; low sensitivity to temperature and moisture; low susceptibility to fogging; wide range of linearity between film blackening and logarithm of exposure time; and high sensitivity to the wavelengths produced by X-ray tubes.

Nowadays it can be difficult to obtain film material of such quality because the medical purposes for which it was primarily intended have mostly vanished. However, the conditions can be relaxed because quantitative intensity measurements are no longer made by using films.

Debye – Scherrer Camera. Figure 11 illustrates the principle of the Debye – Scherrer camera, while Figure 12 shows schematically how the diffraction pattern is generated. A specimen of crystalline powder (ca. 0.1 mg) is placed in a borate glass capillary of diameter 0.3 – 1 mm and length ca. 1 cm, and the capillary is rotated about a specified axis. The film fits snugly inside the camera.

In a camera with internal diameter 114.6 mm, 1 mm on the film corresponds exactly to a diffraction angle of 1°. The radiation passes through the collimator (with three diaphragms S_1, S_2, S_3) to reach the specimen; S_1 and S_2 have the same diameter, while S_3 is somewhat larger. The primary beam is absorbed in the primary beam trap, which carries diaphragm S_4, and can be looked at through the lead-glass window, Pb, on the fluorescent screen F for setting purposes. The position and size of screen S_4 limit the smallest measurable diffraction angle $2\,\theta$, and thus the resolution power of the camera (A – A) in the small angle region. The position of S_3 relative to S_2 determines the range of angles within which scattered radiation from S_2 is observed (B – B). S_3 also masks backscattering, so that no diffracted X rays can reach the region C – C. Figure 13 shows a Debye – Scherrer pattern. The intersections of the diffraction cones in Figure 12 with the film give the Debye – Scherrer rings.

The camera should be evacuated while the pattern is being recorded, to minimize the background due to air scattering.

This method has the advantages that a small specimen can be used, operation is simple, and there is only slight falsification of the intensities due to nonstatistical distribution of crystallite orientations in the specimen. Disadvantages of the method are that exposure times are quite long (up to and beyond 12 h) and interpretation of measurements is inconvenient, especially when there are many specimens, because the results are not in digital form.
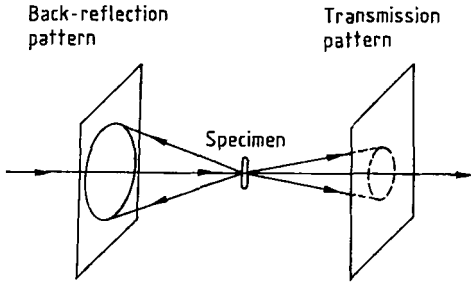
Figure 14. Laue camera

Figure 16. Principle of the Guinier method
a) Seemann – Bohlin circle; b) Monochromator; c) Reflection; d) Specimen
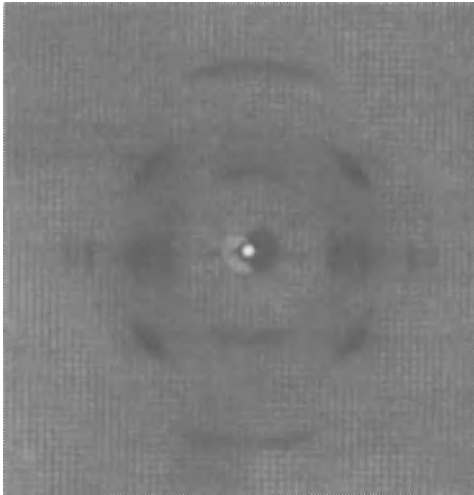
Figure 15. Flat-film diffraction pattern of polypropylene fibers
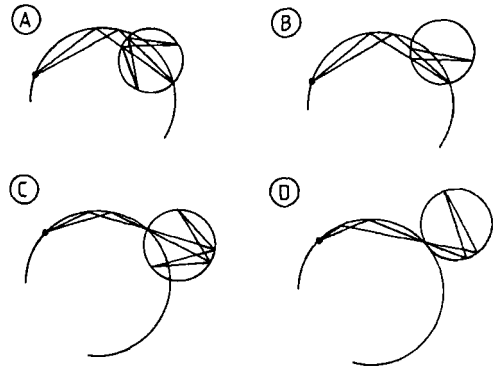
Figure 17. Guinier method, showing the four possible relative positions of specimen and monochromator
A) Symmetrical transmission; B) Asymmetrical transmission; C) Symmetrical reflection; D) Asymmetrical reflection

Laue Camera. This flat-film camera can record both transmission and back-reflection patterns. The principle is illustrated in Figure 14; Figure 15 shows a flat-film pattern. Transmission patterns with polychromatic radiation yield information about the symmetry of a crystal (Laue symmetry) or with monochromatic radiation the orientation of the crystallites (preferred orientation, texture); such patterns are recorded, for example, from fibers. Back-reflection patterns permit the non-destructive examination of workpieces.

In the Laue method, single crystals are irradiated with polychromatic X rays, so that the crystal need not be rotated. All net planes have wavelengths for which the reflection conditions are satisfied. The patterns are employed for the alignment of single crystals and the identification of their symmetries.

Guinier Camera. The Guinier camera is a focusing goniometer; for a certain geometric configuration of specimen and X-ray tube focus F, the X rays are focused, so that higher angular resolution is achieved (Fig. 16) . The radiation is reflected from a bent quartz monochromator (Johansson principle). The focussing circle (Seemann – Bohlin circle) is coupled with it in such a way that both circles have the same focal axis [70]. In addition to the symmetrical transmission configuration shown, three other relative positions of the monochromator and Seemann – Bohlin circles are possible, so that the entire X-ray diffraction spectrum can be recorded; these are the asymmetric transmission position, the symmetric reflection position, and the asymmetric reflection position (Fig. 17).

The specimen, in the form of a crystalline powder, is applied to the substrate in the thinnest possible layer. With strongly scattering inorganic
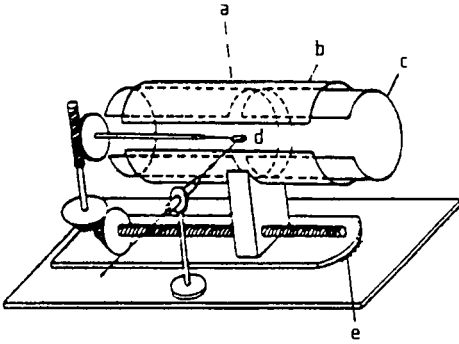
Figure 18. Principle of the Weissenberg camera
a) Equi-inclination axis; b) Film; c) Layer-line screen;
d) Crystal; e) Equi-inclination angle setting
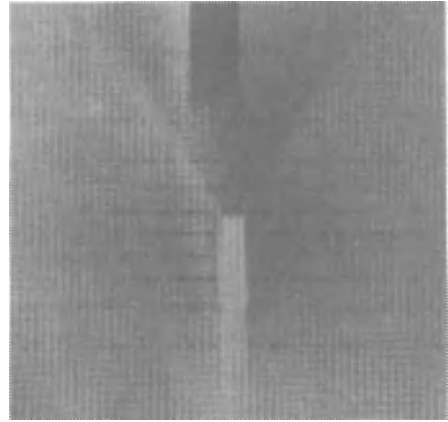
Figure 20. Rotating crystal pattern
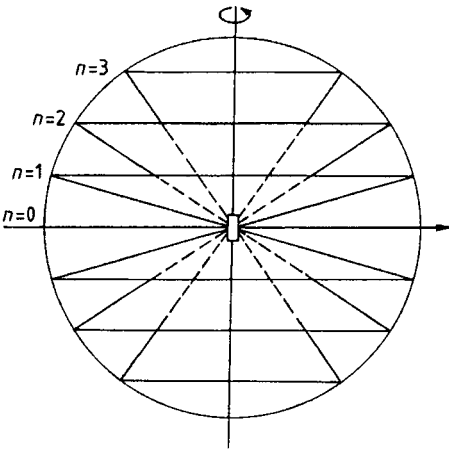
Figure 19. Diffraction cones in the rotating-crystal method
(n = number of layer line)

Figure 21. Weissenberg pattern

substances, the specimen can be dusted onto a film of grease. With organic substances, it is necessary to work with a layer at least $0.1 - 0.3$ mm thick, which may be sandwiched between foils, with the result, that the resolution of the system is somewhat less.

Weissenberg Camera. Figure 18 shows the camera schematically, while Figure 19 illustrates the registration of a diffraction pattern. This camera is used in recording diffraction data from single crystals [10, pp. 221 – 225]. If the layer-line screens are removed, an overall record of the reciprocal lattice is obtained (rotating crystal pattern; Fig. 20). With these screens in place, the individual diffraction cones of Figure 19 (the layer lines) can be isolated. Coupling the crystal rotation with a translation of the film holder in the direc-

tion of the rotation axis spreads the one-dimensional layer line of the rotating crystal pattern into a layer plane of the reciprocal space. If the crystal is rotated about one of the crystallographic axes a, b, c, the number of the layer line gives the index of the corresponding reciprocal axis. The diffraction points can be uniquely indexed, even though the reciprocal lattice is recorded in distorted form (Fig. 21). The method is superbly suited to intensity measurements, since an entire film pack can be placed in the film cassette and exposed to the X-ray beam; the films are then exposed simultaneously, but exposure varies with the penetration depth of the X rays.

The normal technique is the equi-inclination method, in which the direct and diffracted X-ray
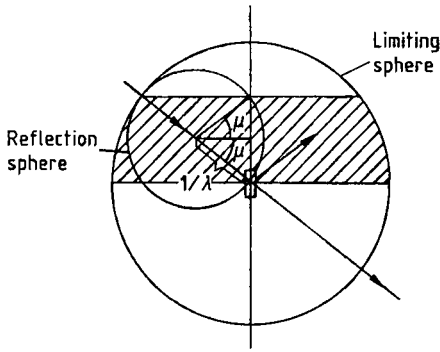
Figure 22. The portion of the limiting sphere that is accessible with the equi-inclination method
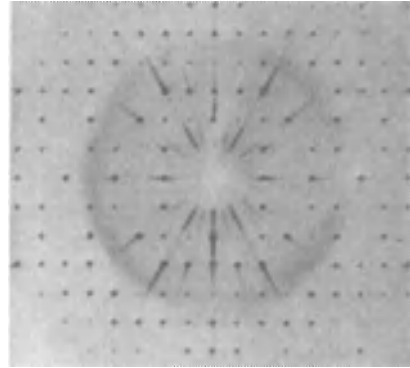
Figure 24. Precession pattern
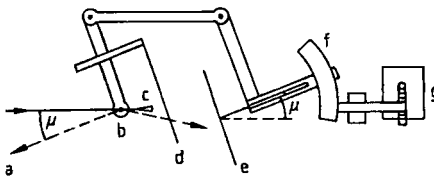
Figure 23. Precession camera
a) Vector in real space; b) Crystal; c) Primary beam stop;
d) Layer-line screen; e) Film plane; f) Precession angle setting; g) Motor

beams make equal angles with the layer line being measured. In addition to increasing the number of measurable reciprocal lattice points, this also simplifies the necessary intensity corrections. Figure 22 shows that the maximum length of the reciprocal lattice vector parallel to the rotation axis is:

$$\zeta_{max} = \frac{2}{\lambda} \sin \mu_{max} \tag{25}$$

The maximum equi-inclination angle μ_{max} is normally 30° ($\zeta_{max} = 1/\lambda$), so that ζ_{max} is the same as in the normal rotating crystal pattern, except that there is no blind region near the rotation axis (Figs. 5 and 22).

Precession (Buerger) Camera. In contrast to the Weissenberg camera, the crystal carrier is now coupled to the film cassette so that both move in exactly the same way while the pattern is being recorded (Fig. 23). A direction in the crystal precesses (precession angle μ) about the incident X-ray beam, so that reciprocal lattice points lying in a plane perpendicular to this crystal direction cut the reflection sphere. A precession pattern represents the reciprocal lattice in undistorted form
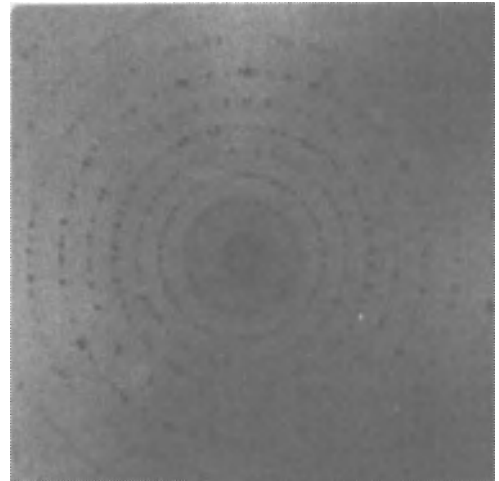
Figure 25. Cone pattern

(Fig. 24); the reflections are easily indexed and the pattern is easily interpreted [9, pp. 15 – 68]. The cone pattern of Figure 25, which shows the layer lines measurable with a given crystal setting, corresponds to the rotating crystal pattern.

Figure 26 illustrates the region of reciprocal space whose lattice points can be measured with a single spindle setting (shaded area). The maximum ζ is $2/\lambda \cos \mu$. One further spindle setting, which should be as close as possible to 90° away from the first, will eliminate most of the blind spot about the ζ-axis.

Space Explorer. The Weissenberg camera records only reciprocal lattice planes perpendicular to the crystal axis that is parallel to the goniometer
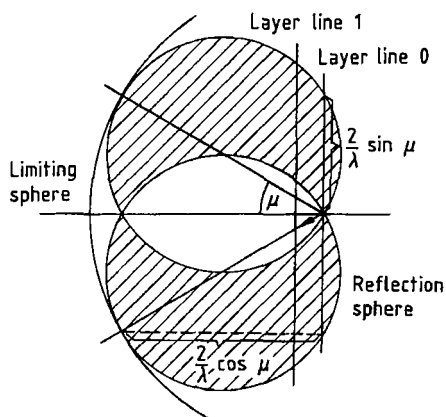
Figure 26. Region of the limiting sphere covered by the precession method

head axis, and the precession camera can measure only lattice planes that have a common reciprocal lattice axis parallel to the goniometer head axis and lattice planes which are parallel to these [24, pp. 98 – 167]. With the Space Explorer both types of series can be recorded. This facilitates the determination of cell constants and space groups in many cases.

These film methods may appear old fashioned, but this is wrong. Any serious X-ray laboratory will have a few of these machines, which are always necessary to give a first impression of the diffraction problem on a relatively cheap apparatus, especially when the expensive computer-controlled diffractometers are not immediately available or when there are problems and a three-dimensional inspection of the reciprocal space is necessary to gain an impression of the difficulties.

Counter Tube Method. The term "counter tube" means a detector that moves on a circle, the equator of the diffractometer. Only point measurements along this circle are possible. If one wishes to measure reflections that do not lie on this plane, the specimen must be oriented to position the reflections to be measured on the equator.

The use of this type of counter tube demands complicated electronics that make it possible to count the absorbed X-ray quanta. In the proportional counter, every X-ray quantum initiates a gas discharge; the voltage surge is electronically amplified and recorded as a pulse. In the scintillation counter, the X-ray quanta absorbed by the scintil-

lator crystal are transformed into light (fluorescence), and amplified by a photomultiplier. The two types of counter tube have very similar properties, and the pulses, once amplified, can be processed by the same electronics. A device called an electronic discriminator can block a given energy range ("window") and thus render the beam more monochromatic, greatly increasing the ratio of peak to background. Unfortunately, the use of a filter or a monochromator crystal to separate the K α and K β radiation cannot be dispensed with, because the resolution of an ordinary discriminator is not sufficient to do so.

The situation is different if a semiconductor detector (lithium-doped silicon crystal) and the proper electronics are available. This setup can separate K α and K β radiation, so that the beam does not have to be passed through a monochromator. The counter tube itself has nowadays become easier to operate, for the semiconductor crystal no longer has to be maintained at liquid nitrogen temperature — the problem has been solved electronically. Efforts continue to find broader applications for this type of detector, since replacement of the monochromator by electronics means that there is no longer a 50 % loss of primary intensity through diffraction by the monochromator crystal or absorption by the β filter.

The diffractometer must have a precision mechanical system for the positioning of specimen and detector, with a reproducibility of at least 0.01°. Digital evaluation makes it relatively easy to perform a profile analysis of the reflections.

In comparison with film cameras, diffractometers are roughly a factor of 10 more costly. They are still more expensive if the point counter is replaced by a one- or two-dimensional position-sensitive detector. The data acquisition time is, however, shortened by a factor of 50 – 100, or even more, depending on the geometry of the crystal lattice measured.

There are four types of position-sensitive detectors: Wire detectors (one- and two-dimensional), video detectors, CCD (charge coupled device chip), and image-plate detectors. The wire detector is based on the proportional counter principle; voltage pulses produced by quanta incident at different points along the wire take different times to reach its two ends. The travel times are derived from the profile of the sum of the two pulses. In the video counter, a phosphor layer generates light flashes where X-ray quanta are incident, and the flashes are picked up by a television camera. The CCD chip is covered by a scintillation foil, which
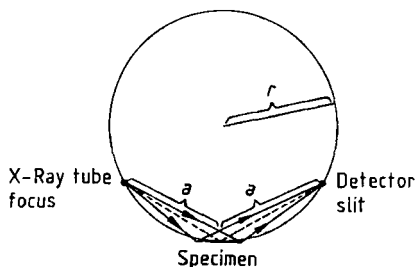
Figure 27. Beam path of the Bragg – Brentano two-circle powder diffractometer

Figure 28. X-ray diffraction pattern of the γ-phase of quinacridone, recorded in a Bragg – Brentano diffractometer with Cu K α radiation

is specific for a certain X-ray wavelength. The generated light flashes are counted by the CCD. The image-plate has a coating of rare-earth oxides, in which metastable states are produced by X-ray quanta. Transitions back to the ground state, induced by laser light, can be observed and quantified.

These position-sensitive detectors permit the use of simpler mechanical systems in diffractometers. A one-dimensional counter makes it possible to leave out one circle; a two-dimensional counter, two circles. In principle, then, a single-crystal diffractometer equipped with a two-dimensional counter needs only one circle, while a powder diffractometer equipped with a one-dimensional counter does not require a circle at all. Experience has shown that the greatest flexibility is achieved by incorporating these counters without reducing the number of circles.

A direct oscillation photograph, for example, can only be made by a four-circle diffractometer equipped with a two-dimensional detector.

Two-circle powder diffractometer fitted with a point counter. There are three powder diffractometer geometries with focusing optics, whereby the monochromator, the specimen, or both, can have a focusing characteristics. Usually, they are two- circle diffractometers.

The Bragg – Brentano two-circle diffractometer has a Seemann – Bohlin focusing circle, on which the X-ray tube anode, the counter tube window, and the specimen are located (Fig. 27). Only the specimen is focusing.

The planar specimen is placed tangentially to the focusing circle. Because theoretically this can be the case only along a line parallel to the line focus, there are residual geometric beam-spreading effects. The radius r of the focusing circle decreases with increasing θ, from infinity at $0°$ to $a/2$ at $90°$. In order to maintain the focusing
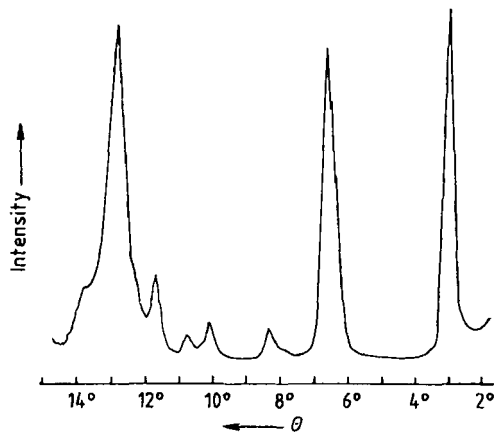
conditions while a diffraction pattern is being recorded, the detector is moved at twice the angular velocity of the specimen carrier. Because the specimen extends in two dimensions, it is possible, and necessary, to work with the line focus of the X-ray tube (perpendicular to the plane of Fig. 27); this makes very effective use of the X-ray tube anode surface.

As an alternative to a β filter, beam monochromatization can be achieved by a graphite monochromator between specimen and detector; this arrangement removes not only K β radiation, but also fluorescence emitted by the specimen. The effect is particularly beneficial in the study of iron-containing specimens with Cu K α radiation.

An $x - y$ plotter connected to the instrument (if it is not inherently computer-controlled) makes it possible to output the scattered X-ray intensity in analog form as a function of diffraction angle (Fig. 28). The intensity of the diffracted beam can also be output in digital form. An especially economical approach is to fit the instrument with an automated specimen changer, record the measured X-ray intensities on a mass storage medium (magnetic tape, diskette, or preferably yet a hard disk), and control the operation of the diffractometer by computer. Data measured, say, overnight or over the weekend can then be interpreted during regular working hours with the same computer.

In the *Debye – Scherrer diffractometer* with monochromator (Fig. 29), the line focus of the X-ray tube, the monochromator crystal, and the detector window all lie on the focusing circle of
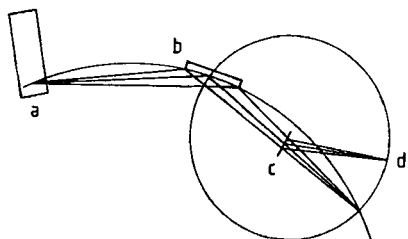
Figure 29. Beam path of the Debye – Scherrer diffractometer
a) X-ray tube; b) Monochromator; c) Specimen; d) Detector and detector slit
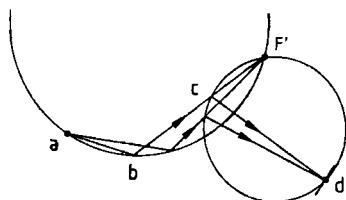
Figure 30. Beam path of the Guinier diffractometer
a) Line focus of X-ray tube; b) Monochromator; c) Specimen; d) Detector slit

Figure 31. X-ray diffraction pattern of the γ phase of quinacridone, recorded by the Guinier method with Cu K α_1 radiation

the curved monochromator crystal (zero setting). The specimen is measured in transmission. For inorganic substances with relatively high scattering power, a specimen dusted onto a foil with the thinnest possible coating of grease is sufficient. This technique is generally not adequate for organic specimens; instead, the specimen is contained between two thin foils, separated by a spacer. It is useful to do a transmittance measurement in the primary beam before performing the diffraction measurement itself. A transmittance of about 70 % should be obtained. For substances that do not absorb too strongly, the specimen can be placed in a capillary; among the advantages of this technique is that the measurement can be carried out in a controlled atmosphere.

Figure 29 further shows that the specimen position does not contribute to focusing. In a fixed specimen position, only absorption varies somewhat with diffraction angle. This setup is thus suitable when a curved one-dimensional position-sensitive detector is employed. There are detectors on the market that can measure up to $2\theta = 120°$ at the same time. When such a detector is used, no diffractometer is needed — only a base plate support. This technique has proven itself useful, particularly with programmed heating of specimens, and with in situ recording of diffraction diagrams.

The principle of the Guinier camera, described in section p. 385, has been developed by replacing the film by a counter tube to give the *Guinier diffractometer* (Fig. 30). The reflection intensities can be digitally evaluated; this is a major advantage for quantitative phase and profile analyses. The resolution obtained is much better than with the Bragg – Brentano method (compare Figs. 28 and 31), but specimen preparation is more complicated, and no automatic sample changer is available.

Single-Crystal Diffractometer. Most intensity measurements are now carried out with a counter tube; measurements are more accurate, and the results are obtained directly in digital form.

A relatively simple and inexpensive diffractometer is the single-crystal instrument with Weissenberg geometry [24, pp. 98 – 167] (see Fig. 18). Measurements can be automatically controlled by computer, but the counter tube must be realigned manually for each layer line. The instrument has the advantage of low cost, but frequent realignment is needed (semiautomatic instrument), and the individual reflections are not all measured in the same geometrical arrangements (Weissenberg geometry).

Four-circle diffractometers (Figs. 32 and 33) are free of these disadvantages. The ϕ and χ circles set the crystal in such a way that the reciprocal lattice vector of the reflection being measured lies in the equatorial plane of the instrument, while the counter tube moves along this circle (2θ circle). The θ and 2θ circles make it possible to set up the Bragg relation. The χ circle, called the Eulerian cradle, can be either closed or open (as in Fig. 32). A closed Eulerian cradle has the advantage that any reciprocal lattice vector can, in principle, be
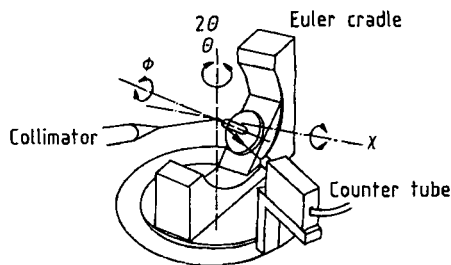
Figure 32. Four-circle diffractometer with open Eulerian cradle (χ circle)

brought into the equatorial plane; with an open 90° Eulerian cradle, this is possible only for half the Ewald sphere, though normally this is sufficient. The open Eulerian cradle permits a much sturdier construction, and "shading" by the closed Eulerian cradle ceases to be a problem. Complete control and monitoring by computer is possible with either form.

An important advantage of the closed Eulerian cradle is that in low-temperature measurements the cold nitrogen stream can keep its lamellar characteristics when it is passing the crystal on the rod, provided the nitrogen stream and the axis of the goniometer head remain always coaxial, what can be easily arranged with a closed Eulerian cradle.

An elegant version of the four-circle diffractometer is the CAD 4 diffractometer made by Nonius. Here the Eulerian cradle or χ circle is replaced by the κ circle, whose axis intersects that of the θ circle at an angle of 50° (Fig. 33). The axis of the ϕ circle likewise makes an angle of 50° with that of the κ circle, so that the ϕ axis can lie anywhere on a 100° cone.

Data Processing by Computer. Because minicomputers have become so inexpensive, it is an obvious step to use the computer not only to control the diffractometer, but also to solve the structure problem. In fact, it is common to use one computer for control, and a second for data processing. While the diffractometer, for example, is at work measuring the diffracted intensities of the computer-adjusted crystal of a compound, another structure can be solved, by either the same minicomputer or a second one. Nowadays, a personal computer is used for the least-squares refinement of the full matrix of the atomic parameters even for large structures. The same holds for the solution of the phase problem.
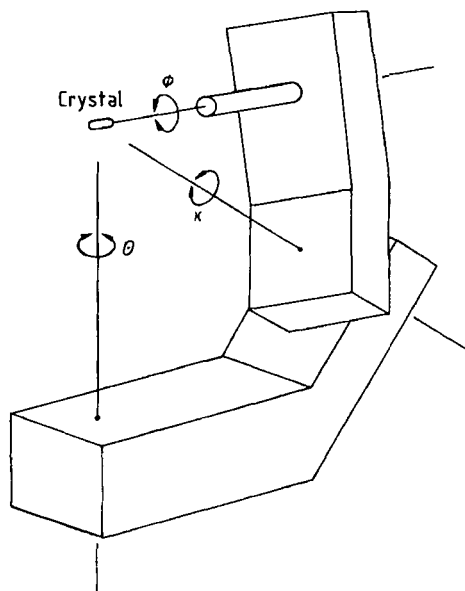
Figure 33. CAD 4 diffractometer

The combination of the rotating anode X-ray generators which are nowadays very stable, the robust one-, two-, three-, or four-circle single-crystal diffractometers, the area detectors (image plate, wire counters, or CCD), and powerful PCs make solving the crystal structure of a medium-sized molecule a job of a few hours, of a few days, or exceptionally of a week for crystals that are relatively small ($<$ 50 µm in one or all dimensions) and/or show greater disorder.

15.2.1.7. Safety

For electron radiation, the very high absorption in air means that no special safety rules have to be followed; but the X rays produced by electrons do call for precautions. The manufacturers of most electron microscopes have solved this problem. Neutron diffraction experiments are not performed in the chemist's own lab, and the host institute therefore has the responsibility of complying with safety regulations. In this section, only practices relating to X rays are discussed.

The regulations, for example, currently in force in Germany are set forth in the "X-Ray Regulation" of 8. January 1987 [71], which supersedes the similarly titled regulation of 1. March 1973. The new regulation is based on recommendations issued in 1958 by the International Committee for

Radiation Protection [72], [73]. The code prescribes, in part, as follows:

1) A radiation protection supervisor must be designated for the operation of X-ray instruments other than fully shielded instruments. If necessary, the radiation protection supervisor designates radiation protection officers, who must demonstrate familiarity with the requisite specialist knowledge.

2) X-ray equipment other than a fully shielded instrument may be operated only in a room enclosed on all sides and designated as an X-ray room. The controlled access area is defined as an area where persons can receive more than 15 mSv (1.5 rem) in a calendar year as whole-body exposure; the corresponding figure for the monitored area is 5 mSv.

3) Workers occupationally exposed to radiation may not exceed a whole-body dose of 50 or 15 mSv in a calendar year. In the first case, much more stringent provisions apply to the medical supervision of the worker. Special provisions apply for pregnant women and persons under 18 years of age.

4) The individual dose is measured by two independent methods. The first method is performed with dosimeters, which are to be submitted, at intervals not to exceed 1 month, to the government agency having jurisdiction. The second measurement must permit the determination of dose at any time, and may be put into use at the direction of the radiation protection supervisor or of the radiation protection officer or at the request of the worker.

Exposure to the primary beam from an X-ray tube (ca. 10^5 R/min) must be avoided under all circumstances.

15.2.1.8. Instrument Manufacturers

What follows is a list of companies trading in Germany, together with the instruments they market. No claim for completeness is made.

Nonius B. V., Röntgenweg 1, P. O. Box 811, 2600 AV Delft, The Netherlands. X-ray sources: Sealed tubes or rotating anode generators, microsource generators; goniometer heads; high- and low-temperature devices; four-circle K-axis goniometers; one- and two-dimensional detector systems, CCD and imaging plate detectors.

Freiberger Präzisionsmechanik GmbH, Hainichener Strasse 2 a, 09599 Freiberg, Germany.

X-ray sources; X-ray tubes; Debye – Scherrer cameras; two-circle powder diffractometers; precession cameras; Weissenberg cameras.

Huber Diffraktionstechnik GmbH, Sommerstraße 4, 83253 Rimsting/Chiemsee, Germany. X-ray diffraction equipment, goniometer heads: Laue, Debye-Scherrer, Buerger-precession, and Weissenberg cameras; imaging-plate Guinier cameras; X-ray diffractometers; high- and low-temperature attachments; monochromators; Eulerian cradles, rotation stages, linear stages, multiaxis goniometers for synchrotron and neutron facilities.

Philips GmbH, Miramstrasse 87, 34123 Kassel, Germany. X-ray sources; X-ray tubes; goniometer heads; high- and low-temperature stages; position-sensitive detectors; texture goniometers; two-circle powder diffractometers; thin-film high-resolution diffractometers.

Molecular Structure Corporation Europe, Unit D2, Chaucer Business Park, Watery Lane, Kemsing, Sevenoaks, Kent TN15 6YU, United Kingdom. Imaging plate detectors; CCD detectors; partial-Chi circle; low-temperature devices; rotating anode generators; peripheral devices.

Bruker AXS GmbH, Östliche Rheinbrückenstrasse 50, 76187 Karlsruhe, Germany. Rotating-anode and sealed-tube sources; ceramic X-ray tubes; multilayer optics; mono/poly capillaries; two-circle powder diffractometer; reflectometer; texture/stress diffractometer; attachments for high/low temperature, high pressure, humidity control, reactive gas; micro diffractometer; 2D SAXS camera; process automation diffractometer; chemical and biological 2D-CCD single-crystal systems.

Rich. Seifert & Co GmbH Röntgenwerk, Freiberger Präzisionsmechanik GmbH, Bogenstraße 41, 22926 Ahrensburg, Germany. Four-circle diffractometer, Theta/Theta powder diffractometer, low- and high-temperature sample stages, phase-texture-stress-diffractometer for laboratory and automotive industry, thin-film texture/stress Eta-diffractometer, reflectometer, high-resolution diffractometer, wafer ingot orientation systems for industrial applications, Rayflex Windows Software, X-ray generators, X-ray tubes, X-ray optics, parabolic multilayers, monochromators, film cameras, position-sensitive detectors PSD, energy dispersive detectors, electroluminescence detectors LUX.

Stoe & Cie GmbH, P.O. Box 101302, Hilpertstrasse 10, 64295 Darmstadt, Germany. Goniometer heads; Debye – Scherrer cameras; Guinier cameras; Weissenberg cameras; precession cameras; Reciprocal Lattice Explorers; powder diffrac-
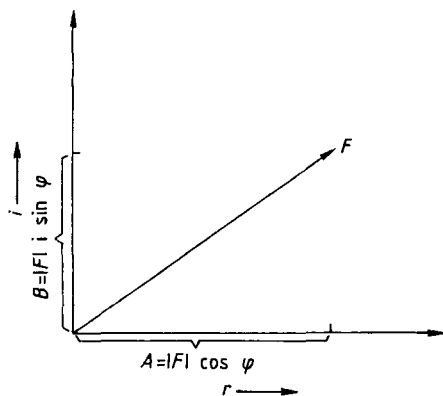
Figure 34. Gaussian plane
The real part of the complex number F is plotted along the r-axis; the imaginary part along the i-axis

Figure 35. Addition of structure factor contributions in the complex plane

tometers with scintillation counter, position-sensitive detectors or imaging plates; special sample holders; high-temperature and texture sample stages; four-circle single-crystal diffractometers also with CCD detector; one- and two-circle single-crystal diffractometers with imaging plates; crystallographic software.

15.2.2. The Phase Problem

In general, $F(hkl)$ is a complex number; when plotted in the Gaussian plane (Fig. 34), it has not only a modulus $|F|$ but also a phase angle φ.

By the Euler–Moivre theorem

$$\|F\| \exp(i\varphi) = \|F\| \cos\varphi + i \|F\| \sin\varphi = A + iB \tag{26}$$

A structure factor is obtained not by simple scalar addition of the scattering contributions atom-by-atom, but by vector addition, as the plot in the Gaussian plane makes clear (Fig. 35). The angle between a scattering contribution and the real axis is given by $2\pi(hx_n + ky_n + lz_n)$ (Eq. 18). The phase angles of the structure factors give the angles by which the sine waves of the individual structure factors in the Fourier synthesis must be shifted with respect to the origin in order to yield the electron density.

An important special case is where B in Equation (26) vanishes, and the only two phase angles possible for $F(h,k,l)$ are 0° and 180°. The phase problem then becomes a matter of adding signed quantities. According to Equations (18) and (26),

this occurs whenever an atom with coordinates x, y, z always has an atom of the same species at coordinates $-x$, $-y$, $-z$ corresponding to it. In other words, the phase problem becomes a signed addition problem in all centrosymmetric space groups.

In the methods now commonly used for measuring diffracted X-ray intensities, it is generally not possible to measure the phase angle of a structure factor. The phase is said to be "lost" in the intensity measurement. In the early days of X-ray structure analysis, it was thought that phases could not be deduced from measured intensities.

15.2.2.1. Patterson (Vector) Methods

Not until A. L. PATTERSON's work [74] did it gradually become clear that the measured intensities also contain phase information. The function:

$$P(uvw) = \frac{1}{V} \sum_{hkl=-\infty}^{+\infty} \|F(hkl)\|^2 \exp[2\pi i(hu + kv + lw)] \tag{27}$$

does not give the electron density at point x, y, z but does yield maxima in a vector space. Each such maximum has a position (relative to the origin) corresponding to the vector between two atoms; its weight is proportional to the product of the electron numbers of the two atoms. The vector space equation requires knowledge only of the structure amplitudes $|F|$, not of the phases. It is possible to solve a structure that is not too complicated, if all the maxima of a Patterson synthesis are known.
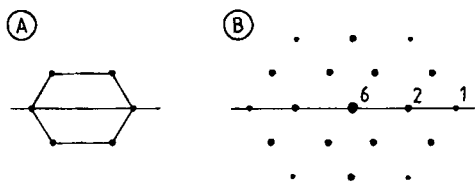
Figure 36. A) Crystal space: B) Vector space

Multiple solutions may exist for one and the same set of vectors in extremely rare cases (homometric structures). No case is known in which two homometric structures are both chemically meaningful.

The complete interpretation of a Patterson synthesis is generally made much more difficult because the maxima are not resolved, and cannot be explicitly identified. Figure 36 illustrates this situation for the two-dimensional case. Corresponding to the 6 maxima in the crystal space, there are 19 maxima in the vector space; these must be placed in the same unit cell volume that accommodates the 6 atoms. What is more, the maxima in the vector space require more volume than atoms in the crystal space, because the indeterminacies of each pair of atoms add at the Patterson maximum. The half-width values of the maxima in the vector space are thus roughly twice those in the crystal space. As a consequence, the peaks overlap strongly. The relative weights of the maxima are indicated in Figure 36.

Patterson synthesis by itself cannot solve even moderately large molecular structures (five or six equally heavy atoms) unless concrete structure information already exists so that the vector space can be systematically searched for a presumed structure or substructure. A search of this kind might be done with a computer, perhaps by the method of pattern-seeking functions, or the convolution molecule method [57], [75]. These techniques, however, are not so often used today, having been supplanted by direct phase-determination methods (Section 15.2.2.2), which make it possible to reach the goal quickly on the basis of relatively little prior knowledge [76]. Patterson synthesis may return to prominence when used to identify sets of starting phases (Section 15.2.2.2).

An exception of that is PATSEE [77], which uses one or two known fragments of the structure, whose orientation is found by a real-space Patterson rotation search and its translation in the cell by direct methods (Section 15.2.2.2), starting from random positions. Figures of merit related to those

of the direct methods are used to refine only promising phase sets.

Heavy Atom Method. The principal use of Patterson synthesis at present is in the heavy atom method. If a few heavy atoms, such as bromine or iodine, are present in a molecule along with light atoms, such as carbon, nitrogen, or oxygen, the Patterson technique makes it possible to analyze the X-ray diffraction pattern into heavy- and light-atom components. For a structure to be solved reliably, however, the light and heavy atoms must have sufficiently different scattering powers, but the sum of the heavy-atom scattering powers (proportional to the sum of the squares of the atomic numbers Z) must not be much different from that of the light atoms:

$$\sum z_H^2 \geq \sum z_L^2 \qquad (28)$$

Given these conditions, the heavy-atom vectors in the Patterson synthesis have markedly higher maxima associated with them than the heavy/light- or light/light-atom vectors. In centrosymmetric structures, this holds even when the ratio of heavy to light atoms is less favorable, since in this case the knowledge of a small structural component is enough to establish with fair confidence the sign trend of a large structure factor.

If, say, a bromine atom has been incorporated into an organic molecule to facilitate X-ray structure analysis, the weights of bromine–bromine, bromine–carbon, and carbon–carbon vectors are in the ratio:

$$Z_{Br}^2:Z_{Br}Z_C:Z_C^2 = 1225 : 210 : 36 \qquad (29)$$

Thus, it should not be hard to identify the bromine–bromine vectors, given the sizes of the maxima. Further, heavy-atom maxima are not expected to overlap, because they have virtually the entire cell available, while carbon–carbon maxima and, to a great extent, bromine–carbon maxima will recede into the background noise.

Example. Vectors between heavy atoms are especially easy to identify if just one heavy atom is located in the asymmetric unit. This is shown here for the centrosymmetric space group $P\,2_1/c$. From the coordinates in the crystal space:

$$x, y, z; x, \tfrac{1}{2} - y, \tfrac{1}{2} + z; -x, -y, -z; -x, \tfrac{1}{2} + y, \tfrac{1}{2} - z$$

Table 2. Patterson maxima* for the space group $P2_1/c$ with one heavy atom in the asymmetric unit

Atomic co-ordinates	x y z	x $\frac{1}{2}-x$ $\frac{1}{2}+z$	$-x$ $-y$ $-z$	$-x$ $\frac{1}{2}+y$ $\frac{1}{2}-z$
x y z	0 0 0	0 $\frac{1}{2}+2y$ $\frac{1}{2}$	$2x$ $2y$ $2z$	$2x$ $\frac{1}{2}$ $\frac{1}{2}+2z$
x $\frac{1}{2}-y$ $\frac{1}{2}+z$	0 $\frac{1}{2}-2y$ $\frac{1}{2}$	0 0 0	$2x$ $\frac{1}{2}$ $\frac{1}{2}+2z$	$2x$ $-2y$ $2z$
$-x$ $-y$ $-z$	$-2x$ $-2y$ $-2z$	$-2x$ $-\frac{1}{2}$ $-\frac{1}{2}-2z$	0 0 0	0 $-\frac{1}{2}-2y$ $-\frac{1}{2}$
$-x$ $\frac{1}{2}+y$ $\frac{1}{2}-z$	$-2x$ $\frac{1}{2}$ $\frac{1}{2}-2z$	$-2x$ $2y$ $-2z$	0 $\frac{1}{2}+2y$ $\frac{1}{2}$	0 0 0

* Vector space coordinates are differences between pairs of atomic coordinates.

the coordinates in the vector space can be found by forming the 16 possible vectors, as shown in Table 2. Such a table should always be prepared for difficult cases.

The space group in the vector space is $P\,2/m$, which is the Laue group of $P2_1/c$. Only 21 such "Patterson groups" are possible with conservation of the Bravais lattice. The space group of the vector space (Patterson group) is derived from that of the crystal by adding a center of symmetry (if none is present) and removing translation from the group of symmetry elements.

Four maxima are located at the origin; four at $0, \frac{1}{2}\pm 2\,y, \frac{1}{2}$; four at $\pm 2\,x, \frac{1}{2}, \frac{1}{2}\pm 2\,z$; two at $\pm 2\,x, \pm 2\,y, \pm 2\,z$; and two at $2\,x, \pm 2\,y, 2\,z$. For the $0, \frac{1}{2}\pm 2\,y, \frac{1}{2}$ maximum, only y is variable; the maximum lies on a "Harker line". At $\pm 2\,x, \frac{1}{2}, \frac{1}{2}\pm 2\,z$, the section cuts through the cell at $y=\frac{1}{2}$ (Harker section). The first condition yields the y coordinate; the second yields the x and z coordinates of the heavy atom. The maxima at $2\,x, 2\,y, 2\,z$ (and the symmetry-related positions) simply provide confirmation of the values found.

Maxima on Harker lines and sections are always doubly occupied, so that they are still easier to find.

If there is more than one heavy atom in the asymmetric unit, there are also vectors between the heavy atoms in the asymmetric unit. A table should be prepared in order to make clear the weighting of individual maxima. The convolution technique can lead to a solution in a very direct way [58].

All of this work is nowadays carried out by computer programs (e.g., SHELX-97 [49]). It is advisable, however, to do at least part of it by hand, so that one is familiar with the theory when there are difficulties.

From the heavy-atom positions obtained by this procedure, the structure factors are calculated from Equation (18), and the signs or phases are assigned to the measured structure amplitudes. The result is a Fourier synthesis of electron density. Because the phases depend only on the positions of the heavy atoms, this first Fourier synthesis often fails to reveal the whole structure of the molecule; instead, it yields only a partial structure, but this is generally enough to lead to better phases for the next Fourier synthesis. The structure should be solved after two or three (at most six or seven) of these successive Fourier syntheses of electron den-

sity. This step can be automated. The maxima of a Fourier synthesis can be analyzed by computer; the only maxima accepted as atoms are those that form a cluster with chemically meaningful geometries [73]. The combination of this technique with direct methods is discussed further on.

Special difficulties arise when the heavy-atom structure, taken separately, contains one more symmetry element (usually a center of symmetry) than the overall structure. The first Fourier synthesis of the electron density then shows two light-atom structures along with the heavy-atom structure. In order to calculate the second Fourier synthesis, the heavy-atom structure is then used together with only that part of the light-atom structure that can be assigned with high confidence to just one of the two light-atom structures indicated in the Fourier synthesis. The additional symmetry element vanishes in the next Fourier synthesis, and the structure analysis proceeds in the usual way.

15.2.2.2. Direct Methods

As Equation (17) implies, the distribution of the electron density in a crystal structure is unambiguously determined once the structure factors have been obtained. A structure factor comprises the structure amplitude and the phase angle; as has already been shown, all that can be determined by experiment is generally the structure amplitude, not the phase angle φ. The preceding section discussed a fairly straightforward way of using the heavy-atom method to gain knowledge of a substructure, and using this to determine the φ values indirectly from the Patterson synthesis.

As the theory of Patterson synthesis shows, it must be generally possible to infer the phases, given the associated structure amplitudes.

The best-known equation for direct phase determination is that of DAVID SAYRE [78]:

$$F(h) = \frac{1}{V}\frac{f}{g} \sum_{h'} F(h')F(h-h') \tag{30}$$

where $(1/V)(f/g)$ can be regarded as merely a kind of scaling factor and h, h' and $h-h'$ are vectors in reciprocal space, i.e., each is a triplet of indices (h,k,l). Equation (30) is not a probability relation, but a true equation; all atoms are assumed identical, and all combinations of structure factors are required. SAYRE used this formula to elucidate a centrosymmetric projection of the crystal structure of hydroxyproline [78].

Multiplying both sides of Equation (30) by $F(-h)$ yields:

$$\|F(h)\|^2 = \frac{1}{V}\frac{f}{g} \sum_h F(-h)F(h')F(h-h')$$ (31)

For large $F(h)$, the right-hand side is also large, real, and positive. If $F(h)$, $F(h')$, and $F(h-h')$ are large, the following probability statements then hold:

$$s(h)s(h')s(h-h') \approx +1$$ (32)

in the centrosymmetric case, and:

$$\varphi(-h) + \varphi(h') + \varphi(h-h') \approx 0 \quad (\text{mod } 2\pi)$$ (33)

in the noncentrosymmetric case. Here, $s(h)$ denotes the sign and $\varphi(h)$ the phase of $F(h)$, and similarly for the other terms. Given the signs or phases of two structure factors, it is thus possible to infer the sign or phase of a third.

The form in which the Sayre equation is commonly used today is due to HUGHES [79]:

$$E(h) = N^{1/2}\langle E(h')E(h-h')\rangle_h$$ (34)

where N is the number of atoms in the unit cell and $\langle \rangle_{h'}$ denotes the average over all h'; $E(h)$ is the normalized structure factor defined as:

$$\|E(h)\|^2 = \frac{\|F(h)\|^2}{\sum\limits_{j=1}^{N} f_j^2} = \frac{I(h)}{\bar{I}}$$ (35)

where the sum of the squares of the atomic form factors f is taken over all N atoms in the unit cell. This is the expected value of the intensity of a reflection; $|F(h)|$ is the structure amplitude after absolute scaling is applied (see Eq. 58); $I(h)$ is the

intensity of the reflection (h) and I the average intensity at the glancing angle of (h).

Equation (32) is a probability relation. The probability that the positive sign is correct is:

$$P_+(h) = \frac{1}{2} + \frac{1}{2}\tanh\left(N^{-1/2}\|E(h)E(h')E(h-h')\|\right)$$ (36)

It should be noted that the probability of the calculated phases depends not on the absolute intensities, but on the relative intensities (Eq. 35) that describe the scattering power of a structure with point atoms. Basically, the highest E values are the reflections containing the largest amount of information. The more atoms in the unit cell, the lower the probability that relation (32) holds. For structures where the unit cell contains atoms differing in atomic number:

$$P_+(h) = \frac{1}{2} + \frac{1}{2}\tanh\left(\sigma_3\sigma_2^{-3/2}\|E(h)\|\sum_h E(h')E(h-h')\right)$$ (37)

for all products $E(h')E(h-h')$ available for $E(h)$, and:

$$\sigma_n = \sum_{j=1}^{N} Z_j^n$$ (38)

Using Equation (37) for the probability of relation (32) takes care of the fact that not all products required for the Sayre equation are available. A strict equation has become a probability relation. One of the corresponding equations for the probability that relation (33) is correct in the noncentrosymmetric case reads as follows:

$$P(\varphi(-h) + \varphi(h-h') + \varphi(h')) =$$

$$\frac{\exp\left[2\sigma_3\sigma_2^{-3/2}\|E(-h)E(h')E(h-h')\|\cos(\varphi(-h) + \varphi(h-h') + \varphi(h'))\right]}{2\pi I_0 2\sigma_3\sigma_2^{-3/2}\|E(-h)E(h')E(h-h')\|}$$ (39)

where I_0 is a Bessel function of the second kind [80].

The Hughes form of the Sayre equation (34) is for the centrosymmetric case; in the noncentrosymmetric case, at least as far as physical information is concerned, it corresponds to the tangent formula. In contrast to Equation (33), here the sum is weighted with the normalized structure factors, and all pairs of structure factors possible in a special case are taken care of:

$$\tan\varphi(h) = \frac{\sum_{h'} K(hh') \sin\left[\varphi(h') + \varphi(h - h')\right]}{\sum_{h'} K(hh') \cos\left[\varphi(h') + \varphi(h - h')\right]} \quad (40)$$

where:

$$K(hh') = 2\sigma_3\sigma_2^{-3/2} \left\| E(-h)E(h')E(h - h') \right\| \quad (41)$$

A measure of the reliability of a phase obtained in this way is $\alpha(h)$, where:

$$[\alpha(h)]^2 = \left[\sum_{h'} K(hh') \cos\left(\varphi(h') + \varphi(h - h')\right)\right]^2$$
$$+ \left[\sum_{h'} K(hh') \sin\left(\varphi(h') + \varphi(h - h')\right)\right]^2 \quad (42)$$

These equations are also used in the centrosymmetric case for determining the signs of structure factors (only the cosine terms are required).

In summary, it is probable that the sum of the phases for two reflections equals the phase of a third reflection, if the sums of the indices of the first two reflections are the indices of the third, and if all three reflections are relatively strong. The probability that this holds can be calculated. It is also possible to determine the probability for one phase when many phase combinations exist.

As Equation (34) shows, phases must be known or assumed at the start of a phase determination, before the equations can be used at all. Depending on the space group or Bravais lattice type, the phases of up to three reflections can be freely chosen; this fixes the origin of the coordinate system [81]–[84]. In the case of noncentrosymmetric space groups or those without mirror symmetry, the phase of one further reflection must be restricted to a 180° angular range in order to fix one of the two enantiomers. Other phases can be represented by letter symbols (symbolic addition) [85], [86] or simply by angle values (multisolu-

tion) [87]–[89]. In the case of a center of symmetry, for every additional reflection with freely selected phase, only two possibilities (0° and 180°) must be taken into account; this is done by letting each resulting phase combination serve as the starting point for determination of a new phase set. In noncentrosymmetric space groups, on the other hand, an infinite number of possibilities between 0° and 360° would have to be examined, but experience has shown that four values per structure factor (45°, 135°, 225°, 315°) are enough to place an adequate restriction on one phase for the determination of other phases. If there are n additional reflections with freely selected phases, the number of phase sets will be 2^n in the centrosymmetric case and 4^n in the noncentrosymmetric case. If the phases have been calculated in symbolic form, the stated possibilities must be permuted over all the symbols to yield the same number of phase sets as in the latter cases (unless relations can be established between the symbols, as is very often the case).

In another direct phasing method, a random number generator is used to supply phases for a relatively large number of reflections. "Figures of merit" are calculated in order to decide whether the calculation of a phase set should be completed. With present-day fast computers, 20 000 or more such phase sets can be calculated. Because one starting phase set contains a great number of reflections, there is no longer any need to fix the origin, or select one enantiomer through the phases of some reflections [90].

Ultimately, the only way to decide whether a phase set is the right one is to look for a chemically meaningful interpretation of the Fourier synthesis of the electron density. But it would be rather tedious to check, say, 64 Fourier syntheses in this way (256 or more in troublesome cases). An attempt is therefore made to find figures of merit [91] that will make one phase set more probable than the others, even before the Fourier synthesis is calculated. Such a figure of merit might refer to the internal consistency of a phase set. The determination of phases or signs by direct methods then involves the following steps:

1) Calculate the magnitudes $|E|$ of the normalized structure factors
2) Rank the E values (decreasing order)
3) List pairs of reflections that can figure in a Sayre relation with a given reflection

4) For the starting phase set, select the reflections defining the origin and a few additional reflections
5) Determine the phases of each phase set
6) Select the phase sets that appear best under given criteria (figures of merit)
7) Calculate the Fourier synthesis of the electron density (E map) and interpret it

Different systems of weighting for phase determination and an almost countless number of figures of merit have been developed [92]–[94].

H is now possible to solve rather large crystal structures, including small proteins with more than 1000 non-hydrogen atoms, by real/reciprocal space Fourier recycling methods [95]–[99]. These are very lenghtly procedures, and a powerful PC may require four weeks or more without breaks. These major problems usually involve measuring synchrotron data.

15.2.2.3. Trial-and-Error Method, R Test

The trial-and-error method used to be employed for simple inorganic structures, but it is no longer important.

The R value:

$$R_1 = \frac{\sum \left| \|F(hkl)\|_{\text{obs}} - \|F(hkl)\|_{\text{calc}} \right|}{\sum \|F(hkl)\|_{\text{obs}}} \, 100 \, [\%] \qquad (43)$$

indicates how close the solution approaches the correct structure. The sums are taken over all measured structure factor magnitudes. The $|F|_{\text{obs}}$ are observed structure amplitudes, while the $|F|_{\text{calc}}$ are the amplitudes calculated from Equation (18). If a least-squares fit is also done, a second R test becomes important; this includes the weights w, a function based on the counting statistics of the measurements, assigned to the squares of the differences between observed and calculated structure amplitudes:

$$R_2 = \left[\frac{\sum w \left(\|F(hkl)\|_{\text{obs}} - \|F(hkl)\|_{\text{calc}} \right)^2}{\sum w \, \|F(hkl)\|_{\text{obs}}^2} \right]^{1/2} \, 100 \, [\%]$$

$$(44)$$

A value of $R < 50\%$ generally indicates that one is on the right path toward a solution of the structure. In particularly simple cases, space-filling considerations and the requirements of a highly symmetric space group may limit the number of possible solutions.

15.2.2.4. Experimental Phase Determination

High-quality crystals, together with intense X-ray sources, such as synchrotron radiation (and, in especially simple cases, rotating anode generators), make it possible to find the phase angles of structure invariants (triple products or reflections with even indices) by experimental means. This supplemental information can be derived from intensity fluctuations that occur when a crystal is rotated about the reciprocal lattice vector of a net plane in reflection position, and another reciprocal lattice point additionally passes through the Ewald sphere. The results can no longer be interpreted by kinematic lattice theory; the dynamic theory must be employed [100], [101].

15.2.3. Least-Squares Refinement

The least-squares method is used to achieve the best possible fit of the structure model, found after phase determination, to the measured data. For every atom, in general, the least-squares calculation refines three positional parameters as well as one (isotropic) or six (anisotropic) temperature coefficients. There are also "overall" parameters such as the absolute scaling factor. Special cases may require further parameters, for example a factor for enantiomorphism.

Formerly the general practice was to minimize the quantity:

$$\sum \left(\|F_{\text{obs}}\| - \|F_{\text{calc}}\| \right)^2 \qquad (45)$$

This approach has the disadvantage that no significance is accorded to "negative" intensities [103], which means that the zero point was too high.

Modern least-squares computer programs minimize:

$$\sum \left(F_{obs}^2 - F_{calc}^2 \right)^2 \qquad (46)$$

Among other advantages of this procedure is that even low-intensity data sets can result in useful parameter refinements [50].

Least-squares refinement yields not only the structure parameters, but also their standard deviations. The R values defined above can serve as figures of merit if w represents the weighting function employed in the refinement. A more important point for assessing the quality of a structure analysis, however, is that the results fit with the wealth of experience of crystallographers and chemists, and deviate from prior knowledge only in justifiable cases.

Spectroscopic data must also be taken into account.

Constraints (fixed parameters) and restraints (partly fixed parameters with a specified allowance of deviation) have become important in the last few years [51]. In this way it is possible to extract as much "truth" as possible out of relatively bad data (originating from the "best" crystal which was available).

Another important problem is twinning, which usually gets apparent during parameter refinement. This "model" is very often more "true" than any model of disorder [102].

15.2.4. Determination of Absolute Configuration

Friedel's law states that:

$$\|F(hkl)\| = \|F(-h-k-l)\| \qquad (47)$$

even for noncentrosymmetric space groups. When there is anomalous dispersion, however, Friedel's law no longer holds rigorously. The atomic form factors adopt complex values:

$$f = f_0 + f' + if'' \qquad (48)$$

where f_0 is the atomic form factor with no allowance for anomalous scattering, f' is the real part of the anomalous scattering, and f'' is its imaginary part.

For the noncentrosymmetric case, the calculated magnitudes of $F(h,k,l)$ and $F(-h,-k,-l)$ are different, so that the absolute configuration

of a molecule or atomic group can be determined. The effect is the greater the more strongly an atom scatters and the closer (on the long-wavelength side) the absorption edge of an atom is to the X-ray wavelength. Normally, when extremely accurate intensity measurements are not possible, an atom heavier than oxygen must be present in the molecule (or must be incorporated synthetically) for the configuration of the molecule to be unambiguously determinable. Naturally, the synthesis must exclude racemization.

There are several methods for determining the absolute configuration. After the structure has been solved, one can calculate which pairs of reflections $(h, k, l/-h, -k, -l)$ differ most in intensity, and then measure these reflections more accurately. This approach makes it relatively easy to decide between enantiomers. D. W. ENGEL has used the technique to determine the absolute configuration for compounds that do not contain atoms heavier than oxygen [104].

A more convenient method that does not call for further measurements is to obtain the best possible least-squares fits of both enantiomers to the measurements. The R_2 values (Eq. 44) are subjected to a χ^2 test, which yields the statistical confidence level for one of the two configurations. The confidences are normally well over 99 %. This method was proposed and developed by W. C. HAMILTON [105].

Another method is to attach a factor to the imaginary part in Equation (48) and to introduce this factor as a parameter in the least-squares fit. If it refines to $+1$, the configuration is correct; to -1, the other configuration is [106]. A new formalism has been introduced [107].

15.2.5. Example of a Single-Crystal X-Ray Structure Analysis Solved by Direct Methods

To illustrate the technique, consider a compound for which two structures have been proposed (Fig. 37). Spectroscopic measurements (NMR, mass spectroscopy) have not yielded decisive results. The coupling constants and chemical shifts are too unusual, especially for proposed structure **1**, which further involves bond angles so far from the norm that no Dreiding model could be built. At this point in the structure analysis, if not earlier, X-ray examination becomes essential.

A glass-clear crystal measuring 0.45×0.45×0.2 mm was mounted on the inside wall of an capillary with a tiny amount of grease, and the capillary was sealed. The capillary was glued to the glass fiber of a goniometer head in such a way that the primary beam avoided the fiber. Alignment patterns of a precession camera showed a very high probability that

the specimen was a single crystal, because it yielded sharp reflections.

It is important to consider what type of X rays to use. Photographic recording of patterns was done with Cu K α radiation; diffractometer measurements were performed with Mo K α radiation. The absorption coefficient μ was calculated as:

$$\mu = [0.68 (\mu/\varrho)_C + 0.06 (\mu/\varrho)_H + 0.26 (\mu/\varrho)_O]\varrho = 0.788\varrho \tag{49}$$

where the weighting corresponds to the mass fractions of the atoms in the compound.

Values of μ/ϱ by element and wavelength are given in the International Tables [29, vol. III, pp. 157–200]. The density ϱ is calculated from the molecular mass M, the number of molecules Z in the cell, Avogadro's number (6.022×10^{23}), and the cell volume V (in 10^{-30} m^3):

$$\varrho = \frac{10 M Z}{6.022 V} (\text{g cm}^{-3}) \tag{50}$$

With a computer-controlled single-crystal diffractometer, the cell constants and the setting matrix for the diffractometer were automatically determined on 98 reflections [63]; an appropriate program for controlling the diffractometer was employed, and the measurements were subjected to least-squares refinement. The result was the triclinic space group P$\bar{1}$ with cell constants:

$a = 10.561$ (5); $b = 6.134$ (3); $c = 10.418$ (6) (10^{-10} m)
$\alpha = 89.37°$ (4); $\beta = 64.08°$ (4); $\gamma = 89.46°$

The numbers in parentheses are the standard deviations derived from the least-squares calculations. The density, determined by suspension in aqueous K_2HgI_4, was 1.35 g/cm^3; this implies that the unit cell contains two molecules, i.e., the X-ray measurements give a density of 1.347 g/cm^3. Accordingly, the absorption coefficient is 1.06 cm^{-1}. The fraction of the radiation not absorbed can be calculated:

$$I/I_0 = \exp(-\mu d) \tag{51}$$

It is 93 % for the maximum path length d (0.67 mm) in the crystal, and 98 % for the minimal path length (0.2 mm). Thus, no absorption correction is required with Mo K α radiation; the lower intensity of the reflections compared with Cu K α radiation must, however, be taken into consideration (Eq. 14).

Diffracted X-ray intensities were measured on a computer-controlled diffractometer with a $\theta/2\theta$ scan mode [108]. The reference reflection was remeasured at the beginning and end of each cycle of 20 reflections. This reflection had to be carefully selected because it formed the basis for the relative scaling. At a maximum glancing angle of 28°, the resolution was 0.76×10^{-10} m [$= \lambda/(2 \sin\theta)$].

Of the 2930 reflections to be measured, 2391 had X-ray intensities significantly above the background.

The five-point measurement used here (Fig. 38) is not so often employed today. Large mass storage devices make it possible to store all the measured points for a reflection profile, so that laborious corrections do not have to be applied during the measurements.

The standard deviation σ of the intensity of a reflection is derived from the counting statistics:

$$\sigma(hkl) = \sqrt{\text{meas}_1 + \text{meas}_3 + \text{meas}_5 + 4\,\text{meas}_2 + 4\,\text{meas}_4} \tag{52}$$

The relative error $r(hkl)$ is given by

$$r(hkl) = \frac{\sigma(hkl)}{I(hkl)_{\text{obs}}} \tag{53}$$

In the least-squares calculations, the reciprocal of $r(hkl)$ is needed for the weight function $w(hkl)$ in order to fit the structure parameters to the measured data.

The Lorentz–polarization factor correction had to be applied to the diffracted intensities:

$$I(hkl) = (\text{LP})^{-1} I(hkl)_{\text{obs}} \tag{54}$$

For the case under consideration:

$$\text{LP} = \frac{1 + \cos^2 2\theta}{2 \sin 2\theta} \tag{55}$$

The absolute structure amplitude is then calculated as:

$$\|F(hkl)\| = c \sqrt{I(hkl)} \tag{56}$$

The use of direct methods for phase determination requires not just the relative structure amplitudes, but also the absolute structure amplitudes (Eq. 56). These can be calculated with the aid of Wilson statistics [109]:

$$\bar{I} = \exp\left(-2B \sin^2\theta/\lambda^2\right) k \sum_{n=1}^{N} f_n^2 \tag{57}$$

$$\ln\left(\bar{I}/\sum_{n=1}^{N} f_n^2\right) = -2B \sin^2\theta/\lambda^2 + \ln k \tag{58}$$

The theory says that the average $\bar{I}$ of the diffracted X-ray intensities for a given glancing angle equals the sum of the squares of the atomic form factors of the N atoms in the unit cell. If the left-hand side of Equation (58) is plotted against $\sin^2\theta/\lambda^2$, the slope of the regression line is twice the overall temperature coefficient B, and the vertical-axis intercept is the natural logarithm of the scaling factor k (Fig. 39). The resulting values were $B = 4.3 \times 10^{-20}$ m^2 and $k = 0.32$. After parameter refinement, a scaling factor of 0.28 and an overall temperature coefficient of 4.8×10^{-20} m^2 were obtained. Factor B can be described by:

$$B = 8\pi^2 \bar{u}^2 \tag{59}$$

where $\bar{u}^2$ is the mean-square displacement of atoms from their average positions perpendicular to the diffracting net planes.

From Equation (35), the normalized structure amplitudes $|E|$ were obtained. These provided the basis for direct phase determination.

Table 3. Maxima in the E map plotted by computer

Peak	Peak height	Initial coordinates			Interpretation	Coordinates after least-squares refinement		
		x	y	z		x	y	z
1	999	0.0914	0.2955	0.1963	O(3)	0.0881	0.3031	0.2044
2	873	0.3008	0.4824	0.1738	C(9)	0.2904	0.4925	0.1766
3	871	−0.1507	0.4079	0.2851	O(2)	−0.1478	0.4049	0.2865
4	844	0.3264	0.3231	0.2409	C(14)	0.3287	0.3208	0.2401
5	842	0.1360	0.6191	0.0820	O(4)	0.1401	0.6100	0.0727
6	790	0.3636	0.6667	0.1563	C(10)	0.3715	0.6809	0.1391
7	771	−0.3413	0.1716	0.4126	O(1)	−0.3412	0.1781	0.4128
8	747	0.1515	0.5000	0.1563	C(8)	0.1671	0.4802	0.1443
9	743	0.4954	0.6820	0.1616	C(11)	0.4886	0.6936	0.1655
10	728	−0.0462	0.2833	0.1889	C(7)	−0.0403	0.2748	0.1869
11	703	−0.0724	0.0370	0.2020	C(1)	−0.0691	0.0334	0.2044
12	672	0.4485	0.3440	0.2635	C(13)	0.4467	0.3348	0.2647
13	561	−0.1418	−0.0400	0.3657	C(2)	−0.1393	−0.0384	0.3634
14	554	−0.2104	−0.0157	0.1946	C(4)	0.2051	−0.0301	0.1941
15	550	0.5288	0.5165	0.2304	C(12)	0.5265	0.5216	0.2272
16	536	−0.2746	−0.0438	0.3560	C(3)	−0.2813	−0.0325	0.3561
17	453	−0.1515	0.1667	0.4688	C(5)	−0.1499	0.1433	0.4641
18	426	−0.2326	0.3038	0.4336	C(6)	−0.2334	0.3031	0.4214
19	322	0.2147	0.3104	0.2263				
20	309	−0.0538	0.4083	0.0974				
21	225	0.6667	0.5556	0.2188				

The subsequent calculations were done with the MULTAN 77 program system [110], which comprises six programs linked through data files. From the corrected intensities $I(hkl)$, NORMAL generates normalized structure amplitudes $|E(hkl)|$. Wilson statistics (or Debye statistics [111] if parts of the molecule are already known) are used to determine the scaling factor and the overall temperature factor, and to work out the statistics of the scattered X-ray intensities by reflection groups. The actual phase determination takes place in the MULTAN program. The 393 largest E values were used for phase determination, the maximum E being 3.5 and the minimum 1.0. First, all combinations of two reciprocal lattice vectors for each reflection were found: $H = H' + (H - H')$, where H stands for an index triplet (h, k, l). Processing continued with the 2000 largest triple products of $|E|$ values.

The critical question in phase determination is which reflections are assigned phases in the starting set. The case under study was centrosymmetric, so that only signs had to be allocated. The origin of the coordinate system was fixed by the reflections $(7, 3, 3)$, $(1, 1, 4)$, and $(0, 1, 2)$. The four reflections $(1, 1, 2)$, $(8, 2, 2)$, $(5, 0, 2)$, and $(0, 5, 2)$ gave 16 sets of signs. The reflections were suggested by the program. The basic criteria for including a reflection in the starting set were the number, size, and type of triple-product relations. In more difficult cases this selection must be performed manually.

The program used the tangent formula (Eq. 40) to generate from the initial sets, sign sets with the fewest contradictions. Statistical criteria (figures of merit) were then applied to choose the most probable set of signs. The most probable set was that in which only $(1, 1, 2)$ among all the starting reflections had negative sign. From these results, and the normalized structure amplitudes $|E|$, the programs EXPAND, FFT, and SEARCH were used to calculate an E map (a Fourier synthesis of the electron density of point scatterers)

and to interpret it. Figure 1 and Table 3 present the result. Maximum 19 in Table 3 is markedly smaller than those which precede it. Comparison with Figure 37 reveals that structural formula **1** is in agreement with the computer suggestion. The 18 nonhydrogen atoms of proposed structure **1** were assigned to the 18 strongest maxima of the Fourier synthesis. A structure factor calculation (Eq. 18) gave $R_1 = 28.6\%$ (Eq. 43).

The structure model obtained is chemically meaningful. The chemist's experience thus comes into play only when a possible solution to the phase problem has already been found. In contrast to indirect structure analysis methods (such as spectroscopy), the analysis does not start with structure model.

In the least-squares refinement the structure parameters (one isotropic temperature coefficient and three position parameters per atom, one scaling factor) had to be fitted to the 2391 observed reflections; unit weights were assigned to all reflections. When R_1 reached 14.3%, every atom was assigned six temperature coefficients. The expression for the anisotropic temperature factor is:

$$\exp - \left(b_{11}h^2 + b_{22}k^2 + b_{33}l^2 + 2b_{12}hk + 2b_{13}hl + 2b_{23}kl\right) \tag{60}$$

After further refinement cycles, hydrogen atoms in chemically meaningful positions were assigned to the largest maxima in the difference Fourier synthesis of electron density. This is a Fourier synthesis in which the Fourier coefficients are the differences between the observed and calculated structure factors. When the hydrogen coordinates had been fitted to the measured data, the isotropic temperature coefficients of the hydrogen atoms were also refined. The R_1 was now 5.3%. Finally, weighting the squared deviations in

(1)

(2)
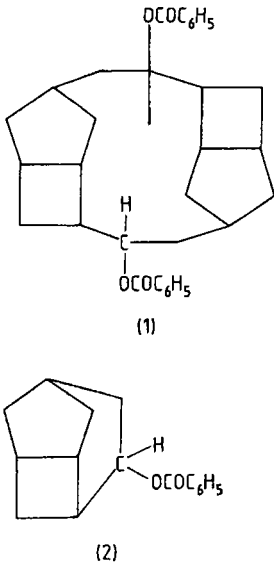
Figure 37. Two suggested structures for the compound under study

Figure 38. Steps in the five-point measurement

Figure 39. Computer plot of results from Wilson statistics

accordance with the counting statistics gave $R_1 = 4.7\%$. The weighted R_2 from Equation (44) was 3.6%.

At this point, the least-squares refinement could be regarded as complete because the largest change in a parameter was smaller than 10% of the corresponding standard deviation. In a difference Fourier synthesis of the electron density, the values for the 10 largest maxima were $0.12-0.16$ electrons/10^{-30} m^3.

The software used was MULTAN 77 [110]. There are a number of more up-to-date program systems:

SIR 92 [42]
DIRDIF [43]
MULTAN 88 [44]
NRCVAX [46]
XTAL [47]
ORTEP II [48]
SHELXS 86 [49]
SHELXL 92 [50]
SHELX 97 [51]
CRYSTALS [52]

Representations of the molecule are shown in Figures 40 (configuration and bond lengths), 41 (stereoscopic view), and 42 (projection of unit cell content along the a-axis). If the bond and torsion angles (not shown here) are considered, along with the bond lengths, it becomes clear that major difficulties arise in determining the molecular structure by spectroscopic means alone. The results of the structure analysis have been published elsewhere [112].

15.2.6. Diffraction by Polycrystalline Specimens

A substance is characterized as polycrystalline if its crystalline particles cannot be distinguished with the naked eye (< 0.01 mm). Larger particles, if any, must be ground before the measurement, provided this does not adversely affect the crystal structure of the substance.

Figure 40. Molecular structure of (1 RS, 3 SR, 4 RS, 6 SR, 8 SR)-5,9-dioxatricyclo[4.2.1.0^{3,8}]-non-4-yl benzoate. Lengths in 10^{-10} m. Standard deviations $\bar{\sigma}$: 0.002×10^{-10} m for O – C, 0.002×10^{-10} m for C – C, 0.01×10^{-10} m for C – H
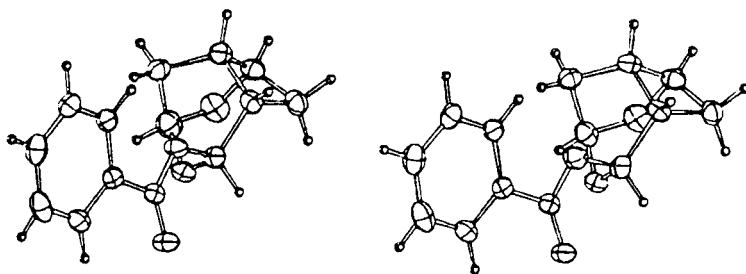
Figure 41. Stereoscopic view

15.2.6.1. General Considerations

If the apparatus ordinarily used to record diffraction patterns of polycrystalline specimens is compared with that used for single-crystal diffractometry, the following essential differences are seen:

1) Instruments for powder diffractometry are much simpler. If an "ideally polycrystalline" specimen is assumed, and if a one- or two-dimensional detector with appropriately large angular aperture is employed, there is no need to rotate either the specimen or the detector. Given an ideal statistical distribution of crystallites in the specimen, together with not too small a specimen size, it can be assumed that every orientation of a crystallite relative to the incident X-ray beam occurs in the specimen (as a statistical average). A rotation about one axis is enough to get a quite substantial increase in the scattered X-ray intensities, because then a large number of crystallites contribute to the intensity for a selected net plane. Single-axis goniometers used for this purpose are the Debye – Scherrer and the transmission diffractometers.

2) The specimen sizes employed are much larger than those in single-crystal studies. Because ideally there must be a statistical distribution of crystallites, only a small fraction of the specimen can contribute to a given reflection.

One of the problems in X-ray diffraction by a polycrystalline powder is that the specimen is not "uniformly polycrystalline". The crystallites vary widely in size and their relative orientations are not random. The predominance of one habitus or
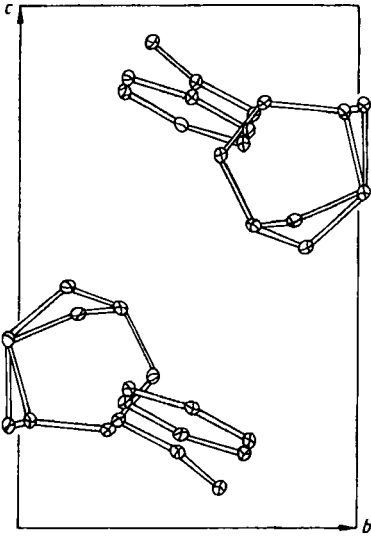
Figure 42. Projection of unit cell along the a-axis

another is crucial in the preparation of the specimen; for example, needles line up parallel to one another, while plates align themselves parallel to the specimen surface. In principle, this effect can be diminished by grinding the specimen and screening out the coarse particles, but there are two disadvantages to this:

1) Organic substances, in particular, are so sensitive that the crystal structure is disrupted by grinding, the specimen undergoes a change in degree of crystallinity, or there may even be a phase transformation

2) The specimen thus treated is no longer in the as-delivered condition that was to be characterized

If one has a phase mixture, demixing is possible, because the components often differ in their grinding characteristics.

Far better than excessive size reduction is to rotate the specimen appropriately and integrate over all directions in space. These ideal conditions can be approximated if an area detector is employed.

15.2.6.2. Qualitative and Quantitative Analyses of Powder Specimens

The X-ray diffraction pattern of a polycrystalline specimen is a plot of diffracted X-ray intensity against diffraction angle. Because the diffracted

intensity is a single-valued function of the unit cell contents (Eqs. 15, 18), and the diffraction angle is a single-valued function of the unit-cell dimensions (Eq. 2), an identical unit cell with identical contents must also yield an identical diffraction pattern. The diffraction pattern is thus a genuine fingerprint.

But the fingerprint also contains information on the macroscopic properties of a specimen. The half-value widths of the reflections are inversely proportional to the crystallite sizes, and directly proportional to the concentration and magnitude of structural defects. The coherent scattered component of the background results from amorphous constituents in the specimen, while the incoherent scattering is chiefly due to inelastic atomic scattering.

Determination of Cell Constants, Crystal System, and Space Group. The requirements on specimen preparation are a bit less demanding when a powder pattern is used to determine the cell constants only as long as lines are not fully missing, because preferred orientation is so heavy. Unfortunately, an unknown unit cell cannot always be determined in this simple way. The method is most suitable for the cubic case. Squaring the Bragg equation gives:

$$\sin^2\theta = \frac{\lambda^2}{4d^2} \tag{61}$$

For the cubic crystal system:

$$1/d^2 = \frac{h^2 + k^2 + l^2}{a^2} \tag{62}$$

where a is the cell constant. It follows that:

$$\sin^2\theta = \frac{\lambda^2}{4a^2}\left(h^2 + k^2 + l^2\right) \tag{63}$$

For a cubic substance, the constant $\lambda^2/4\,a^2$ can normally be identified directly if the $\sin^2\theta$ values of the reflections are ranked in ascending order of magnitude. The value of $\lambda^2/4\,a^2$ can be recognized and all $\sin^2\theta$ values can be obtained by multiplying by a sum of at most three squares.

Unambiguous results are, however, obtained only if the errors of measurement are small. It can never be excluded with utter confidence that slight deviations from, say, the cubic system occur. Absolutely certain results come only from single-crystal patterns, which additionally reveal the symmetry.

The powder patterns of substances belonging to other crystal systems can be indexed on the basis of analogous regularities, and thus assigned to a unit cell. The method of T. Ito even makes it possible to index the diffraction spectra of triclinic compounds [113]. The precondition is that many discrete reflections must be present, and their glancing angles must be determinable with great accuracy. Organic compounds commonly do not fulfill these conditions. Large cell constants lead to overlapping of reflections at relatively small glancing angles; many crystal structure defects and large temperature factors have the same effect because they increase the half-value widths of the reflections.

Special difficulties arise in determining the cell constants of high polymers. Using single crystals instead of crystalline powders is not ordinarily an option because in most cases there are principally no single crystals. In most cases, however, filaments can be drawn from the substance; after appropriate heat treatment, such filaments tend to take on a single-crystal character. The reflections are then located on layer lines in flat-film patterns. With some luck, trial and error will lead to not only the translational period in the fiber direction, but also the other cell constants [1, pp. 405–406].

If provisional cell constants have been determined for a specimen, all reflections can be indexed. There must be no common divisor for all the indices h, k, and l (primitive cell). Now the reduced cell (whose edges are the three shortest noncoplanar translational periods of the lattice) can be calculated. From the Niggli matrix:

$$\begin{pmatrix} a \cdot a & b \cdot b & c \cdot c \\ b \cdot c & c \cdot a & a \cdot b \end{pmatrix} \tag{64}$$

which is formed from the scalar products of the cell constants, it is possible to deduce the crystal system and Bravais lattice [4, pp. 124–159].

With the final cell constants, the crystal system derived from them, and the Bravais lattice, the reflections can be indexed. The space group can be identified from systematic absences [19, pp. 21–53].

As the preceding discussion implies, the determination of cell constants from powder patterns involves large uncertainties, even more so for the determination of the crystal system and space group, because the diffraction pattern does not directly lead to the symmetries, which are instead deduced from measured angles which include experimental errors, e.g., 90° for one of the cell constants. The case is similar with systematic extinctions: Lack of resolution can result in uncertainty or a wrong conclusion.

A better procedure is the use of powder patterns for crystalline phases whose cell constants are known. Changes in the cell under given conditions can be tracked in a convenient and reliable way. The cell data can be fitted to the measured data by least-squares methods.

Phase Analysis. If a specimen is crystalline, an obvious way to study and identify it is by using X-ray diffraction patterns. The X-ray diffraction pattern of a crystalline compound functions as a fingerprint that should permit unambiguous identification. Isomorphous crystalline compounds often create major problems; elemental analysis can be of aid here.

The case is even more difficult when the specimen for analysis is a mixture of crystalline phases differing markedly in the content of the individual components. This problem may be so aggravated that one of the crystalline phases is represented solely by its strongest reflection. Other methods of investigation are then needed in order to verify this crystal phase. All that X-ray diffraction can do in such instances is to make the presence of a crystalline phase probable.

If two isomorphous crystalline compounds cannot be distinguished from the X-ray diffraction pattern, neutron diffraction is an option. The atomic form factors for neutron diffraction and X-ray diffraction are completely different, so that the reflections can be expected to differ in intensity.

The successful use of X-ray diffraction to analyze unknown substances and mixtures of substances requires a knowledge of many X-ray diffraction patterns. The International Center for Diffraction Data (ICDD, 1601 Park Lane, Swarthmore, PA 19081, United States) has taken on the job of compiling X-ray diffraction patterns. The tabulated data can be obtained in the form of data cards, microfiches, compact disc (CD) or magnetic tape, together with a search program. In the first two cases, the X-ray diffraction data can be searched by means of tables ordered via lattice plane distances and relative intensities. The continually updated compilation concentrates on inorganic substances. The collection of data for organic compounds is far less complete. Furthermore, analysis is made more difficult by the poorer crystal quality and larger cell constants of organic compounds. Both these factors tend to

push the principal reflections to relatively small glancing angles, so that the net plane distances must be assigned relatively large errors.

Quantitative Analysis of Crystalline Phases. The intensity of a reflection is directly proportional to the quantity of the crystalline phase in a mixture, but it is also a function of the crystal structure. Comparing the intensities of two reflections is therefore not enough to determine the mass ratio of two phases in a mixture; the constant of proportionality must also be found.

The following technique can be used in the case of a two-phase mixture in which the percentage content of one phase is to be determined. In a number of specimens of known composition, the intensity ratio for the two crystalline phases is plotted against the mass ratio. The dependence is linear only if both crystalline phases have similar absorption coefficients, which is roughly so for, say, crystal modifications of the same compound.

A method for the absolute quantitative determination of crystalline phases in a multiphase mixture is to mix a standard substance into the specimen. Calibration curves are necessary for mixtures of the standard with each of the phases to be determined. The components can then be determined independently of each other. The critical point is to determine accurately the integral intensity of a sufficiently strong reflection that does not have any other reflection superposed on it. With a diffractometer linked to a computer, it is possible to separate overlapping reflections and determine their intensities, although some idealizations are required.

When a phase mixture contains just a few percent ($< 8\%$) or even fractions of a percent of the phase being analyzed a more accurate result can be obtained by recording X-ray diffraction patterns for specimens that lie in the composition range of interest. These allow the content of the strongly deficient phase to be estimated. The results are often distorted because it proves quite difficult to find a suitable standard for the specimen being measured. The standard should have the same crystal quality and crystallite size as the substance to be determined; otherwise, errors in the results are inevitable. These idealized conditions are scarcely ever fully achievable.

15.2.6.3. Solid Solutions

Compounds are normally miscible in all proportions in the gas phase, but in the liquid phase two or more substances must have certain chemical properties in common if they are to be miscible. The situation is much more complicated in the solid phase. Here, miscibility demands not just chemical affinity, but also a similar manner of space filling. For the formation of a continuous series of solid solutions, a molecule or atom of one phase must be arbitrarily replaceable by a molecule or atom of the other one. The formation of solid solutions has a relatively great effect on the cell volume, and this effect can be used for the very exact determination of a compound in trace element analysis (e.g., quantitative determination of trace impurities in semiconductor crystals).

Very often, solid solutions can be formed only over a limited range. The host lattice tries to adapt, more or less, to the guest until a point is reached where the stability limit of the host lattice is passed. Variations of the lattice parameters in the stability range are often controlled by the Vegard rule, which says that variations of parameters are proportional to the content of foreign components. The validity of this rule can, conversely, be regarded as evidence of solid-solution formation. It is generally sufficient if proportionality can be demonstrated for a few net plane distances.

The use of computers to control the diffractometer and interpret the measurements has proven itself here in particular. Not only can glancing angles be determined accurately, but any new reflections that appear can be identified with sufficient statistical reliability.

15.2.6.4. Rietveld Method

Measurements on crystalline powders lead to a solution of only relatively small and simple structures: Most of the reflections must be resolved, so that intensities can be associated with them unequivocally. Even theoretically, this is often impossible, e.g., when several reflections occur at precisely the same glancing angle. In the cubic crystal system this is so when the sum of the squares of the indices does not permit a single indexing; e.g., (221) and (300), where the sum of the squares of the indices is 9 for each case. Measurements on crystalline powders that give unambiguous values for all reflections should be inherently preferable to single-crystal measurements, because many experimental difficulties (e.g., extinction and absorption) either do not occur at all, or are easier to handle.

This is theoretically correct, but this ideal case never occurs. The atomic resolution is always low-
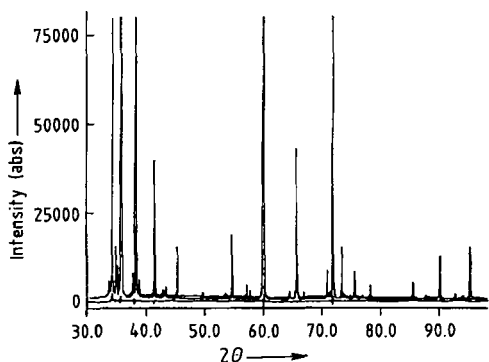
Figure 43. Diffraction pattern of a polytypic mixture of SiC

Figure 44. Diffraction pattern of a solid solution and a mixed phase: 12.9 % $Nd_{0.4}Ce_{0.6}O_{1.8}$, 87.1 % $Nd_{1.85}Ce_{0.15}CuO_4$

er. In general a single-crystal structure analysis is always better and gives more significant data, nevertheless the single crystal might be very poor. The powder method is used when it is not possible to obtain sufficiently good single crystals.

The strength of the Rietveld method lies in the refinement of the structure parameters of the whole powder sample not just of one single crystal. RIETVELD [114] recognized that it is by no means necessary to allocate the measured intensities to individual reflections in a one-dimensional powder pattern. Instead, it is better to measure a diffraction pattern as exactly as possible and use it in its entirety as the basis for refinement of the structure parameters. Parameters such as the atom coordinates, temperature coefficients, and scaling factors are fitted to the diffraction data by a least-squares procedure.

This method is particularly important for neutron diffraction analysis, where in general a previous single-crystal X-ray structure analysis has led to a structure model that is satisfactory within certain limits. The powder method can then be used, for example, to examine the magnetic properties of a substance in more detail, without having to grow large single crystals.

Rietveld analysis has recently proved especially valuable for the quantitative determination of components in polytypic mixtures. If the structures of the polytypes are known, the only parameters to be determined are the temperature coefficients, the scaling factors, and the cell constants. Figure 43 shows an analysis of SiC. The specimen contains 11.2 % of polytype 4 H, 85.2 % of 6 H, and 3.6 % of 15 R. The profile R value is ca. 6.7 %. Another very important application is the study of
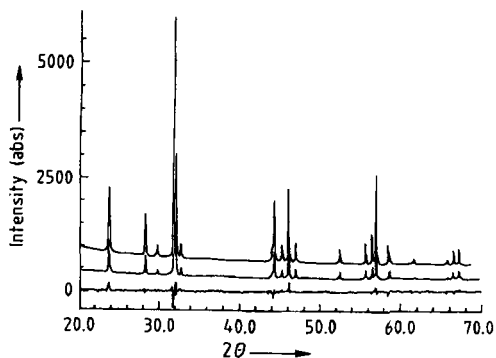
solid-solution systems. As Figure 44 shows, the content of an additional phase can be estimated when there are gaps in the series of solid solutions [115]. The profile R value in this case was 2.67 %.

In the past this method could only be used when the overall crystal structure was known and only special characteristics had to be explored. Now many methods have been developed to perform quasi-a-priori crystal structure analyses of crystal powder samples. Computer programs for single-crystal work often have an option for powder structure analysis (e.g., [42], [50]). They need the successful extraction of intensities of as many as possible single reflections out of the powder diagram, what is often performed by the method of Le Bail [117]. This method is also used by POSIT, in which only low-order reflections are necessary to determine a model from known fragments of the unknown structure for the Rietveld refinement [118]. The computer program Powder Cell is a complete program system for solving crystal structures from powder data [53]. Fully another approach is Monte Carlo/simulated annealing [119]. Only an approximate starting model with a few internal degrees of freedom is necessary. Thousands of diffraction diagrams are calculated and compared with the experimental one. The software package Powder Solve is fully integrated within the Cerius^2 molecular modeling environment [54]. A further approach, quite different from the above, is the ab initio prediction of possible molecular crystal structures by energy minimization. No cell determinations, no indexing, and no extraction of intensities of single reflections are necessary [53], [120]–[122], [124].

15.2.6.5. Texture

Up to this point, two distinct types of specimen, single crystals and polycrystalline powders, have been discussed. In the analysis of a single crystal, it is essential that the specimen contains no polycrystalline material; in polycrystalline powders, one condition for X-ray examination is that the directions of the net planes have a uniform statistical distribution in space, otherwise more or less complicated corrections must be applied, perhaps by preparation or by computer programs. An intelligent use of preferred orientation for structure analysis is recording the diffraction diagram by an area detector. The reflections are partly spread out into two dimensions, many more reflections are resolved, more reflections can be used for determining the cell constants, and overlap of intensities of reflections is diminished [116].

The properties of many substances, however, exhibit a direction dependence, which is similar to a single crystal, but at the same time indicates that the substance has only partial single-crystal character. Such a solid is said to have texture. The individual crystallites are fully or partly aligned; they are said to have preferred orientations.

Texture in a polycrystalline structure can arise in quite varied ways, e.g., deformation texture, casting texture, recrystallization texture, plate texture, or fiber texture.

Texture studies on wires and plates are very important. Initial information can be acquired from transmission and reflection patterns recorded with a flat-film camera (see Fig. 15). The type and degree of texture can be determined more or less exactly, depending on the object of investigation and the type of instrument used. One of the best-developed instruments is the Lücke texture goniometer [125]. With this apparatus, the distribution of given net planes can be directly imaged in a pole figure.

The fully automated, computer-controlled X-ray texture goniometer sold by Seifert allows reflection and transmission measurements; the background is measured and the appropriate correction applied automatically. Other corrections, e.g., absorption, are also performed automatically. The pole figure is output to a terminal. Several reflections can be measured in a fully automated programmed sequence.

15.2.7. Crystallographic Databanks

X-ray studies typically generate large amounts of data. Specimen synthesis and preparation, data acquisition, and interpretation of results are generally so expensive that repeating a structure determination is a waste of resources, unless there are compelling reasons to do so. For this reason, databanks have been set up and made available to scientists worldwide. Unfortunately, using these databanks generally incurs high costs; the importance of structure studies to the advance of the natural sciences suggests that the scientific community ought to be trying to make such data freely available to all. The following databanks deserve mention:

Powder Patterns. PDF (Powder Diffraction File) Databases [38]

Crystal Structures of Organic and Organometallic Compounds. CSDS (Cambridge Structural Database System) [37]

Crystal Structures of Inorganic Compounds. Inorganic Crystal Structure Database [39]; MDAT (minerals) [40]

Crystal Structures of Proteins. Protein Data Bank [41]

15.2.8. Noncrystallinity

The diffraction of X rays by solid matter is only possible when a certain regularity of structure exists. Scattering does not become diffraction unless the material contains equal and repeated distances. These conditions are fulfilled by a crystal lattice. If the crystallites are too small, or the crystals contain structural defects, the coherent scattering domains become smaller and the reflections become broader; this broadening may be so great that it leaves only an amorphous halo, often with a half-value width of several degrees, as is the case with quartz glass, for example. In addition, the halo is often centered near the highest intensity reflection of the corresponding crystalline phase. A similar effect occurs in solids that do not have this continuous gradation from crystalline to amorphous, but in which similar atomic distances recur. Some amorphous substances give no halo at all. If the particles in a uniform host matrix are large $(20-1000 \times 10^{-10}$ m and above) scattering takes

place at small angles ($< 0.5°$), where profile forms can yield information about the particle shape. The precondition for such studies is that the particles and the enclosing matrix differ sufficiently in electron density. Several problem areas are dealt with in somewhat more detail.

15.2.8.1. Crystal Quality Analysis

The relations discussed above hold for an ideal crystal; in practice, crystals are ideal only to a limited extent. In the following section, an attempt is made to indicate where a particular specimen should be placed in the range from ideal crystal to amorphous matter. Along with structural defects in the interior of the crystallites, or at their surfaces, the crystallite size also has to be estimated. A crystal may also contain stresses, or it may be that the single-crystal ordering is only partial, or applies in just one or two dimensions. Each of these factors affects the X-ray diffraction pattern.

15.2.8.2. Crystallite Size

A real crystal is not infinitely large. In the case of X rays and neutrons, size does not begin to affect the diffraction pattern until the crystallites (or more precisely the coherent scattering domains) become smaller than ca. 0.5 μm. When electrons are used, this limit is nearly two orders of magnitude smaller (0.01 μm).

SCHERRER, in 1918, derived the following equation between the "physical half-value width" β (in degrees 2 θ), i.e., that part of the broadening of a reflection due solely to inadequate crystallite size, and the crystallite size L:

$$\beta = \frac{K\lambda}{L\cos\theta} \tag{65}$$

Theoretically, L is the dimension of the crystallite perpendicular to the diffracting net planes; K is a constant, close to unity (often taken as 0.9); and λ is the wavelength of the radiation employed. The observed half-value width B is commonly assumed to be the sum of the instrumental broadening b and the physical broadening β. Some authors also give the formula:

$$B = \sqrt{\beta^2 + b^2} \tag{66}$$

It should be kept in mind that the error of determination is 20 – 30 %.

The instrumental broadening b can be obtained by calibrating with a substance with a large crystallite size (>0.5 μm) [126].

15.2.8.3. Defects

Like excessively small crystallite size, crystal structure defects also increase the half-value width of a reflection. Such defects make the coherent scattering domains smaller. The amount of broadening depends on the type of defect, because in general newly generated scattering domains remain coherent with one another across the new disruptions.

It is very difficult to separate the broadening of a reflection into components due to small crystallite size and to structural defects. Equation (65) implies that $\beta\cos\theta$ should be constant for all glancing angles, if the broadening results entirely from small crystallite size. Because L is the dimension perpendicular to the diffracting net planes, this statement cannot hold rigorously, except for the diffraction orders of one and the same net plane. If line broadening is the consequence of crystal structure defects alone, the width increases with θ.

15.2.8.4. X-Ray Crystallinity

The two-phase model generally accepted for high polymers has amorphous and crystalline domains occurring next to each other. If the crystallinity obtained by X-ray examination isstated as a percentage, this figure should not be regarded as a serious value; it serves only for comparison within a series of experiments. In no case should comparison be made between the X-ray crystallinities of high polymers derived from different monomers.

It has become customary to extend the term X-ray crystallinity to compounds other than high polymers, in which case numerical figures should be treated with even more caution. For example, the properties of organic pigments depend very strongly on the crystal quality of the primary particles, i.e., the coherent scattering domains. Figure 45 presents X-ray diffraction patterns of organic pigments consisting of the same compound. Between A and C, a continuous transition (such as case B) is possible. It therefore appears meaningful to speak of a crystalline component, but numerical values can never be assigned in such a case, as they can for high polymers, where por-
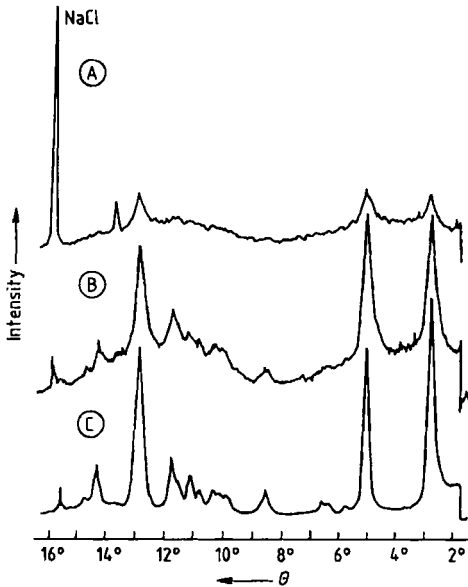
Figure 45. Organic pigments consisting of the same compound, but differing in crystal quality (patterns recorded with Cu K α radiation)
A) "Over-ground" pigment; B) Moderately well-finished pigment; C) Well-finished pigment

tions of the X-ray diffraction pattern can be attributed to amorphous and crystalline parts (Fig. 46).

15.2.8.5. Elastic Stress

The measurement of elastic stress is especially important in the examination of steels. Elastic stress (residual macroscopic stress) expands the lattice in the direction of the tensile stress and contracts it in the perpendicular direction. The corresponding shifts toward lower and higher glancing angles, respectively, can be measured by X-ray methods. These lattice changes can be correlated with the elastic moduli E:

$$\sigma_\varphi = \frac{d_\psi - d_0}{d} \frac{E}{1 + \nu} \frac{1}{\sin^2\psi} \tag{67}$$

where d is the net-plane distance in the stress-free material; d_ψ is the distance in the stressed material, where the lattice planes make an angle of ψ with the surface, and similarly d_0 for $0°$; ν is the Poisson ratio (ratio of transverse contraction to longitudinal extension); and σ_φ is the stress in direction φ in the specimen surface (the X-ray beam direction projected onto the specimen).
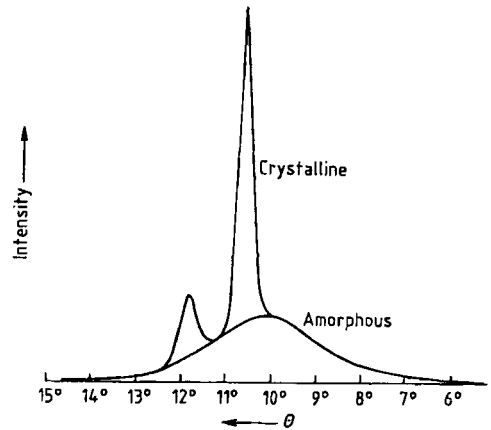
Figure 46. Amorphous and crystalline components in the X-ray diffraction pattern of a high polymer, recorded with Cu K α radiation

The macrostress should be distinguished from the residual microstress, which is due to internal stresses caused by factors such as the hardening of steels.

In contrast to other methods of measuring elastic stress, the X-ray method is not subject to distortion of the results by plastic deformation. Furthermore, highly inhomogeneous stress fields can be measured, because the irradiated surface can be as small as $1 - 10$ mm^2. The technique has the disadvantage that X rays penetrate only some 10^{-2} cm into steels, so that only stresses at the surface can be determined.

The elastic stress in a given direction can be obtained from Equation (67), by making measurements at a range of incident angles. If a photographic method is used, objects such as bridge piers can be tested nondestructively in situ.

15.2.8.6. Radial Electron Density Distribution

The systems discussed up to this point have been crystalline in the common meaning of that term. As a rule of thumb, there must be more than 10 translational periods in every direction in space before a material can be deemed crystalline (a one- or two-dimensional state of order does not qualify). The expression "noncrystalline" here means the same as "X-ray amorphous".

The more noncrystalline a sample becomes, the less possible it is to give precise statements about individual atomic distances. In these cases, only a radial distribution function (RDF) of den-

sity $\varrho(r)$ of electrons or atoms can be given, according to the equation [17, p. 795]

$$4\pi r^2 \varrho(r) = 4\pi r^2 \varrho_0 + \frac{2r}{\pi} \int_0^\infty S\, i(S)\, \sin r S\, dS \qquad (68)$$

where ϱ_0 is a function of the density of the solid, S equals $4\pi \sin \theta/\lambda$, and $i(S)$ is a function of the diffraction pattern. This equation can only be applied to samples with only one kind of atom.

15.3. Synchrotron Radiation

Synchrotron radiation is generated when electrons travel along curved trajectories at speeds close to that of light. The curvature is produced by magnetic fields. The nature and intensity of these fields determine the spectral distribution of the electromagnetic radiation produced. In general, a spectrum ranging from short-wavelength X rays ($< 0.5 \times 10^{-10}$ m) up to the ultraviolet (200 nm) is available. In contrast to X rays from rotating anode tubes, the synchrotron source offers high intensity, but also a broad spectrum, only slight divergence, and well-defined state of polarization. The radiated power (whole spectrum) obtained from an electron current of 100 mA is around 50 kW, several orders of magnitude greater than the output of conventional X-ray sources [127].

For elastic interactions at large angles ($>5°$), scattering by an amorphous material (short-range density fluctuations) is distinguished from diffraction, where interference effects play a key role. Studies in this field employ the powder, rotating crystal, and diffractometer methods, but the principal method and the most developed is also the oldest — the Laue method. Protein crystal structures can be examined in the time-resolved mode (submillisecond range) with this technique. Because of the high degree of polarization of the radiation, magnetic structures can also be investigated. At small angles, the Bragg reflections of structures with very large cells (including long periods) are accompanied by scattering, which yields information about domain sizes, defects, microcracks, and clusters and segregations of all kinds — provided there are regions differing in electron density. Time-resolved studies are possible in all these cases because of the high intensity of the X-ray beam used.

The inelastic interaction between synchrotron radiation and solid matter is also utilized by crystallographers, although spectroscopic transitions rather than diffraction phenomena are involved (EXAFS, XANES, UPS, XPS). These techniques measure the fine structure of absorption edges, to yield information about the bonding of the atom under study.

15.4. Neutron Diffraction

Many experiments can be performed with either X rays or neutrons. Neutron radiation is some orders of magnitude more expensive and so is not a candidate for studies that can be done equally well or better (because of the higher intensity) with X rays [127], [128].

An important application of neutron diffraction is based on the fact that the dependence of scattering power on atomic number is different for neutrons and X rays. For X rays, the atomic form factor is directly proportional to the atomic number, but the relationship is more complicated for neutron radiation, and there is a correlation with the energy states of isotopes. Between hydrogen and deuterium, for example, there is an appreciable difference in scattering power. The fact that neutrons, unlike X-ray quanta, are scattered by atomic nuclei means that the scattering power is virtually independent of the diffraction angle. A further advantage of the use of neutrons is the lower absorption.

By virtue of these factors, diffraction of neutrons by polycrystalline matter yields far more information than that of X rays. Thicker layers of material can be penetrated, and preferred orientations can be eliminated. Useful intensities can be measured out to large diffraction angles.

For most structure problems investigated with neutrons, an X-ray structure analysis already exists. The task is to improve or refine certain parameters. There are four essential problems:

1) To determine the coordinates of hydrogen atoms
2) To distinguish between atoms of similar atomic number
3) To perform Rietveld profile analyses on powder samples to determine special structure parameters or achieve structure refinement
4) To analyze magnetic structures
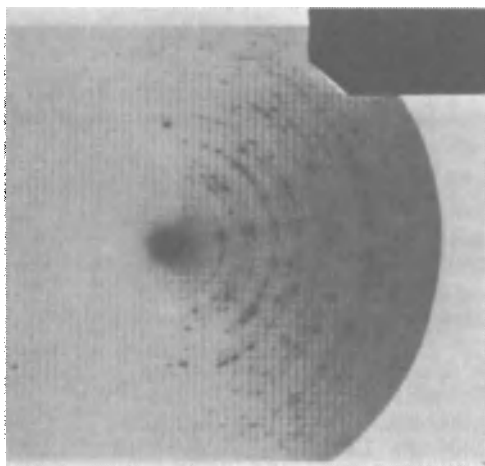
Figure 47. Electron diffraction pattern of titanium sheet, measured in reflection immediately after annealing

15.5. Electron Diffraction

15.5.1. Diffraction in the Electron Microscope

Because of their strong interaction with matter, electrons have little penetrating power, and are thus particularly suited to surface studies. No line broadening is expected up to a crystallite size of 100×10^{-10} m for electrons with a wavelength of 0.05×10^{-10} m (accelerating voltage 60 kV). If the crystallite orientation has a statistical distribution, a diffraction spectrum comparable to the Debye – Scherrer diagram is obtained. This can be utilized for phase analysis (like an X-ray powder pattern); in the case of reflection measurements, the analysis is valid for layers near the surface (Fig. 47).

Surface regions just a few atomic layers thick can be examined by diffraction of electrons that have been accelerated by a low voltage (150 V). Epitaxy can be dealt with in this way. Because of absorption, deeper layers contribute nothing to the diffraction.

Often, the distribution of the vectors of the net planes in a powder specimen is not statistical. Because the curvature of the Ewald sphere is small relative to the dimensions of the reciprocal lattice, many net planes are unequally well in their reflection position, resulting in a correspondingly unequal degree of error in the diffracted electron intensities. Electron diffraction results should therefore be verified by comparative X-ray measurements whenever possible.

If the X-ray diffraction pattern offers a fingerprint with two parameters, the glancing angle and the reflection intensity, the electron diffraction pattern gives only the glancing angle, and one cannot speak of a fingerprint, especially as even strong reflections may be absent altogether. Transmission electron diffraction data, however, can be very similar to the corresponding X-ray patterns, and may contain a similar level of information (see Figs. 48 and 13). This is only the case when the distribution of net plane vectors is statistical.

15.5.2. Electron Diffraction in the Gas Phase

The discussion of this technique has been kept short because the diffraction spectra are very similar to Debye – Scherrer patterns; the method is becoming very important for molecular structure analysis. An electron beam in a high vacuum (0.1 – 10 Pa) collides with a molecular beam, and the electrons are diffracted by the molecules. The film reveals washed-out Debye–Scherrer rings, from which a radial electron density distribution function for the molecule can be derived. Together with spectroscopic data, the distribution makes it possible to infer the molecular structure [129].

15.6. Future Developments

Many of the techniques discussed here have become routine. X-ray diffraction data of organic pigments, for example, are employed for product surveillance in industry. Not just the fingerprint as such, but also the half-value widths of the reflections and the level of the background, are utilized. The situation is similar in the fiber industry, as well as many other areas of chemical production. (Fig. 48)

The most spectacular successes of X-ray methods, however, are in molecular and crystal structure analysis. Examples are the structures of insulin [130], hemoglobin [131], and vitamin B_{12} [132]. Today, single-crystal X-ray structure analyses of relatively complex compounds (up to ca. 40 – 200 nonhydrogen atoms) are performed with computer-controlled diffractometers whose computers can be used simultaneously to solve the phase and structure problem within a few days or just a few hours.. The method has gained parity in investigation time with spectroscopic methods
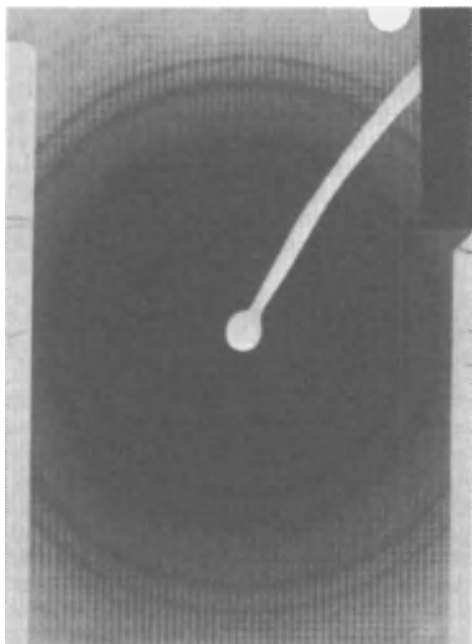
Figure 48. Transmission electron diffraction pattern of LiF

for the solution of relatively complex structure problems as far as the compounds are crystallizable. What is more, X-ray structure analysis is far superior to all other methods with respect to the reliability of the results. A single-crystal X-ray structure analysis of a compound yields everything that one could "see" if it were possible to examine a single molecule under the microscope. But in the end it is very important to remove all the discrepancies between diffraction and spectroscopic methods.

However, it can be quite laborious to solve the phase problem, even for relatively simple crystal structures. There is not yet any general method; a special solution of the phase problem must be found for every structure. The methods have been developed to the point where the solution can often be achieved by a computer alone; if this fails, however, it is usually not clear why, considerable ingenuity may be needed to reach the solution.

The fundamental reason for this situation is that the relationships in direct phase-determination methods are nothing more than probability relations. If these can be turned into true equations, single-crystal X-ray structure analyses would become a routine analytical procedure. H. HAUPT-

MANN, co-creator of the direct methods of phase determination, expressed the dilemma as "One good idea and we are out of a job".

One attempt to replace the probability relations with true equations for phase determination was that of R. ROTHBAUER [133], based on earlier work by SAYRE [78], and WOOLFSON [134]. The future must show whether the equations derived can be used for practical structure determination [135]. Their use for the determination of the absolute scaling factor and the overall temperature factor has already been tested in practice [136]. A publication by R. ROTHBAUER shows, that solving the phase problem generally is still up to date [0]

15.7. References

General References

Text Books

[1] L. E. Alexander: *X-Ray Diffraction Methods in Polymer Science*, Wiley, New York 1969.

[2] K. W. Andrews, D. J. Dyson, S. R. Keown: *Interpretation of Electron Diffraction Patterns*, Hilger & Watts, London 1967.

[3] U. W. Arndt, B. T. M. Willis: *Single Crystal Diffractometry*, University Press, Cambridge 1966.

[4] L. V. Azaroff, M. J. Buerger: *The Powder Method in X-Ray Crystallography*, McGraw-Hill, New York 1958.

[5] G. E. Bacon: *Neutron Diffraction*, 3rd ed., Clarendon Press, Oxford 1975.

[6] T. L. Blundell, L. N. Johnson: *Protein Crystallography*, Academic Press, New York 1976.

[7] M. J. Buerger: *Vector Space and Its Application in Crystal-Structure Investigation*, Wiley, New York 1959.

[8] M. J. Buerger: *Crystal-Structure Analysis*, Wiley, New York 1960.

[9] M. J. Buerger: *The Precession Method in X-Ray Crystallography*, Wiley, New York 1964.

[10] M. J. Buerger: *X-Ray Crystallography*, 7th ed., Wiley, New York 1966.

[11] M. J. Buerger: *Kristallographie, eine Einführung in die geometrische und röntgenographische Kristallkunde*, De Gruyter, Berlin 1977.

[12] J. D. Dunitz: *X-Ray Analysis and the Structure of Organic Molecules*, Cornell University Press, Ithaca 1979.

[13] C. Giacovazzo et al.: "Fundamentals of Crystallography," in C. Giacovazzo (ed.): *International Union of Crystallography*, Oxford University Press, Oxford 1992.

[14] J. P. Glusker, K. N. Trueblood: *Crystal Structure Analysis, a Primer*, Oxford University Press, New York 1985.

[15] E. F. Kaelble (ed.): *Handbook of X-Rays*, McGraw-Hill, New York 1967.

[16] A. J. Kitaigorodsky: *Molecular Crystals and Molecules*, Academic Press, New York 1973.

[17] H. P. Klug, L. E. Alexander: *X-Ray Diffraction Procedures for Polycrystalline and Amorphous Materials*, J. Wiley & Sons, New York 1974.

[18] H. Krebs: *Grundzüge der anorganischen Kristallchemie*, Enke-Verlag, Stuttgart 1968.

[19] H. Lipson, W. Cochran: *The Determination of Crystal Structures*, 3rd ed., G. Bell & Sons, London 1966.

[20] P. Luger: *Modern X-Ray Analysis on Single Crystals*, De Gruyter, Berlin 1980.

[21] D. McKie, C. McKie: *Essentials of Crystallography*, Blackwell Scientific Publ., Oxford 1986.

[22] H. Neff: *Grundlagen und Anwendungen der Röntgenfeinstrukturanalyse*, 2nd ed., R.-OldenbourgVerlag, München 1962.

[23] G. H. Stout, L. H. Jensen: *X-Ray Structure Determination*, Macmillan, New York 1968.

[24] E. R. Wölfel: *Theorie und Praxis der Röntgenstrukturanalyse*, Vieweg, Braunschweig 1975.

[25] M. M. Woolfson: *An Introduction to X-Ray Crystallography*, University Press, Cambridge 1970.

[26] W. Borchardt-Ott: Kristallographie, eine Einführung für Naturwissenschaftler; 5. Auflage; Springer-Verlag, Berlin 1997.

[27] J. P. Glusker, M. Lewis, M. Rossi: Crystal Structure Analysis for Chemists and Biologists; New York, Weinheim, Cambridge (UK) 1994

[28] J. S. Rollet (ed.): *Computing Methods in Crystallography*, Pergamon Press, Oxford 1965.

General References

Reference Works

[29] *International Tables for X-Ray Crystallography*, Kynoch Press, Birmingham, N. F. M. Henry, K. Lonsdale (eds.): "Symmetry Groups," vol. I, 1952;
J. S. Kasper, K. Lonsdale (eds.): *"Mathematical Tables,"* **vol. II**, 1959;
C. H. MacGillavry, G. D. Rieck (eds.): *"Physical and Chemical Tables,"* **vol. III**, 1962;
J. A. Ibers, W. C. Hamilton (eds.): *"Revised and Supplementary Tables,"* **vol. IV**, 1974.

[30] International Tables for Crystallography, Kluwer Academic Publishers, Dordrecht: Th. Hahn (ed.), Vol. A: Space-Group Symmetry, fifth revised edition, 2000; U. Shmueli (ed.), Vol. B: Reciprocal Space, second edition, 2000; A. J. C. Wilson, E. Prince (eds.), Vol. C: Mathematical, Physical & Chemical Tables, second edition, 1999.

General References

Databanks

[31] J. D. H. Donnay, H. M. Ondik (eds.): *Crystal Data: Determinative Tables*, 3rd. ed., **vol. 1, "Organic Compounds"; vol. 2, "Inorganic Compounds,"** U.S. Dep. of Commerce – Nat. Bur. of Standards – Joint Committee on Powder Diffraction Standards, 1972.

[32] *Landolt-Börnstein*, 7th ed., **III,** 5; III 6; III 7; III 8.

[33] A. R. Hölzel: *Systematics of Minerals, Systematik in der Mineralogie*, Ober-Olm, Germany 1990.

[34] O. Kennard, F. H. Allen, D. G. Watson (eds.): *Molecular Structures and Dimensions, Guide to the Literature 1935–1976, Organic and Organometallic Crystal Structures*, Crystallographic Data Centre, Cambridge 1977.

[35] O. Kennard et al.: *Interatomic Distances, 1960–1965, Organic and Organometallic Crystal Structures*, Crystallographic Data Centre, Cambridge 1972.

[36] R. W. G. Wyckoff: *Crystal Structures*, 2nd ed., vols. 1–6, Wiley, New York 1963–1971.

[37] F. H. Allen, et al., *J. Chem. Inf. Comput. Sci.* **31** (1991) 187–204.

[38] L. G. Berry (ed.): *Powder Diffraction File (Organic and Inorganic)*, International Centre for Diffraction Data (ICDD), Swarthmore, USA.

[39] G. Bergerhoff et al., ICSD, Inorganic Crystal Structure Database, University Bonn.

[40] A. R. Hölzel: MDAT, Datenbank der Mineralien, Systematik in der Mineralogie, Ober-Olm, Germany 1992.

[41] Protein Data Bank, Brookhaven National Laboratory, USA.

General References

Computer Programs

[42] A. Altomare et al., SIR 92, User's Manual, University of Bari and Perugia, Italy 1992 [138]

[43] P. T. Beurskens et al., DIRDIF, a Computer Program System for Crystal Structure Determination by Patterson Methods and Direct Methods applied to Difference Structure Factors, 1992.

[44] T. Debaerdemaeker et al., MULTAN 88, a System of Computer Programs for the Automatic Solution of Crystal Structures from X-Ray Diffraction Data, Universities of Ulm (Germany), Louvain (Belgium), and York (England) 1988.

[45] A. L. Spek, PLATON, a Multi-Purpose Crystallographic Tool, University of Utrecht, the Netherlands (1998)

[46] E. J. Gabe, A. C. Larson, F. L. Lee, Y. Lepage: NRCVAX, Crystal Structure System, National Research Council, Ottawa, Canada.

[47] S. R. Hall, J. M. Stewart (eds.): XTAL 3.0, Reference Manual, University of Western Australia, Australia, and Maryland, USA.

[48] C. K. Johnson: ORTEP II, Report ORNL 5138, Oak Ridge National Laboratory, Tennessee, USA.

[49] G. M. Sheldrick: SHELXS 86, Program for the Solution of Crystal Structures, University of Göttingen, Germany 1986.

[50] G. M. Sheldrick: SHELXL 93, Program for the Refinement of Crystal Structures, University of Göttingen, Germany 1993.

[51] G. M. Sheldrick: SHELX-97, Computer program for the solution and refinement of crystal strucutres. University of Göttingen, Germany 1997.

[52] D. J. Watkin, J. R. Carruthers, P. W. Betteridge: Crystals User Guide, Chemical Crystallography Laboratory, University of Oxford, England 1985.

Specific References

[53] R.-G. Kretschmer, G. Reck, POSIT in W. Kraus, G. Noke, Computer program PowderCell, BAM Berlin (1996).

[54] G. E. Engel, S. Wilke, O. König, K. D. M. Harris, F. J. J. Leusen, PowderSolve (1999) in Cerius² (Molecular Simulations Inc., 9685 Scranton Road, San Diego, CA 92121-3752, USA).

[55] E. Keller, *Chem. unserer Zeit* **16** (1982) 71–88, 116–123.

[56] R. Bonart, R. Hosemann, *Makromol. Chem.* **39** (1960) 105–118.

[57] W. Hoppe, *Z. Elektrochem.* **61** (1957) 1076–1083.

[58] W. Hoppe, E. F. Paulus, *Acta Crystallogr.* **23** (1967) 339–342.

[59] H. Fuess, *Chem. Br.* **14** (1978) 37–43.

[60] H. Diehls: *Die Fragmente der Vorsokratiker*, 8th ed., **vol. 2**, Weidmannsche Verlagsbuchhandlung, Berlin 1952, p. 97.

[61] T. Kottke, D. Stalke, *J. Appl. Crystallogr.* **26** (1993) 615–619.

[62] *Sci. Ind. (Eindhoven)* **18** (1972) 22–28.

[63] D. Kobelt, E. F. Paulus, *Siemens Power Eng.* **1** (1979) 269–271.

[64] R. A. Sparks, *Abstracts of Internat. Summer School on Crystallography Computing*, Prague, July, 28–Aug. 5, 1975, pp. 452–467.

[65] R. Boese, M. Nussbaumer: "In situ Crystallization Techniques," in D. W. Jones (ed.). *Organic Crystal Chemistry*, IUCR, Cryst. Symposia, vol. VI, Oxford University Press, Oxford 1993.

[66] E. Pohl et al., *Acta Crystallogr.* **D 49** (1993) in press.

[67] G. M. Sheldrick, E. F. Paulus, L. Vértesy, F. Hahn, Acta Crystallogr. B51, 89–98.

[68] B. Baumgartner et al.: *Modern Fast Automatic X-Ray Powder Diffractometry (Stoe & Cie GmbH, company brochure)*, Darmstadt 1988.

[69] Commission of Crystallographic Apparatus of the Internat. Union of Crystallography, *Acta Crystallogr.* **9** (1956) 520–525.

[70] Robert Huber Diffraktionstechnik: *Manual Guinier-System 600 (company brochure)*. Rimsting, Germany.

[71] Verordnung über den Schutz vor Schäden durch Röntgenstrahlen (Röntgenverordnung) January 8, 1987 (BGBl. I S. 114). Textausgabe mit Anm. von E. Witt. C. Heymanns-Verlag, Köln 1987.

[72] Permissible Dose from External Sources of Ionising Radiations, NBS Handbook 59, pp. 2, 3. addendum, April 15, 1958, Government Printing Office, Washington.

[73] Recommendations of the International Commission on Radiological Protection, Pergamon Press, New York 1959.

[74] A. L. Patterson, *Phys. Rev.* **46** (1934) 372–376.

[75] R. Huber, *Acta Crystallogr.* **19** (1965) 353–356.

[76] J. Karle, *Adv. Struct. Res. Diffr. Methods* **1** (1964) 55–88.

[77] E. Egert, G. M. Sheldrick, *Acta Crystallogr.* **A43** (1985) 262–268.

[78] D. M. Sayre, *Acta Crystallogr.* **5** (1952) 60–65.

[79] E. W. Hughes, *Acta Crystallogr.* **6** (1953) 871.

[80] G. N. Watson: *The Theory of Bessel Functions*, 2nd ed., University Press, Cambridge 1958, p. 77.

[81] H. Hauptmann, J. Karle: "Solution of the Phase Problem. I, The Centrosymmetric Crystal," *Monogr.* no. 3, Polycrystal Book Service, Pittsburgh 1953.

[82] H. Hauptmann, J. Karle, *Acta Crystallogr.* **12** (1959) 93–97.

[83] H. Hauptmann, J. Karle, *Acta Crystallogr.* **9** (1956) 45–55.

[84] J. Karle, H. Hauptmann, *Acta Crystallogr.* **14** (1961) 217–223.

[85] I. L. Karle, J. Karle, *Acta Crystallogr.* **16** (1963) 969–975.

[86] I. L. Karle, J. Karle, *Acta Crystallogr.* **17** (1964) 835–841.

[87] G. Germain, M. M. Woolfson, *Acta Crystallogr.* **B 24** (1968) 91–96.

[88] M. M. Woolfson, *Acta Crystallogr.* **A 33** (1977) 219–225.

[89] P. Main, *Acta Crystallogr.* **A 34** (1978) 31–38.

[90] G. M. Sheldrick, *Acta Crystallogr.* **A 46** (1990) 467–473.

[91] G. Germain, P. Main, M. M. Woolfson, *Acta Crystallogr.* **A 27** (1971) 368–376.

[92] T. Debaerdemaeker, C. Tate, M. M. Woolfson, *Acta Crystallogr.* **A41** (1985) 268–290.

[93] M. M. Woolfson, *Acta Crystallogr.* **A43** (1987) 593–612.

[94] H. Hauptman, *Science* **233** (1986) 178–183.

[95] R. Miller, G. T. DeTitta, R. Jones, D. A. Langs, C. M. Weeks, H. A. Hauptman, *Science* **259** (1993) 1430–1433.

[96] R. Miller, S. M. Gallo, H. G. Khalak, C. M. Weeks, *J. Appl. Cryst.* **27** (1994) 613–621.

[97] G. M. Sheldrick in S. Fortier (ed.): *Direct Methods for Solving Macromolecular Structures*, Kluwer Academic Publishers, Dordrecht 1998, pp. 401–411.

[98] I. Usón, G. M. Sheldrick, *Curr. Op. Struct. Biol.* **9** (1999) 643–648.

[99] M. Schäfer, Th. R. Schneider, G. M. Sheldrick, *Structure* **4** (1996) 1509–1515.

[100] B. Post, *Acta Crystallogr.* **A 39** (1983).

[101] E. Weckert, W. Schwegle, K. Hümmer, *Proc. R. Soc. London A* **424** (1993) 33–46.

[102] R. Herbst-Irmer, G. M. Sheldrick, *Acta Crystallogr.* **B54** (1998) 443–449.

[103] F. L. Hirshfeld, D. Rabinovich, *Acta Crystallogr.* **A 29** (1973) 510–513.

[104] D. W. Engel, *Acta Crystallogr.* 28 (1972) 1496–1509.

[105] W. C. Hamilton, *Acta Crystallogr.* **18** (1965) 502–510.

[106] D. Rogers, *Acta Crystallogr.* **A 37** (1981) 734–741.

[107] H. D. Flack, *Acta Crystallogr. Sect. A: Found Crystallogr.* **A 39** (1983) 876–881.

[108] W. Hoppe, *Angew. Chem.* **77** (1965) 484–492.

[109] A. J. C. Wilson: *Nature (London)* **150** (1942) 151–152.

[110] P. Main et al.: *MULTAN 77, a System of Computer Programmes for the Automatic Solution of Crystal Structures from X-Ray Diffraction Data*, University of York, England 1977.

[111] P. Debye, *Ann. Phys. (Leipzig)* **46** (1915) 809–823.

[112] D. Reuschling, E. F. Paulus, H. Rehling, *Tetrahedron Lett.* 1979, 517–520.

[113] T. Ito: *X-Ray Studies on Polymorphism*, Maruzen Co., Tokyo 1950, pp. 187–228.

[114] H. M. Rietveld, *Acta Crystallogr.* 22 (1967) 151–152.

[115] E. F. Paulus et al., *Solid State Commun.* **73** (1990) 791–795.

[116] R. Hedel, H. J. Bunge, G. Reck, *Material Science Forum* **157–162** (1994) 2067–2074.

[117] A. Le Bail, Structure Determination from Powder Diffraction-Database (1993–1999). http://www.cristal.Org/iniref.html.

[118] G. Reck, R.-G. Kretschmer, L. Kutschabsky, W. Pritzkow, *Acta crystallogr.* **A44** (1988) 417–421.

[119] G. E. Engel, S. Wilke, O. König, K. D. M. Harris, F. J. J. Leusen, *J. Appl. Cryst.* **32** (1999) 1169–1179.

[120] R. J. Gdanitz in A. Gavezzotti (ed.): *Theoretical Aspects and Computer Modeling*, John Wiley & Sons 1997, pp. 185–201.

[121] M. U. Schmidt, R. E. Dinnebier, *J. Appl. Cryst.* **32** (1999) 178–186.

[122] M. U. Schmidt in D. Braga, G. Orpen (eds.): *Crystal Engineering: From Molecules and Crystals to Materials*. Kluwer Academic Publishers, Dordrecht 1999, pp. 331–348.

[123] K. D. M. Harris, M. Tremayne, *Chem. Mater.* **8** (1996) 2554–2570.

[124] P. Erk in D. Braga, G. Orpen (eds.): *Crystal Engineering: From Molecules and Crystals to Materials*, Kluwer Academic Publishers, Dordrecht 1999, pp. 143–161.

[125] G. Ibe, K. Lücke, *Texture (London)* **1** (1973) 87–98.

[126] R. C. Rau, *Adv. X-Ray Anal.* **5** (1962) 104–115.

[127] A. K. Cheetham, A. P. Wilkinson, *Angew. Chem.* **104** (1992) 1594–1608.

[128] H. Fuess: "Application of Neutron Scattering in Chemistry. Pulsed and Continuous Sources in Comparison," in M. A. Carrondo, G. A. Jeffrey (eds.): *Chemical Crystallography with Pulsed Neutrons and Synchrotron X-Rays*, D. Reidel Publ., Dordrecht 1988, pp. 77–115.

[129] L. Schäfer, *Appl. Spectrosc.* **30** (1976) 123–149.

[130] T. Blundell, G. Dodson, D. Hodgkin, D. Mercola, *Adv. Protein Chem.* **26** (1972) 279–402.

[131] M. F. Perutz, H. Muirhead, J. M. Cox, L. G. C. Goaman, *Nature (London)* **219** (1968) 131–139.

[132] C. Brink-Shoemaker et al., *Proc. R. Soc. London A* **278** (1964) 1–26.

[133] R. Rothbauer, *Acta Crystallogr.* **A 33** (1977) 365–367.

[134] M. M. Woolfson, *Acta Crystallogr.* **11** (1958) 277–283.

[135] R. Rothbauer, *Z. Kristallogr.* **209** (1994) 578–582.

[136] R. Rothbauer, *Acta Crystallogr.* **A 34** (1978) 528–533.

[137] R. Rothbauer, *Z. Kristallogr.* **215** (2000) 158–168.

[138] A. Altomare, G. Cascarano, C. Giacovazzo, A. Guagliardi, M. C. Burla, G. Polidori, M. Camalli, *J. Appl. Cryst.* **27** (1994) 435–436.

16. Ultraviolet and Visible Spectroscopy

GÜNTER GAUGLITZ, Universität Tübingen, Institut für Physikalische und Theoretische Chemie. Tübingen, Federal Republic of Germany

Abbreviations

a	linear (decadic) absorption coefficient, concentration
A	absorption
b	matrix of signals
B	magnetic field vector
c	velocity of light, concentration
d	thickness, pathlength
E	energy, absorbance, extinction
E	electric field vector
f	spectral distribution
F	spontaneous emission
g	grating constant
h	Planck's constant
H	magnetic field
I	fluorescence intensity
k	Boltzmann constant, absorption coefficient
l	path length
n	refractive index
N	number of molecules
p	polarization
P	photochemistry
Q	inverse of the matrix of the absorption coefficient
R	reflectance of the sample
s	scattering coefficient
S	singlet state
T	triplet state
v	velocity
$[\alpha]$	specific rotation
α	angle, absorptance
β	angle
Γ	geometric factor
δ	phase shift
ε	extinction coefficient, molar (decadic) absorption coefficient
η	fluorescence quantum yield
ϑ	angle
Θ	ellipticity
λ	wavelength
μ	dipole moment
ν	frequency
$\tilde{\nu}$	wavenumber
τ	turbidity coefficient
φ	phase shift
Φ	angle
$[\Phi]$	molecular rotation
ω	Verdet constant

16.1. Introduction

16.1.1. Comparison with Other Spectroscopic Methods

Spectroscopy in the ultraviolet and visible regions of the spectrum (UV – VIS spectroscopy) involves observation of the excitation of electrons and is, therefore, often referred to as "electron spectroscopy." The term electron spectroscopy does not mean spectroscopy in which excitation is performed by electrons, but spectroscopy of the electronic states (excitation of the electrons).

An electron is excited if the frequency of the incident electromagnetic radiation matches the difference in energy between two electronic states. This energy difference depends on the electronic structure of the molecule being investigated and its environment. In addition, vibrational and rotational transitions can be excited. Therefore, bands in UV – VIS spectroscopy are generally broad. These bands provide little information about the molecular structure and functional groups present, especially for spectroscopy of liquids and solids [1] – [4].
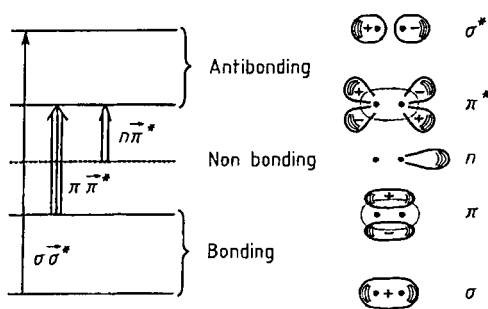
Figure 1. Bonding and antibonding orbitals, showing symmetries (+/–: sign of the wave function) and positions of atomic nuclei (•) for a diatomic molecule (HOMO = π or n; LUMO = π*)
Thickness of vertical arrows indicates intensity of the transition

16.1.2. Development and Uses

The present state of UV–VIS measurement technology has developed from its starting point *colorimetry*. In colorimetry, the intensity of a color was first taken as a measure of concentration, the sample being compared visually with standards, indicator paper being a typical example. This mainly qualitative procedure was then improved by the use of filter-photometers. These measure the attenuation of light by the sample in a narrow spectral region (photometry) and compare it with solutions containing only pure solvent. Replacement of the filter by a monochromator with a wavelength scanning mechanism then resulted in *spectrometry*.

Whereas IR, NMR, and mass spectroscopy are used mainly for the elucidation of structure and the identification of substances, UV–VIS spectroscopy enables quantitative determinations to be carried out much more precisely and reproducibly. Therefore, its primary areas of application are in quantitative analysis and clinical medicine, in the determination of drug concentrations, in the quantification of pharmaceuticals and as detectors in chromatographic processes (HPLC, TLC) (→ Liquid Chromatography) [4], [5]. Furthermore, mixtures as well as pure substances can be studied and the components determined quantitatively by methods of multicomponent analysis [6]. Since modern spectrometers operate very rapidly and can be constructed in the form of photodiode arrays, they have the advantage over other analytical methods of being usable not only for observing

stationary systems, but also for carrying out repeated determinations very rapidly, sometimes within milliseconds [3], [4].

16.2. Theoretical Principles

16.2.1. Electronic States and Orbitals

The energy of a molecule is given by its movement through space (translational energy), its rotary motion (rotational energy), the oscillatory movement of the atoms with respect to each other (vibrational energy), and the distribution of its electron density (electronic energy) [7] – [9]. The corresponding energy states can be calculated as eigenvalues by the Schrödinger equation through use of the electronic, vibrational, and rotational eigenfunctions, and the boundary conditions given by the molecular structure [10].

The energy levels or states in the UV–VIS spectral region correspond to electron density distributions (orbitals). Some typical orbitals for an organic molecule are shown schematically in Figure 1 [3]. Also shown are possible transitions of varying intensity corresponding to the symmetry of the electron density distributions of the initial and final states.

16.2.2. Interaction Between Radiation and Matter

In electromagnetic radiation, either the particle (photons) or the wavelike character is more noticeable. In the ultraviolet and visible wavelength region, the latter is the case. The electric field vector of the radiation induces charge separation (polarization) within the molecules. In the UV–VIS region, the sign of the electric field vector changes at ca. 10^{15} Hz, causing the electron density to be polarized at this frequency. The extent of this polarization depends on the dielectric properties of the molecules, their environment, and the wavelength of the radiation [11] – [13]. The Bohr–Einstein relationship

$$\Delta E = E_2 - E_1 = h\nu$$
$$h\nu = hc\bar{\nu} = hc/\lambda \tag{1}$$

connects the discrete atomic or molecular electron energy states with the frequency ν of the exciting electromagnetic radiation, where $\bar{\nu} = 1/\lambda$ is the
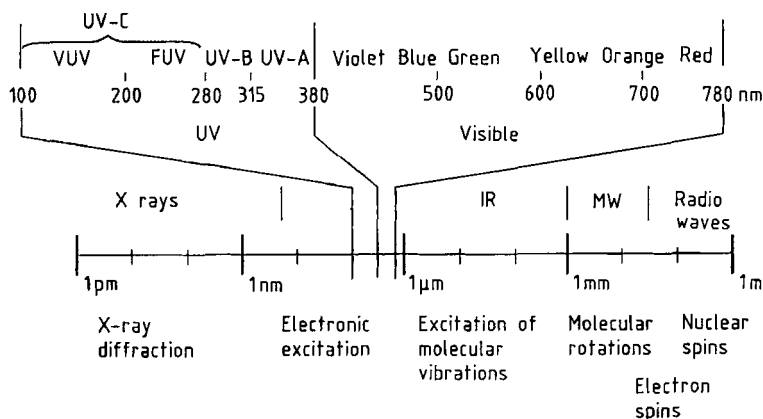
Figure 2. Location of UV and visible regions within the complete electromagnetic radiation spectrum, showing visible colors and types of excitation
VUV = Vacuum UV ; FUV = Far UV ; MW = Microwaves
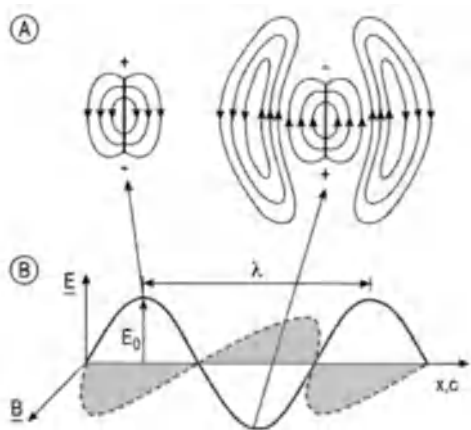
Figure 3. A) Electrical fields for a Hertz dipole ($E \parallel$ *dipole moment μ*); B) Electrical and magnetic field vectors in the propagation of electromagnetic radiation

wavenumber, c is the velocity of light in a vacuum ($c = 2.99 \times 10^8$ m/s), and h is Planck's constant (6.62×10^{-34} J $\cdot$ s). The location of this spectral region within the complete electromagnetic radiation spectrum is shown in Figure 2, which also indicates the colors of the visible range.

16.2.2.1. Dispersion

The simple model of a harmonic oscillator can account for the processes of polarization, dispersion, absorption, and scattering. In *polarization*, interaction of the radiation with matter (dielectric)

induces a reversing dipole in each molecule. The effect is proportional to the number of particles, so that polarization is a volume-based quantity. The induced dipole moment represents a Hertz dipole reversing with the excitation frequency, thus behaving as a radiation source. Because of the fluctuation of the charge distribution that fluctuates with the dipole's eigenfrequency. As shown in Figure 3, this radiation is emitted perpendicularly to the axis of polarization.

When radiation propagates through a medium, its frequency remains the same due to the conservation of energy, but its velocity decreases [11], [13]. The ratio of propagation rates is the refractive index n ($n \geq 1$). The variation of refractive index with frequency or wavelength is the dispersion [11], [13]. Two cases must be distinguished – absorption and scattering.

16.2.2.2. Absorption

If the frequency of the radiation corresponds closely to the energy difference of the transition between two energy states, as given by Equation (1), this leads to a resonance excitation, a change in the electron density distribution, and an electronic transition (see Fig. 1) from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO). In solids and metals, HOMO corresponds to the valence, and LUMO to the conduction band [12], [14], [15]. The energy required for this transition is provided by the radiation. This process is known
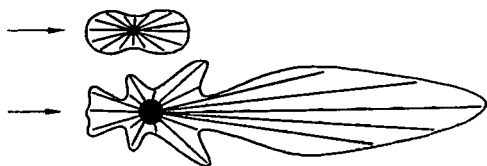
Figure 4. Intensity distribution of Rayleigh (above) and Mie (below) scattered radiation projected onto a plane. Excitation radiation is in the direction of the arrow. The black spot represents the superposition of many molecular dipoles in all spatial directions.

as (induced) absorption. In the range of an absorption band, anomalous dispersion is observed (refractive index increases with wavelength). Depending on the molecular environment and possible pathways of deactivation, the new excited state can exist for 10^{-13} to 10^{-3} s.

Only gaseous atoms give line spectra. Molecules in the gaseous state and in solution produce bands (sometimes with vibrational structure), because the energy for the electronic transition is very high, and this transition is usually accompanied by vibrational and rotational excitation. Line broadening in solution is caused by interaction between the solvent and the molecules, leading to a reduction in the lifetime of the excited state. According to Heisenberg's uncertainty principle

$$\Delta E \Delta \tau \geq h/2\pi \tag{2}$$

the energy uncertainty of the excited state increases at the expense of its lifetime. The relationship between refractive index and absorption coefficient can be expressed by the Kramers – Kronig equations [16].

16.2.2.3. Scattering

If the frequency of incident radiation does not correspond to the energy difference between two electronic levels, the polarization does not lead to a new stable (stationary) electron density distribution. After a very short delay, the molecule emits the energy that it absorbed previously. In this frequency range normal dispersion is observed (i.e., refractive index decreases with wavelength). When refractive index is plotted as a function of wavelength a dispersion curve is obtained. It characterizes the molecule together with the polarizability. With the aid of the dispersion curve, relationships between the molecular structure and the spectrum can be established. This gives good

results in the infrared, but only poor results in the UV – VIS region [17].

Since the energy causing the polarization is reemitted immediately in the nonresonant case (see Fig. 3), no energy is absorbed. Even the smallest molecules show scattering effects. However, since the intensity of scattered radiation increases proportionally to the sixth power of the length of the induced dipole moment, the radiation becomes observable only for molecular diameters of $\geq \lambda/10$ (Rayleigh scattering) and/or extremely high radiation intensities (lasers) [18], [19].

In the case of Mie scattering (larger molecules with a diameter of $\lambda/4 - \lambda/2$), a molecule contains several scattering centers. Thus, the different scattering centers emit waves that interfere to varying extents in various spatial directions, depending on the location of the scattering centers. Hence, the interference pattern contains information about the size and shape of the molecule. The intensity distribution of Rayleigh and Mie scattering assuming many molecules with statistical orientations overlap is shown in Figure 4 [18].

16.2.2.4. Reflection

Whereas in gases or liquids the particles are disordered and cause random scattering, atoms or molecules in the surface layer of solids (especially crystals and metals) are very highly ordered. Especially metals exhibit extremely polarizable electron density distributions. The many Hertz dipoles show constructive interference in certain directions. At solid surfaces this leads to directional reflection, which encloses the same angle with the optical normal to the surface as the incident radiation [11], [20] (Fig. 5). Reflectance R is given by the Fresnel equation, which, for the simplest case of normal incidence and a transparent medium, is [11], [13]

$$R = \left[\frac{(n_1 - n_2)}{(n_1 + n_2)} \right]^2 \tag{3}$$

When radiation enters a material with a higher refractive index n, it is refracted toward the optical axis (Snell's law):

$$n_1 \sin\alpha = n_2 \sin\beta \tag{4}$$

For a material with a lower refractive index, the angle between reflected radiation and the optical axis becomes larger. Under these conditions, total internal reflection occurs at the critical angle β_{cr}
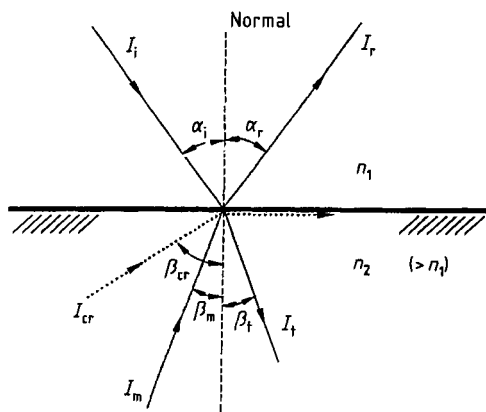
Figure 5. Ray paths for internal, external, and total internal reflection
Indices i, r, t = Incident, reflected, transmitted; I = Intensity; cr = Critical angle; m = Ray out of medium 2 corresponding to incident ray I_i, forms I_r

(see Fig. 5) [20]. In recent years, techniques based on reflection have become very important in analysis.

16.2.2.5. Band Intensity

Regarding transitions from ground to excited state and vice versa, the interaction of the radiation with the molecules are in principle the same. For this reason, the Einstein coefficients, which are a measure of probability of the two transitions, are the same for both (induced) absorption and induced emission [11], [12], [14], [15]. Which of the two transitions is more effective for the interaction with radiation depends only on the relative distribution of the molecules between the two states.

In the normal case, according to Boltzmann's law,

$$N_2 = N_1 e^{-\Delta E/kT} \tag{5}$$

where N_1 is the number of molecules in the ground state and N_2 that in the excited state; the ground state is occupied to a considerably greater extent than the excited state (not strictly true for rotational levels) so absorption mainly occurs (ΔE = energy difference, k = Boltzmann's constant, T = temperature in K). However, if intermediate energy levels are produced by high-intensity radiation (optical pumping), leading to an inversion in the populations of the two energy levels, induced

emission predominates. This principle is utilized in lasers [21].

16.2.3. The Lambert – Beer Law

Many years ago, BOUGUER, LAMBERT, and BEER discovered a relationship between the number of particles in a sample, their properties, the path length of the sample, and the observed attenuation of light [22], [23].

16.2.3.1. Definitions

When radiation penetrates a medium, an exponential relationship exists among the "extinction" (attenuation of transmitted radiation), the concentration of absorbing particles in the medium, and path length. The extinction is the sum of the effects of absorption and scattering. The proportionality constant is the *molar* (decadic) *absorption coefficient* ε_λ in the case of absorption and the *turbidity coefficient* in the case of scattering [24], [25]. By taking the logarithm of the ratio of the radiant powers, i.e., of the nonabsorbing reference [$\Phi(\lambda,0)$] to that observed passing a sample [$\Phi(\lambda,d)$], a linear relationship is obtained:

$$E(\lambda) = \log \frac{\Phi(\lambda, 0)}{\Phi(\lambda, d)}$$
$$= \varepsilon_\lambda c \, d$$
$$= a_\lambda d \tag{6}$$

In the definition of extinction (E, decadic internal absorbance), the sample thickness d is expressed in centimeters, and the concentration c of the solutions in moles per liter. IUPAC definitions of these quantities are given in Table 1, with their meanings and formal relationships [23], [26], [27].

For investigation of solid bodies or thin films, molar concentrations such as those used for solutions cannot be quoted. In these cases, absorption coefficients a, with dimensions of cm^{-1}, are used instead, which correspond to the product of the molar absorption coefficient and the concentration of the particles.

16.2.3.2. Deviations from the Lambert – Beer Law

The Lambert – Beer law assumes that monochromatic radiation falls perpendicularly onto the sample. Since the absorption coefficient is a

Table 1. Physical quantities relating to radiation, with symbols and units in accordance with the IUPAC standard

Quantity	Symbol	Unit
Radiant power (radiant energy Q)	$\Phi_i = dQ/dt$	W
Radiant power transmitted	Φ_t	W
Radiant power reflected	Φ_r	W
Radiant power absorbed	Φ_a	W
Radiant intensity	I	W/sr, W
Reflectance or reflection factor	$\varrho = \Phi_r / \Phi_0$	1
Absorptance or absorbance factor	$x = 1 - \tau$	1
Internal transmittance	τ_i	1
Internal absorptance	α_i	1
(Decadic) internal absorbance	$A = -\log_{10} \tau_i(\lambda)$	1
Linear (decadic) absorption coefficient	$a(\lambda) = A/d$	cm^{-1}
Molar (decadic) absorption coefficient	$\varepsilon = A/(c\,d)$	$L\ mol^{-1}\ cm^{-1}$
Molar concentration of absorber	c	mol/L
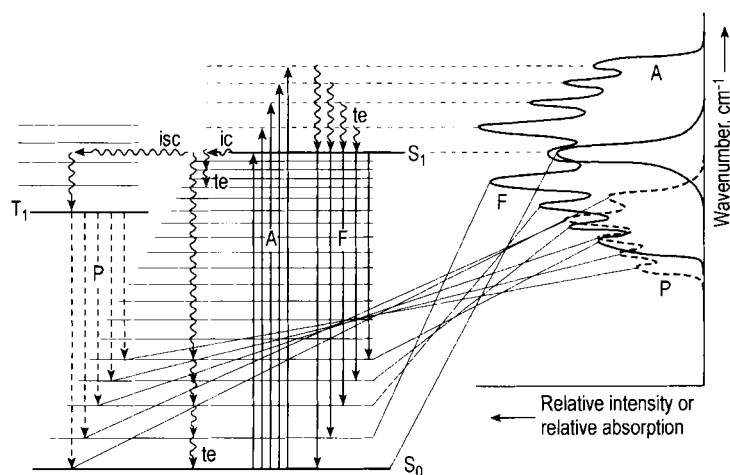
Figure 6. Schematic diagram of energy levels of an organic molecule with singlet ground state (S_0), singlet excited state (S_1), and triplet excited state (T_1) with possible photophysical transitions
The spectra resulting from these transitions are shown on the right. A = Absorption; F = Fluorescence; P = Phosphorescence

molecule-specific quantity, any chemical change of the molecule or interaction with the solvent, with other components or dissociation/association leads to false results when the Lambert – Beer law is applied (chemical deviations). On the other hand physical deviations are observed if the natural bandwidth of the absorption spectrum of the components in the sample is narrower than the spectral bandwidth (see Section 16.3.3) of the "monochromatic" radiation produced by the equipment [3], [22]. If more than one component in the sample absorbs radiation, the total effect is the sum of the molar absorption coefficients of the different components times their concentration.

16.2.4. Photophysics

16.2.4.1. Energy Level Diagram

The energy level diagram for a typical organic molecule is shown in Figure 6, which illustrates the energy levels and transitions described above. In the absorption (A) of electromagnetic radiation at the resonance wavelengths, various vibrational states are reached in the excited electronic state (S_1), and are coupled with the electronic transition (within 10^{-15} s). Many vibrational and rotational levels exist in each electronic state (S_n), but these are not shown for reasons of clarity. Normally, only the first excited electronic state is important in spectroscopy, because all higher states have very short lifetimes (relaxation times).

Table 2. Possible deactivation processes from the vibrational ground state of the first excited singlet state (see also Fig. 6)

Process type	Abbreviation	Name of process	Process description
Radiationless (rd)	te	thermal equilibration	relaxation from high vibrational level in the same electronic state lifetime of excited vibrational level determined by: – external effects (collisions, medium) – inner deactivation (transfer of energy in torsional vibrations to heavy substituents)
	ic	internal conversion	isoelectronic transition within the same energy level system from the vibrational ground state of a higher electronic state into the very high energy vibrational state of a lower electronic state
	isc	intersystem crossing, intercombination	isoelectronic transition into another energy level system $(S \leftrightarrow T)$, usually from the vibrational ground state such a radiative transition is forbidden because of the spin inversion prohibition (except for heavy nuclei)
Radiation transfers (spontaneous emission)	F	fluorescence	without spin inversion from S_n (provided that lifetime is 10^{-8} s) within singlet system
	P	phosphorescence	out of triplet into singlet system (provided that lifetime is 10^{-3} s) very low probability, only possible at low temperatures, or in a matrix
Photochemical reactions			photoinduced reaction leading to ionization, cleavage, or a bimolecular step, provided that lifetime in excited state is relatively long

In organic molecules in the ground state, electron spins are usually paired (singlet states). However, in inorganic transition-metal complexes in the ground state, triplets, and other states with a larger number of unpaired electrons occur, in accordance with Hund's rule. Therefore, several energy level systems usually exist side by side as solutions of the Schrödinger equation. A direct excitation of the singlet electronic ground state to the excited triplet state (T_1 parallel unpaired spins) is forbidden in principle because of spin inversion, so that the process of induced absorption from singlet states always leads to excited singlet states (S_n) [3], [5], [7], [28].

16.2.4.2. Deactivation Processes

Figure 6 shows deactivation as well as excitation processes. After excitation to a higher level, the molecule usually relaxes to the vibrational ground state of the S_1 level by thermal equilibration (te). This S_1 level is the starting point for several competitive processes, the preferred one depending on the type of molecule, temperature, and its environment [29]. These include (1) radiationless deactivation (rd), which consists of internal conversion (ic) followed by thermal equilibration (te); (2) intercombination or intersystem crossing (isc); (3) spontaneous emission (F, P); and (4) photochemical processes (see survey in

Table 2) [3], [28]–[31]. The vibrational ground states in electronically excited states have a relatively long lifetime (lifetime $S_1 = 10^{-8}$ s, lifetime $T_1 = 10^{-3}$ s).

In competition with radiationless deactivation, energy can also be lost in the form of radiation (F, P) [3], [28]. Fluorescence occurs with molecules that either (1) have extended π-systems (such as polycondensed aromatics); (2) do not permit deactivation by torsional or rotational motion of parts of the molecule; or (3) have no heavy atoms as substituents [32], [33]. In addition to these molecular properties, the environment also plays a part. Thus, the fluorescence intensity increases at low temperature and in solid matrices. This is even more important in phosphorescence, where the $T_1 \rightarrow S_0$ transition is in fact spin-forbidden. If a higher vibrational level ($v' > 0$) is occupied in T_1, in accordance with the Boltzmann equation (Eq. 5), so-called delayed fluorescence [28] can occur by backward intersystem crossing $[T_1 \rightarrow S_1 (v = 0) \rightarrow S_0]$.

Alternatively, during this relatively long lifetime, the molecule can undergo an internal structural change or form new reaction products by an activated collision (photochemistry). This deactivation route is promoted by high light intensities and is used in photochemical synthesis, but it can lead to undesired side effects (e.g., in fluorescence spectroscopy) because photochemical processes
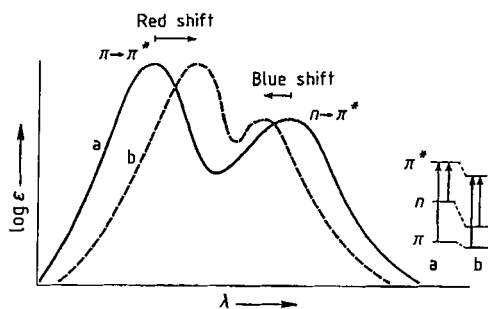
Figure 7. Schematic diagram showing effect of the solvent polarity on positions of the maxima of the $\pi \to \pi^*$ and $n \to \pi^*$ transitions
a) Methanol; b) Cyclohexane

reduce the number of molecules that can fluoresce (photobleaching) [5].

16.2.4.3. Transition Probability and Fine Structure of the Bands

A wave function can be assigned to each energy state. In the induced transition between two energy levels, the overlapping of these wave functions is important and strongly influences the intensity. For this reason, all transitions do not have the same intensity; rather, they have varying transition moments, so that absorption spectrum A on the right of Figure 6 shows a vibrational fine structure with individual peaks of varying intensities [14], [31].

Depending on the positions of the nuclei at their equilibrium distances in the molecule in the ground and electronically excited states, these wave functions overlap to varying extents and lead to differing intensity distributions for the vibrational transitions within the entire band. Since atomic nuclei hardly move during the electronic transition (Born–Oppenheimer approximation), the shape of the entire absorption band, according to the Franck–Condon principle, is given mainly by the relative positions of the equilibrium distances of the atoms in the electronic ground and excited states. The lower the degree of interaction between the molecules and their environment, the larger the vibrational splitting becomes. Absorption, fluorescence, and phosphorescence spectra are shown schematically in the right-hand side of Figure 6.

16.2.5. Chromophores

Unlike IR spectra (normal vibrations), electron density distributions in UV–VIS spectra do not depend so much on molecular structure or on interaction with solvent molecules. However, changes in the electronic ground or excited state lead to a shift in the relative position of the energy levels and to a change in polarizability. Therefore, even in the UV–VIS region, molecules may be identifiable by the position and height of the absorption bands. Both are influenced, e.g., by extension of conjugated π-systems by substitution or, in the case of aromatics, by electron-attracting or electron-repelling substituents in the *ortho*, *para*, or *meta* position and by the effects of solvents and of pH. This is especially important in pharmaceutical chemistry [5].

As shown in Figure 7, $\pi \to \pi^*$- and $n \to \pi^*$-transitions often behave differently. If cyclohexane (b) is replaced by methanol (a) as the solvent, the polarity increases considerably, so that the $\pi \to \pi^*$ band is shifted by 10–20 nm towards the red (bathochromic) and the $n \to \pi^*$-band by 10 nm towards the blue (hypsochromic). This shift gives an indication of the effect not only of the solvent polarity on the behavior of the bands, but also of the structure of the system. A change in intensity of the absorption band is described as hyperchromic (intensity increase) or hypochromic (intensity decrease) [5], [31].

16.2.6. Optical Rotatory Dispersion and Circular Dichroism

16.2.6.1. Generation of Polarized Radiation

For substances that are optically active or for molecules that are arranged in thin films or with a given orientation, additional information can be obtained with polarized electromagnetic radiation [11], [34]. Light can be plane, circular, or elliptic polarized. Plane-polarized light can be obtained with either polarizing films or special anisotropic crystals. Other possibilities are shown in Figures 8 and 9 [11], [13], [35]. Plane-polarized light is obtained from two circular-polarized beams that rotate in opposite directions and have the same intensity and phase (see Fig. 8).

As shown in Figure 9 A plane-polarized light is formed from two plane-polarized beams whose phase difference $\Delta \varphi$ equals zero. If a phase shift of 90° ($\Delta \varphi = \pi/2$) is chosen, as shown in Fig-
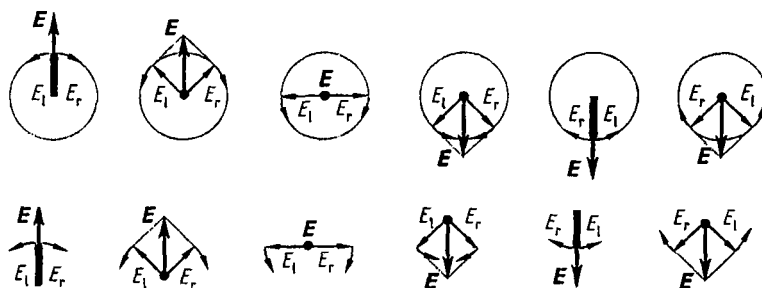
Figure 8. Production of plane-polarized light from two superposed beams of left and right circular-polarized light
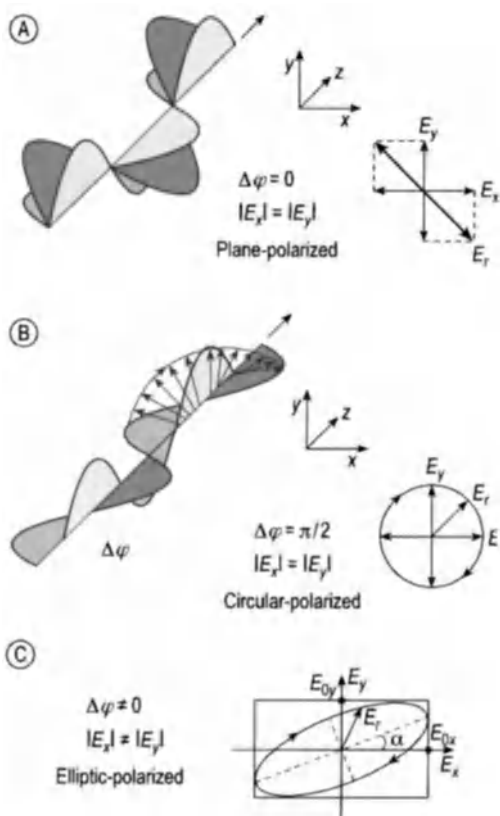
Figure 9. Production of plane- (A), circular- (B), and elliptic- (C) polarized light

ure 9 B, spiral propagation of the resulting field vector is produced, giving circular-polarized light. If the phase difference between two plane-polarized light beams is chosen arbitrarily, or if the amplitudes of the two field vectors are different, elliptic-polarized light (Fig. 9 C) is obtained, which represents the most general form of polarization.

16.2.6.2. Interaction with Polarized Radiation

The plane of polarization of plane-polarized light can be rotated to the left or the right by interaction with optically active substances. By convention, an observed rotation in a counter clockwise direction when viewed against the direction of propagation of radiation is called levorotatory and is given a negative sign (the opposite holds true for clockwise – dextrorotatory – rotation). Optical activity reveals structural properties usually referred to as chirality. These structural properties remain unchanged in molten materials, in solution, and also in complexes. For chiral substances, *quantitative polarimetry* is a very sensitive technique: it is used mainly in pharmacy and medicine to determine concentration.

The two circular-polarized beams of the incident radiation (l, levorotatory; r, dextrorotatory) not only can be influenced with respect to their direction of rotation, but also, in the region of an absorption band, can be absorbed to a different extent by the sample, so that in addition to so-called optical rotatory dispersion (ORD), circular dichroism (CD) or the Cotton effect is observed [35] – [38]. Along with these classical methods of analysis, modern methods for the analytical investigation of surfaces and boundary layers have become very important, including the use of polarized light in ellipsometry (→ Surface Analysis) and surface plasmon resonance. Under the influence of external forces (e.g., a magnetic field), even optically inactive substances can be caused to produce magnetooptic rotation.

16.2.6.3. Optical Rotatory Dispersion

The rotation α caused by optically active bodies is proportional to the thickness d (in millimeters) of the substance through which light
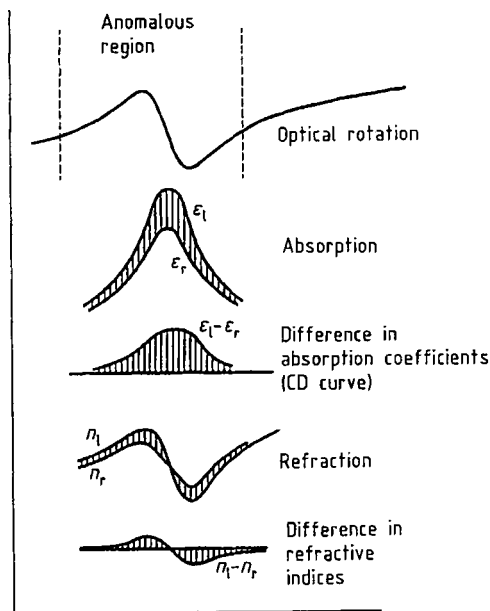
Figure 10. Rotation and circular dichroism optical rotatory dispersion in normal and anomalous spectral regions [38]

passes. The specific rotation $[\alpha]$ for wavelength λ is defined by [36], [38]:

$$[\alpha]^\lambda = \frac{\alpha^\lambda}{d} \tag{7}$$

If an optically active substance is dissolved in an inactive solvent, α is given by Biot's law:

$$\alpha^\lambda = [\alpha]^\lambda \frac{c\,d}{100} \tag{8}$$

Here, because of the relatively small rotation of solutions, sample thickness is given in decimeters, and concentration c in grams per 100 mL of solvent. The sodium D line ($\lambda = 589.3$ nm) is often used as incident radiation, and measurement is performed at 25 °C. These parameters are indicated by quoting indices in the form $[\alpha]_{25}^D$. The *molar* or *molecular rotation* $[\Phi]^\lambda$ (sometimes $[\alpha]^\lambda_M$) is given by

$$[\Phi]^\lambda = [\alpha]^\lambda \frac{M}{100} = \frac{100\alpha^\lambda}{d\,c} \tag{9}$$

where M is the molar mass, c concentration in moles per liter, and d path length in centimeters. Specific rotation is a characteristic quantity for a substance, but it also depends on experimental

conditions (temperature, concentration, solvent, wavelength of light) and perhaps time. Along with concentration and temperature, intermolecular forces also play a part. A change of solvent can even lead to a reversal of the direction of rotation.

If optically active substances undergo chemical conversion (e.g., sucrose in acid solution), the specific rotation becomes time dependent (*mutarotation*). For example, when dextrorotatory sucrose reacts until chemical equilibrium is reached, a levorotatory mixture of glucose and fructose is formed (inversion of sucrose).

The specific rotation as a function of wavelength is known as *optical rotatory dispersion* [39]. The angle of rotation usually increases as wavelength decreases. This *normal rotatory dispersion* is found in broad spectral regions that are sufficiently remote from absorption bands of the optically active substance. Rotation can be represented by Drude's equation

$$\alpha^\lambda = \sum_i \frac{A_i}{\lambda^2 - \lambda_i^2} \tag{10}$$

where A_i is a constant and λ_i is the wavelength of the nearest absorption maximum. This region corresponds to the normal dispersion curve.

16.2.6.4. Circular Dichroism and the Cotton Effect

The different refractive indices n_l and n_r for left- and right-polarized light lead to a specific rotation. In the region of an absorption band the left- and right-hand polarized beam are absorbed to different extents, and the originally plane-polarized light beam becomes ellipticpolarized ($\varepsilon_l \neq \varepsilon_r$). This phenomenon is called circular dichroism. Circular dichroism is coupled with peaks and troughs in the ORD curve (anomalous dispersion) [35], [38]. Both anomalous ORD and CD are grouped under the term *Cotton effect* [34].

Figure 10 shows the relationships between the spectral behavior of the two polarized beams with respect to rotation, absorption, and refraction. Only optically active substances interact differently with two polarized beams, so that this difference leads to observable effects. Whereas in the region of anomalous rotatory dispersion the ORD curve shows a point of inflection, the CD curve has a maximum or minimum [38]. The extreme of the CD curve coincides approximately with the point of inflection of the ORD curve. In analogy

to molar rotation, the molar ellipticity of the CD curve is given by

$$[\Theta]^{\lambda} = 100 \, \Theta^{\lambda} \frac{d}{c} \tag{11}$$

Where Θ is the ellipticity of the transmitted beam in angular degrees, d is expressed in centimeters, and c in moles per liter. In the SI system, molar ellipticity should be expressed in deg m^2 mol^{-1}.

The relationship between rotation and refractive index or between ellipticity and absorption coefficient can be represented, in analogy to dispersion theory, by the model of coupled linear oscillators or by quantum mechanical methods. ORD and CD are related to one another by equations analogous to the Kramers – Kronig equations [34].

16.2.6.5. Magnetooptical Effects

If plane-polarized radiation enters a transparent isotropic body parallel to a magnetic field, the plane of polarization is rotated (magnetooptical rotation, MOR, or Faraday effect) [40]. The angle of rotation α is given by

$$\alpha = \omega l H \tag{12}$$

where H is the magnetic field strength, l the path length, and ω the Verdet constant [38]. The degree of rotation depends on the substance, the temperature, and the wavelength of the radiation. The wavelength dependence is known as *magnetooptical rotatory dispersion* (MORD). Since the Faraday effect is practically instantaneous, it is used in the construction of rapid optical shutters and for the modulation of optical radiation paths. If Zeeman splitting of atomic and molecular terms takes place in the magnetic field, this leads to *magnetic circular dichroism* (MCD) [38], [40].

All of these effects are very important in laser technology for modulation and, as an extension of absorption spectroscopy, for the assignment of electronic transitions.

16.3. Optical Components and Spectrometers

16.3.1. Principles of Spectrometer Construction

Photometers or spectrometers in the ultraviolet and visible spectral region either are built only for the measurement of light absorption, fluorescence, reflection, or scattering, or have a modular construction so that they can be used for several measuring operations. Spectrometers always include a polychromatic light source, a monochromator for the spectral resolution of the beam, a sample holder, and a detector for measuring the radiation that has been modified by its passage through the sample. Depending on the measuring problem, these components can be modified, and the light path changed [3], [4].

16.3.1.1. Sequential Measurement of Absorption

In the measurement of absorption, the radiation is measured against a reference for each wavelength (monochromatic) after attenuation by a sample. By this reference measurement, such effects as reflections at the cell windows or other optical surfaces can be corrected for, if sample and reference differ only with respect to the substance being examined and are otherwise optically identical. The equipment can be of the single- or double-beam type. In *single-beam equipment*, radiation of the selected wavelength is passed alternately through the reference cell and the sample-containing cell before it strikes the detector. In older, more simple equipment, the position of the sample holder in the cell compartment is adjusted manually. Usually, only a small number of wavelengths are measured because a spectrum cannot be recorded automatically.

Photometers can be of the *broadband type* (broad spectral band light source + filter) or the *narrow-band type* (line spectrum light source + filter) [41], [42]. In more expensive equipment, continuous radiation sources (broadband) are used, and the filter is replaced by a monochromator, so that a complete spectrum can be observed point by point. Only a very good monochromator can give resolution (monochromaticity) as good as that from a combination of a line source and filter since in the latter system, the very small width of the
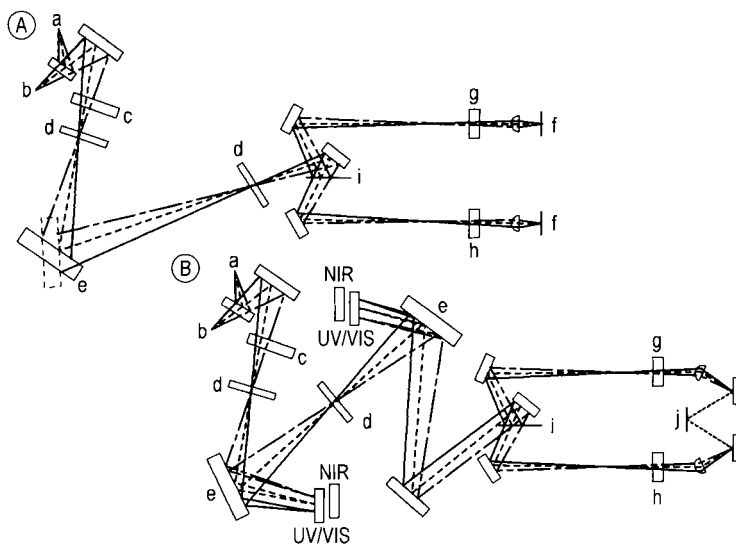
Figure 11. Optical light paths in double beam spectrometers [43]

A) Apparatus with single monochromator and two detectors; B) Research apparatus with double monochromator and beam splitter with one detector

a) Halogen lamp; b) Deuterium lamp; c) Filter wheel; d) Slit; e) Monochromator; f) Photodiode; g) Reference; h) Sample; i) Beam splitter; j) Detector

natural line determines the bandwidth of the radiation used [41], [42].

Single-Beam Equipment. Modern single-beam equipment first records a spectrum of the standard in the reference optical path and then the spectrum of the sample. The internal microprocessor serves not only for control but also for calculating the intensity ratio according to the Lambert – Beer law (see Eq. 6). The optical light path of such equipment is shown in Figure 11 A [43]. This equipment has fewer optical components and is less mechanically complex (no beam splitter) than the double-beam spectrometer, but the time required for each sample measurement is longer (successive measurements of dark, reference, and test spectra) [44]. Good stability of the light source and of the electronics is necessary to avoid any drift between measurements.

Double-Beam Equipment. In double-beam equipment, the two light paths (through the sample and the reference) are automatically continuously interchanged during the wavelength scan. A rotating mirror, oscillating mirror, or half-transparent mirror splits the beam after it emerges from the monochromator. However, for accurate measure-

ment, at each wavelength, the "dark" signal must be observed as well as that for the reference cell and the sample. The optical paths in equipment with single and double monochromators are shown in Figure 11 A and 11 B.

High recording rates lead to errors in the spectral resolving power (which is fixed by the slit width of the monochromator) because sample and reference values are no longer determined at the same wavelength [3]. In many types of equipment, a rotating mirror, which is divided into four segments, is used to measure two dark periods, one reference, and the sample. Usually the mirror wheel rotates at 50 Hz, so that a maximum of 50 points per second is observed. Wavelength scan rates today can be as high as > 1200 nm/min or > 20 nm/s. Thus, at 50 Hz and this extreme scan rate, the reading is taken only every 0.4 nm (with a very short signal integration time); in addition, the wavelength differs from the reference wavelength by 0.2 nm (restricting spectral resolution to ca. 1 nm). Therefore, a considerably slower recording rate is used for high-resolution spectra. Alternatively, it is possible to stop at each measured wavelength ("stop and go") [45], [46].

In double-beam spectrometers, fewer problems of stability of the light source and the electronics
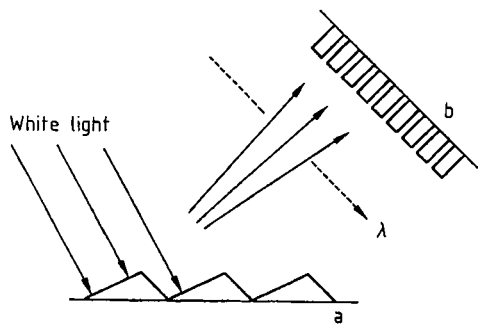
Figure 12. Diagram of multiplex detector (diode array) combined with a grating for spectral dispersion of white light
a) Grating; b) Diode array

arise. Also, a spectrum can be recorded considerably more rapidly than with single-beam equipment. Disadvantages include the greater number of optical components (each optical component – i.e., mirrors and lenses – reduces the available light intensity due to reflection); the polarization of the light beams because of movement of the optical components; and the cost and complexity of construction [3].

Radiation in the UV – VIS region is of high energy. Therefore, in sequential spectrometers the monochromator is positioned before the sample. In IR equipment, in which the sample itself emits radiation, the reverse arrangement is used [3].

16.3.1.2. Multiplex Methods in Absorption Spectroscopy

With multiplex detectors, the radiation of the various wavelengths no longer need to be measured sequentially (one after the other). Instead, by parallel imaging of many wavelengths on a spatially resolving detector they can be measured simultaneously so that a complete spectrum is recorded. This can be done by using linear photodiode arrays or, more recently, so-called charge-coupled devices (CCDs) [2], [3], [47], [48].

For this equipment, the classical single-beam arrangement is used. However, the light striking the sample is not monochromatic, but white. After passing through the sample, light is spectrally dispersed by a monochromator without an exit slit (polychromator) and then strikes an array of diodes. A possible arrangement of such a polychromator with a multiplex detector is shown in Figure 12. The resolution is determined by the

specification of the grating and the number of diode elements in the array (256 – 1024).

By this method, complete spectra can be recorded in fractions of a second (1 – 100 ms, depending on the equipment). Within these times, a reference measurement is impossible so, as with modern microprocessor-controlled single-beam equipment, the reference and sample can only be observed consecutively (except when using a double diode array configuration). The sample is exposed to white light, which can result in photoinduced processes. However, this type of equipment has the advantage of recording hundreds of spectra within a few seconds, so many dynamic processes can be observed in real time. This requires that neither the diode array nor the light source drifts within the time necessary for measurement, which is true to only a limited extent. In fact, the problems are the same as those with microprocessor-controlled single-beam equipment. The application of this type of equipment for routine measurements, in kinetics, and as a detector in chromatography is continually increasing [49].

Hadamard Spectroscopy. If, instead of the exit slit, an irregular arrangement of slits of various widths is used, and an arrangement in the form of a comb-like second slit is moved in front of this, for each relative position a different summation of intensities at various wavelengths is obtained on a nonmultiplex detector. By using convolution functions, spectra can be computed rapidly for a cycle of different slit positions. Since for mathematical deconvolution so-called Hadamard matrices are used, the principle is known as Hadamard spectroscopy. It has gained some importance in waste gas analysis [50].

Fourier Transform Spectroscopy. After the development of diode arrays, the use of Fourier transform spectroscopy, which is very successful in the IR, NIR, and FIR regions, has not been very widespread for routine analyis in the UV–VIS region. In Fourier transform UV spectroscopy [11], the polychromatic light beam is split in a Michelson interferometer, 50% falling on a fixed mirror and 50% on a second moving mirror, and then recombined by the beam splitter. Each position of the moving mirror gives a phase relationship between the two partial beams. After recombination of the component beams the modulation of intensity caused by the position change depending on time can be transformed by Fourier analysis to spectral information (wavenumber) [51]–[54].
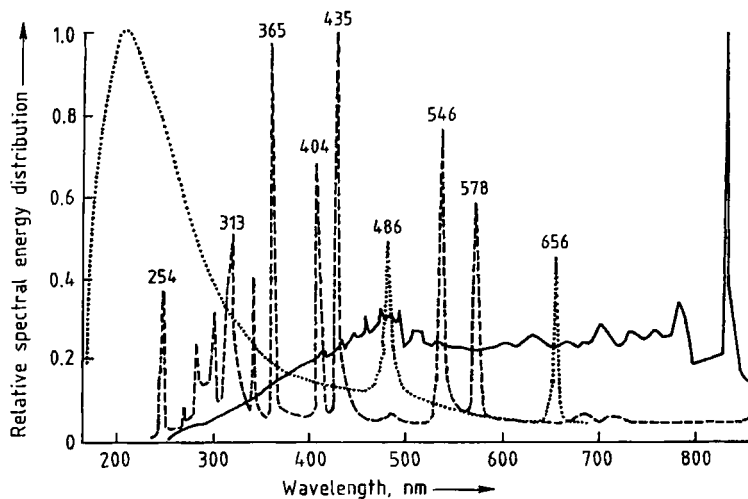
Figure 13. Line and continuous spectra from various light sources
——— = Xenon lamp; – – – = Mercury vapor arcs; · · · · · = Hydrogen lamp

Although modern computers calculate such spectra in fractions of a second, they can give high resolution only if the moving mirror is positioned exactly by using laser interference ($\Delta x \approx 1/\Delta\lambda$).

16.3.2. Light Sources

16.3.2.1. Line Sources

In filter photometers, the spectral bandwidth can be improved considerably if line sources are combined with filters. Mercury line sources [56] emit various wavelengths, as shown in Figure 13. Such equipment is used mainly in clinical analysis, because in general, single wavelengths are only used for photometry or in simple kinetic experiments. Mercury vapor arcs are also employed in polarimeters. By using these, rotations at ca. 10 wavelengths can be measured. In polarimetry, the use of sodium vapor arcs is also common. These emit the doublet 589.0 and 589.6 nm (sodium D line). In the past, only these wavelengths were used to determine optical rotation. The sodium lamp is rarely used today for routine measurements; however, the sodium D line is still indispensable for calibration and reference measurements.

16.3.2.2. Sources of Continuous Radiation

In spectrometers, a light source supplying continuous radiation must be available. Normally, a high-radiant power source that emits in the visible range (tungsten – halogen lamp) is combined with an ultraviolet emitter (hydrogen or deuterium lamp) to provide a wider spectral range.

Tungsten – Halogen Lamps. By addition of iodine to incandescent lamps a higher filament temperature can be reached, and consequently a higher radiation density at the shortwave end of the spectrum. The emission maximum of a tungsten – halogen lamp is in the green part of the spectrum. At wavelengths < 300 nm, the intensity decreases rapidly unless the lamps are operated at high current (high filament temperature), because the intensity maximum is shifted toward shorter wavelengths by an amount proportional to T^4 (Wien's law). However, the lifetime of these light sources is thereby decreased drastically [57], [58].

Hydrogen and Deuterium Lamps. Deuterium lamps, whose emission maximum is at ca. 200 nm, are used mainly in sequential spectrometers. The emission maximum of hydrogen lamps is at a longer wavelength. At > 400 nm, both types have rather low intensities, and show the familiar Fraunhofer lines at 486 and 656 nm whereby the lines of the two isotopes differ by fractions of a

nanometer [57], [59]. These lines can be used for wavelength calibration. The user must realize that equipment that is fitted only with deuterium lamps can be used only to a limited extent for measurement in the visible range.

This is a very important consideration, particularly for diode array spectrometers, since these, like sequential routine equipment, must often cover the range between 200 and 800 nm. Despite the higher costs and greater complexity of the high-voltage supply, *xenon lamps* are preferable because they cover the entire 200–800-nm range with sufficient intensity, as shown in Figure 13. However, in xenon lamps, a marked structure at 490 nm is present in the continuum so that balancing the two optical paths is difficult in this range. These light sources are also used in ORD and CD equipment.

16.3.2.3. Lasers

Lasers represent a special type of light source [16], [21], [60], [61]. They are used in trace analysis by fluorescence measurement or laser-induced fluorescence (LIF) (→ Laser Analytical Spectroscopy) [62]–[64], in high-resolution spectroscopy, and in polarimetry for the detection of very small amounts of materials. Lasers can be of the gas, solid, or dye type [21]. In dye lasers, solutions of dyes are pumped optially by another laser or a flash lamp and then show induced emission in some regions of their fluorescence bands. By tuning the resonator the decoupled dye laser line can be varied to a limited extent, so that what may be termed sequential laser spectrometers can be constructed [65]. In modern semiconductor lasers, pressure and temperature can also be used to "detune" the emission wavelength by 20–30 nm [66], [67].

16.3.3. Selection of Wavelengths

In simple equipment, only relatively wide transmission bands are selected by means of colored glass or interference filters. A working beam of narrow bandwidth is obtained by combining this with a mercury source. Interference filters give spectral bandwidths in the 3–10-nm range [42], [68]. These operate on the principle of multiple reflections between layers of dielectric material that have been produced in a definite sequence by vacuum deposition. The wave character of light leads to interference of the reflected beams at many interfaces. A suitable choice of thicknesses of the evaporated films enables a narrow band containing a small number of wavelengths to be selected.

16.3.3.1. Prism Monochromators

Prisms are relatively expensive and very difficult to manufacture in the highest optical quality. Furthermore, dispersion by diffraction is not linear, so that so-called cam disks must be used to give a wavelength scaling that is as linear as possible over the entire range. Prism monochromators are mechanically very complex and therefore very expensive. They were at one time used widely in the UV region. Today they are often used as premonochromators (predispersion devices) since the decomposition of white light into rainbow colors by a prism gives only one spectrum order [42].

16.3.3.2. Grating Monochromators

In the dispersion spectrum produced by gratings, several wavelengths of different orders appear at the same place. As shown in Figure 14, incident white light is reflected at a number of grooves that are either ruled mechanically on a plane surface or produced holographically by superposition of two laser rays. The reflected beams interfere in different ways, depending on their wavelength, the distance between lines of the grating, and the angle of reflection, so that constructive or destructive interference can occur. Thus, reinforcement at a given wavelength takes place only in a certain direction. However, several orders are produced (see Fig. 14) so that, e.g., the first-order reflection of 400 nm and the second-order reflection of 200 nm form an image at the same place. The greater the number of grooves, the more do the partial beams interfere, and the more significant is the destructive interference at certain wavelengths [11], [42].

The dispersion of the grating is almost linear. However, as already mentioned, individual orders overlap so that either an order filter must be fitted or a predispersion device (prism) must be used. If the grooves are blazed, reflected intensities of mainly one order can be collected. This is at a maximum for a particular angle (blaze angle) and a particular wavelength (blaze wavelength).

An inlet slit is usually located in front of the prism or grating symmetrical to the exit slit. If the inlet and exit slits are coupled and have the same width, the slit function (see below) forms a trian-
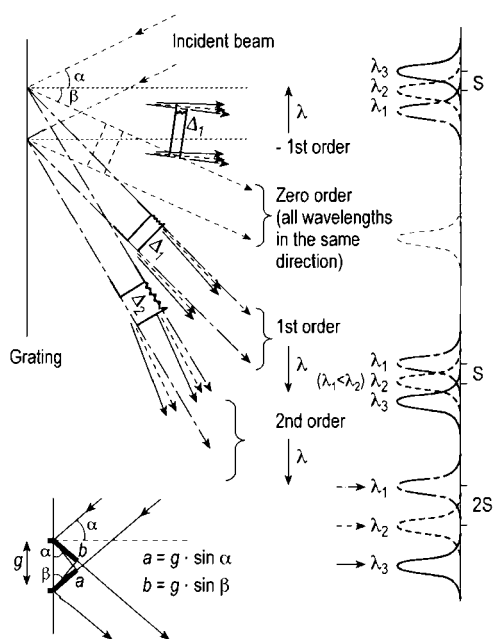
Figure 14. Principle of the grating
Beams reflected at the grating grooves overlap, giving constructive interference at an angle β for each wavelength except the zero-order case. The separation (resolution) of wavelength maxima S on a projection surface for the second order is twice that for the first order.

gular intensity distribution. In high-quality equipment, the first monochromator is followed by a second. These double monochromators have considerably better resolving power, although less light intensity is available since part of the light is cut off by the nonparallel beam through each slit. However, in a qualitatively high-grade double monochromator, a central slit must also be present, adjusted synchronously with the inlet and outlet slits; otherwise slit function is impaired [42], [69].

The slit function can be observed well in the imaging of line sources when the inlet and outlet slits (and the middle slit if present) are moved synchronously. If the slit is too wide the line of a mercury lamp does not appear as the expected Gaussian-shaped curve intensity distribution, but as a triangle. If the inlet and exit slits are different, a trapezoid is obtained. Triangular shapes in the spectrum indicate defective adjustment of slit widths in the equipment. These effects are observed mainly in spectrometers in which only a small number of preset slit widths can be selected. Distortion of a spectral band is negligible only if

the spectral bandwidth of the monochromator (slit width) is not more than one-tenth the natural bandwith of the sample [3], [42], [70].

16.3.4. Polarizers and Analyzers

The polarizer and analyzer separate electromagnetic radiation into an ordinary and extraordinary beam. The so-called Glan – Thompson prism is suitable for this. The path of the ordinary beam, whose mode is parallel to the plane of incidence (incident beam normal to prism edge), is refracted at the edge of the prism, while the extraordinary beam continues along the axis [11], [34]. Simple polarization filters (polarizing films) are also often used, in which two mutually perpendicular, plane-polarized beams are absorbed to different extents. Such polarizing films can be obtained by dyeing highly stretched (oriented) plastic film with dichroic dyes. Additional information (e.g., phase dependence) can be obtained by modulation. This technique is very important, especially in equipment for spectral measurements, particularly ellipsometry, because automatic balancing is possible. Modulation can be achieved by either rotational or oscillatory movement of the analyzer or polarizer, or by superposition with the inclusion of a Faraday modulator, which utilizes the Faraday effect described in Section 16.2.6.5 [11], [38]. Since spectral polarimeters (ORD equipment) are also used in the shortwave UV region, prism assemblies that include Canada balsam adhesive (which absorbs in this spectral range) must be excluded. Calcite can also cause problems because it absorbs at wavelengths below ca. 220–250 nm [38]. For these reasons, quartz Rochon prisms are used mainly today [34], [71].

16.3.5. Sample Compartments and Cells

16.3.5.1. Closed Compartments

Light from the monochromator is usually convergent and is focused on a minimum beam diameter at the position of the cell. The less convergent beam, the easier can measurements be performed in cells having long optical paths (10 cm) without the light grazing the sidewall of the cell. To prevent this grazing, the position of the cell in the path of the beam should be checked. In the visible region of the spectrum, a simple sheet of paper is suitable for this (monochromator set at ca. 550 nm and with a very wide slit) [3]. It should be placed
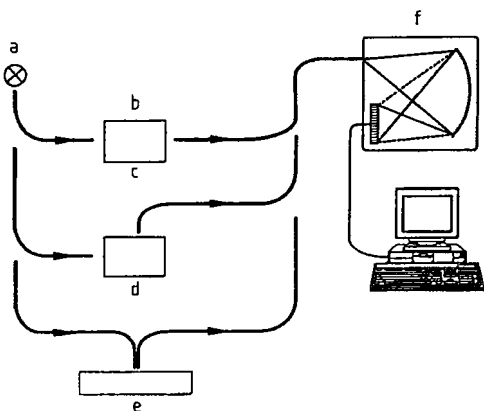
Figure 15. Combination of diode array and glass fibers in various arrangements for measuring absorption, fluorescence, or reflection

perpendicularly to the beam to monitor the green spot with respect to the cell position. Furthermore, the edges of the cell and the regions of adhesive at the edges should be inspected for any radiation. In the UV region, paper containing optical brighteners that emit blue light can be used.

Depending on spectral region, the correct material must be selected from appropriate catalogues, for example, because cells of glass or polymer are not transparent to UV radiation. The plastic cells often used in routine analysis differ in their optical properties. This holds true for cells from the same batch and from different batches [72]. Because of the manufacturing process used, their interior surfaces are not parallel, so the reproducibility and potential applications of these cells are limited.

In the sample compartment, as much space as possible should be available to allow for thermostatically controlled sample holders or other equipment (Ulbricht globes, praying mantis devices, multicell holders, sippers). A secondary cooling circuit is useful to protect the optics and electronics from high temperature. Large sample compartments require well-thought-out, carefully calculated optical paths. Photochemical reactions can be followed directly in a spectrometer by special accessories [45].

16.3.5.2. Modular Arrangements

In modular construction, monochromators, light sources, and detectors can be arranged in

various ways adjacent to the sample compartment, so that absorption, fluorescence, and reflection measurements are possible by using the same set of equipment components.

16.3.5.3. Open Compartments

Today, diode arrays combined with optical fiber arrangements are very well suited to this modular concept. The possible variety of such arrangements is illustrated by the combination of a diode array with optical fibers in Figure 15 [55]. The optical path between the light source (a) and the sample (b), and between the sample and the detection system (f), is realized by means of glass or quartz fibers, so that with one experimental arrangement, absorption (c), fluorescence (d), and reflection (e) measurements are possible by a variable arrangement of optical components. This type of equipment is very important for color measurements and becomes increasingly applied in optical sensor technology [55], [73]. In principle, the use of sequential recording equipment with designs based on optical fibers is also possible [74], [75].

16.3.6. Detectors

Detectors are of mainly two types, based on either the external or the internal photoelectric effect [59], [76]–[78].

Photomultipliers. In photomultipliers, the external photoelectric effect is utilized. Photons incident on a photocathode, which have an energy greater than the work function of cathode material expel electrons, which are captured by an anode that is positively charged relative to the cathode. The number of electrons produced is proportional to light intensity. They can be measured by the current flowing between electrodes.

Dynodes are often placed between the photocathode and the anode as intermediate electrodes. High-quality photomultipliers have 10–14 dynodes (see Fig. 16). This internal amplification is critical because in addition to electrons from the light-induced process, other thermally activated electrons give rise to a dark current. Photomultipliers are therefore often cooled for the measurement of low light intensities.

The head-on arrangement (Fig. 16 A) has a larger inlet angle and is extended lengthwise, with a correspondingly greater distance separating the

Figure 17. Linear photodiode array [3], [81]
The *p*-doped islands can be found in the *n*-silicon under a layer of SiO_2. In the region of these *p*-islands, photons can produce charges that discharge the capacitors arranged in parallel.

Figure 16. Types of photomultiplier
A) Head-on arrangement; B) Side-on arrangement (on a larger scale)
a) Dynodes; b) Anode; c) Focusing electrode; d) Focusing ring; e) Conductive internal coating; f) Front plate; g) Semitransparent photocathode; h) Envelope; i) Grill; j) Opaque photocathode; k) Shield

Figure 18. Charge-coupled device (CCD) [55]
A) Free charges are produced by incident light and localized in potential wells (countervoltage); B) Periodic application of different voltages causes charges to be displaced from the potential well like a shift register
a) Metallic electrode (gate); b) SiO_2 (insulator); c) *p*-Type silicon

dynodes (lower field strengths). It can accommodate more dynodes than the side-on arrangement, and therefore has a greater range of amplification. However, response times are longer. The side-on arrangement (Fig. 16 B) is considerably more compact, the distances between dynodes are shorter and the response times are relatively short. However, the "capture angle" is smaller. For detecting only a small number of photons, the more slowly reacting head-on photomultiplier is preferable [79].

Photodiodes, on the other hand, use the internal photoelectric effect [77]. When these diodes are arranged in arrays, they are known as diode arrays or diode matrices. Multiplex detectors are used in simultaneous spectrometers [80].

An example is the Reticon array shown in Figure 17. Also, CCD arrays (charge coupled devices), which were developed for television cameras and fax machines, have also become impor-

tant in spectroscopy (Fig. 18) [4], [49], [81] – [83]. Apart from the advantage of being multiplex devices, these charge-transfer components have characteristics more similar to those of photomultipliers than normal photodiodes.
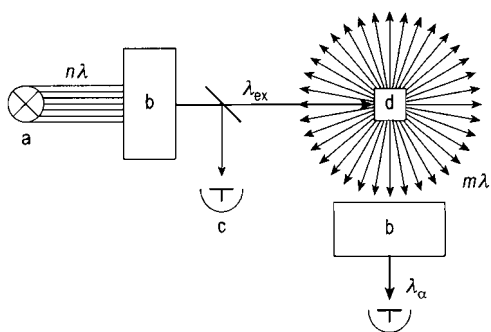
Figure 19. Diagram of optical light paths in a fluorescence experiment
a) Light source; b) Monochromator; c) Photomultiplier; d) Sample

16.3.7. Optical Paths for Special Measuring Requirements

16.3.7.1. Fluorescence Measurement

Unlike transmission spectroscopy, in fluorescence measurement the sample is observed perpendicularly to the incident excitation beam. In modern equipment, monochromators are placed both in the path of the excitation beam and the observation beam (see Fig. 19). Since no direct comparative measurement is possible in fluorescence, the intensity of the excitation light beam is usually controlled at least by a semitransparent mirror and a photodiode. To measure a *fluorescence spectrum*, the wavelength of the excitation beam is kept constant and the wavelength of the observation monochromator is varied, so that the excitation monochromator can be replaced by a filter in simple equipment.

For observation of *fluorescence excitation spectra*, the monochromator at the observation end is adjusted to a wavelength of highest possible fluorescence intensity, and the wavelength of the monochromator on the excitation side is varied. Hence, for these spectra, a shape analogous to an absorption spectrum is obtained [84]. If the absolute fluorescence spectra must be measured, the equipment must be calibrated for all wavelengths by using fluorescence standards [85]. This operation is highly complex and can be performed only with special equipment and appropriate software. Commercial equipment of this type is illustrated in Figure 20 [43].

16.3.7.2. Measuring Equipment for Polarimetry, ORD, and CD

Simple *polarimeters* consist of a polarizer, several slits, a light source (usually the sodium D line), and a detector, which in the simplest case is the eye of the observer. To adjust an optimal minimum of intensity passing the analyzer half-shaded apparatus are used. The measuring field is separated into two adjacent areas which are adjusted to equal brightness for an adequate analyzer position. Equipment with self-equalizing polarimeters is more complex. Self-equalizing polarimeters contain an additional Wollaston prism and a Faraday modulator, as shown in Figure 21. The polarizer (a) and analyzer (e) are crossed with respect to each other. Radiation from the polarizer is split in the Wollaston prism into two beams of plane-polarized light, whose planes of polarization are perpendicular to each other. The prism (c) is positioned such that the plane of polarization of the incident light forms an angle of 45° to the two beams passing the prism (without sample). The two split beams are of equal intensity under such conditions. Passage through Faraday modulator causes the direction of rotation of the two partial beams to oscillate, and they combine to give a constant radiation flux. This situation is disturbed when an optically active substance is placed in front of the Wollaston prism. An additional oscillating radiation flux falls on the photomultiplier. By use of a feedback system, the polarizer is rotated until this signal disappears.

In *spectral polarimeters*, another monochromator is introduced in front of the polarizer. In this type of ORD equipment, the Wollaston prism and polarizer are often combined by pivoting the polarizer to and from in a defined angle. CD equipment is often provided as an additional device for classical absorption spectrometry. With these devices, absorption of left and right circular-polarized light can be measured independently. For this, combinations of Rochon prisms (as polarizers) and Fresnel quartz rhombic prisms (for production of circular-polarized light) are used [34]. Equipment of this type is shown in Figure 22.

16.3.7.3. Reflection Measurement

Reflection can be either regular (at smooth surfaces) or diffuse (by material whose small particle size causes scattering). *Diffuse reflection* is used mainly in thin-film chromatography
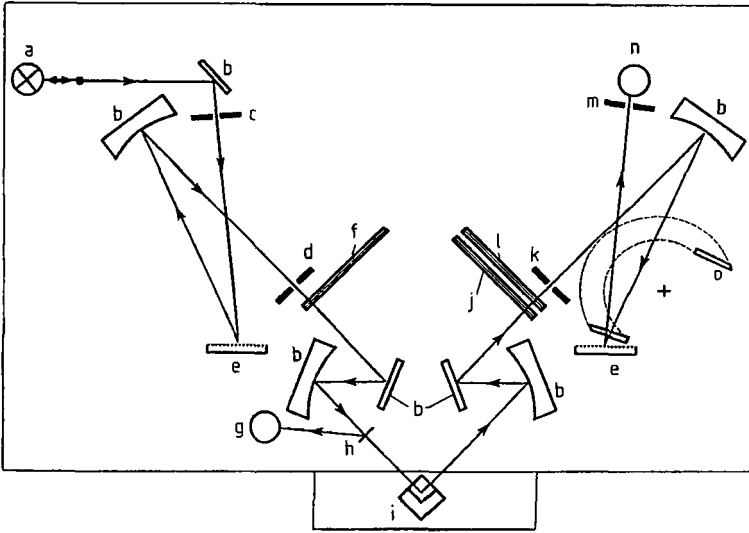
Figure 20. Light paths in commercial fluorimeter from Perkin – Elmer [43]
a) Light source; b) Mirror; c) Slit excitation entry; d) Slit excitation exit; e) Grating; f) Excitation filter wheel with excitation polarizer accessory; g) Reference photomultiplier; h) Beam splitter; i) Sample; j) Emission polarizer accessory; k) Slit emission entry; l) Emission filter wheel; m) Slit emission exit; n) Sample photomultiplier; o) Zero-order mirror accessory
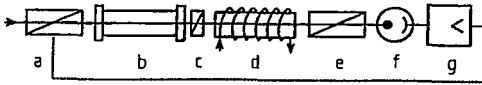
Figure 21. Schematic arrangement of self-balancing polarimeter [38]
a) Polarizer (rotatable); b) Sample cell; c) Wollaston prism;
d) Faraday modulator; e) Analyzer; f) Photomultiplier;
g) Amplifier
— = Feedback between detector and polarizer

[87] – [89], to evaluate the Kubelka – Munk function at an angle of $< 60°$ from the Lambert cosine squared law [4], [90]. Diode arrays can also be used as detector [91].

To observe *regular (specular) reflection*, goniometer equipment is used in combination with optical fibers and spectrometers. A modular system of this type is shown in Figure 15.

16.3.7.4. Ellipsometry

If measuring equipment for ORD and CD (Section 16.3.7.2) is combined with a reflection system, either ellipsometry at one wavelength or spectral ellipsometry can be carried out at defined angles (Fig. 23). In this apparatus, a rotating polarizer is used to produce rotating plane-polarized light. After reflection, an analyzer is used to meas-

ure the phase and attenuated relative intensity of the two beams (i.e., ellipticity). From the modulated signal, Hadamard matrices can be used to calculate the refractive index and absorption coefficient of the sample as a function of wavelength [93], [94], [95].

16.3.8. Effect of Equipment Parameters

In many spectrometers, some parameters (e.g., slit width, time constant, and amplification) cannot be chosen freely. Hence, their effects on measurement must be known. These settings depend on the type of spectrum being observed (e.g., a spectrum for routine analysis; a rapid scan spectrum, or high-resolution spectrum).

The following properties of the sample or the equipment can affect measurement [3], [5], [70], [96], [97]:

1) After being switched on, all equipment requires time to warm up until the electronics, light sources, and optical components are in thermal equilibrium.
2) The number of digits that can be read correctly from the display is affected by noise. The number of correct digits does not necessarily correspond with those displayed.

Figure 22. Light path of CD apparatus from Jasco company [86]
a) Mirror; b) Light source; c) Slit; d) Prism (vertical optical axis); e) Prism (horizontal optical axis); f) Ordinary ray;
g) Extraordinary ray; h) Lens; i) Filter; j) CD modulator; k) Sample compartment; l) Photomultiplier
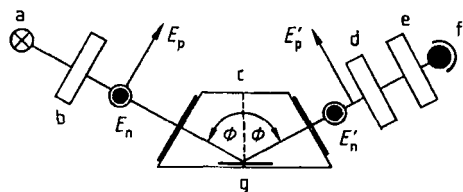
Figure 23. Design of MOSS ES4G ellipsometer from
SOPRA [92]
a) Light source; b) Rotating polarizer; c) Measuring cell;
d) Analzyer; e) Monochromator; f) Detector; g) Sample

3) The cell compartment as well as the cell holder should be temperature controllable to protect the electronics and optics from high measuring temperature.

4) To prevent any effects of fluctuation in voltage supply on the "bright value" (100% signal), such fluctuations should be controllable even under load.

5) Measurements in the spectral range < 190 nm can be performed accurately, only if the entire optical light path and sample compartment can be spilled by nitrogen.

6) Blaze wavelengths of ca. 320 nm ensure good measurements down to 200 nm even if the transmission of the optical components and sensitivity of the detector decrease in this range. However, this short blaze wavelength cannot usually be compensated for by higher detector sensitivity if wavelengths of 800–1000 nm are measured.

7) If cells with a long path length are used, they must be filled to the minimum level, and the geometry of the measuring beam must be correct.

8) If the intensity of the measuring beam in the UV is too high, it can lead to photochemical reactions in the sample.

9) In principle, the wavelength scan should be stopped for each measurement to prevent integration of the changing signal with the wavelength. Some equipment includes this facility, but the overall scan rate is reduced [45], [46].

Damping (time constant), amplification, spectral bandwidth, and scan rate have a considerable effect on the signal obtained and are interdependent. For these reasons, depending on the spectral characteristics of the sample and the purpose of the experiment (exploratory, routine, or high-resolution spectrum), parameters must be optimized relative to each other. Some criteria are given below. Other information on standards and calibration methods can be found in the literature [98], [99].

Reproducibility and Accuracy of Wavelengths. In both grating and prism equipment, mechanical wandering of the wavelength scale can occur over time. Therefore, the accuracy of the wavelength must be monitored, and adjusted if necessary, using holmium perchlorate solution, holmium chloride solution, or filters, or by using the lines of a mercury high pressure arc. The simplest method is to use the lines of the hydrogen or deuterium lamp that is part of the equipment.

Wavelength reproducibility can be simply checked by an overlay of repetitively recorded line emitter or holmium perchlorate solution spectra. Since wavelength reproducibility is especially important in kinetic investigations, a chemical reaction can also be used as standard, for which the
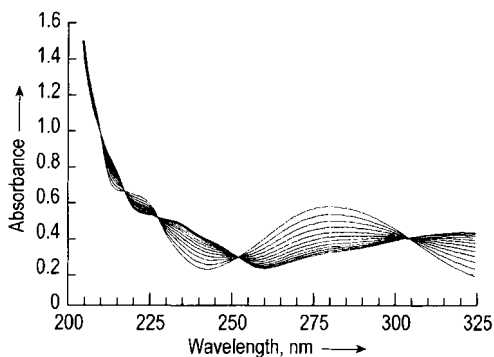
Figure 24. Sulfone reaction spectrum for characterization of wavelength reproducibility of a spectrometer

Figure 26. Cutoff filter properties for various totally absorbing salt solutions
a) NaI (cutoff at 260 nm); b) NaBr (cutoff at 220 nm); c) KCl (cutoff at 200 nm)

Figure 25. Effect of slit width on spectrum for various ratios of spectral slit width to natural bandwidth
a) 1 : 20 (1 nm); b) 1 : 4 (5 nm); c) 1 : 2 (10 nm); d) 1 : 1 (20 nm)

solution of potassium dichromate in 0.005 mol/L sulfuric acid [100], [101]. The Lambert – Beer law requires a linear relationship between the signal and concentration. Deviations can be due to several reasons provided chemical effects can be excluded: (1) the recording rate or time constant may be too high; (2) the spectral bandwidth chosen may be too large compared to the natural bandwidth of the sample; (3) too much stray light may be present. As shown in Figure 25, if the spectral bandwidth used is too large, this leads not only to a reduction in the peak height, but also to a displacement of the maximum. A good rule of thumb is that the spectral bandwidth of the equipment should be only one-tenth the natural bandwidth of the substance [3]; otherwise, peak height will be $> 0.5\%$ less than the true value. Therefore, equipment with fixed slit widths is of limited use.

absorbance at certain wavelengths does not change during reaction (isosbestic points). An example is shown in Figure 24, in which repeated measurements were carried out on a thermal reaction of 2-hydroxy-5-nitrobenzylsulfonic acid [3].

Photometric Reproducibility and Accuracy. Standards for absolute determination of photometric accuracy are available from, for example, national standards offices (Physikalisch-Technische Bundesanstalt, PTB; NBS). However, the accuracy of these standards (both filters and standard solutions) is not greater than ca. 0.3%, and even that can be achieved only by expert operators.

A simpler determination of photometric reproducibility is often carried out with the aid of a

Stray Light. In addition to radiation of the desired wavelength, every monochromator passes a proportion of so-called stray light (light of different wavelength). If radiation at the extremes of the wavelength range (< 250 nm or > 500 nm) is strongly attenuated owing to high absorption properties of the sample, even stray light of low intensity can lead to severe signal distortion because it usually has a wide bandwidth, is scarcely attenuated by the sample, and strikes the detector, which is not wavelength selective. This effect is most noticeable at absorbances > 2. The transmission of the sample appears higher, a false value of absorbance is recorded, and the absorption maxima are reduced. As shown in Figure 26, totally absorbing salt solutions with cutoff filter properties can be used to reveal stray light [3], [102].
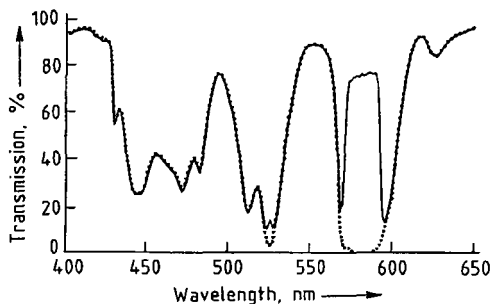
Figure 27. Determination of amount of stray light by a difference measurement of colored glass of various thicknesses
····· = Colored glass 1 mm thick with respect to silica glass 1 mm thick (without chromophore); ——— = Colored glass 4 mm thick measured against colored glass 3 mm thick

Figure 28. Benzene vapor spectrum recorded with the double monochromator equipment Specord M-500 (C. Zeiss) at various settings of spectral slit widths
··· = > 2.0 mm; – – – = 0.5 mm; ——— = 0.2 mm

Another possibility, which is successful mainly in the visible region, is provided by comparative measurements of colored glasses of different thicknesses. As shown in Figure 27, the bands have a tendency to even out in the high-absorption region. In an extreme case, this can be manifest by an intermediate minimum in the resulting spectrum which can lead to misleading interpretations [3], [103].

In principle, the following rules of thumb can be given for recording various types of spectra:

1) For exploratory spectra (recorded at high speed), high amplification and small time constants are used, with a large spectral slit width.
2) For routine spectra, a medium-sized spectral slit width is used, depending on the natural bandwidth, and the scan rate selected should not be too high. A medium time constant with a medium degree of amplification avoids noisy spectra.
3) For high-resolution spectra, a very low recording rate with large time constants, high amplification, and an extremely small slit width is used.

The difference between these three types of spectra is illustrated in Figure 28, which shows results obtained on benzene vapor under various conditions by using a spectrometer capable of high resolution.

Cleaning the Optical Cell. The type of material used for the cell and the cleaning method are very important for measurement quality. Standard cleaning procedures recommended by cell manu-

facturer should be used [104]. Under no circumstances should finger marks be left on the cell because this seriously impairs the refractive properties of the cell walls. Alteration of the optical path and the refractive properties of a cell with scratched walls are revealed by a change in the signal of an empty or solution-filled cell when the cell is rotated through 180°.

Since plastic cells generally do not have two plane-parallel walls, fluctuations can occur even in different cells from the same batch [72]. Cells with polytetrafluoroethane stoppers should be used for quantitative measurements because, with other types of cells, the organic solvent can creep over the edges or evaporate.

16.3.9. Connection to Electronic Systems and Computers

Modern spectrometers have digital readouts and integrated microprocessors for purposes of calibration, control of the wavelength drive, slit adjustment, and sometimes for the interchange of the light sources. They also include interfaces to personal computers sometimes called data stations. The trend is now toward spectrometers with a limited amount of built-in intelligence, which are controlled by personal computers that can also readout data. For this purpose, software for control, data collection, and evaluation is available, operating under commonly used operating systems. In this way, available additional development costs are reduced. Built-in analog-to-digital converters should be provided with preampli-

fiers to ensure that a wide dynamic range is available.

In the UV – VIS range, personal computers are used in data evaluation, especially in multicomponent analysis. When connected to detection systems in chromatography, they are also used for peak detection, peak integration, and eluent optimization [105]. Although in UV – VIS spectrometry, libraries of spectra do not play an important role, standard spectra are used in chromatography for identification [106].

In larger laboratories, personal computers are linked via a local area network (LAN) to a laboratory information and management system (LIMS) [107].

16.4. Uses of UV – VIS Spectroscopy in Absorption, Fluorescence, and Reflection

16.4.1. Identification of Substances and Determination of Structures

Although the main strength of UV – VIS spectroscopy does not lie in the identification of substances, it can often be used as a quick, inexpensive method for identifying certain classes of materials [108] – [111]. This is especially true in the area of pharmaceutical chemistry, where many otherwise difficult-to-identify substances can be identified with certainty, especially at different pH values and via color reactions. Other examples are aromatic carboxylic acids [111], where the position of the absorption maxima is characteristic and depends on the substituents, for example, with benzadiazepines [112]; and also pregnenones and androstenones [108], [109].

Since steroids contain α, β-unsaturated keto groups absorption maxima can even be predicted by increment rules [108]. In this case, identification is easier by UV – VIS spectroscopy than by the more sophisticated NMR method. The position of methyl or hydrogen substituents relative to the keto groups in the ring enables the wavelength of absorption peaks to be calculated. The spectra of antidiabetics such as carbutamide, tolbutamide, etc. [111], can be calculated and classified readily, although they sometimes differ only in the *para*-substituents on the aromatic sulfonamide ring. Since carbutamide contains an amino group, a

significant pH effect occurs that is not observed with the other compounds.

16.4.2. Quantitative Analysis

16.4.2.1. Determination of Concentration by Calibration Curves (→ Chemometrics)

Molar decadic absorption coefficients can be determined from the slope of the absorbance-versus-concentration curve. This slope is known as the sensitivity. Any intersections with axes or nonlinearities can be caused either by impurities inside or outside the cell, by limited photometric linearity of the instrument, or by deviations from the Lambert – Beer law. Influences on the spectrum of a pure substance are shown in Figure 29 (A – C). These changes are caused by the presence of an impurity, fluorescence of the absorbing substance, or scattering. Three measuring wavelengths (λ_1, λ_2, λ_3) are also indicated in the figure. *Scattering* is apparent at all three measuring wavelengths, although to different extents. The *impurity* affects only the spectrum of the pure substance at wavelengths λ_2 and λ_3, and fluorescence appears only at wavelength λ_3. By drawing extinction – concentration diagrams for the three cases (see Fig. 30), instead of the "true" absorbance marked by the symbol dot in circle, a false signal is used. Calibration at one wavelength can lead to incorrect determination. The measured absorbance of a sample can lead to various apparent concentrations, depending on what is causing the false reading.

To ensure correctness of calibration, this operation must always be carried out at several wavelengths that are carefully selected from the spectrum. A simple graphical method of detecting errors in calibration and measurement, and also of recognizing their causes, is given in Figure 31 (A – C). Absorbances are compared for various wavelength combinations by using the same sample. These show a pattern that depends on the type of influence, thereby enabling various types of disturbance to be distinguished. The diagram obtained (see Fig. 31) is known as an extinction diagram [3], [30]. If the influence is due not to the presence of an additional substance in solution, but to an impurity on the cell wall (e.g., finger marks or cells that are not equally clean), this appears in the calibration diagram as an intersection with an axis (i.e., a parallel shift of the cal-
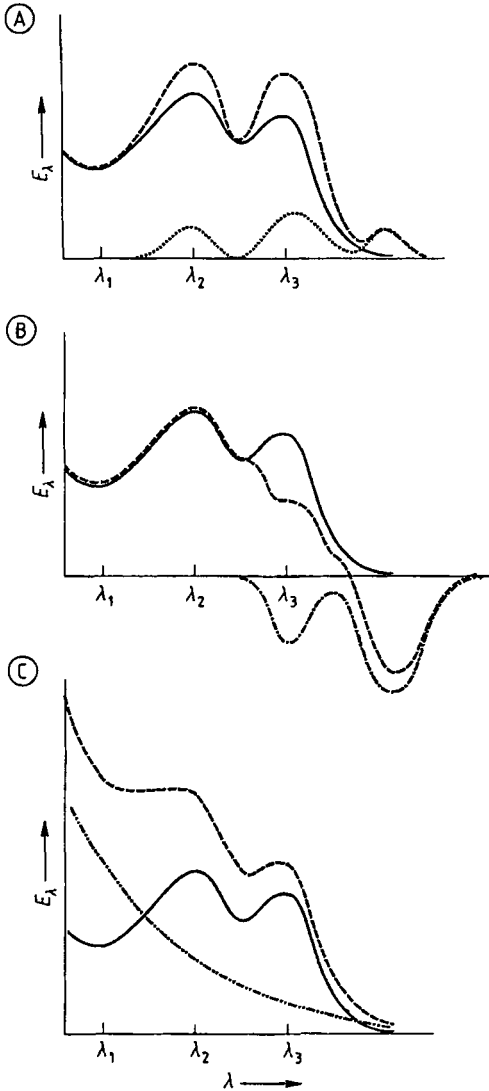
Figure 29. Diagram of absorption spectrum of a pure substance (—)
A) Superimposed on spectrum of an impurity (· · ·); B) With a fluorescent component (— · — · — · —); C) With a scattering substance (— — — — — — — — —)
— — — = Measured spectrum

Figure 30. Absorbance plotted as a function of concentration to give a calibration line
Various signals are observed for the same concentration. Same absorbance value corresponds to different concentrations.
a) Substance + impurity ; b) Substance + scattering ; c) Pure substance ; d) Substance + fluorescence ; e) Calibration curve if, e.g., finger marks occur on outside faces of cell
○○○ = Measured points

ibration line with offset), as shown by curve e in Figure 30 [3].

16.4.2.2. Classical Multicomponent Analysis

Calibration at several wavelengths also enables multicomponent analysis to be carried out in mixtures of substances by using the spectra of pure substances or defined mixtures for calibration (see Fig. 32) [113]. In this case, the Lambert – Beer law for a pure substance is modified, giving the relationship [3]

$$E_\lambda = d \sum_i \varepsilon_{j,i} c_i \tag{13}$$

Thus, for three components A, B, C, a linear equation system is obtained, which could take the form

$$E_{500} = d(\varepsilon_{500,A}\, a + \varepsilon_{500,B}\, b + \varepsilon_{500,C}\, c)$$
$$E_{400} = d(\varepsilon_{400,A}\, a + \varepsilon_{400,B}\, b + \varepsilon_{400,C}\, c)$$
$$E_{300} = d(\varepsilon_{300,A}\, a + \varepsilon_{300,B}\, b + \varepsilon_{300,C}\, c)$$

$$E = \varepsilon c\, d \tag{14}$$

if three wavelengths are being measured. In classical multicomponent analysis, this linear equation system is solved by matrix operations and normal equations. However, in this so-called K-matrix method [6], [114], errors that arise in calibration and measurement must be taken into account. These are represented by the error vector e, so that the following matrix equation is obtained

$$E = \varepsilon c\, d + e \tag{15}$$

The number of components that can be determined independently in such multicomponent analysis is normally limited to three or four if individual
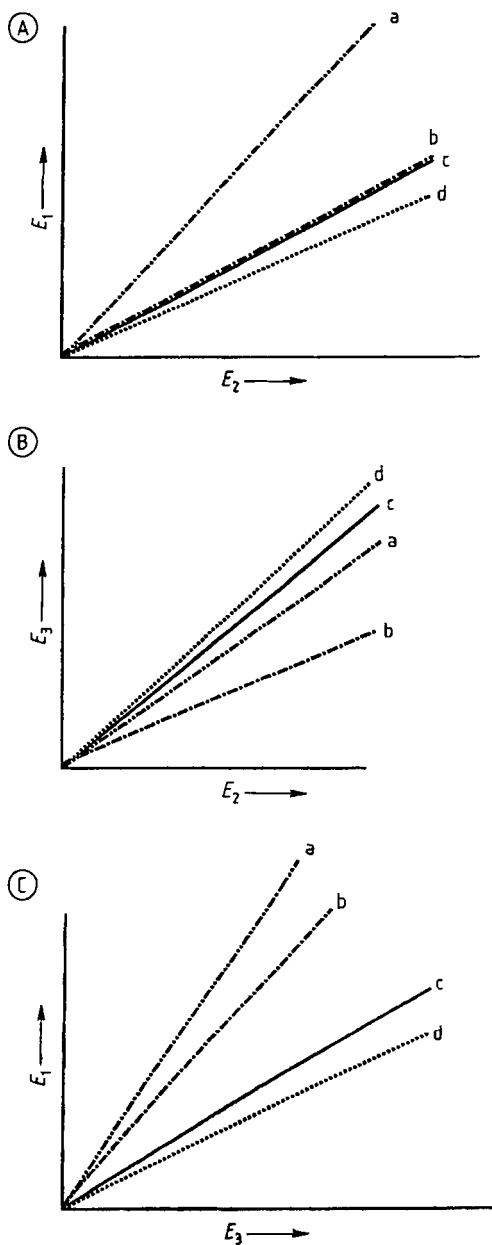
Figure 31. Extinction diagram in various combinations for detection of errors in calibration measurement
a) Substance + scattering; b) Substance + fluorescence; c) Pure substance; d) Substance + impurity

components are not situated favorably in the spectrum with respect to their wavelengths and bandwidths. The ε-matrix is obtained from the calibration measurements. Inversion of this matrix often leads to numerical problems [3], [6], [115].

16.4.2.3. Multivariate Data Analysis

To take account of interactions between individual components (association, nonlinearities), calibration using multivariate data analysis is often also carried out with mixtures rather than pure substances. Despite this fact, limitations to this method of assessment are encountered quickly. Therefore, the so-called inverse method using the Q-matrix is employed, and either principal component regression (PCR) or the partial least squares (PLS) method is used [6], [114], [116]. In both methods, calibration is carried out not with pure substances, but with various mixtures, which must cover the expected concentration range of all components. Within limits, this can allow for nonlinearities:

$$c = Qb \qquad (16)$$

Principal component analysis first gives the number of significant components, which avoids a typical error of classical multicomponent analysis, which distributes an unconsidered component among the others. However, principal component analysis reveals it as an additional "impurity." Then a large number of regression steps is carried out with respect to each of the mixtures. In PCR, all possible combinations of factors must be considered since the algorithm does not produce these factors in systematic order whichis the intended aim (quantitative multiple component determination with minimal error). By using the Lanczos bidiagonalization method (PLS), this validation is simplified considerably, since only combinations of the first sequence of factors need be considered [113], [117].

Another method is the so-called Kalman filter [118], [119]. Iterative calculations enable the spectrum of an unknown component to be determined using a sufficiently large number of calibrating wavelengths. Methods of multicomponent analysis by fuzzy logic have not proved very successful [106]. In contrast, a method using Fourier transforms followed by multicomponent analysis is available commercially [120]. The most recent methods, using neuronal networks [6], [121], are used much less in classical multicomponent analysis [122] than in sensor arrays with relatively low selectivity, where they are employed to determine relative concentrations in gas mixtures or liquids [123], [124]. Another wide field of application, in the infrared, is in foodstuffs analysis and pharmaceutical chemistry [125].
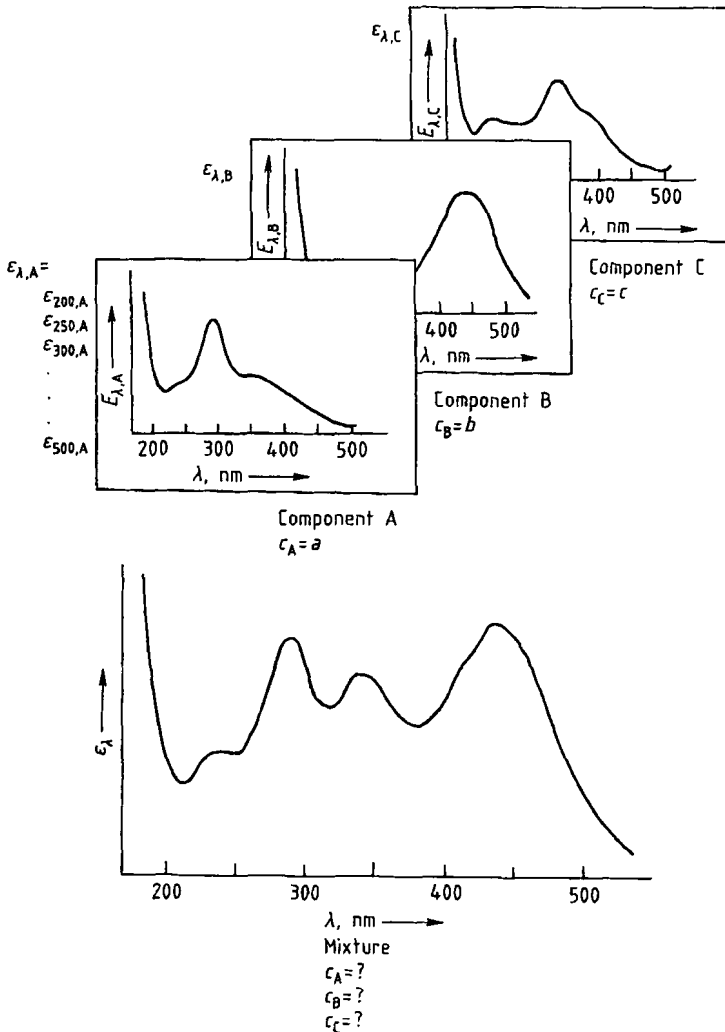
Figure 32. Diagram of pure spectra of three components A, B, and C for calibration, and spectrum of an unknown mixture

16.4.2.4. Use in Chromatography

In principle, multicomponent analysis can also be used in chromatography if peaks cannot be separated completely. By using a diode array while recording a chromatogram, components hidden within a peak can be detected by factor analysis if they are spectrally distinguishable [126]. If their calibration spectra are known, their relative concentrations can be determined.

Every multicomponent analysis should be assessed critically by the analyst, and different algorithms should be used on the same data to improve the level of significance of the evaluation.

16.4.3. Fluorimetry

As explained in Section 16.2.4, fluorescence takes place in competition with radiationless deactivation. Fluorescent emission is normally spherically isotropic and of longer wavelength than the exciting radiation. It is extinguished by heavy atoms in the molecule, by the excitation of free rotation about chemical bonds and torsional vibrations, by high concentrations because of increased collision probabilities, or by the presence of paramagnetic oxygen in the sample solution. This can affect reproducibility in quantitative measurements [28], [84], [127].

The observed fluorescence intensity I_α^F depends not only on the absorbing power of the molecule at the excitation wavelength $\varepsilon(\lambda)$, but also on the observed wavelength α within the fluorescence spectrum, depending on its spectral distribution f_α. Measuring fluorescence does not involve any reference measurement, so any fluctuations in the lamps or the electronics, especially long-term drifts, distort the measured signal. An advantage of fluorescence measurement is that by increasing the intensity of the excitation light source, trace analysis can be carried out even with samples that fluoresce only slightly [28], [84], [127]. However, high intensities of the excitation light source can also cause photochemical decomposition reactions in the sample. Even laser dyes have such poor photo stability that two fluorescence spectra of these dyes observed consecutively can give different results [128].

16.4.3.1. Inner Filter Effects

At high analyte concentrations, inner filter effects can occur. These lower fluorescence intensity by self-absorption if the observed wavelength is incorrectly chosen (within the absorption band). To prevent this quenching due to high concentrations, measurements are taken at the lowest possible concentrations (usually $< 10^{-7}$ mol/L). A rule of thumb is that absorbance at the excitation wavelength should be < 0.02 to prevent errors $> 2.5\,\%$. Under these conditions, fluorescence intensity I_α^F is proportional to the concentration c of the fluorophor [3]:

$$I_\alpha^F = \Gamma I_0(\lambda)\eta f_\alpha c\varepsilon(\lambda)\frac{1-10^{-E(\lambda)}}{E(\lambda)} \tag{17}$$

Here, η is the fluorescence quantum yield, $E(\lambda)$ the absorbance, and Γ a geometric factor. In practice (cuboid-shaped cell), this factor corrects for non-isotropic intensity distribution. The factor $(1-10^{-E(\lambda)})/E(\lambda)$ describes the finite absorption and becomes E^{-1}, if $E \ll 1$ [3], [28]. A linear relationship exists between observed fluorescence intensity and concentration [3], [28].

16.4.3.2. Fluorescene and Scattering

If fluorescence is measured in the direction of excitation radiation and analyte concentration is low, much unattenuated excitation radiation falls on the detector. Therefore, measurements are usually made at 90° to the path of the excitation beam. Despite this, the fluorescence signal is usually mixed with scattered light, especially if the solution is cloudy and the substances show only weak fluorescence.

Particularly in trace analysis, artifacts are recorded that occur because of the so-called Raman bands of the solvent. These result in incoherent scattering by the solvent at high excitation intensities ($\rightarrow$ Infrared and Raman Spectroscopy). To avoid this, in good-quality equipment, excitation radiation is produced with the aid of a monochromator rather than a filter, so that these additional Raman lines can be cut off by the second (observation) monochromator. In fluorescence measurements, the spectrum of pure solvent should always be recorded as a "blank" before the sample is investigated.

16.4.3.3. Excitation Spectra

Whereas absorption spectra describe the relative position of the vibrational level of the first excited state, fluorescence spectra give the position of the vibrational level of the ground state.

If the observation monochromator is kept at the wavelength of the fluorescence maximum and the wavelength of the excitation monochromator is varied, a so-called excitation spectrum is obtained that resembles the absorption spectrum in its spectral distribution. If a mixture of substances contains only one component capable of fluorescence, a fluorescence excitation spectrum can be used to obtain spectral information about this specific component [28], [33].

16.4.3.4. Applications

Fluorometry is used to determine elements such as boron, silicon, aluminum, beryllium, and zirconium [7], [129], [130], as well as organic compounds [131] (e.g., vitamins [132]). An important application in the field of trace analysis is the determination of aromatic hydrocarbons in wastewater [133]. Here, the limits of detection can be improved considerably [64] if fluorescence is excited by a laser (LIF). This is very important for the measurement of polycyclic aromatic hydrocarbons [134].

Other applications, especially those involving the use of sensors, have been reviewed recently [135], [136].

16.4.4. Reflectometry

16.4.4.1. Diffuse Reflection

Depending on the surface of the substrate, incident electromagnetic radiation can be reflected either diffusely or regularly (see Fig. 5). If the surface is rough or consists of many small crystallites, diffuse reflection is observed. For a pulverized sample with a very large layer thickness (so that no appreciable reflection of substrate material on which the powder is deposited is observable), the reflectance R_∞ can be related to both the scattering coefficient s and the absorption coefficient k. The reflection behavior of a powdered sample can be described by the so-called Kubelka–Munk function [90], [137]:

$$F(R_\infty) = \frac{k}{s} = \frac{(1 - R_\infty)^2}{2 R_\infty} \tag{18}$$

Diffuse reflection has a wide range of uses, e.g., in research, where it is used in the observation of molecules adsorbed on a surface [138] (especially in catalysis), and also in analysis (e.g., thin-film chromatography [89], [139]). To "capture" as much of the diffusely reflected radiation as possible, an Ulbricht sphere is often used. This type of arrangement, in conjunction with optical fibers, is currently used in production lines in the automobile industry for quality control of bodywork paint [140].

16.4.4.2. Color Measurement

An important use of reflection spectroscopy is in color measurement [141]. Visual comparisons are replaced by methods based on spectral quantities and the C.I.E. standard colorimetric system (DIN 5033). These techniques are very important in the paint industry [142]. Subtractive and additive mixing of several components can be represented in the color space by means of vectors. Proportions of standard color values based on standard spectral curves, together with the physiological color stimulus specification, give a color stimulus corresponding to the color equation [143]. This method of measuring paint colors is very important, especially in the automobile manufacturing industry.

16.4.4.3. Regular Reflection

Depending on the change in refractive index at an interface, internal and external reflection can occur. In *external reflection* (transition to a medium of higher refractive index, i.e., into an optically denser medium), reflection occurs at all angles of incidence. The reflection can be calculated from Fresnel's laws (see Eq. 3).

In *internal reflection*, at angles of incidence larger than the critical angle, electromagnetic radiation is totally reflected (attenuated total reflectance, ATR, see Section 16.2.2.4 and Fig. 5). This special case is very important in analysis for two approaches. First, simple transportation of radiation within the fiber (or a waveguide). Second, in total reflection, an evanescent field appears in which the electrical field vector decays exponentially in the optically less dense medium. Every change within the medium with lower refractive index influences the field vector coupled to the field in the optically denser medium. Therefore, the totally reflected radiation contains information about effects on the other side of the phase boundary (the medium with lower refractive index) [20], [144]. Various principles to interrogate this effect are known and used in evanescent field sensors.

If ATR experiments are combined with dichroitic measurements, additional information on orientation and lateral spacing can be obtained. This is especially important if molecules are oriented perpendicularly or parallel at interfaces owing to self-organizing effects [145].

Both approaches of internal reflection (transportation of radiation, application of evanescent fields) are discussed in the following.

16.4.4.4. Determination of Film Thickness

If reflection occurs at not just one interface but at least two adjacent boundaries, the regularly reflected beams can superimpose and interfere. Interference occurs in certain preferred directions, at certain wavelengths, and with certain substrate and superstrate materials (see Fig. 33 A). If the partial beams extinguish each other, it is called destructive interference; if they reinforce each other, constructive interference [11]. If white light is used, well-resolved interference spectra are obtained with film thicknesses of $1-5$ μm [146]. Bifurcated fibers are used for incident radiation and its collection after reflection at the phase boundaries.
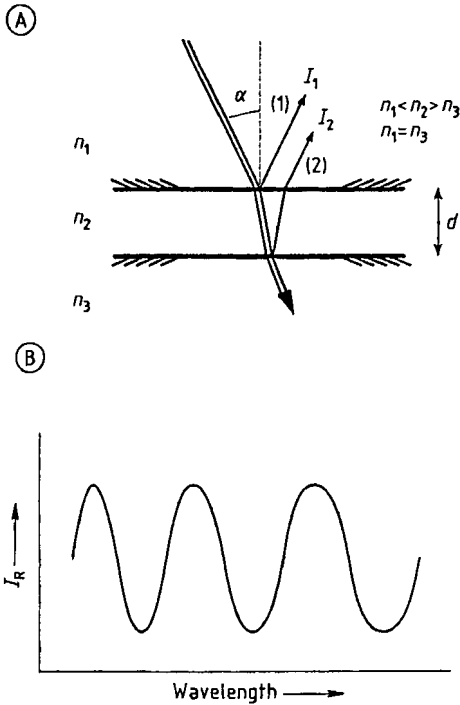
Figure 33. A) Reflected beams at a multilayer system (upper diagram) in which two beams (1) and (2) interfere constructively or destructively; B) Resulting interference spectrum (lower diagram)

An interference pattern of this type is shown in Figure 33 B; the modulation is given by a $\cos(1/\lambda)$ function. Reflection at the phase boundaries leads to reflected intensities I_1 and I_2 which are of the order of a few percent of the total intensity. According to wave theory, these intensities superimpose as follows [11]:

$$I_R = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos\left(2\pi \frac{2 n_2 d}{\lambda} + \delta\right) \qquad (19)$$

where δ corrects for a phase shift on reflection at the optically denser medium; n_2 is the refractive index of the film; and d is the thickness of the film at whose phase boundaries the reflections take place. According to Equation (19), maxima in the interference spectrum occur where $m = 0, 1, 2, \ldots$

$$2\pi \frac{2 n_2 d}{\lambda} = 2\pi m \qquad (20)$$

Whereas absolute film thicknesses can be determined only by complete curve fitting of the many

distances between maxima of the calculated dispersion curve [124], relative film thickness changes can be determined simply and accurately. According to [11] and [147], from the equation ($\Delta m = 1$)

$$2 n_2 d = \frac{\lambda_1 \lambda_2}{\lambda_1 - \lambda_2} \qquad (21)$$

the optical film thicknesses can be determined from the spacing of the extrema at the two wavelengths. Minimal changes in refractive index lead to a shift of extrema in the interference pattern.

Improvement of the signal-to-noise ratio and refinement of the evaluation algorithm [148] enable film thickness changes of ca. 1 pm to be observed. The method is dynamic and has response times of < 1 s if diode arrays are used to obtain interference spectra [149]. The method is called reflectometric interference spectroscopy (RIfS) [147]. Similar procedures are also followed in the IR region [116].

16.4.4.5. Ellipsometry

The method described in Section 16.3.7.4 for performing ellipsometric measurements using polarized light (see Fig. 23) [150] is very important in microchip production [151]. This method is also used in the characterization of alloys and the investigation of biological films or membranes [152].

The disadvantage of ellipsometry is undoubtedly that it does not represent a direct analytical method, but attempts to match functional relationships as well as possible to the experimental data by the development of models [93]. On the other hand, ellipsometry is one of the few nondestructive analytical methods that allows measurement of the monolayer or submonolayer. However, even when a "microspot" of measuring light is used, it represents several square millimeters in area. In this surface area, the product nd (optical film thickness) is averaged laterally.

Compared to analyses carried out with monochromatic light at a fixed angle, the result is improved by measurement either as a function of angle or as a function of wavelength at a fixed angle [94]. Samples with weak absorbing properties cause problems [153]. Complex refractive indices must then be used for calculations.

However, in contrast to simple reflection measurements or by surface plasmon spectroscopy at one angle or wavelength [153], spectral meas-
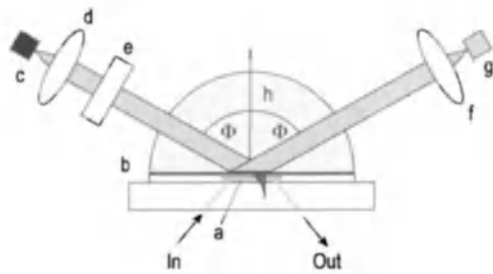
Figure 34. Diagram of system for measuring spectral surface plasmon resonance with a flow cell and resonance point
a) Flow cell; b) Silver layer; c) Light source; d) Collimator lens; e) Polarizer; f) Lens; g) Optical fiber (to the diode array); h) Glass cylinder

Figure 35. Principle of grating coupler [160]
a) Grating; b) Flow cell; c) Carrier; d) Waveguide

urements with ellipsometry enable n and d to be determined independently.

16.4.5. Resonance Methods

16.4.5.1. Surface Plasmon Resonance

Surface plasmon resonance (SPR) is a typical method to interrogate evanescent field effects. The measuring system consists of a thin metal film produced by vacuum deposition onto a glass prism (or half cylinder). Depending on wavelength, angle of incidence, and refractive index at the phase boundary, resonance conditions are produced if electromagnetic radiation leaving the prism excites plasmons in the metal to coupled vibrations [153]–[156].

Generally, monochromatic radiation is used and the angle of incidence is varied. Resonance leads to attenuation of the reflected beam because some of its energy is used for plasmon excitation. The second possibility, spectral excitation by white light at a preset angle, is also of interest. Here, the resonance wavelength is detected by a diode array [94]. An apparatus of this type, with a flow cell in the path of the signal, is shown in Figure 34. Gold, silver, and occasionally copper are used to form the films.

SPR is used to characterize biological membranes and Langmuir–Blodget films, as well as a sensor in bioanalysis [157], [158]. SPR application to sensors is reviewed in [159].

16.4.5.2. Grating Couplers

Another resonance technique involves the use of grating couplers, introduced by Lukosz [160],

[161]. These can be obtained from the ASI company in Zürich. Grating patterns are imprinted on thin films of highly refractive material that is supported by a carrier material. Depending on the refractive index ratios at the grating, a resonance condition is produced for the film waveguide under the grating, leading to coupling or decoupling of the interfering beams at the grooves of the grating. The principle is shown in Figure 35. Both modes of electromagnetic radiation (transversal electric, TE, and transversal magnetic, TM) can be used, so that a difference is produced [162]. A large number of uses have been published in the field of gas sensors and for rapid analysis with biosensors [163].

16.4.6. On-Line Process Control

16.4.6.1. Process Analysis

By use of flow cells, normal spectrometers can be employed for routine measurement. Since recording a spectrum by sequential spectrometers required a certain amount of time, the absorption was usually observed continuously at one wavelength. The development of diode arrays has enabled this system of measurement using photometers to be superseded. Also, these new types of measuring equipment use optical fibers, enabling measuring cells to be installed at a considerable distance from the spectrometer, so that continuous measurements are possible in dangerous or explosion-protected spaces. Equipment designs such as those shown in Figure 15 are very suitable for such tasks, enabling absorption, reflection, and fluorescence measurements to be carried out with the aid of glass fibers.

By using optical fibers, biotechnological processes in fermentors can be monitored continuously with spectral measurements. In simple cases, transmission measurements are performed [164]; or for monitoring reactions, scattering phenomena are employed. These techniques are also used in

wastewater analysis by continuous measurement of turbidity. If multicomponent systems are to be observed (e.g., in the analysis of foods), methods using the near IR are often more suitable [125].

Many analytical problems can be solved by the use of flow cells. If these are combined with flow injection analysis equipment [165], the limits of detection in water analysis can be improved [166] or the scope of biochemical analysis increased [147].

16.4.6.2. Measurement of Film Thicknesses

When beams of radiation reflected from the two surfaces of a film superimpose, constructive or destructive interference occurs when visible light is used if the film thickness is in the region of a few micrometers, since one partial beam has a longer path length to travel than the other. The resulting phase difference depends on the physical distance between film interfaces and the refractive index of the film (see also Fig. 33). From the interference pattern, the so-called optical film thickness $n\,d$ can be calculated. Use of the diode arrays previously described, in combination with optical fibers, enables rapid measurements of optical film thicknesses to be performed even on films produced at high speed in coating machines. The possibility of monitoring coating thickness at various parts of the strip by means of glass fibers, and of reacting to production problems within a few seconds, makes this technique extremely useful in process control.

16.4.6.3. Optical Sensors

For a general definition of sensors, see → Chemical and Biochemical Sensors. The term sensor denotes a small specialized device that operates selectively on several analytes [167]. Optical sensors are certainly competitive with ion-selective electrodes because they are now less complex than in the past and miniaturization promises further success in the future. Furthermore, optical sensors have advantages [168] in remote sensing and by utilization of the spectral information, especially in the case of sensor arrays.

Along with simple optical sensors that measure transmission, so-called optodes [169] play an important role. In these, either the color change of an indicator dye in a membrane is monitored by using diffuse reflection [170], or the effect of an analyte on the fluorescence spectrum is observed. In addition, many other effects are used, such as surface

plasmon resonance (see Section 16.4.5.1), and grating couplers (see Section 16.4.5.2). In most cases, interaction of the analyte with the evanescence field of a glass fiber, a slab or strip waveguide plays the important part, influencing the wave propagation of the coupled electromagnetic field by changing the refractive index of the superstrate [171]. Furthermore, determination of the thickness of thin films by interferometric methods (see Section 16.4.4.4) can be used for analysis of both gases [146], [172], [173] and liquids [147], [174]. Here, the amount of swelling of polymer film due to the permeation by the analyte is determined. Accumulations of material caused by antigen – antibody interactions [147] can also be observed by methods sensitive to changes in layer thickness. Label-free detection of biomolecular interaction by optical sensors has been reviewed recently [175].

Direct optical detection of DNA hybridisation by SPR [176] or RIfS [177] has been achieved. In routine sensing, assays with labeled compounds are preferred using different fluorescent techniques [159].

Optical sensors have become very important, are used more and more widely, and have been reviewed comprehensively in several monographs [178].

16.4.7. Measuring Methods Based on Deviations from the Lambert – Beer Law

Effects that lead to deviations from the Lambert – Beer law are given below. On the left-hand side, the factors that influence the spectrum are listed, on the right the chemical and physical aspects are given:

Changes to the spectrum due to	Caused by
Concentration	self-association, dissociation
pH change	acids/bases proteolytic reaction, equilibrium
Solvent	interaction between solvent and solute
Presence of additional component B	complex formation
Time	thermal chemical reaction (including solvolysis)
Incident light	photochemical reaction, photochromism
Temperature	thermochromism

The most important of these are the temperature dependence of the molar absorption coefficient (due to density change); the effect of association
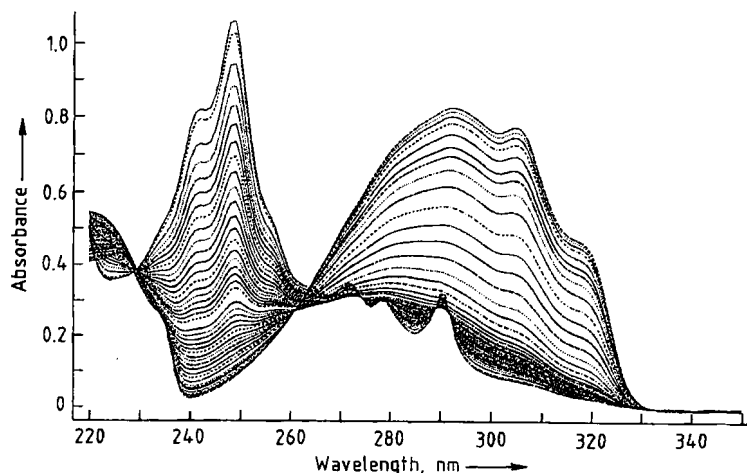
Figure 36. Reaction spectrum of the photoreaction of *trans*-stilbene [183]

and dissociation equilibria, which change with temperature, concentration, and pH; and, not least, concentration changes caused by chemical and photochemical reactions.

Like typical titrations in which electrodes are used, photometric titrations can be carried out by using indicator dyes. Here, the pH is plotted not against the concentration of anions or dissociated components, but against the absorbance at one or more wavelengths [179]. From these diagrams, pK values of even polyvalent acids and bases or of complex equilibria can be determined. Furthermore, it is possible to linearize the diagram, which allows better control of the model of the number of steps of titration [180], [181].

UV – VIS spectroscopy has also found wide use in monitoring thermal and photochemical reactions. Absorbance is measured at several wavelengths during the reaction, and dynamic multicomponent analysis is carried out by using the Lambert – Beer law (see Section 16.4.2). In the simplest case, wavelengths can be found at which just one component absorbs. The absorbance signal is then directly proportional to the concentration of one reactant and can be evaluated by the usual kinetic equations [182], [183]. An example is given in Figure 24, in which case the thermal reaction has also been used to check the reproducibility of the wavelength of a spectrometer.

The interpretation of photochemical reactions is more complicated if they have to be followed spectroscopically. An example is given in Figure 36, which illustrates the photochemical isomerization of *trans*-stilbene to *cis*-stilbene followed

by ring closure to form the end product phenanthrene via dihydrophenanthrene, an intermediate that is not detectable by conventional methods. The diagram shows a photochemical reaction spectrum. The measured data can be evaluated by using various methods [184]. When investigating such reactions, a process-controlled system is usually employed in which the irradiation equipment and the spectrometer are combined [185].

16.5. Special Methods

16.5.1. Derivative Spectroscopy

If absorption bands of two compounds overlap (one broad and the other narrow), derivative spectroscopy can be used for resolution. Herewith even trace amounts can be analyzed quantitatively [186]. Derivative spectroscopy is also used to investigate turbid solutions such as industrial wastewaters or biological samples. This method is also suitable for the analysis of shoulder-shaped spectra and overlapping absorption bands. The derivative of transmission with respect to wavelength is not proportional to concentration, but the derivative of absorbance with respect to wavelength is. Generally, the first and second derivatives are used [115]. Higher derivatives present noise problems, as the amount of noise increases with each derivative. Higher derivatives have been claimed to be useful for special applications [187].

Table 3. Various methods of obtaining derivatives, with their advantages and disadvantages

Method	Details of method	Advantages	Disadvantages
Optical	frequency modulation of the light beam by oscillating gratings, slits, or mirrors determination of Fourier coefficients equipment with constant wavelength distance	no effect on the spectrum of variations in the intensity of the light source very good signal to noise ratio	higher derivatives not obtainable special optical construction necessary
Electronic	derivative obtained from the analog signal by differentiation and attenuation circuits	direct on-line spectra higher derivatives possible	derivative as a function of time and not of wavelength
Computed	the digitalized data are stored and the derivative is calculated	direct derivative as a function of wavelength original spectrum is not lost; use of various algorithms optimization of the selected parameters	necessity for an integrated computer or data station

Table 4. Various mathematical algorithms for smoothing curves and producing their derivatives, with their advantages and disadvantages

Numerical algorithm	Advantages	Disadvantages
Differential coefficient $\Delta E / \Delta \lambda$	simple, quick	signal-to-noise ratio is made worse by subtraction of two measured values that are subject to error
Polynomial smoothing	derivative obtained directly from polynomial coefficients number of restart points determines the degree of smoothing	loss of measured points at the beginning and end of the spectrum problems with sharp peaks
Cubic spline	treatment of the entire spectrum in one step good linear dependence on concentration	large influence of the parameters

Optical, electronic, and mathematical methods are available [3], [188], whose advantages and disadvantages are listed in Table 3. The algorithms generally used for derivation are listed in Table 4, along with their advantages and disadvantages. For three overlapping Gaussian peaks, the normal spectra and the first, second, third, and seventh derivatives are shown schematically in Figure 37. For both the first and the second derivative, by forming the difference for a neighboring minimum and maximum, a linear relationship can be found between these values and concentration. On differentiation, the peaks become narrower and the slopes steeper. Thus, in the second derivative, a steep slope occurs in the calibration curve, and the extreme values are further apart. The derivative signals are less marked if the distance between supporting points is comparable to, or wider than, the width at half peak value.

However, Figure 37 shows not only that the number of peaks increases with higher derivatives, but that when several bands overlap, the number of minima and maxima increases so that the shape of the curve becomes unclear. If real, noisy, raw data are used instead of simulated data, interpretation

with the aid of derivatives becomes much more critical (see Fig. 37 B). If too much smoothing is used, information is lost. If too little smoothing is employed, significantly higher derivatives are impossible.

Each differentiation step leads in principle to an increase in noise, so that no additional information can be obtained by this mathematical procedure, although in some cases this form of "spectral manipulation" offers a possible way of revealing information hidden in raw data. However, as implied by the word "manipulation," the final result is affected by the procedure used.

The potential and limitations of derivative spectroscopy are described in many literature references, which also include several examples of its use [7], [186], [189].

Derivative spectroscopy can be applied to fluorescence and fluorescence excitation spectra [188]. It is also used in chromatography in the search for peak maxima, and in interferometric measurements for determining minima and maxima in interference spectra [124]. A typical application is the determination of phenol in water (Fig. 38) [190].
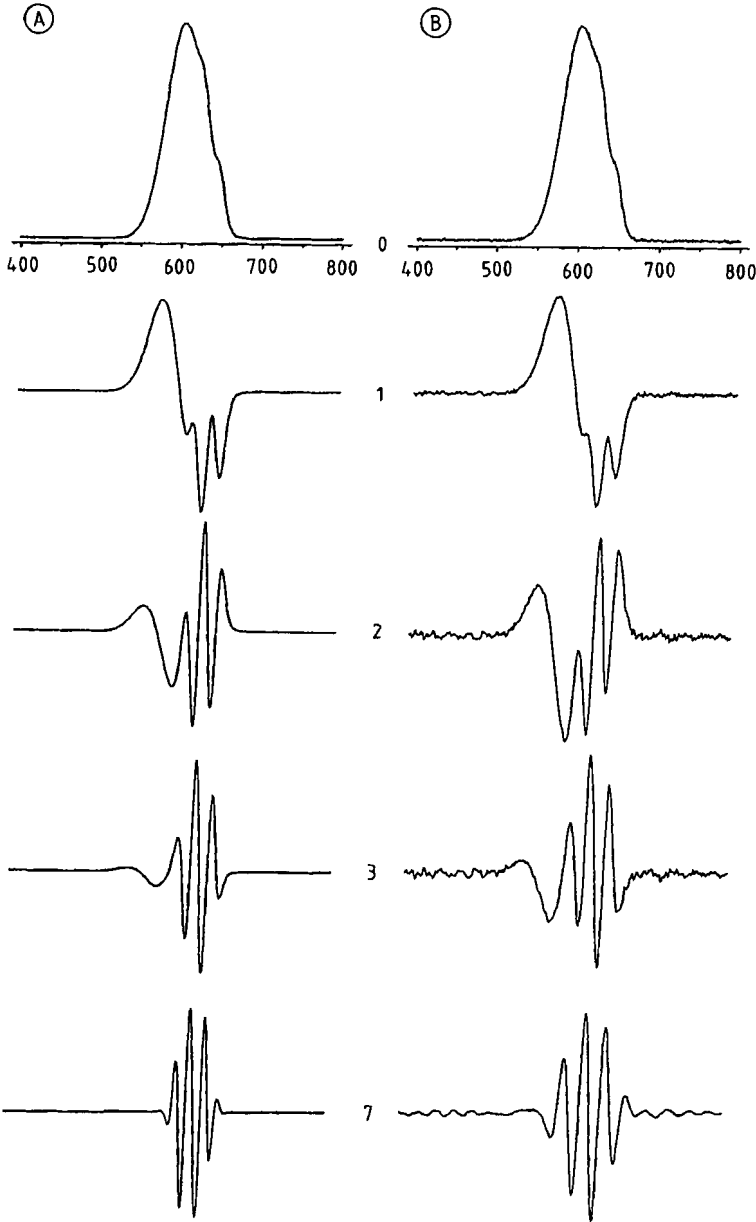
Figure 37. Schematic diagram of various derivatives of three overlapping Gaussian peaks
A) Simulated spectra; B) Spectra with 0.3% errors superimposed

16.5.2. Dual-Wavelength Spectroscopy

Another possible method of investigating biological samples that cause light scattering is by dual-wavelength spectroscopy [191], [192], although this method is limited to very special applications. The number of types of equipment used is therefore also limited. First, normal spectrometers can be used, the signals for two wavelengths being taken from the observed spectra. Second, so-called bifrequency equipment can be used, which operates with two beams at different wavelength at the same time. This is possible either with two monochromators, which cause light of
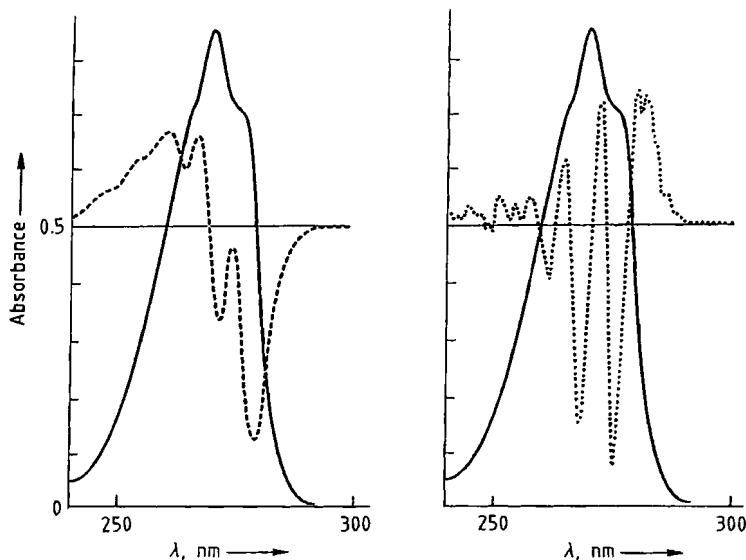
Figure 38. Absorption spectrum of phenol in water (—), showing first (– – –) and second (· · · ·) derivatives of the spectrum [190]

two wavelengths to fall on the sample simultaneously, or with one monochromator and a modulated slit, which provide light of two wavelengths that alternate continuously.

This method has the advantage of avoiding problems from differences in cell positions, differences in the cells themselves due to spurious absorptions or impurities inside or outside the cell, scattering effects, or turbidity. Two techniques are used, first the so-called *equiabsorption method* in which two wavelengths are chosen for which one of the components has the same absorption. The second component must absorb at only one of the two wavelengths. This enables sodium nitrite to be determined, for example, in the presence of sodium nitrate (see also Fig. 39).

The *signal amplification method* is of general application. It is used either if no region occurs in which only one component absorbs or if the interfering component has no absorption maximum. With this type of equipment, the absorption can be corrected by a factor at one wavelength to give the same value as that obtained at the other wavelength (the uncharacteristic spectrum is "evened out"). This enables the concentration of the desired component to be determined from the absorbance difference at the two wavelengths if its absorbance coefficients at the two measuring wavelengths are known. This method has been used in medicine and for biological samples [192]. However, it has recently decreased in importance.

16.5.3. Scattering

As described in Section 16.2.2.3, scattering of aqueous solutions increases considerably when larger suspended particles such as soot, mist droplets, fat droplets, or bacteria are present. In analogy to the Lambert – Beer law, attenuation of the incident light by a scattering medium is described in the ideal case by

$$I = I_0 e^{-\tau l} \tag{22}$$

where I is the intensity after passage through the medium, I_0 is the incident intensity, l is the optical path length, and τ is the so-called turbidity coefficient. This equation is valid only for scattering particles whose diameter is small compared to the wavelength of the light. The intensity of scattered light depends on the number of scattering particles per unit of volume, the angle at which the scattered light is observed, the size and shape of the scattering particles, and the wavelength of the light [18], [193]. These various relationships can be utilized in analysis [194].

16.5.3.1. Turbidimetry

Measurement of the attenuation of an incident light beam on passing through a scattering medium is known as turbidimetry and can be carried out with simple colorimeters or photometers.
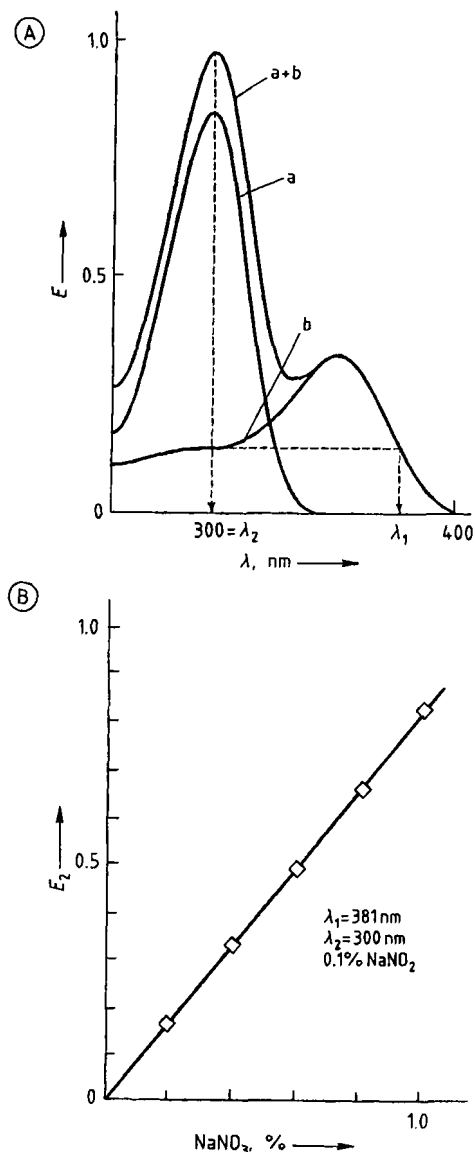
Figure 39. Determination of sodium nitrate and sodium nitrite in the presence of each other [192]
A) Spectra of 1 % NaNO₃ (a) and 0.1 % NaNO₂ (b);
B) Calibration curve

However, spectrometers in which the wavelengths of the incident radiation can be chosen are used for more exact determinations. These are preferable for measuring either relatively low or very high turbidity. Measuring errors can become very large in both cases.

16.5.3.2. Nephelometry

Nephelometry is used if turbidity is slight. Here, the intensity of scattered radiation is measured, not the difference between incident and attenuated radiation. Under these conditions, errors due to absorption are less likely to occur in nephelometry. The intensity of light deflected to the side is determined; this can be carried out by using fluorometers. Since measurement at an angle of 90° to the incident light direction does not always give the best results (even with particles having a diameter of 1 μm, ca. 90 % of the scattered light is in a cone at an angle of only 10° to the incident light beam), the equipment must include a goniometer so that measurements can be carried out at various angles [195], [196].

Synthetic turbidity standards are generally used in the quantitative measurement of turbidity. Alternatively, calibration curves can be obtained by precipitating known amounts of the substance to be measured. However, grain size depends not only on precipitation conditions, but also on the time delay between precipitation and measurement, so conditions must be standardized carefully to ensure reproducible results. Nephelometric and turbidimetric methods have also become important in recent years in continuous automatic on-line process control (e.g., of wastewater).

16.5.3.3. Photon Correlation Spectroscopy

Turbidimetry and nephelometry give no information on particle shape. However, if the principle of nephelometry is combined with time-dependent measurement, photon correlation spectroscopy can provide information on the shape of the scattering particle. This dynamic light scattering has become very important in colloid chemistry [197], [198].

16.5.4. Luminescence, Excitation, and Depolarization Spectroscopy, and Measurement of Lifetimes

If the molecules and their direction of polarization are not distributed statistically in the sample (anisotropy), a preferred spatial direction for the fluorescence exists. If polarized light is used for excitation [199], only those molecules whose axis of polarization is parallel to the axis of polarization of incident light are excited substantially. This results in a preferred direction of polarization for the fluorescence radiation. However, if molecules
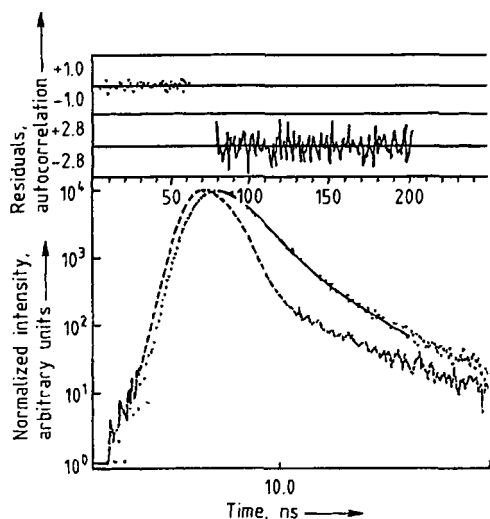
Figure 40. Logarithmic plot of measured intensity of fluorescence against time in a single photon counting experiment

Top diagram shows autocorrelation and residuals between fluorescence intensity decay curve ($\cdots$) and excitation light pulse (– – –) [138]

can move or rotate freely within the lifetime of the excited state, this preferred direction of fluorescence polarization is disturbed (depolarization effects) [200]. These effects enable statements to be made about the environment of the molecules. The degree of polarization is given by

$$p = \frac{I_\parallel - I_\perp}{I_\parallel + I_\perp} \qquad p = \frac{3\cos^2\alpha - 1}{\cos^2\alpha + 3} \qquad (23)$$

where α is the angle between the transition moments during absorption and emission processes.

Since the lifetime of the excited state is also characteristic of individual substances, and particularly of the environment of the molecule, time-dependent fluorescence measurements with polarized light, with variation of the excitation and polarization wavelength, can be performed. They give important information about the orientation and dynamics of the sample. Many of these methods are successfully applied in biosensing and bioanalytics [159].

The method of *single photon counting* [201], [202] is generally used for such lifetime determinations. A weak pulsed light source causes excitation of a small number of molecules. Spread over a period of time (spontaneous fluorescence is a statistical process), single fluorescence photons

are emitted and are "counted" only in a small time window which proceeds between pulses by a ramp technique. By following thousands of pulses, the statistically emitted photons are counted in the channels (time windows), giving the decay curve shown in Figure 40. The time constants are determined by curve fitting.

If several deactivation processes occur at the same time, an incorrectly chosen number of overlapping exponential functions for fitting can lead to physical misinterpretation. Figure 40 shows the intensity of the excitation light source, the decay of fluorescence intensity, and curve fitting. If application of the residuals leads to systematic deviation from the mean zero value, this can indicate a poor curve fit or an incorrectly chosen number of exponential curves.

16.5.5. Polarimetry

16.5.5.1. Sugar Analysis

Saccharimetry is very important in the sugar industry. An aqueous solution of sugar is analyzed by determining the optical rotation α, using the green mercury line (546 nm) or yellow sodium line (589 nm). However, the effects of evaporation, the unstable nature of sugar solutions, and temperature effects can easily lead to errors in the determination of optical rotation.

In sugar manufacture, automatic quality control is important, and in yeast production the molasses stream must be determined to have optimum sugar content. Here, too, the optical rotation of molasses is measured automatically in a continuous flow cell [38].

16.5.5.2. Cellulose Determination

Whereas conventional gravimetric cellulose determination is complicated and time-consuming, cellulose concentrations can be determined rapidly in ammoniacal copper solution by measuring the optical rotation of the levorotatory solution. However, the solutions must be completely free of turbidity [138].

16.5.5.3. Stereochemical Structural Analysis

Optical rotatory dispersion can be used as a routine procedure, especially for steroids, since the asymmetry of spatial orientation of a molecule is also affected by its conformation. Cholestan-
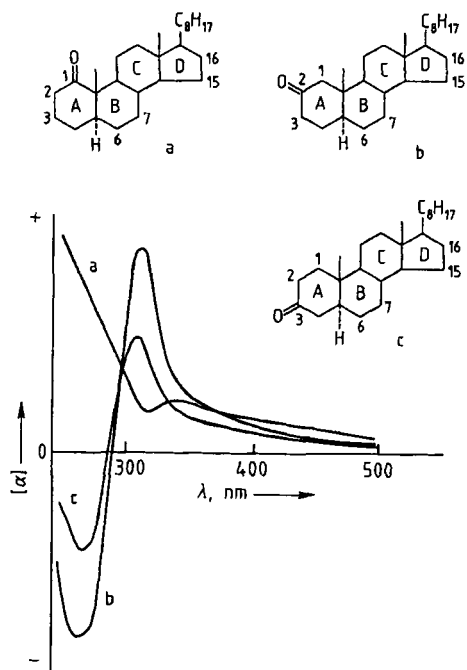
Figure 41. ORD curves of cholestan-1-one (a), cholestan-2-one (b), and cholestan-3-one (c), all dissolved in methanol [38]

1-one, cholestan-2-one, and cholestan-3-one, for example, can be distinguished because of their different ORD curves (see Fig. 41). Molecular asymmetry (dissymmetry) is a prerequisite of optical activity. A dissymmetric disturbance of the electron system of the keto group caused by a change in the stereochemical arrangement is the case of other ketosteroids (such as androsterone). Here, ORD and CD curves often give information only on relative configuration.

Such ORD and CD curves are used for purity control of optically active substances in pharmaceutical chemistry.

16.5.5.4. Use of Optical Activity Induced by a Magnetic Field

In solid-state inorganic chemistry, MCD is an important technique for the determination of color centers in alkali-metal halide crystals, rare earths, and highly symmetric compounds such as hexahalides of platinum or ruthenium. In organic chemistry, MCD is observed with, e.g., metal porphyrins, pyrimidine bases, and nucleoside derivatives.

16.5.6. Photoacoustic Spectroscopy (PAS)

Photophysical processes can take place in combination with the excitation of molecules to more highly excited electronic states (see Section 16.2.4.2). In the subsequent radiationless transitions, the energy of excited molecules is taken up by the environment (neighboring molecules). Thermal equilibrium is thus reestablished and the temperature of the environment is increased. This is associated with a volume change, which is appreciable for a gas but less pronounced for liquids and solids.

If periodically modulated visible radiation is used, the periodic volume changes cause pressure variations and sound waves that can be detected (e.g., by a microphone). The photoacoustic or optoacoustic effect is based on this principle [203], [204]. Although this has been known for a long time, it has become more important since the late 1970s as an adjunct to reflection spectroscopy if samples are either very strongly or very weakly absorbing. The "heating" of the sample or its environment depends on the extent to which the modulated electromagnetic radiation can be absorbed. This corresponds to the spectral distribution of the molar absorption coefficients. If the wavelength of excitation radiation is varied, the amplitudes of pressure fluctuations are proportional to dynamic changes in the temperature of the system, and thus to the absorption spectrum [205].

Photoacoustic spectra can be observed with a simple microphone and optical apparatus. A diagram of a modern piece of equipment is given in Figure 42, in which there is a blackbody sample at the second input of a look-in amplifier. Corrected photoacoustic spectra can be recorded with this type of monochromator equipment. Frequency-modulated xenon lamps are used as white light sources, so that a chopper is not needed for modulation.

Photoacoustic spectroscopy is used with gases [206]–[208], liquids [209], and solid samples [210]–[212]. Simple solutions of the thermal diffusion equation can be used only under certain conditions. Areas of use [213] include investigation of waste gases (see Fig. 43), absolute measurements of fluorescence quantum yields [214], determination of dye concentrations in films, investigation of thin films in catalysis [210], [215], and biological and physiological problems such as the investigation of leaves and samples of skin or
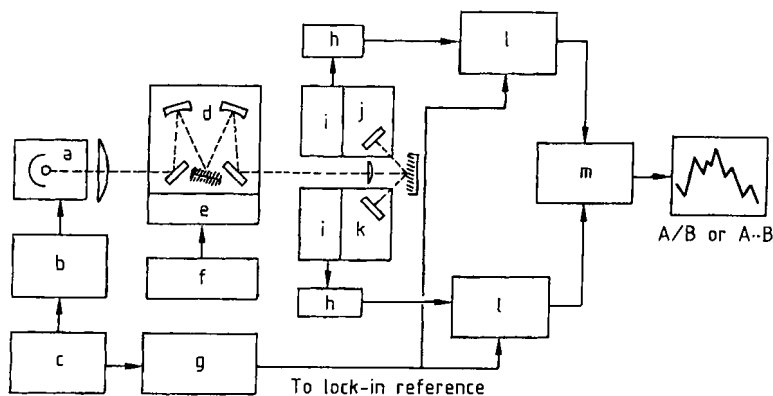
Figure 42. Block diagram of commercial photoacoustic equipment
a) 300-W Xenon lamp; b) Lamp power supply; c) Sine-wave modulator; d) Monochromator (two gratings); e) Wavelength drive stepper motor; f) Monochromator controls; g) Modulation frequency display; h) Preamplifier; i) Microphone; j) Cell A; k) Cell B; l) Lock-in amplifier; m) Ratiometer/difference amplifier
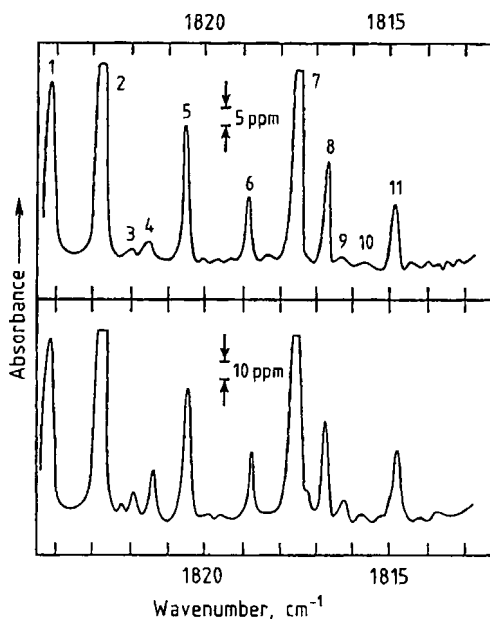
Figure 43. Absorption spectra of gas samples
Upper diagram shows calibration spectra for 20 ppm NO in N_2. Lines 1, 5, 6, 8, and 11 are due to NO, and other lines to water vapor; lower diagram shows spectrum of automobile exhaust gas

blood. Despite the wide area of application, use of this equipment is very limited so few types of equipment are available on the market. However, individual instances of the application of PAS in sensor technology have been described [216].

16.6. References

[1] J. D. Ingle, Jr., S. R. Crouch: *Analytical Spectroscopy,* Prentice-Hall, Englewood Cliffs, NJ, 1988.

[2] G. Svehla (ed.): "Analytical Visible and Ultraviolet Spectrometry," in: *Comprehensive Analytical Chemistry,* vol. XIX, Elsevier, Amsterdam 1986.

[3] G. Gauglitz: *Praktische Spektroskopie,* Attempto Verlag, Tübingen 1983.

[4] D. A. Skoog, J. J. Leary: *Principles of Instrumental Analysis,* Saunders College Publishing, Fort Worth 1992.

[5] G. Gauglitz, K.-A. Kovar in E. Nürnberg, P. Surmann (eds.): *Hagers Handbuch der pharmazeutischen Praxis,* 5th ed., **vol. 2,** Methoden, Springer Verlag, Berlin 1991, p. 157.

[6] D. L. Massart, B. G. M. Vandeginste, S. N. Demmig, Y. Michotte, L. Kaufman: Data Handling in Science and Technology 2: *Chemometrics: a Textbook,* Elsevier, Amsterdam 1988.

[7] H.-H. Perkampus: *UV-VIS Spectroscopy and its Applications,* Springer Verlag, Berlin 1992.

[8] H. H. Jaffé, M. Orchin: *Theory and Applications of Ultraviolet Spectroscopy,* Wiley, New York 1962.

[9] M. Klesinger, J. Michl: *Lichtabsorption und Photochemie organischer Moleküle,* VCH Verlagsgesellschaft, Weinheim 1989.

[10] P. W. Atkins: *Molecular Quantum Mechanics,* 2nd ed., Oxford University Press, Oxford 1983.

[11] L. Bergmann, C. Schaefer: *Lehrbuch der Experimentalphysik,* vol. 6: Optik, De Gruyter, Berlin 1987.

[12] M. Born, E. Wolf: *Principles of Optics,* Pergamon, New York 1980.

[13] E. Hecht, A. Zajac: *Optics,* Addison-Wesley, Reading 1974.

[14] P. W. Atkins: *Physikalische Chemie,* VCH Verlagsgesellschaft, Weinheim 1990.

[15] R. P. Feynman, R. B. Leighton, M. Sands: *The Feynman Lectures on Physics,* Bilingna Edition, R. Oldenbourg, München, Addison-Wesley, Reading 1977.

[16] W. Demtröder, "Laser Spectroscopy," Springer Series in Chemical Physics 5, Springer Verlag, Berlin 1981.

[17] H. A. Stuart: *Molekülstruktur.* Springer Verlag, Berlin 1967.

[18] H. C. van der Hulst: *Light Scattering by Small Particles,* Wiley, New York 1957.

[19] K. A. Stacey: *Light-Scattering in Physical Chemistry,* Butterworths, London 1956.

[20] N. J. Harrick: *Internal Reflection Spectroscopy,* Harrick Scientific Corporation, New York 1979.

[21] H. Weber, G. Herziger: *Laser,* Physik-Verlag, Weinheim 1972.

[22] G. Kortüm: *Kolorimetrie, Photometrie und Spektrometrie,* Springer Verlag, Berlin 1962.

[23] M. Krystek in W. Erb (ed.): *Leitfaden der Spektroradiometrie.* INSTAND vol. 6, Springer Verlag 1989, p. 321.

[24] H.-H. Perkampus: *Lexikon der Spektroskopie,* VCH Verlagsgesellschaft, Weinheim 1993.

[25] K. Laqua, W. H. Melhuish, M. Zander: "Molecular Absorption Spectroscopy, Ultraviolet and Visible (UV/VIS)," *Pure Appl. Chem.* **60** (1988) 1449–1460.

[26] V. A. Fassel et al., *Pure Appl. Chem.* **45** (1976) 105.

[27] J. Inczedy, T. Lengyel, N. Ure: *Compendium of Analytical Nomenclature.* Blackwell Science, Oxford 1998.

[28] C. A. Parker: *Photoluminescence of Solutions,* Elsevier, Amsterdam 1969.

[29] W. H. Melhuish, *Pure Appl. Chem.* **56** (1984) 231.

[30] H. Mauer, G. Gauglitz: "Photokinetics; Theoretical Fundamentals and Applications," in R. G. Compton, G. Hancock (eds.): *Comprehensive Chemical Kinetics,* **vol. 36,** Elsevier, Amsterdam 1998.

[31] H. A. Staab: *Einführung in die theoretische organische Chemie,* Verlag Chemie, Weinheim 1966.

[32] G. G. Guilbault: *Practical Fluorescence; Theory, Methods, and Techniques,* Marcel Dekker, New York 1973.

[33] Th. Förster: *Floureszenz organischer Verbindungen,* Vandenhoek, Göttingen 1951.

[34] J. Michl, E. W. Thulstrup: *Spectroscopy with Polarized Light,* VCH Verlagsgesellschaft, Weinheim 1986.

[35] H. Rau in H. Naumer, W. Heller (eds.): *Untersuchungsmethoden in der Chemie; Optische Aktivität und Polarimetrie,* Thieme-Verlag, Stuttgart 1990.

[36] C. Djerassi: *Optical Rotatory Dispersion Applications to Organic Chemistry,* McGraw Hill, New York 1960.

[37] G. Snatzke: *Optical Rotatory Dispersion and Circular Dichroism in Organic Chemistry,* Heyden & Son, London 1967.

[38] *Ullmann,* 4th ed., **5,** 628.

[39] F. Ciardelli, P. Salvadori (eds.): *Fundamental Aspects and Recent Developments in Optical Rotatory Dispersion and Circular Dichroism,* Heyden & Son, London 1973.

[40] J. Michl: "Magnetic Circular Dichroism of Cyclic π-Electron Systems," *J. Am. Chem. Soc.* **100** (1978) 6801, 6812, 6819.

[41] DIN Norm 5030 Part 3 (1983), DIN Norm 32 635, Beuth-Verlag, Berlin 1982.

[42] A. Reule in W. Erb (ed.): *Leitfaden der Spektroradiometrie,* INSTAND vol. 6, Springer Verlag, Berlin 1989, p. 179.

[43] Perkin-Elmer, company information.

[44] W. Kaye, D. Barber, R. Marasco: "Design of a Microcomputer-Controlled UV–VIS Spectrophotometer," *Anal. Chem.* **52** (1980) 437A.

[45] G. Gauglitz, T. Klink, A. Lorch, *Fresenius Z. Anal. Chem.* **319** (1984) 364.

[46] Application information by Varian, Darmstadt, Germany.

[47] P. M. Epperson, J. V. Sweedler, R. B. Bilhorn, G. R. Sims, M. B. Denton, *Anal. Chem.* **60** (1988) 327A.

[48] J. V. Sweedler, R. B. Bilhorn, P. M. Epperson, M. B. Denton, *Anal. Chem.* **60** (1988) 282A.

[49] D. G. Jones, *Anal. Chem.* **57** (1985) 1057A.

[50] J. A. Decker: "Hadamard Transform Spectrometry: a New Analytical Technique," *Anal. Chem.* **46** (1974) 1803.

[51] R. Geick, *Chem. Labor Betr.* **23** (1972) 193, 250, 300.

[52] L. Genzel, *Fresenius Z. Anal. Chem.* **273** (1975) 391.

[53] A. G. Marshall, F. R. Verdon: *Fourier Transform in NMR. Optical, and Mass Spectrometry,* Elsevier, New York 1990.

[54] P. R. Griffiths: *Chemical Fourier Transform Spectroscopy,* Wiley, New York 1975.

[55] O. S. Wolfbeis, G. Boisdé, G. Gauglitz in W. Göpel (ed.): *Sensors and Biosensors,* VCH Verlagsgesellschaft, Weinheim 1992, p. 575.

[56] J. Kiefer (ed.): *Ultraviolette Strahlen,* De Gruyter, Berlin 1977.

[57] L. Endres, H. Frietz in W. Erb (ed.): *Leitfaden der Spektroradiometrie:* Strahler, INSTAND vol. 6, Springer Verlag, Berlin 1989, p. 23.

[58] H. A. Strobel, W. R. Heineman: *Chemical Instrumentation: a Systematic Approach,* Wiley & Sons, New York 1989.

[59] J. F. Rabeck: *Experimental Methods in Photochemistry and Photophysics,* Wiley, Chichester 1982.

[60] J. Wilson, J. F. B. Hawkes: *Lasers, Principles and Applications,* Prentice-Hall, Englewood Cliffs, NJ, 1987.

[61] D. Engelage: *Lichtwellenleiter in Energie und Automatisierungsanlagen.* Hüthig-Verlag, Heidelberg 1986.

[62] J. Pfab, *Anal. Proc. (London)* **28** (1991) 415.

[63] K. Niemax, *Fresenius J. Anal. Chem.* **337** (1990) 551.

[64] U. Panne, F. Lewitzka, R. Nießner, *Analusis* **20** (1992) 533.

[65] F. P. Schäfer in F. P. Schäfer (ed.): *Dye Lasers. Topics in Applied Physics,* **vol. 1,** Springer Verlag, Berlin 1973, pp. 1–90.

[66] F. K. Kneubühl, M. W. Sigrist: *Laser,* Teubner Studienbücher Physik, Stuttgart 1991, p. 331.

[67] J. Eichler, H.-J. Eichler: *Laser,* Springer Verlag, Berlin 19990, p. 173.

[68] H. A. Macleod: *Thin-film Optical Filters.* Adam Hilger Ltd., Bristol 1986.

[69] A. G. Reule: "Errors in Spectrometry and Calibration, Procedures to Avoid Them," *J. Res. Nat. Bur. Stand. Sect. A* **80** (1976) 609.

[70] Optimum Parameters for Spectrophotometry, Varian Instruments, OPT-720A.

[71] L. Velluz, M. Legrand, M. Grosjean: *Optical Circular Dichroism,* Verlag Chemie, Weinheim 1965.

[72] U. B. Seiffert, D. Janson: "Zur optischen Qualität von Photometerküvetten aus Plastik," *J. Clin. Chem. Clin. Biochem.* **19** (1981) 41.

[73] G. Gauglitz: "Opto-Chemical and Opto-Immuno Sensors," in H. Baltes, W. Göpel, J. Hesse (eds.): *Sensors Update,* **vol. 1,** VCH Verlagsgesellschaft, Weinheim 1996.

[74] W. Böhme, H. W. Müller, W. Liekmeier, *Int. Lab.* (1991) Jan/Feb.

[75] W. Böhme, K. Horn, D. Meissner, *LaborPraxis* **11** (1987) 628.

[76] E. L. Dereniak, D. G. Crowe: *Optical Radiation Detectors*, Wiley, New York 1984.

[77] K. Möstl in W. Erb (ed.): *Leitfaden der Spektroradiometrie*, INSTAND vol. 6, Springer Verlag, Heidelberg 1989, p. 101.

[78] F. Grum, C. J. Bartleson: *Optical Radiation Measurements*, vol. 4: Physical Detectors of Optical Radiation, Academic Press, New York 1983.

[79] Photomultiplier Handbook, Burle Ind., Lancaster 1989.

[80] C. Zeiss: Produktbeschreibung Simultanspektrometer, Oberkochen.

[81] Firma Reticon, München, data sheets.

[82] Y. Talmi: *Multichannel Image Detectors*, **vols. 1, 2,** Am. Chem. Soc., Washington, DC, 1979, 1982.

[83] J. V. Sweedler, K. L. Ratzlaff, M. B. Denton (eds.): *Charge-Transfer Devices in Spectroscopy*, VCH Verlagsgesellschaft, Weinheim 1994.

[84] E. L. Wehry (ed.): *Modern Fluorescence Spectroscopy*, Plenum, New York 1981.

[85] E. Lippert, W. Nägele, I. Seibold-Blankenstein, U. Steigert, W. Voss, *Fresenius Z. Anal. Chem.* **170** (1959) 1.

[86] Operation manual CD-spectrometer Jasco.

[87] R. Hamilton, S. Hamilton: *Thin Layer Chromatography*, Wiley, New York 1987.

[88] D. Oelkrug, A. Erbse, M. Plauschinat, *Z. Phys Chem.* **96** (1975) 283.

[89] U. Hezel: "Direkte quantitative Photometrie an Dünnschichtchromatogrammen," *Angew. Chem.* **85** (1973) 334; *GIT Fachz. Lab.* **21** (1977) 694.

[90] P. Kubelka, *J. Opt. Soc. Am.* **38** (1948) part 1, 448.

[91] S. Bayerbach, G. Gauglitz, *Fresenius J. Anal. Chem.* **335** (1989) 370.

[92] Operation manual of MOSS ES4G ellipsometer, SOPRA Paris, France.

[93] R. M. A. Azzam, N. M. Bashera: *Ellipsometry and Polarized Light*, North Holland, New Amsterdam 1977.

[94] Ch. Striebel, A. Brecht, G. Gauglitz: "Characterization of Biomembranes by Spectral Ellipsometry, Surface Plasmon Resonance and Interferometry with regard to Biosensor Application," in *Biosen. Bioelectr.*, **9** (1994) 139.

[95] H. Arwin: "Ellipsometry," in A. Baszkin, W. Norde (eds.): *Physical Chemistry of Biological Interfaces*, Marcel Dekker, New York 2000.

[96] "Maintaining Optimum Spectrophotometer Performance," Perkin-Elmer, Application Data Bulletin; *Am. Soc. Clin. Pathol. Techn. Improv. Ser.* **27.**

[97] Recommended Practices for General Techniques of Ultraviolet Quantitative Analysis, E169-63, ASTM in Manual of Recommended Practices in Spectroscopy, Philadelphia: American Society of Testing Materials Committee E13, 1969, 10–34.

[98] C. Burgess, K. D. Mielenz (eds.): *Advances in Standards and Methodology in Spectrophotometry*, Elsevier, Amsterdam 1987.

[99] A. Knowless, C. Burgess: *Practical Absorption Spectrometry*, Chapman and Hall, London 1984.

[100] C. Burgess, A. Knowless: *Standards in Absorption Spectrometry*, Ultraviolet Spectrometry Group, vol. 1, Chapman and Hall, London 1981.

[101] R. W. Burke, E. R. Deardorff, O. Menis: "Liquid Absorbance Standards," *J. Res. Nat. Bur. Stand. Sect. A* **76** (1972) 469.

[102] R. E. Poulson: "Test Methods in Spectrophotometry: Stray Light Determination," *Appl. Opt.* **3** (1964) 99.

[103] W. Luck, *Z. Elektrochem.* **64** (1960) 676.

[104] Instruction manual, Hellma, Mühlheim/Baden.

[105] S. Tesch, K. H. Rentrop, M. Otto, *Fresenius J. Anal. Chem.* **344** (1992) 206.

[106] H. Bandemer, M. Otto, *Mikrochim. Acta* 1986, no. 2, 93.

[107] K. Schuchardt, *Nachr. Chem. Tech. Lab.* **40** (1992) M1–M20;H. J. Majer, *GIT Fachz. Lab.* **37** (1993) 881.

[108] D. H. Williams, I. Flemming: *Strukturaufklärung in der organischen Chemie*, Thieme-Verlag, Stuttgart 1985.

[109] G. Rücker, M. Neugebauer, G. G. Williams: *Instrumentelle pharmazeutische Analytik*, Wissenschaftliche Verlagsgesellschaft, Stuttgart 1988.

[110] M. Hesse, H. Meier, B. Zeeh: *Spektroskopische Methoden in der organischen Chemie*, Thieme Verlag, Stuttgart 1985.

[111] G. Gauglitz, K.-A. Kovar in E. Nürnberg, P. Surmann (eds.): *Hagers Handbuch der pharmazeutischen Praxis*, 5th ed., **vol. 2**, Methoden, Springer Verlag, Berlin 1991, p. 471.

[112] I. Barret, W. F. Smyth, I. E. Davidson, *J. Pharm. Pharmacol.* **25** (1973) 387.

[113] H. Martens, T. Naes: *Multivariate Calibration*, J. Wiley & Sons, Chichester 1993.

[114] E. R. Malinowski, D. G. Howery: *Factor Analysis in Chemistry*, Wiley, New York 1980.

[115] S. Ebel, E. Glaser, S. Abdulla, U. Steffens, V. Walter, *Fresenius Z. Anal. Chem.* **313** (1982) 24.

[116] H. M. Heise, *Fresenius J. Anal. Chem.* **345** (1993) 604.

[117] R. Marbach, H. M. Heise, *Chemometr. Intel. Lab. Syst.* **9** (1990) 45.

[118] S. D. Brown, *Chemometr. Intel. Lab. Syst.* **10** (1991) 87.

[119] M. Mettler, G. Gauglitz, *TrAc Trends Anal. Chem.* **11** (1992) 203.

[120] J. A. Weismüller, A. Chanady, *TrAc Trends Anal. Chem. (Pers. Ed.)* **11** (1992) 86.

[121] T. Kohonen: *Self-Organization and Associative Memory*, Springer Verlag, Berlin 1989.

[122] C. Schierle, M. Otto, *Fresenius J. Anal. Chem.* **344** (1992) 190.

[123] J. Göppert et al., Proc. Neuro Nimes 1992.

[124] G. Kraus, G. Gauglitz, *Fresenius J. Anal. Chem.* **346** (1993) 572.

[125] K. Molt, M. Egelkraut, *Fresenius J. Anal. Chem.* **327** (1987) 77.

[126] W. Windig, *Chemometr. Intel. Lab. Syst.* **16** (1992) 1.

[127] J. Eisenbrand: *Fluorimetrie*, Wissenschaftliche Verlagsgesellschaft, Stuttgart 1966.

[128] G. Gauglitz, R. Goes, W. Stooss, R. Raue, *Z. Naturforsch. A Phys. Phys. Chem. Kosmophys.* **40 A** (1985) 317.

[129] G. Elliott, A. Radley, *Analyst (London)* **86** (1961) 62.

[130] S. Udenfried: *Fluorescence Assay in Biology and Medicine*, Academic Press, New York, vol. I, 1962, vol. II, 1969.

[131] I. B. Berlman: *Handbook of Fluorescence Spectra of Aromatic Compounds*, Academic Press, New York 1965.

[132] R. Strohecker, H. M. Hemming: *Vitaminbestimmungen*, Verlag Chemie, Weinheim 1963.

[133] F. P. Schwarz, St. P. Wasik, *Anal. Chem.* **48** (1976) 524.

[134] U. Panne, R. Nießner, *Vom Wasser* **79** (1992) 89.

[135] O. S. Wolfbeis (ed.): *Fluorescence Spectroscopy, New Methods and Applications.* Springer Verlag, Berlin 1993.

[136] U. Schobel, C. Barzen, G. Gauglitz: "Immunoanalytical techniques for pesticide monitoring based on fluorescence detection," *Fresenius J. Anal. Chem.* **366** (2000) 646–658.

[137] G. Kortüm: *Reflexionsspektroskopie,* Springer Verlag, Berlin 1969.

[138] H. Wilsing, *Leitz-Mitt. Wiss. Tech.* **1** (1960) 133.

[139] A. Zlatkis, R. E. Kaiser, *J. Chromatogr. Lib.* **9** (1977).

[140] *Accessory Lambda-Series* Perkin-Elmer, Überlingen.

[141] H. Arens: *Farbmetrik,* Akademie Verlag, Berlin 1957.

[142] M. Richter: *Einführung in die Farbmetrik.* De Gruyter, Berlin 1981.

[143] H. Naumer, W. Heller (eds.): *Untersuchungsmethoden in der Chemie,* Thieme-Verlag, Stuttgart 1990, p. 348.

[144] R. Th. Kersten: *Einführung in die optische Nachrichtentechnik,* Springer Verlag, Berlin 1983.

[145] V. Hoffmann et al., *J. Mol. Struct.* **293** (1993) 253.

[146] G. Gauglitz, A. Brecht, G. Kraus, W. Nahm, *Sens. Actuators B* **11** (1993) 21.

[147] A. Brecht, G. Gauglitz, W. Nahm, *Analusis* **20** (1992) 135.

[148] G. Kraus, G. Gauglitz, *Fresenius J. Anal. Chem.* **344** (1992) 153.

[149] G. Gauglitz, J. Krause-Bonte, H. Schlemmer, A. Matthes, *Anal. Chem.* **60** (1988) 2609.

[150] D. E. Aspnes: "The Accurate Determination of Optical Properties by Ellipsometry," in: E. D. Palik (ed.): *Handbook of Optical Constants of Solids,* Academic Press, London 1986, pp. 89 ff.

[151] D. E. Aspnes, A. A. Studna, *Phys. Rev. B Condens. Matter* **27** (1983) 985.

[152] H. Arwin, D. E. Aspnes, *Thin Solid Films* **138** (1986) 195.

[153] G. Kraus, A. Brecht, G. Gauglitz, Ch. Striebel, Lecture held at the Bunsentagung, Leipzig 1993.

[154] H. Raether: Surface Plasmons on Smooth and Rough Surfaces and on Gratings, Springer Tracts in Modern Physics, vol. 111, Springer Verlag, Berlin 1988.

[155] E. Kretchmann, H. Raether, *Z. Naturforsch. A Astrophys. Phys. Phys. Chem.* **23A** (1968) 2135.

[156] A. Otto, *Z. Phys.* **216** (1968) 398.

[157] B. Liedberg, C. Nylander, I. Lundström: "Surface Plasmon Resonance for Gas Detection and Biosensing," *Sens. Actuators* **4** (1983) 299.

[158] P. B. Daniels, J. K. Deacon, M. J. Edowes, D. G. Pedley: "Surface Plasmon Resonance Applied to Immunosensing," *Sens. Actuators* **15** (1988) 11.

[159] J. Homola, S. S. Yee, G. Gauglitz: "Surface plasmon resonance sensors: review," *Sensors & Actuators* **B 54** (1999) 3–15.

[160] W. Lukosz, K. Tiefenthaler, *Opt. Lett.* **8** (1983) 537.

[161] W. Lukosz, Th. Brenner, V. Briguet, Ph. Nellen, P. Zeller, *Proc. SPIE Int. Soc. Opt. Eng.* **1141** (1989) 192.

[162] Ch. Stamm, W. Lukosz, *Sens. Actuators* **B11** (1993) 177.

[163] D. Clerc, W. Lukosz, *Sens. Actuators* **B11** (1993) 177.

[164] W. Böhme, P. Pospisil, K. Horn, H. Morr, *Spectrosc. Int.* **3** (1991) 36.

[165] U. Ruzicka, E. H. Hansen: *Flow Injection Analysis,* John Wiley, New York 1981.

[166] H. Müller, B. Frey, W. Böhme, *Fresenius J. Anal. Chem.* **341** (1991) 647.

[167] W. Göpel et al. (eds.): *Sensors.* VCH Verlagsgesellschaft, Weinheim 1992.

[168] O. S. Wolfbeis, *Fresenius J. Anal. Chem.* **325** (1986) 387.

[169] D. W. Lübbers, N. Opitz, *Z. Naturforsch. C. Biosci.* **C 30** (1975) 532.

[170] G. Gauglitz, M. Reichert, *Sens. Actuators* **B6** (1992) 83.

[171] G. Gauglitz, J. Ingenhoff, *Ber. Bunsen-Ges. Phys. Chem.* **95** (1991) 1558.

[172] M. A. Butler, R. Buss, A. Galuska, *J. Appl. Phys.* **70** (1991) 2326.

[173] W. Nahm, G. Gauglitz, *GIT Fachz. Lab.* **34** (1990) 889.

[174] G. Kraus, A. Brecht, V. Vasic, G. Gauglitz, *Fresenius J. Anal. Chem.* **348** (1994) 598.

[175] H.-M. Haake, A. Schütz, G. Gauglitz, "Label-free detection of biomolecular interaction by optical sensors," *Fresenius J. Anal. Chem.* **366** (2000) 576–585.

[176] C. Mischiati et al.: "Interaction of the human NF-kappaB p52 transcription factor with DNA-PNA hybrids mimicking the NF-kappaB binding sites of the human immunodeficiency virus type 1 promoter," *J. Biol. Chem.* **274** (1999) 33114–33122.

[177] M. Sauer et al.: "Interaction of Chemically Modified Antisense Oligonucleotides with Sense DNA: A Label-Free Interaction Study with Reflectometric Interference Spectroscopy," *Analytical Chemistry* **71** (1999) 2850–2857.

[178] W. Göpel, J. Hesse, J. N. Zemel (eds.): "Optical Sensors," in *Chemical and Biochemical Sensors,* VCH Verlagsgesellschaft, Weinheim 1992.

[179] J. Polster, H. Lachmann: *Spectrometric Titrations.* VCH Verlagsgesellschaft, Weinheim 1989.

[180] R. Blume, H. Lachmann, H. Mauser, F. Schneider, *Z. Naturforsch. B Anorg. Chem. Org. Chem.* **B 29** (1974) 500.

[181] H. Lachmann, *Z. Anal. Chem.* **290** (1978) 117.

[182] G. Gauglitz: "Physical Properties of Photochromic Substances," in H. Dürr, H. Bois-Laurant (eds.): *Photochromic Systems,* Elsevier, Amsterdam 1990.

[183] G. Gauglitz in H. G. Seiler, A. Sigel, H. Sigel (eds.): *Handbook on Metals in Clinical Chemistry,* Marcel Dekker, New York, in press.

[184] H. Mauser, G. Gauglitz: *Principles and Applications of Photokinetics,* Elsevier, Amsterdam, in preparation.

[185] G. Gauglitz, E. Lüddecke, *Z. Anal. Chem.* **280** (1976) 105.

[186] T. C. O'Haver, G. L. Green, *Anal. Chem.* **48** (1976) 312. *Am. Lab. (Fairfield, Conn.)* March (1975) 15.

[187] G. Talsky, L. Mayring, H. Kreuzer, *Angew. Chem.* **90** (1978) 840.

[188] G. L. Green, T. C. O'Haver, *Anal. Chem.* **46** (1974) 2191.

[189] S. Shibata, M. Furukawa, T. Honkawa, *Anal. Chim. Acta* **81** (1976) 206.

[190] *Ullmann,* 4th ed., **5**, 269.

[191] S. Shibata, *Angew. Chem.* **88** (1976) 750.

[192] A. Schmitt: *Labor-Praxis,* vol. 9, Vogel-Verlag, Würzburg 1979.

[193] G. Stella, *Chem. Phys. Lett.* **39** (1976) 176.

[194] H. P. Engelbert, R. Lawaczek, *Chem. Phys. Lipids* **38** (1985) 365.

[195] J. C. Sternberg, *Clin. Chem.* **23** (1977) 1456.
[196] G. Träxler, *Medizintechnik (Stuttgart)* **99** (1979) 79.
[197] K. Oka et al., *Int. Lab.* March (1986) 24.
[198] E. Lüddecke, D. Horn, *Chem. Ing. Tech.* **54** (1982) 266.
[199] F. Dörr: "Polarized Light in Spectroscopy and Photochemistry," in A. A. Lamola (ed.): *Creation and Detection of the Excited State,* vol. 1, Part A, Dekker Inc., New York 1971, pp. 53 ff.
[200] Th. Förster: *Fluoreszenz organischer Verbindungen,* Vandenhoeck G. Ruprecht, Göttingen 1982.
[201] D. V. O'Connor, D. Phillips: *Time-Correlated Single Photon Counting.* Academic Press, New York 1984.
[202] A. E. W. Knight, B. K. Selinger, *Aust. J. Chem.* **26** (1973) 1.
[203] R. B. Somoano, *Angew. Chem.* **90** (1978) 250.
[204] Yok-Han Pao (ed.): *Optoacoustic Spectroscopy and Detection.* Academic Press, New York 1977.
[205] L. B. Kreuzer, *J. Appl. Phys.* **42** (1971) 2934.
[206] L. G. Rosengreen, *Appl. Opt.* **14** (1975) 1960.
[207] C. F. Dewey, R. D. Kamm, C. E. Hackett, *Appl. Phys. Lett.* **23** (1973) 633.
[208] K. Kaya et al., *J. Am. Chem. Soc.* **97** (1975) 2153.
[209] J. F. McClelland, R. N. Knisely, *Appl. Phys. Lett.* **28** (1976) 467 ;*Appl. Opt.* **15** (1976) 2658.
[210] A. Rosencwaig: *Optoacoustic Spectroscopy and Detection.* Academic Press, New York 1977, p. 193.
[211] J. C. Murphy, L. C. Aamodt, *Appl. Phys. Lett.* **31** (1977) 728.
[212] A. Rosencwaig, A. Gersho, *J. Appl. Phys.* **47** (1976) 64.
[213] W. R. Harshberger, M. B. Robin, *Acc. Chem. Res.* **6** (1973) 329.
[214] M. G. Rockley, K. M. Wangle, *Chem. Phys. Lett.* **54** (1978) 597.
[215] A. Rosencwaig, *Science (Washington, D.C.)* **181** (1973) 657.
[216] A. Petzold, R. Niessner, SPIE 1716 (1992) 510.

17. Infrared and Raman Spectroscopy

HANS-ULRICH GREMLICH, Novartis AG, Basel, Switzerland

17.1. Introduction

The roots of infrared spectroscopy go back to the year 1800, when WILLIAM HERSCHEL discovered the infrared region of the electromagnetic spectrum. Since 1905, when WILLIAM W. CO-BLENTZ ran the first infrared spectrum [1], vibrational spectroscopy has become an important analytical tool in research and in technical fields. In the late 1960s, infrared spectrometry was generally believed to be a dying instrumental technique that was being superseded by nuclear magnetic resonance and mass spectrometry for structural determination, and by gas and liquid chromatography for quantitative analysis.

However, the appearance of the first research-grade Fourier transform infrared (FT-IR) spectrometers in the early 1970s initiated a renaissance of infrared spectrometry. After analytical instruments (since the late 1970s) and routine instruments (since the mid 1980s), dedicated instruments are now available at reasonable prices. With its fundamental multiplex or Fellgett's advantage and throughput or Jacquinot's advantage, FT-IR offers a versatility of approach to measurement problems often superior to other techniques. Furthermore, FT-IR is capable of extracting from samples information that is difficult to obtain or even inaccessible for nuclear magnetic resonance and mass spectrometry. Applications of modern FT-IR spectrometry include simple, routine identity and purity examinations (quality control) as well as quantitative analysis, process measurements, the identification of unknown compounds, and the investigation of biological materials.

Raman and infrared spectra give images of molecular vibrations which complement each other; i.e., the combined evaluation of both spectra yields more information about molecular structure than when they are evaluated separately. The 1990s witnessed a tremendous growth in the capabilities and utilization of Raman scattering measurements. The parallel growth of Fourier transform Raman (FT-Raman) with charge-coupled device (CCD) based dispersive instrumentation has provided the spectroscopist with a wide range of choice in instrumentation. There are clear strengths and weaknesses that accompany both types of instrumentation. While FT-Raman systems virtually eliminate fluorescence by using near-IR laser sources, this advantage is paid for with limited sensitivity. On the other hand, flu-orescence interferences always play a role in dispersive measurements with higher sensitivity.

As an alternative to infrared spectroscopy, Raman spectroscopy can be easier to use in some cases; for example, whereas water and glass are strong infrared absorbers they are weak Raman scatterers, so that it is easy to produce a good-quality Raman spectrum of an aqueous sample in a glass container.

General References. Details of vibrational theory are given in the books by HERZBERG [2], [3]; WILSON, DECIUS, and CROSS [4]; STEELE [5]; KONINGSTEIN [6]; and LONG [7]. The technique and applications of FT-IR spectroscopy are discussed in detail by GRIFFITHS and DE HASETH [8], and by MACKENZIE [9]. An introduction to infrared and Raman spectroscopy, dealing with theoretical as well as experimental and interpretational aspects, is provided by COLTHUP, DALY, and WIBERLEY [10] and by SCHRADER [11]. Books about Raman spectroscopy in particular are those of KOHLRAUSCH [12]; BARANSKA, LABUDZINSKA, and TERPINSKI [13]; and HENDRA, JONES, and WARNES [14]. Practical FT-IR and Raman spectroscopy is covered by FERRARO and KRISHNAN [15]; PELLETIER [16]; and GRASSELLI and BULKIN [17]. For the interpretation of IR and Raman spectra of organic compounds, see SOCRATES [18], BELLAMY [19], [20], DOLLISH, FATELEY, and BENTLEY [21], and LIN-VIEN, COLTHUP, FATELEY, and GRASSELLI [22]. Vibrational spectroscopy of inorganic and coordination compounds is described by NAKAMOTO [23]. Biological applications of IR and Raman spectroscopy are covered by GREMLICH and YAN [24]. Various collections of vibrational spectra are available. Both Raman and IR spectra of about 1000 organic compounds are offered as an atlas by SCHRADER [25]. The most comprehensive reference-spectra libraries, digital and hardcopy, can be obtained from SADTLER [26]. A book on modern techniques in applied optical spectroscopy was edited by MIRABELLA [27].

17.2. Techniques

17.2.1. Fourier Transform Infrared Technique

The basis of Fourier transform infrared (FT-IR) spectroscopy is the two-beam interferometer, designed by MICHELSON in 1891 [28], and shown
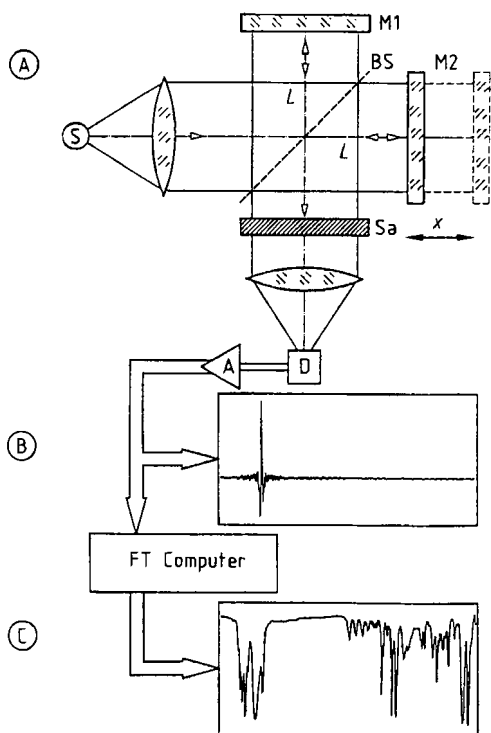
Figure 1. A) Schematic diagram of a Michelson interferometer; B) Signal registered by the detector D, the interferogram; C) Spectrum obtained by Fourier transform (FT) of the interferogram
S = Radiation source; Sa = Sample cell; D = Detector; A = Amplifier; M1 = Fixed mirror; M2 = Movable mirror; BS = Beam splitter; x = Mirror displacement

schematically in Figure 1 A. Broadband infrared radiation is emitted by a thermal source (globar, metal strips, Nernst glower) and falls onto a beam splitter which, in the ideal case, transmits half the radiation and reflects the other half. The reflected half, after traversing a distance L, falls onto a fixed mirror M1. The radiation is reflected by M1 and, after traversing back along distance L, falls onto the beam splitter again. The transmitted radiation follows a similar path and also traverses distance L; however, the mirror M2 of the interferometer can be moved very precisely along the optical axis by an additional distance x. Hence, the total path length of the transmitted radiation is $2(L+x)$. On recombination at the beam splitter the two beams possess an optical path difference of $\Delta = 2x$. Since they are spatially coherent, the two beams interfere on recombination. The beam, modulated by movement of the mirror, leaves the interferometer,

passes through the sample cell and is finally focussed on the detector. The signal registered by the detector, the interferogram, is thus the radiation intensity $I(x)$, measured as a function of the displacement x of the moving mirror M2 from the distance L (Fig. 1 B). The mathematical transformation, a Fourier transform [8], of the interferogram, performed by computer, initially provides a so-called single-beam spectrum. This is compared with a reference spectrum measured without the sample to obtain a spectrum analogous to that measured by conventional dispersive methods. This spectrum is stored digitally in the computer, from which it can be retrieved for further use (Fig. 1 C). In addition to Michelson interferometers, which often have air bearings for the moving mirror and thus require a source of dry compressed air, mechanical interferometer concepts such as frictionless electromagnetic drive mechanisms or refractively scanned interferometers have been developed by instrument manufacturers.

Two major kinds of detector are used in mid-infrared FT-IR spectrometers, the deuterated triglycine sulfate (DTGS) pyroelectric detector and the mercury cadmium telluride (MCT) photodetector. The DTGS type, which operates at room temperature and has a wide frequency range, is the most popular detector for FT-IR instruments. The MCT detector responds more quickly and is more sensitive than the DTGS detector, but it operates at liquid-nitrogen temperature and has both a limited frequency range and a limited dynamic range; therefore, it is used only for special applications [8].

The resolution of an interferometer depends mainly on two factors: firstly, the distance x the moving mirror travels or the maximum optical path difference $\Delta = 2x$ (Fig. 1), and secondly, the apodization function used in computing the spectrum.

The first point is derived from the Rayleigh criterion of resolution, which states that to resolve two spectral lines separated by a distance d, the interferogram must be measured to an optical path length of at least $1/d$ [29].

The second factor is related to the fact that in real life the interferogram is truncated at finite optical path difference. In addition, in the fast Fourier transform (FFT) algorithm, according to COOLEY and TUKEY [30], which is used to perform the Fourier transform faster than the classical method, certain assumptions and simplifications are made. The result is that the FFT of a monochromatic source is not an infinitely narrow line,
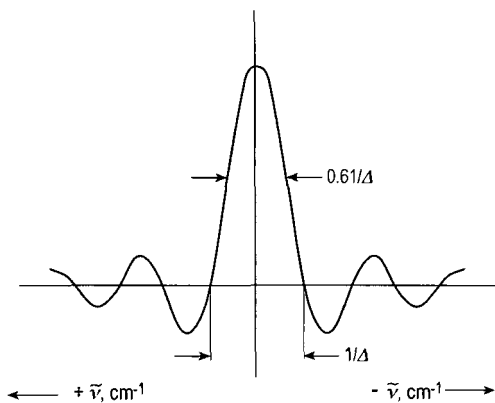
Figure 2. The sinc x function: instrumental line shape function of a perfectly aligned Michelson interferometer with no apodization

$\tilde{v}$ = wavenumber, cm^{-1}; Δ = maximum retardation, cm

corresponding to infinite optical path difference and hence to infinite resolution (see above), but a definable shape known as instrumental line shape (ILS) function [8]. The ILS of a perfectly aligned Michelson interferometer has the shape of a $(\sin x)/x$ or sinc x function (Fig. 2), exhibiting strong side lobes or feet, which may even have negative intensities. To attenuate these spurious feet, so-called apodization functions are used as weighting functions for the interferogram. Numerous such functions exist (e.g., triangular, Happ – Genzel, or Blackman – Harris) which produce instrumental line shapes with lower intensity side lobes, but also with broader main lobes than the sinc function. Hence, side-lobe suppression is only possible at the cost of resolution, because the full width at half-height of the ILS defines the best resolution achievable with a given apodization function [31]. In contrast to dispersive spectroscopy, where the triangular apodization function is determined by the slit of the grating spectrometer, in FT-IR spectroscopy, a choice between highest resolution (no apodization) and best side lobe suppression (Blackman – Harris apodization function) is possible.

In comparison with conventional spectroscopy, the FT-IR method possesses significant advantages. In grating spectrometers, the spectrum is measured directly by registering the radiation intensity as a function of the continuously changing wavelength given by a monochromator [10]. Depending on the spectral resolution chosen, only a very small part (in practice less than 0.1 %) of the radiation present in the monochromator reaches

the IR detector. In the FT-IR spectrometer, all frequencies emitted by the IR source reach the detector simultaneously, which results in considerable time saving and a large signal-to-noise ratio advantage over dispersive instruments; this is the most important advantage, known as the multiplex or Fellgett's advantage [8]. A further advantage stems from the fact that the circular apertures employed in FT-IR spectrometers have larger surface areas than the linear slits of grating spectrometers and hence permit radiation throughputs at least six times greater; this is known as the throughput or Jacquinot's advantage [8]. In FT-IR spectroscopy, the measuring time is the time required to move the mirror M2 the distance necessary to give the desired resolution; since the mirror can be moved very rapidly, complete spectra can be obtained within fractions of a second. By contrast, the measuring time for a conventional spectrum is on the order of minutes. In an FT spectrum, the accuracy of each wavenumber is coupled to the accuracy with which the position of the moving mirror is determined; by using an auxiliary HeNe laser interferometer the position of the mirror can be determined to an accuracy better than 0.005 μm. This means that the wavenumbers of an FT-IR spectrum can be determined with high accuracy (< 0.01 cm^{-1}). In other words, FT-IR spectrometers have a built-in wavenumber calibration of high precision; this advantage is known as accuracy or Connes' advantage [29]. As a result, it is possible to carry out highly precise spectral subtractions and thus very reliably detect slight spectral differences between samples.

17.2.2. Dispersive and Fourier Transform Raman Techniques

In modern dispersive Raman spectrometers, arrays of detector elements such as photodiode arrays or charge-coupled devices (CCDs) are arranged in a polychromator. Here, each element records a different spectral band at the same time, thus making use of the multichannel advantage. In dispersive instruments, the very strong radiation at and near the exciting line, the Rayleigh radiation, may produce stray radiation in the entire spectrum with an intensity much higher than that of the Raman lines. In order to remove this unwanted radiation, line filters in the form of so-called notch filters or holographic filters are used which specifically reflect the Rayleigh line. In conventional, dispersive Raman spectroscopy visible lasers are
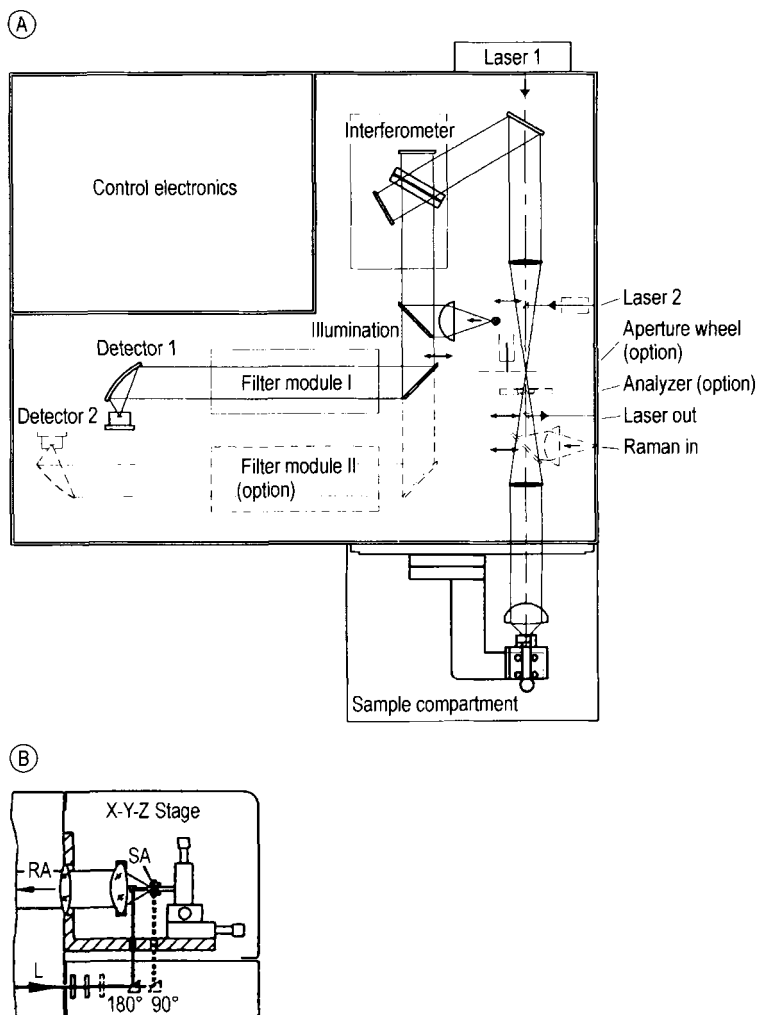
Figure 3. A) Schematic diagram of an FT-Raman spectrometer (Bruker RFS 100); B) Schematic diagram of the sample compartment
L = laser; RA = Raman; SA = Sample
(Reproduced by permission of Bruker Optik GmbH, D-76275 Ettlingen, Germany)

usually used to excite the Raman effect (see Section 17.3.1), while in FT-Raman spectroscopy near-IR lasers are employed. Excitation with visible lasers often produces strongly interfering fluorescence originating from the sample or impurities; however, this is eliminated by using a neodymium yttrium aluminum garnet (Nd : YAG) laser providing a discrete line excitation at 1.064 μm, as at this wavelength in the near-IR region no electronic transitions are induced.

An FT-Raman spectrometer is shown in schematic form in Figure 3 A. The beam of the laser

source passes to the sample compartment where it is focused, typically to a spot of 100 μm diameter, onto the sample, either in 90° or 180° geometry (Fig. 3 B). The scattered Raman light is collected by an aspherical lens and passed through a filter module to remove the Rayleigh line and Rayleigh wings [32]. It is then directed through the infrared input port into the interferometer which is optimized for the near-IR region as is the liquid-nitrogen-cooled germanium or room-temperature indium gallium arsenide (InGaAs) detector [33]. The procedure to obtain the FT-Raman spectrum from

the interferogram, the signal registered by the detector, is essentially the same as that described above for an FT-IR spectrum.

Before FT-Raman spectrometers could be successfully used several difficulties had to be overcome. As the intensity of Raman scattering is proportional to the fourth power of the frequency of excitation a loss in sensitivity by a factor between 20 and 90 must be compensated for when the exciting radiation is shifted from the visible to the near-IR region [32]. Moreover, the noise equivalent power (NEP) of near-IR detectors is greater by about one to two orders of magnitude than that of a photomultiplier tube, the usual detector for conventional Raman spectroscopy. However, both disadvantages are compensated for by the throughput (Jacquinot) and multiplex (Fellgett) advantages of Fourier transform spectroscopy (see Section 17.2.1). Another major problem is the fact that Raman scattering is always accompanied by very strong radiation, the Rayleigh line and Rayleigh wings, occurring at and near the frequency of excitation with six to ten orders of magnitude higher intensity than the Raman signal. The shot noise of the Rayleigh line and Rayleigh wings produces a background noise which would completely mask all Raman lines. Therefore, an effective filter that removes the Rayleigh line and Rayleigh wings in the range $750 - 100$ cm^{-1} is one of the most important components of a powerful FT-Raman instrument [32]. In addition, optimization of the sampling technique is indispensable in order to convert the flux of exciting radiation into a maximum flux of Raman radiation to be received by the optical system of the spectrometer [34]. The 180° backscattering arrangement (Fig. 3 B), rather than the 90° arrangement used in visible, dispersive Raman technique, is undoubtedly the most efficient way of collecting the Raman signal and provides the best quality data in the near-IR FT-Raman experiment.

17.3. Basic Principles of Vibrational Spectroscopy

17.3.1. The Vibrational Spectrum

To explain the origins of a vibrational spectrum the vibration of a diatomic molecule is considered first. This can be illustrated by a molecular model in which the atomic nuclei are represented by two point masses m_1 and m_2, and the inter-

atomic bond by a massless spring. According to Hooke's law, the vibrational frequency v (in s^{-1}) determined by classical methods is given by

$$v = \frac{1}{2\pi}\sqrt{\frac{f}{\mu}} \qquad (1)$$

where f is the force constant of the spring in N/m, and μ is the reduced mass in kg:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \qquad (2)$$

Thus, the vibrational frequency is higher when the force constant is higher, i.e., when the bonding of the two atoms is stronger. Conversely, the heavier the vibrating masses are, the smaller is the frequency v. The frequency of fundamental vibrations is on the order of magnitude of 10^{13} s^{-1}. Quantum-mechanical methods give the following expression for the vibrational energy of a Hooke's oscillator in a state characterized by the vibrational quantum number v:

$$E_{\text{vib}} = \frac{h}{2\pi}\sqrt{\frac{f}{\mu}}\left(v + \frac{1}{2}\right) = hv_0\left(v + \frac{1}{2}\right) \qquad (3)$$

where f and μ have the same definition as in Equation (1) and v_0 is the vibrational frequency of the ground state. This relation is valid with good accuracy for vibrational transitions from the ground state ($v = 0$) to the first excited state ($v = 1$). In free molecules (gas state), vibrational transitions are always accompanied by rotational transitions. The corresponding energies are given by:

$$E_{\text{vib,rot}} = hv_0\left(v + \frac{1}{2}\right) + BhJ(J + 1) \qquad (4)$$

where J is the rotational quantum number and B is the rotational constant:

$$B = \frac{h}{8\pi^2 I} \qquad (5)$$

where I is the moment of inertia. In the real world, molecular oscillations are inharmonic. Therefore, the transition energy decreases with increasing vibrational quantum number until the molecule finally dissociates. The quantum mechanical energy of a diatomic inharmonic oscillator is given in good approximation by:

Table 1. Infrared and microwave regions of the electromagnetic spectrum

Region	Wavelength λ, μm	Wavenumber $\tilde{v}$, cm^{-1}	Frequency v, s^{-1}
Infrared			
near	7.8×10^{-1} to 2.5	12 800 to 4000	3.8×10^{14} to 1.2×10^{14}
mid	2.5 to 5.0×10^{1}	4000 to 200	1.2×10^{14} to 6.0×10^{12}
far	5.0×10^{1} to 1.0×10^{3}	200 to 10	6.0×10^{12} to 3.0×10^{11}
Microwave	1.0×10^{3} to 1.0×10^{6}	10 to 0.01	1.0×10^{12} to 3.0×10^{8}

$$E_{\text{vib}} = hv_0\left[\left(v + \frac{1}{2}\right) - x_0\left(v + \frac{1}{2}\right)^2\right] \qquad (6)$$

where x_0 is the inharmonicity constant.

A complex molecule is considered as a system of coupled inharmonic oscillators. If there are N atomic nuclei in the molecule, there will be a total of $3N$ degrees of freedom of motion for all the nuclear masses in the molecule. Subtracting the pure translations and rotations of the entire molecule leaves $3N-6$ internal degrees of freedom for a nonlinear molecule and $3N-5$ internal degrees of freedom for a linear molecule. The internal degrees of freedom correspond to the number of independent normal modes of vibration; their forms and frequencies must be calculated mathematically [10]. In each normal mode of vibration all the atoms of the molecule vibrate with the same frequency and pass through their equilibrium positions simultaneously. The true internal motions of the molecule, which constitute its vibrational spectrum, are composed of the normal vibrations as a coupled system of these independent inharmonic oscillators. Thus, a molecule is unambiguously characterized by its vibrational spectrum. Real-world molecules are best described by the model of inharmonic oscillators because here transitions are allowed which are forbidden to the harmonic oscillator: transitions from the ground state ($v=0$) to states with $v=2$, 3, ... (overtones), transitions from an excited level of a vibration (difference bands), and transitions between states belonging to different types of normal mode (combination bands).

In vibrational spectroscopy, instead of the frequency v (in s^{-1}) the so-called wavenumber $\tilde{v}$ expressed in cm^{-1} (reciprocal centimeters) is mostly used; this signifies the number of waves in a 1 cm wavetrain:

$$\tilde{v} = \frac{v}{(c/n)} = \frac{1}{\lambda} \qquad (7)$$

where c is the velocity of light in a vacuum (2.997925×10^{10} cm/s), c/n is the velocity of light in a medium with refractive index n in which the wavenumber is measured (the refractive index of air is 1.0003), and λ is the wavelength in centimeters. In the infrared region of the electromagnetic spectrum the practical unit of wavelength is 10^{-4} cm or 10^{-6} m (i.e., μm), and wavenumber and wavelength are related as follows:

$$\tilde{v}/\text{cm}^{-1} = \frac{10^4}{\lambda/\mu\text{m}} \qquad (8)$$

In infrared spectroscopy, both wavenumber and wavelength are used, whereas in Raman spectroscopy only wavenumber is employed.

The infrared and microwave regions of the electromagnetic spectrum are shown in Table 1. In the near-infrared region, overtones and combination bands are observed. Since these transitions are only possible because of inharmonicity, their probability is diminished. A good rule of thumb is that the intensity of the first overtone is about one-tenth that of the corresponding fundamental vibration. The fundamentals occur in the mid-infrared range, which is the most important infrared region. In the far-infrared region, the fundamentals of heavy, single-bonded atoms and the absorptions of inorganic coordination compounds [23] are found.

Infrared spectroscopy is based on the interaction of an oscillating electromagnetic field with a molecule, and it is only possible if the dipole moment of the molecule changes as a result of a molecular vibration. While the absorption frequency depends on the molecular vibrational frequency, the absorption intensity depends on how effectively the infrared photon energy is transferred to the molecule. It can be shown [3] that the intensity of infrared absorption is proportional to the square of the change in the dipole moment μ, with respect to the change in the normal coordinate q describing the corresponding molecular vibration:
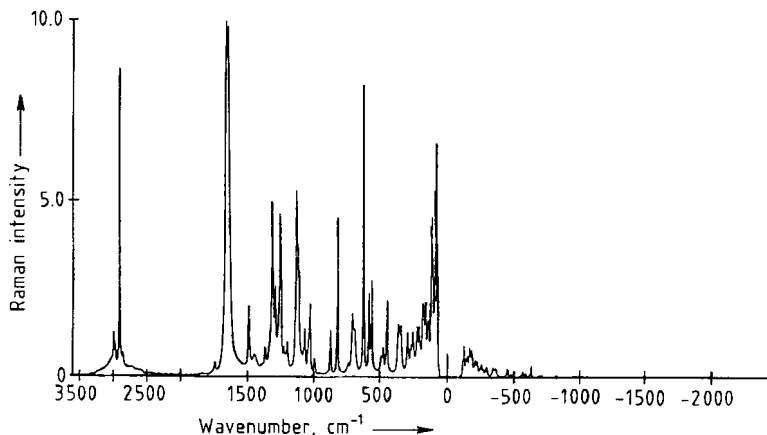
Figure 4. FT-Raman spectrum of ascorbic acid recorded on a Bruker FT-Raman accessory FRA 106 directly interfaced to a Bruker FT-IR spectrometer IFS 66
The accessory was equipped with a germanium diode cooled by liquid nitrogen. Resolution was 2 cm^{-1} and laser output power was 300 mW. Stokes shift: 3600 to 50 cm^{-1}, anti-Stokes shift: -100 to -2000 cm^{-1}.

$$I_{IR} \sim \left(\frac{\partial \mu}{\partial q}\right)^2 \qquad (9)$$

A second way to induce molecular vibrations is to irradiate a sample with an intense source of monochromatic radiation, usually in the visible or near-infrared region, this leads to the Raman effect, which can be regarded as an inelastic collision between the incident photon and the molecule. As a result of the collision, the vibrational or rotational energy of the molecule is changed by an amount ΔE_m. For energy to be conserved, the energy of the scattered photon $h\nu_s$ must differ from that of the incident photon $h\nu_i$ by an amount equal to ΔE_m:

$$h\nu_i - h\nu_s = \Delta E_m = h\nu_m \qquad (10)$$

If the molecule gains energy, then ΔE_m is positive and ν_s is smaller than ν_i, giving rise to so-called Stokes lines in the Raman spectrum. If the molecule loses energy, ΔE_m is negative and ν_s is larger than ν_i, giving rise to so-called anti-Stokes lines in the Raman spectrum. The dipole moment μ induced by the Raman effect can be related to the electric field E of the incident electromagnetic radiation as follows:

$$\mu = \alpha E \qquad (11)$$

where α is the polarizability of the molecule. It can be derived from classical Raman theory [3] that for a molecular vibration to be Raman active it must

be accompanied by a change in the polarizability of the molecule. The intensity of Raman absorption is proportional to the square of the change in polarizability α, with respect to the corresponding normal coordinate q:

$$I_{RA} \sim \left(\frac{\partial \alpha}{\partial q}\right)^2 \qquad (12)$$

Although it follows from classical theory that Stokes and anti-Stokes lines should have the same intensity, according to quantum theory and in agreement with experiment (Fig. 4) anti-Stokes lines are a much less intense since the number of molecules in the initial state $v = 1$ of anti-Stokes lines is only $e^{-(h\nu_m/kT)}$ times the number of molecules in the initial state $v = 0$ of the Stokes lines [3]. The Raman shift in cm^{-1}, i.e., the difference $\Delta\tilde{\nu}$ between the wavenumber $\tilde{\nu}_e$ of the exciting laser and the wavenumber $\tilde{\nu}_s$ of the scattered Raman light, is indicated on the abscissa of each Raman spectrum.

17.3.2. Quantitative Analysis

Quantitative analysis is well established not only in ultraviolet/visible and in near-infrared spectroscopy, but it is also very important in mid-infrared measurements. The general prerequisite for spectrometric quantitative analysis is defined as follows [35]: information derived from the spectrum of a sample is related in mathematical terms

to changes in the level(s) of an individual component, or several components within the sample or series of samples, i.e., the spectral response of an analyte can be related by a mathematical function to changes in concentration of the analyte. In the ideal situation, the measured spectroscopic feature varies linearly with concentration. In reality, however, true linearity is not always obtained, but this is not important provided the measured function is reproducible. Most practical analyses are not absolute measurements, and normally measurements are made on a given instrument within a fixed working environment; under such set circumstances, reproducibility and consistency of the measurement are the most important factors.

If I_0 is the intensity of monochromatic radiation entering a sample and I is the intensity transmitted by the sample, then the ratio I/I_0 is the transmittance T of the sample. The percent transmittance ($\%T$) is $100\,T$. If the sample cell has thickness b and the absorbing component has concentration c, then the fundamental relation governing the absorption of radiation as a function of transmittance is:

$$T = \frac{I}{I_0} = 10^{-abc} \tag{13}$$

The constant a is called the absorptivity and is characteristic of a specific sample at a specific wavelength. As the transmittance T does not vary linearly with the concentration of an absorbing species the equation above is usually transformed by taking the logarithms of both sides of the equation and replacing I/I_0 with I_0/I to eliminate the negative sign:

$$\log \frac{I_0}{I} = abc \tag{14}$$

The term $\log(I_0/I)$ is the absorbance A, which changes linearly with changes in concentration of an absorbing species:

$$A = abc \tag{15}$$

This fundamental equation for spectrometric quantitative analysis is known as Beer – Lambert – Bouguer law, sometimes shortened to Beer's law or Beer – Lambert law.

In practice, several steps are involved in the development of a quantitative method [35]. The first and probably most crucial for the ultimate success of the analytical method is to understand the system, i.e., to obtain reference spectra of the analyte and all other components. In the second step, the best method of sampling should be determined. Then, with standards having been prepared, the system must be calibrated. The last step before analyzing samples is to prepare validation samples and to evaluate the method.

With modern sampling techniques, good quantitative infrared analysis with virtually every type of sample is practicable; however, liquids are ideal for this purpose, being measured in a liquid cell of fixed thickness, either as 100 % sample or diluted with solvent. In this connection it should be taken into account that there are no ideal solvents for infrared spectroscopy [35]. In addition, because absorption bands and path length can be influenced by the temperature of the transmission cell, it is advisable to control the temperature.

Polymers are usually analyzed as pressed films, and solids are often difficult to examine quantitatively by infrared methods; in these cases, absorption band ratios [10] give the best results. In fact, the main application of multicomponent quantitative infrared analysis is gas analysis. If the difficulty in handling the gases is overcome, multicomponent analysis can be readily performed.

A simple chemical system can consist of a pure single component, or of a single component or more than one component in a mixture with no spectral interference; it is assumed that the radiation absorption by one component is not affected by the presence of other components. In simple systems, absorbance peak height measurements, directly or by using a selected baseline [35], are often employed for calibration and analysis. Because of intrinsic instrumental errors the practical limit for usable absorbance values is about three. Peak height measurements are also sensitive to changes in instrumental resolution and can vary considerably from instrument to instrument. To circumvent these problems, an alternative method is the use of integrated absorbance or peak area [10].

Complex chemical systems are composed of one or more components in a mixture with a significant degree of spectral interference, or of several components with a large amount of mutual physical and/or chemical interaction. In these cases, quantitative analysis is best performed by statistical methods such as principal component regression (PCR) or partial least squares (PLS) [36]; these are offered in the software packages of instrument manufacturers and software suppliers. Artificial neural networks (ANNs) should be primarily used when a data set is nonlinear [37].
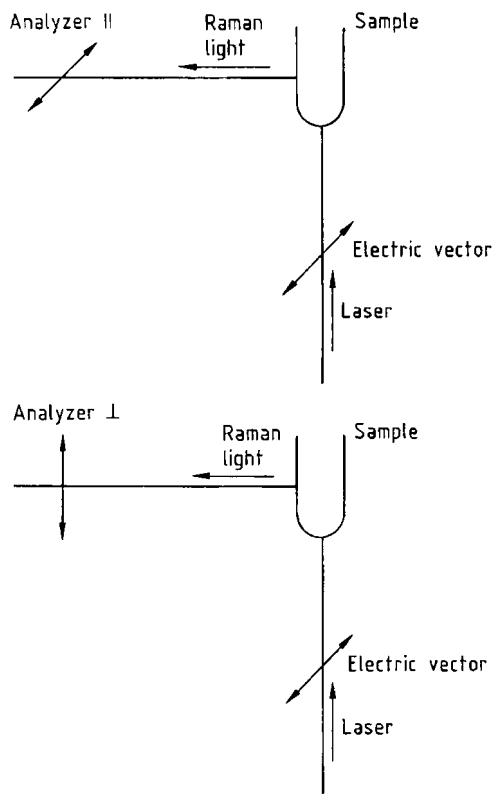
Figure 5. Orientation of the analyzer in relation to the direction of the electrical field of the exciting laser

17.3.3. The Symmetry of Molecules and Molecular Vibrations; Selection Rules

All molecules show symmetry properties and they all possess at least one (trivial) symmetry element, the identity. The symmetry of a molecule is important in spectroscopy because changes in symmetry during molecular vibration determine whether a vibrational dipole moment μ occurs or not. As mentioned, a vibration is infrared active if μ changes, if not, it is infrared inactive; analogously, this is true of the polarizability α and Raman activity.

In infrared and Raman spectroscopy, the symmetry of molecules is usually discussed in terms of five symmetry elements and their corresponding five symmetry operations [38]. If a wide variety of molecules is investigated, it will be found, as can be proved by mathematical group theory [39], that only certain combinations of symmetry elements are possible. Such restricted combinations of symmetry elements, known as point groups, are used

for the classification of molecules. Each point group is associated with a set of normal vibrations. These in turn are classified according to the symmetry of vibration. From this it is possible to predict whether a normal vibration is infrared or Raman active or neither. These are the so-called selection rules. Many vibrations of molecules with low symmetry are both infrared and Raman active, and it is chiefly in the band intensities that the two types of spectra differ, sometimes markedly so. On the other hand, when a center of symmetry is present, i.e., for molecules with a high degree of symmetry, bands that are infrared active are Raman inactive and vice versa. By using laser excitation (i.e., linearly polarized light) (see Section 17.2.2), Raman spectroscopy also provides a means of recognizing totally symmetrical vibrations [3]. The intensity of a Raman line depends on the direction of the exciting electric field in relation to the orientation of the analyzer (Fig. 5). The latter can be either perpendicular or parallel to the direction of the electric field. The depolarization ratio ϱ is the ratio between the intensities of scattered light measured at each of these two positions:

$$\varrho = \frac{I_\perp}{I_\parallel} \tag{16}$$

The depolarization ratio of a Raman line depends on the symmetry of the molecular vibration involved (i.e., the change in molecular symmetry induced by the corresponding vibration); the maximum value of depolarization observed with linearly polarized light is $\varrho_l^{max} = 3/4$. If a Raman line shows this extent of depolarization, it is said to be depolarized, whereas, if the degree of depolarization is smaller, the line is polarized. It can be shown [3] that only Raman lines corresponding to totally symmetric vibrations can have a degree of depolarization smaller than the maximum value, that is, can be polarized.

17.4. Interpretation of Infrared and Raman Spectra of Organic Compounds

17.4.1. The Concept of Group Frequencies

A complex molecule can be considered as a system of coupled inharmonic oscillators. Empir-

Table 2. Commonly used symbols and descriptions or different vibrational forms

Symbol	Designation	Example	Illustration
s	symmetric	ν_s	
as	antisymmetric (asymmetric)	ν_{as}	
ip	in-plane (ip)	δ	
oop	out-of-plane (oop)	γ	
v	stretching		
δ	ip deformation		
γ	oop deformation		
ω	wagging		
τ	twisting		
ϱ	rocking		

Table 3. Usual abbreviations for the classification of vibrational absorption bands

Abbreviation	Signification
vs	very strong
s	strong
m	medium
w	weak
vw	very weak
sh	shoulder
b	broad
sr	sharp
v	variable

ically it is found that vibrational coupling is restricted to certain submolecular groups of atoms. This coupling is relatively constant from molecule to molecule, so that the submolecular groups produce bands in a characteristic frequency region of the vibrational spectrum. These bands — the characteristic group frequencies — are readily predictable and so form the empirical basis for the interpretation of vibrational spectra. Their position, intensity, and width are decisive for the correlation of an absorption band with a certain submolecular group. For example, the vibrational spectra of n-heptane, n-octane, and n-nonane have a number of bands in common; these are the group frequencies for normal alkanes. This concept of group frequencies corresponds to the concept of chemical shift in nuclear magnetic resonance spectroscopy. These spectra also have a number of bands which are not in common; these so-called fingerprint bands are characteristic of the individual chemical compound and are used to distinguish one compound from another. In Table 2, commonly used symbols and descriptions of different vibrational forms are given, and Table 3 lists usual abbreviations for the classification of vibrational absorption bands. For practical applications, spectral interpretation is mainly based on personal experience.

In the following sections, dealing with spectrum – structure correlations, characteristic group frequencies having diagnostic value are given in the form of charts. These correlation charts show the positions of the characteristic group frequencies which are the same in infrared and Raman spectra. The two types of spectra mainly differ in the band intensities, which, however, are not indicated in the correlation charts. To provide this information, examples of infrared and Raman spectra of most of the functional groups discussed are shown. Fourier transform Raman spectra (Fig. 4) are usually recorded between 3600 and 50 cm^{-1} (Stokes shift) and -100 to -2000 cm^{-1} (anti-Stokes shift). For comparison with mid-infrared spectra, in the following sections the Fourier transform Raman spectra are presented in the range between 4000 and 400 cm^{-1}. The relative intensities of vibrational absorption bands, which are evident in the model spectra presented, are important for appropriate spectral interpretation. Specific differences between infrared and Raman spectra are given special mention.

In the figures in this chapter (Figs. 8 – 44), the upper curve is the infrared spectrum, with intensity increasing from the bottom to the top of the diagram. The lower curve is the Raman spectrum with an ordinate linear in relative intensity units increasing from the bottom to the top of the diagram. The Raman spectra have not been corrected for fluctuations in the sensitivity of the spectrometer. The infrared spectra were recorded with a Bruker IFS 66 FT-IR spectrometer; for recording of the Raman spectra a Bruker FRA 106 FT-Raman accessory was used.

17.4.2. Methyl and Methylene Groups

Figure 6 shows characteristic group frequencies, and Figure 7 illustrates typical vibrations of
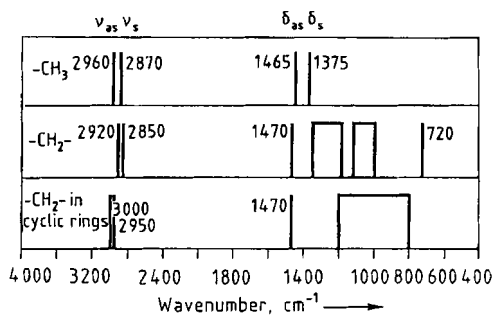
Figure 6. Characteristic group frequencies of alkanes

CH₃ stretching vibrations   CH₃ deformation vibrations

v_{as} 2960 cm^{-1} v_s 2870 cm^{-1} δ_{as} 1465 cm^{-1} δ_s 1375 cm^{-1}

CH₂ stretching vibrations   CH₂ deformation vibration

v_{as} 2920 cm^{-1} v_s 2850 cm^{-1} δ 1470 cm^{-1}

CH₂ wag, twist, and rock vibrations

Wagging, ω,   Twisting, τ,   Rocking, ρ,
1350–1180 cm^{-1} 1300 cm^{-1} 720 cm^{-1}
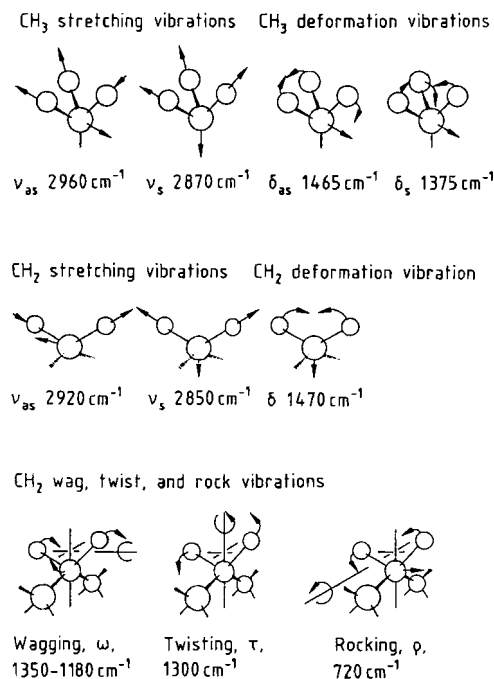
Figure 7. Typical vibrations of alkanes

alkanes. In Figure 8, infrared and Raman spectra of 2-methylbutane (isopentane) are given as an example of alkane spectra. The most intense bands are those of the antisymmetric and symmetric CH₃ or CH₂ stretching vibrations between 2960 and 2850 cm^{-1}. Typical of isopropyl and tertiary butyl groups is that the symmetric CH₃ deformation band, usually observed at 1375 cm^{-1}, is split into two bands near 1390 and 1370 cm^{-1}. In the Raman spectra, the symmetric CH₃ deformation band near 1375 cm^{-1} is generally relatively weak in alkanes, but when the CH₃ group is next to a double or
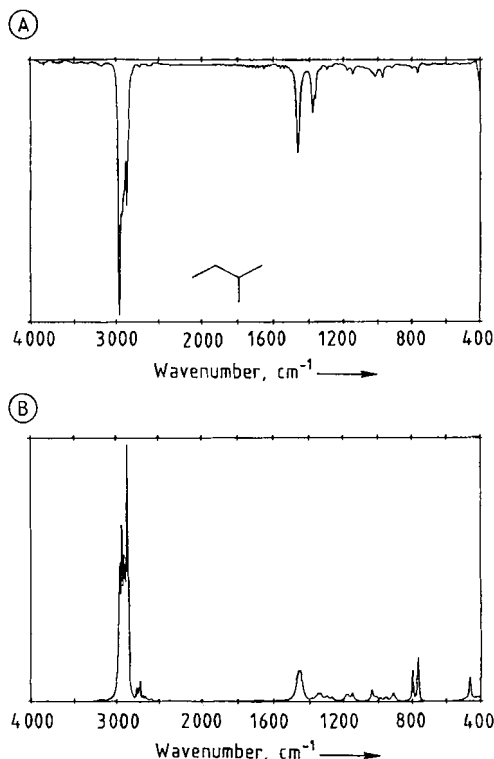
Figure 8. FT-infrared (A) and FT-Raman (B) spectra of 2-methylbutane

triple bond or an aromatic ring the Raman intensity is noticeably enhanced [10]. This is demonstrated in Figure 9, which shows the spectra of 2-methyl-2-butene.

The CH₂ wagging bands are spread over a region between 1350 and 1180 cm^{-1}, occurring as a characteristic progression of weak bands. They are best seen in the solid-phase spectra of long straight-chain compounds such as fatty acids [10].

The CH₂ twisting vibrations in CH₂ chains have frequencies dispersed over the same region as the wagging bands, as can be seen in the spectra of polyethylene [25], for example. The infrared intensity of these bands is weak, whereas in the Raman spectrum at about 1300 cm^{-1} the in-phase CH₂ twist vibration is a useful group frequency.

The C–C stretching vibration of CH₂ chains is seen only in the Raman spectrum, at 1131 and 1061 cm^{-1} [40], while the CH₂ rocking vibration at 720 cm^{-1}, which is characteristic for longer (CH₂)$_n$ chains with $n \geq 4$, is observed only in the infrared spectrum (e.g., polyethylene) [25].
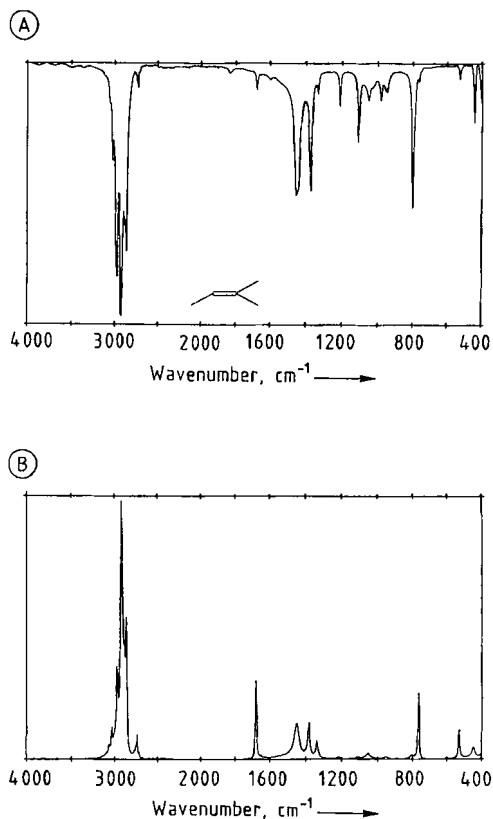
Figure 9. FT-infrared (A) and FT-Raman (B) spectra of 2-methyl-2-butene

Aliphatic ring compounds are best characterized by their Raman spectra, since in the infrared only very weak characteristic ring frequencies due to C–C ring stretching vibrations are observed near 1000 cm^{-1} (Fig. 10). Raman spectra, in contrast, show prominent lines due to ring stretching vibrations: cyclopropane 1188 cm^{-1} [25], cyclobutane 1001 cm^{-1} [40], cyclopentane 886 cm^{-1} [25], and cyclohexane 802 cm^{-1} (Fig. 10).

17.4.3. Alkene Groups

Characteristic group frequencies of alkenes are shown in Figure 11. In alkenes, the =CH stretching vibration generally occurs above 3000 cm^{-1}. In symmetrical *trans* or symmetrical tetrasubstituted double-bond compounds, the C=C stretching frequency near 1640 cm^{-1}, usually a medium intensity band, is infrared inactive because, in this

case, no change of dipole moment occurs as a result of the vibration. However, the C=C stretching vibration gives rise to a strong Raman signal in the region between 1680 and 1630 cm^{-1} in all types of alkenes (Fig. 12).

In vinyl and vinylidine groups, the =CH$_2$ in-plane deformation gives rise to a medium intensity band in both the infrared and Raman spectrum near 1415 cm^{-1}. In the case of the vinyl group, an additional infrared and Raman deformation band is observed near 1300 cm^{-1} [10]. In *cis*-1,2-dialkyl ethylenes, the in-plane deformation band appears near 1405 cm^{-1} in the infrared and at ca. 1265 cm^{-1} in the Raman spectrum. In the corresponding *trans* compounds, the in-plane bending vibrations occur at 1305 cm^{-1} in the Raman and at 1295 cm^{-1} in the infrared spectrum. In these *trans* compounds, the =CH out-of-plane bending vibration gives rise to a strong infrared band, of high diagnostic value, occurring between 980 and 965 cm^{-1}.

17.4.4. Aromatic Rings

Figure 13 shows characteristic group frequencies of aromatic rings. The bands between 1600 and 1460 cm^{-1} are ring modes involving C–C partial double bonds of the aromatic ring. The band at 1500 cm^{-1} is usually the strongest of these bands, which generally show medium to strong infrared but weak Raman intensities. The use of both infrared and Raman spectra gives reliable information about the type of substitution of an aromatic ring: in infrared spectra, intense bands in the 850–675 cm^{-1} region, due to out-of-plane CH wagging and out-of-plane sextant ring bending vibrations, are indicative of the type of substitution [10]. In the Raman spectra, there are very strong lines at ca. 1000 cm^{-1} involving ring stretching and ring bending vibrations. As shown in Figures 14–17, a very strong signal at 1000 cm^{-1} occurs for mono and *meta* substitution, a strong line is observed between 1055 and 1015 cm^{-1} for *ortho* substitution, while no signal is observed around 1000 cm^{-1} for *para* substitution. In infrared spectra, the out-of-plane CH wagging vibrations give rise to relatively prominent summation bands in the 2000–1650 cm^{-1} region. As the summation band patterns are approximately constant for a particular ring substitution, they can also be used to determine the type of substitution [10].
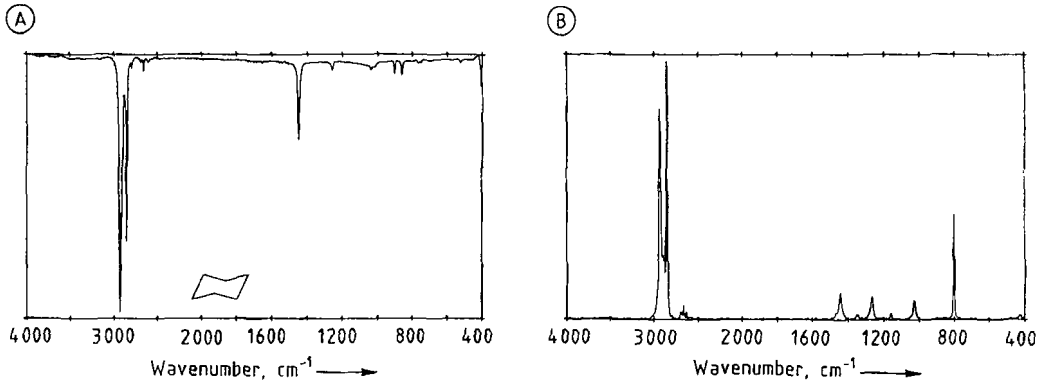
Figure 10. FT-infrared (A) and FT-Raman (B) spectra of cyclohexane

Figure 12. FT-infrared (A) and FT-Raman (B) spectra of 2,3-dimethyl-2-butene
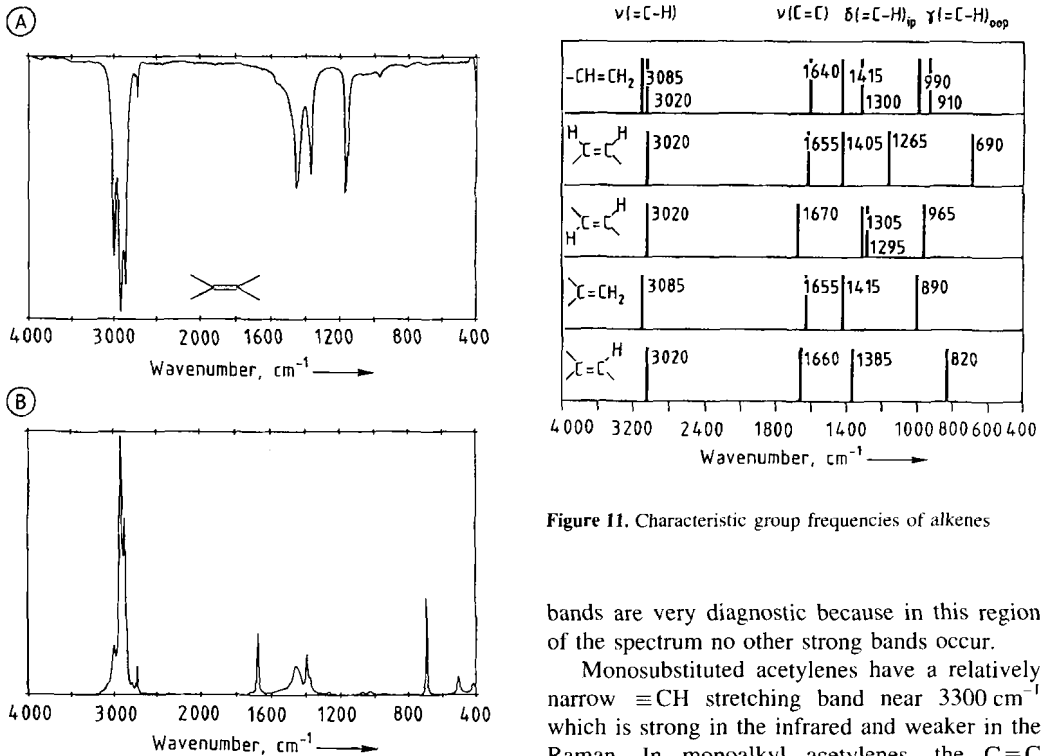
Figure 11. Characteristic group frequencies of alkenes

17.4.5. Triple Bonds and Cumulated Double Bonds

Triple bonds $(X \equiv Y)$ and cumulated double bonds $(X=Y=Z)$ absorb roughly in the same region, namely, $2300-2000$ cm^{-1} (Table 4). The

bands are very diagnostic because in this region of the spectrum no other strong bands occur.

Monosubstituted acetylenes have a relatively narrow $\equiv$CH stretching band near 3300 cm^{-1} which is strong in the infrared and weaker in the Raman. In monoalkyl acetylenes, the $C \equiv C$ stretching vibration shows a weak infrared absorption near 2120 cm^{-1}, while no infrared $C \equiv C$ absorption is observed in symmetrically substituted acetylenes because of the center of symmetry. In Raman spectra, this vibration always occurs as a strong line. Nitriles are characterized by a strong $C \equiv N$ stretching frequency at $2260-2200$ cm^{-1} in the infrared and in the Raman (Fig. 18). Upon electronegative substitution at the α-carbon atom, the infrared intensity of the $C \equiv N$ stretching vi-

Figure 13. Characteristic group frequencies of aromatic rings

bration is considerably weakened so that only the Raman spectrum, where a strong signal is always observed, can be used for identification (Fig. 19).

17.4.6. Ethers, Alcohols, and Phenols

17.4.6.1. Ethers

Characteristic group frequencies of ethers are shown in Figure 20. The key band for noncyclic ethers is the strong antisymmetric C–O–C stretching vibration between 1140 and 1085 cm^{-1} for aliphatic, between 1275 and 1200 cm^{-1} for aromatic, and between 1225 and 1200 cm^{-1} for vinyl ethers. With the exception of aliphatic – aromatic ethers, the symmetric C–O–C stretching band at ca. 1050 cm^{-1} is too weak to have diagnostic value. In cyclic ethers, both the antisymmetric (1180 – 800 cm^{-1}) and the symmetric (1270 – 800 cm^{-1}) C–O–C stretching bands are characteristic. The antisymmetric C–O–C stretch is usually a strong infrared band and a weak Raman signal, the symmetric C–O–C stretching vibration generally being a strong Raman line and a weaker infrared band (Fig. 21).

17.4.6.2. Alcohols and Phenols

Characteristic group frequencies of alcohols and phenols are shown in Figure 22. Because of the strongly polarized OH group, alcohols generally show a weak Raman effect (and therefore are excellent solvents for Raman spectroscopy). In the liquid or solid state, alcohols and phenols exhibit strong infrared bands due to O–H stretching vibrations. Due to hydrogen bonding of the OH groups, these bands are very broad in the pure liquid or solid, as well as in concentrated solutions or mixtures. The position of the similarly intense C–O stretching band permits primary, secondary, and tertiary alcohols, as well as phenols to be distinguished between.
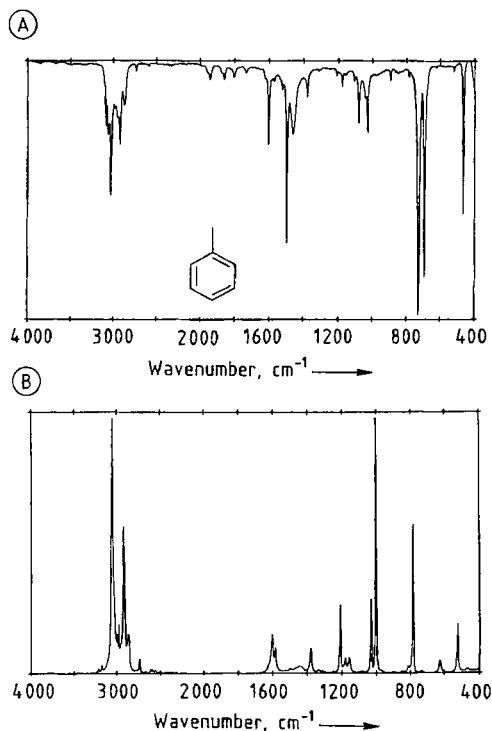
Figure 14. FT-infrared (A) and FT-Raman (B) spectra of toluene

17.4.7. Amines, Azo, and Nitro Compounds

17.4.7.1. Amines

Characteristic group frequencies of amines are shown in Figure 23. In the infrared and Raman, key bands are the NH stretching bands, although they are often not very intense. Primary amines exhibit two bands due to antisymmetric (3550 – 3330 cm^{-1}) and symmetric (3450 – 3250 cm^{-1}) stretching, while for secondary amines only one NH stretching band is observed. The intensities of NH and NH$_2$ stretching bands are generally weaker in aliphatic than in aromatic amines. The C–N stretching vibration is not necessarily diagnostic, because in aliphatic amines it gives rise to absorption bands of only weak to medium intensity between 1240 and 1000 cm^{-1}. In aromatic amines, strong C–N stretching bands are observed in the range 1380 – 1250 cm^{-1}.
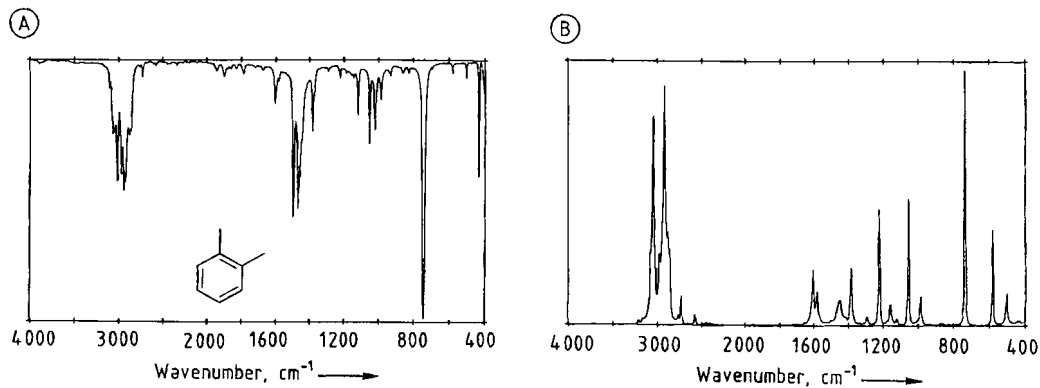
Figure 15. FT-infrared (A) and FT-Raman (B) spectra of *o*-xylene
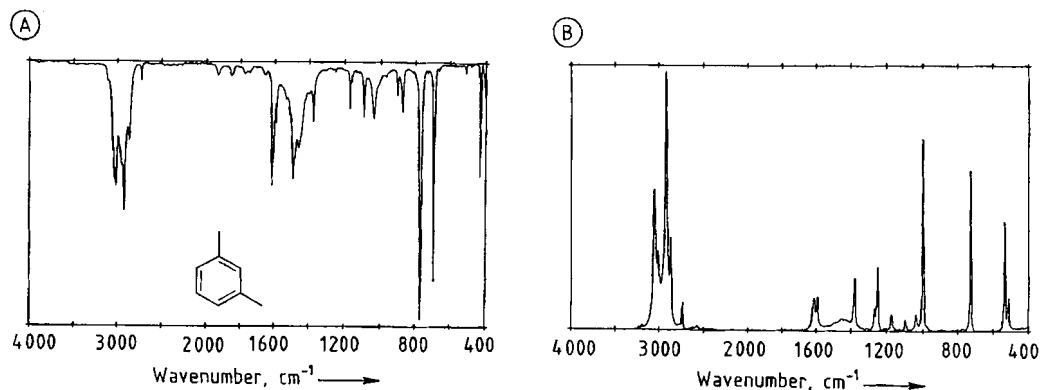
Figure 16. FT-infrared (A) and FT-Raman (B) spectra of *m*-xylene
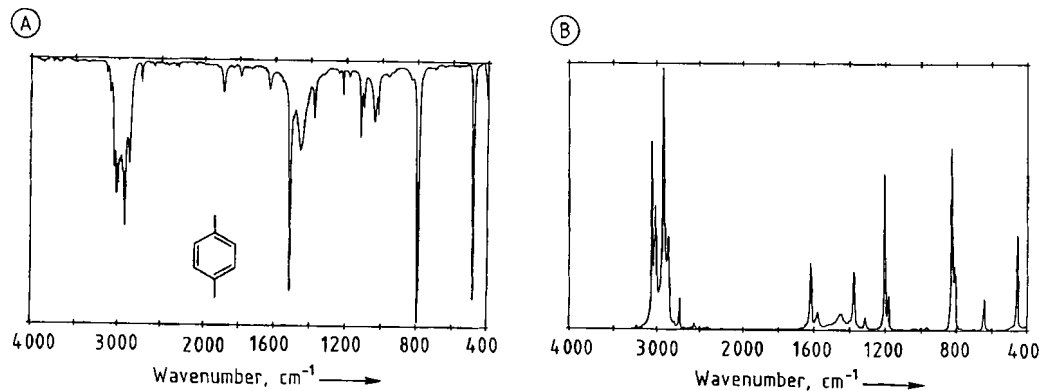
Figure 17. FT-infrared (A) and FT-Raman (B) spectra of *p*-xylene
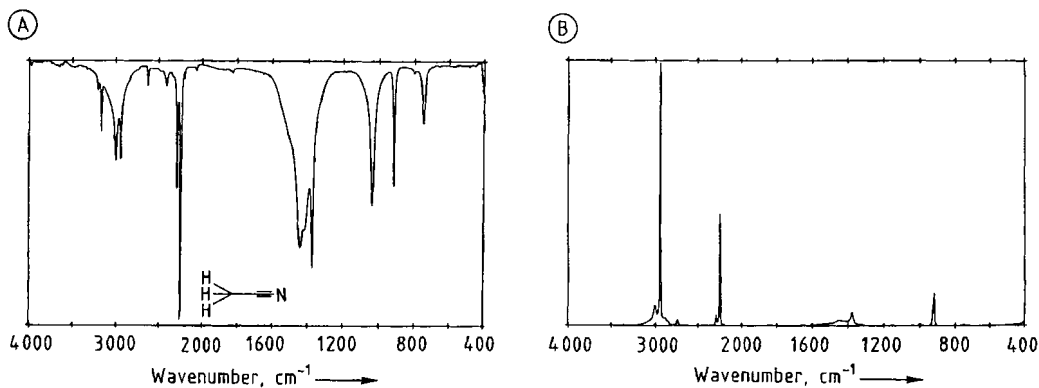
Figure 18. FT-infrared (A) and FT-Raman (B) spectra of acetonitrile
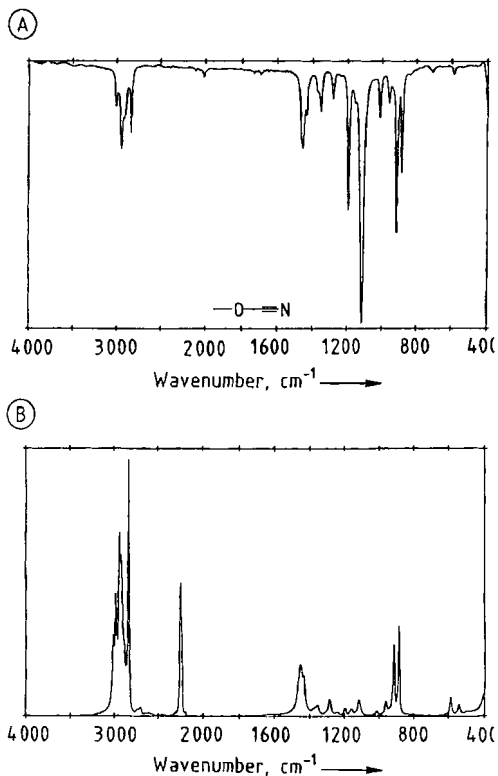
Figure 19. FT-infrared (A) and FT-Raman (B) spectra of methoxy-acetonitrile

Table 4. Characteristic group frequencies of triple bonds and cumulated double bonds

Bond	$v(\equiv C\text{-}H)$, cm^{-1}	$v(X \equiv Y)$ or $v_{as}(X=Y=Z)$, cm^{-1}
$-C \equiv C\text{-}H$	3300	2260–2100
$-C \equiv N$		2260–2200
$-N \equiv C$		2165–2110
$-C \equiv N \rightarrow O$		2300–2290
$-N^+ \equiv N$		2300–2140
$-S\text{-}C \equiv N$		2180–2140
$-N=C=S$		2150–2000
$-N=C=O$		2275–2230
$-N=N^+=N^-$		2250–2100
$-N=C=N-$		2150–2100
$-C=C=O$		2155–2130
$-C=N^+=N^-$		2100–2010
$-CO\text{-}C=N^+=N^-$		2100–2050
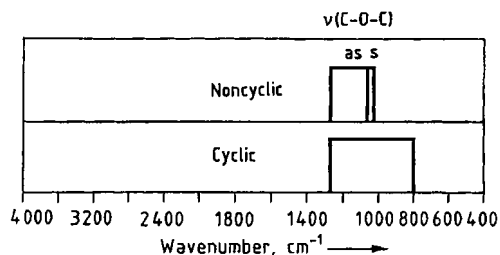
Figure 20. Characteristic group frequencies of ethers

17.4.7.2. Azo Compounds

Characteristic group frequencies of azo compounds are shown in Figure 24. In the infrared spectrum, similarly to symmetrical *trans*-substituted ethylenes, the N=N stretching vibration of the *trans* symmetrically substituted azo group does not occur, whereas all azo compounds show a medium (aliphatic compounds) or strong (aromatic compounds) N=N stretching Raman line. This band occurs between 1580 and 1520 cm^{-1} for purely aliphatic or aliphatic–aromatic, and

Ⓐ

Ⓑ
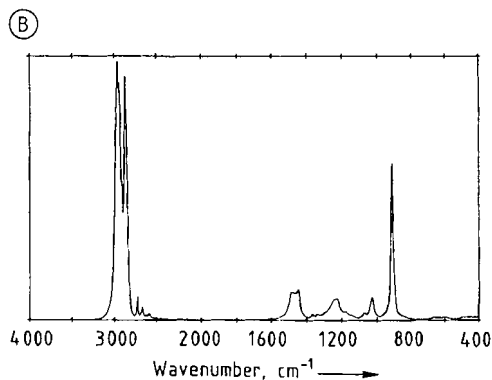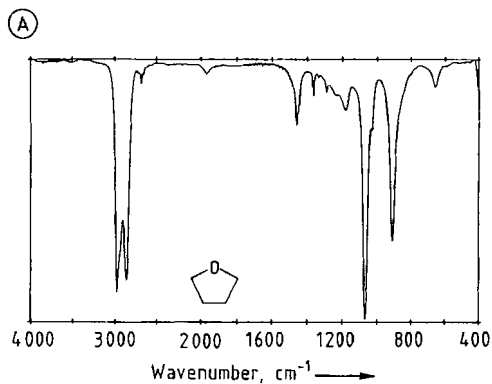
Figure 21. FT-infrared (A) and FT-Raman (B) spectra of tetrahydrofuran

v(O-H)   v(C-O)

3650   2400

Primary alcohols   1075 1000

Secondary alcohols   1150 1030

Tertiary alcohols   1170 1100

Phenols   1230 1140
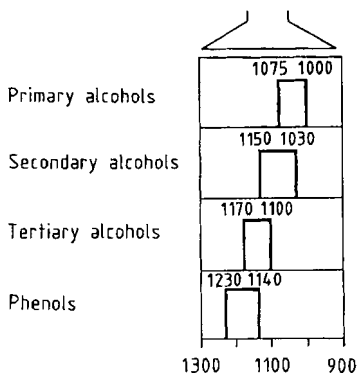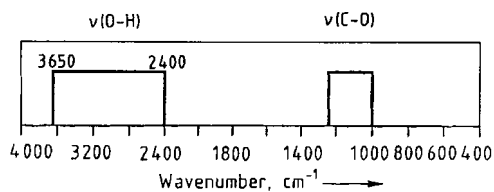
Figure 22. Characteristic group frequencies of alcohols and phenols

v(N-H)   v(C-N)

3550 3250 Primary   1380 1000

3450 3310 Secondary   1380 1000

Tertiary   1380 1000
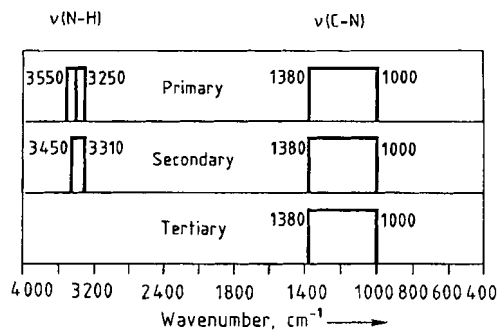
Figure 23. Characteristic group frequencies of amines

v(N=N)

Aliphatic Aromatic   1140
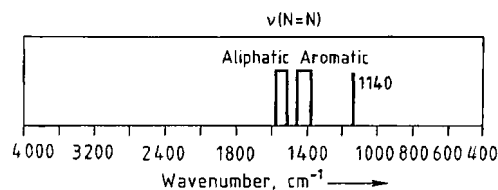
Figure 24. Characteristic group frequencies of azo compounds

between 1460 and 1380 cm^{-1} for purely aromatic azo compounds. In azoaryls, an additional strong Raman line involving aryl – N stretch is observed near 1140 cm^{-1}. As an example, the spectra of azobenzene are shown in Figure 25.

17.4.7.3. Nitro Compounds

Characteristic group frequencies of nitro compounds are shown in Figure 26. The nitro group, with its two identical N–O partial double bonds, gives rise to two bands due to antisymmetric and symmetric stretching vibrations. In the infrared, the antisymmetric vibration causes a strong absorption at 1590 – 1545 cm^{-1} in aliphatic, and at 1545 – 1500 cm^{-1} in aromatic nitro compounds. In comparison, the infrared intensities of the symmetric stretching vibration between 1390 and 1355 cm^{-1} for aliphatic, and 1370 and 1330 cm^{-1} for aromatic nitro compounds are somewhat weaker. However, a very strong Raman line due to symmetric vibration is observed in this region (see Fig. 27).
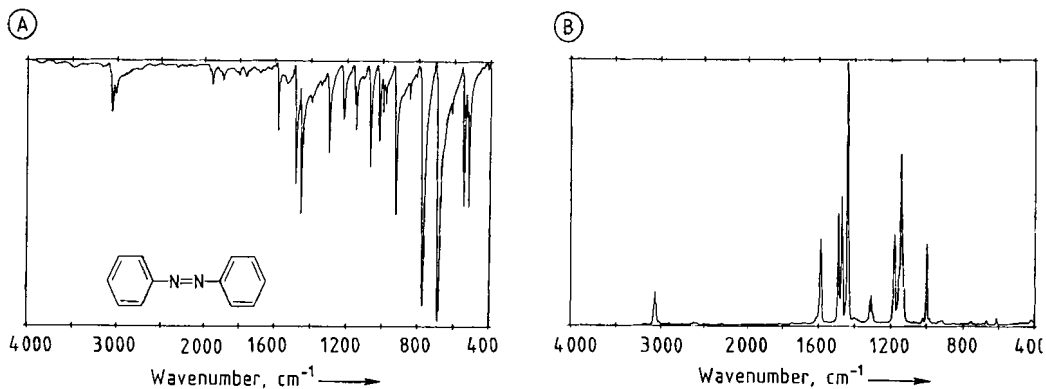
Figure 25. FT-infrared (A) and FT-Raman (B) spectra of azobenzene
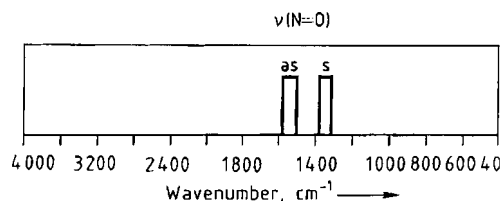
Figure 26. Characteristic group frequencies of nitro compounds

17.4.8. Carbonyl Compounds

Because of the stretching of the C=O bond, carbonyl compounds generally give rise to a very strong infrared and a medium to strong Raman line between 1900 and 1550 cm^{-1}. The position of the carbonyl band is influenced by inductive, mesomeric, mass, and bond-angle effects [10].

17.4.8.1. Ketones

Characteristic group frequencies of ketones are shown in Figure 28. Ketones are characterized by a strong C=O band near 1715 cm^{-1}, often accompanied by an overtone at ca. 3450 cm^{-1}. Ketones in strained rings show considerably higher C=O frequencies, whereas in di-*tert*-butyl ketone, for example, the carbonyl frequency is observed at 1687 cm^{-1}; a steric increase in the C–C–C angle causes this lower value.

17.4.8.2. Aldehydes

Characteristic group frequencies of aldehydes are shown in Figure 29. Together with the car-

bonyl band, the most intense band in the infrared spectrum, aldehydes are characterized by two CH bands at 2900–2800 cm^{-1} and 2775–2680 cm^{-1}. These bands are due to Fermi resonance [3], i.e., interaction of the CH stretch fundamental with the overtone of the O=C–H bending vibration near 1390 cm^{-1}.

17.4.8.3. Carboxylic Acids

Characteristic group frequencies of carboxylic acids are shown in Figure 30. Because carboxylic acids have a pronounced tendency to form hydrogen-bonded dimers, in the condensed state they are characterized by a very broad infrared OH stretching band centered near 3000 cm^{-1}. As the dimer has a center of symmetry, the antisymmetric C=O stretch at 1720–1680 cm^{-1} is infrared active only, whereas the symmetric C=O stretch at 1680–1640 cm^{-1} is Raman active only; this is shown in Figure 31 using acetic acid as an example. In the infrared, a broad band at 960–880 cm^{-1} due to out-of-plane OH$\cdots$O hydrogen bending is diagnostic of carboxyl dimers, as is the medium intense C–O stretching band at 1315–1280 cm^{-1}. On the other hand, monomeric acids have a weak, sharp OH absorption band at 3580–3500 cm^{-1} in the infrared, and the C=O stretching band appears at 1800–1740 cm^{-1} in both infrared and Raman.

17.4.8.4. Carboxylate Salts

Characteristic group frequencies of carboxylate salts are shown in Figure 32. In carboxylate salts, the two identical C–O partial double bonds give rise to two bands: the antisymmetric CO_2
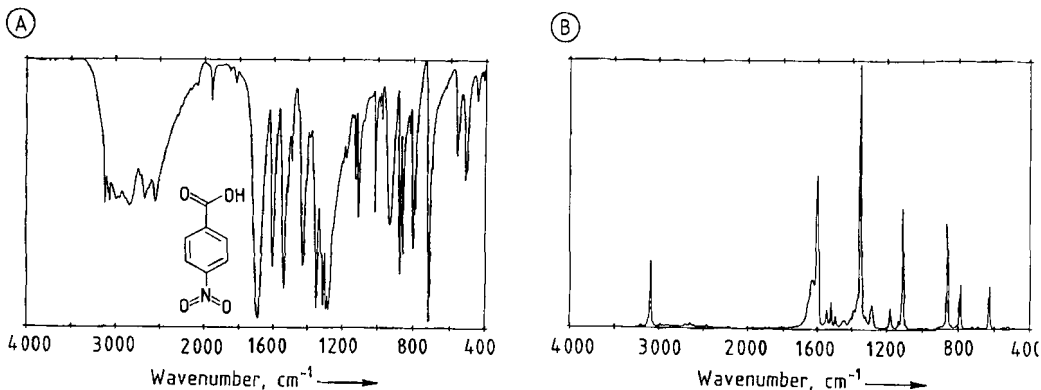
Figure 27. FT-infrared (A) and FT-Raman (B) spectra of 4-nitrobenzoic acid
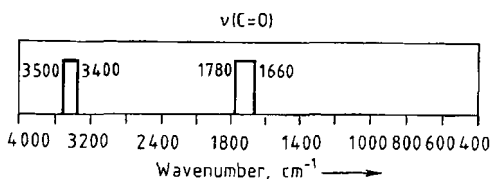
Figure 28. Characteristic group frequencies of ketones
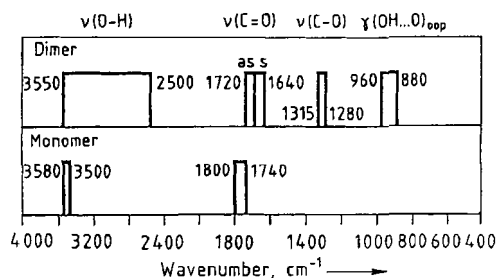
Figure 30. Characteristic group frequencies of carboxylic acids
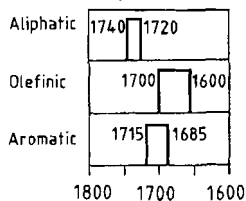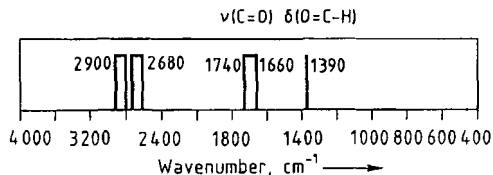
Figure 29. Characteristic group frequencies of aldehydes

17.4.8.5. Anhydrides

Characteristic group frequencies of anhydrides are shown in Figure 33. Anhydrides are characterized by two C=O stretching vibration bands, one of which occurs above 1800 cm^{-1}. These bands are strong in the infrared but weak in the Raman. In noncyclic anhydrides, the infrared C=O band at higher wavenumber is usually more intense than the C=O band at lower wavenumber, whereas in cyclic anhydrides the opposite is observed. In the Raman, the C=O line at higher wavenumber is generally stronger. Unconjugated straight-chain anhydrides have a strong infrared absorption, due to the C–O–C stretching vibration, at 1050–1040 cm^{-1} (except for acetic anhydride: 1125 cm^{-1}), while cyclic anhydrides exhibit this infrared band at 955–895 cm^{-1}. In Figure 34, the spectra of acetic anhydride are shown as an example.

stretching vibration at 1650–1550 cm^{-1} is very strong in the infrared and weak in the Raman, while the corresponding symmetric vibration at 1450–1350 cm^{-1} is somewhat weaker in the infrared but strong in the Raman.
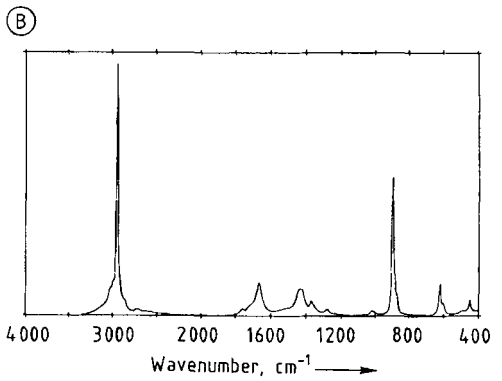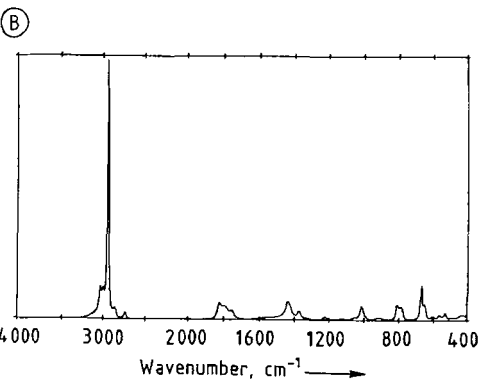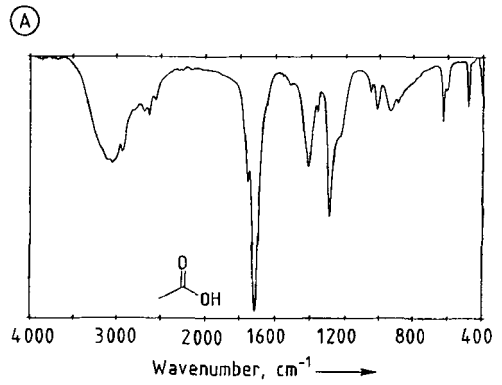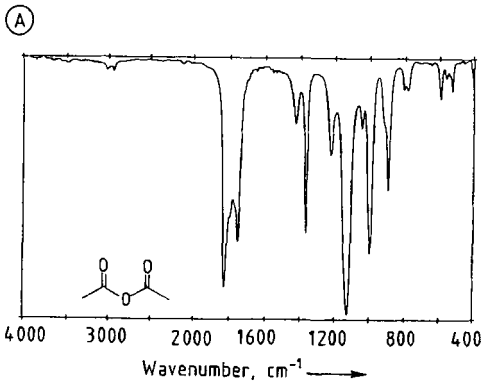
Figure 34. FT-infrared (A) and FT-Raman (B) spectra of acetic anhydride

Figure 32. Characteristic group frequencies of carboxylate salts

Figure 33. Characteristic group frequencies of anhydrides

Figure 31. FT-infrared (A) and FT-Raman (B) spectra of acetic acid

17.4.8.6. Esters

Figure 35 shows the characteristic group frequencies of esters. In esters, the C=O stretching vibration gives rise to an absorption band at 1740 cm^{-1} which is very strong in the infrared and medium in the Raman. Its position is strongly influenced by the groups adjacent to the ester group. The C–O stretching band at 1300–1000 cm^{-1} often shows an infrared intensity similar to the C=O band.

17.4.8.7. Lactones

In lactones, the position of the C=O band is strongly dependent on the size of the ring (Table 5). In lactones with unstrained, six-membered and larger rings the position is similar to that of noncyclic esters.

$v(C=O)$ $v(C-O)$
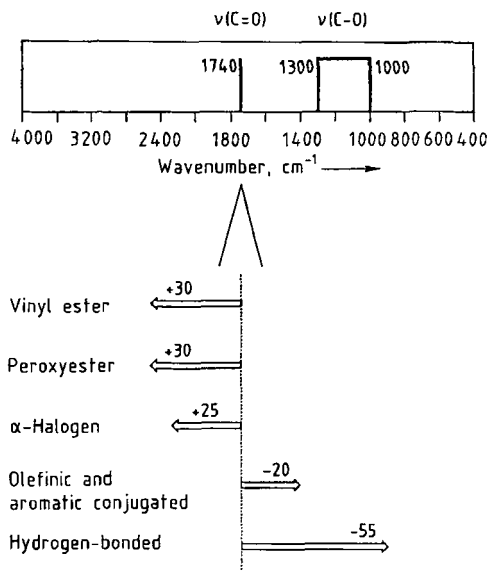
Figure 35. Characteristic group frequencies of esters
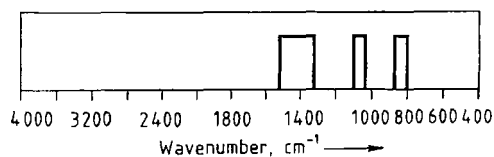
Figure 36. Characteristic group frequencies of carbonate salts

17.4.8.8. Carbonate Salts

Characteristic group frequencies of carbonate salts are shown in Figure 36. In inorganic carbonate salts, the antisymmetric CO_3 stretching vibration gives rise to a very broad absorption band at $1520-1320$ cm^{-1}, which is very strong in the infrared but only weak in the Raman. The symmetric CO_3 stretching vibration, on the other hand, gives rise to a very strong Raman line at $1100-1030$ cm^{-1}, which is normally inactive in the infrared. Another characteristic medium sharp infrared band at $880-800$ cm^{-1} is caused by out-of-plane deformation of the CO_3^{2-} ion (Fig. 37).

17.4.8.9. Amides

Figure 38 shows characteristic group frequencies of amides. In amides, the C=O stretching vibration gives rise to an absorption at
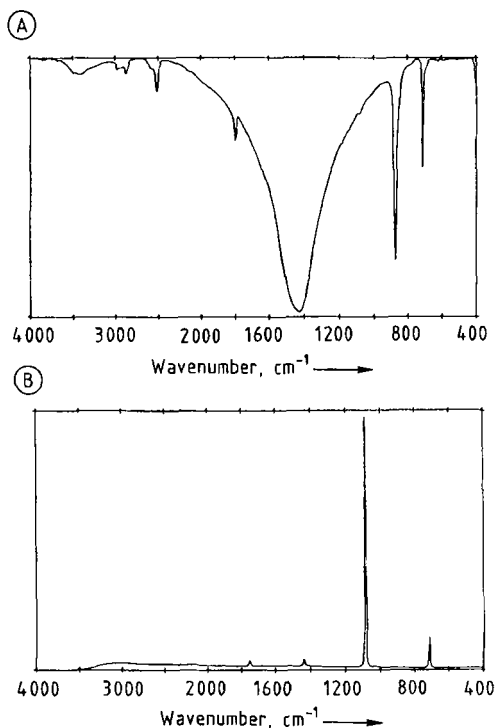
Figure 37. FT-infrared (A) and FT-Raman (B) spectra of calcium carbonate
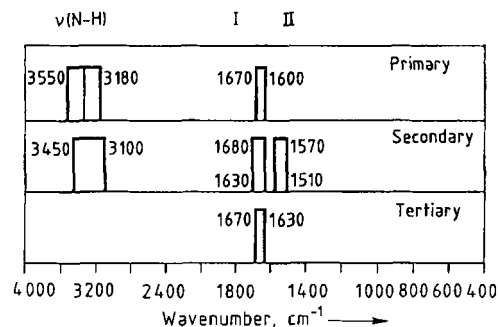
Figure 38. Characteristic group frequencies of amides

$1680-1600$ cm^{-1} known as the amide I band; it is strong in the infrared and medium to strong in the Raman. In primary amides, a doublet is usually observed in the amide I region, the second band being caused by NH_2 deformation. A characteristic band for secondary amides, which mainly exist with the NH and C=O in the *trans* configu-

Table 5. Characteristic group frequencies of lactones

Lactone		Structure	$v\,(C=O)$, cm^{-1}
β-Lactones			1840
γ-Lactones:	saturated		1795 – 1775
	α, β-unsaturated		1785 – 1775 and 1765 – 1740
	β, γ-unsaturated		1820 – 1795
δ-Lactones:	saturated		1750 – 1735
	α, β-unsaturated		
	β, γ-unsaturated		1760 – 1750 and 1740 – 1715

ration, is the amide II band at 1570 – 1510 cm^{-1}, involving in-plane NH bending and C – N stretching. This band is less intense than the amide I band.

As in amines, the NH stretching vibration gives rise to two bands at 3550 – 3180 cm^{-1} in primary amides and one near 3300 cm^{-1} in secondary amides; these are strong both in infrared and Raman. In secondary amides a weaker band appears near 3100 cm^{-1} due to an overtone of the amide II band.

17.4.8.10. Lactams

Characteristic group frequencies of lactams are shown in Figure 39. As in lactones, the size of the

ring influences the C=O frequency in lactams. In six- or seven-membered rings the carbonyl vibration absorbs at 1680 – 1630 cm^{-1}, the same as the noncyclic *trans* case. Lactams in five-membered rings absorb at 1750 – 1700 cm^{-1}, while β-lactams absorb at 1780 – 1730 cm^{-1}. These bands are strong in the infrared and medium in the Raman.

Because of the cyclic structure the NH and C=O groups are forced into the *cis* configuration so that no band comparable to the amide II band in *trans* secondary amides occurs in lactams. A characteristically strong NH stretching absorption near 3200 cm^{-1} occurs in the infrared, which is only weak to medium in the Raman. A weaker infrared band near 3100 cm^{-1} is due to a combination band of the C=O stretching and NH bending vibrations.
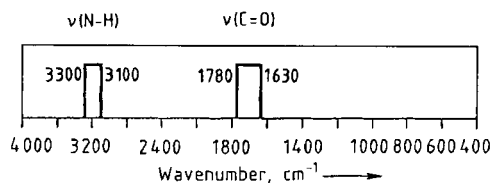
Figure 39. Characteristic group frequencies of lactams
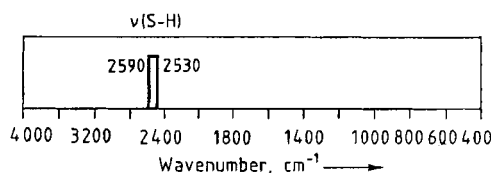
Figure 40. Characteristic group frequencies of thiols
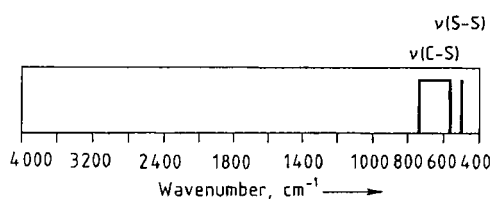
Figure 41. Characteristic group frequencies of sulfides and disulfides

Figure 42. FT-infrared (A) and FT-Raman (B) spectra of L-cystine

17.4.9. Sulfur-Containing Compounds

Sulfur-containing functional groups generally show a strong Raman effect, and groups such as C–S and S–S, for example, are extremely weak infrared absorbers. Hence, Raman spectra of sulfur-containing compounds have a much greater diagnostic value than their infrared spectra.

17.4.9.1. Thiols

Characteristic group frequencies of thiols are shown in Figure 40. The S–H stretching vibration of aliphatic and aromatic thiols gives rise to a medium to strong Raman line at $2590-2530$ cm^{-1}; this band is weak or very weak in the infrared.

17.4.9.2. Sulfides and Disulfides

Figure 41 shows characteristic group frequencies of sulfides and disulfides. The stretching vibration of C–S bonds gives rise to a weak infrared but a strong Raman signal at $730-570$ cm^{-1}. Similarly, the S–S stretching vibration at 500 cm^{-1} is a very strong Raman line but very weak infrared band (Fig. 42). Both Raman signals are very diagnostic in the conformational analysis of disulfide bridges in proteins.

17.4.9.3. Sulfones

Similar to nitro groups and carboxyl salts, sulfones are characterized by two bands due to antisymmetric and symmetric stretching of the SO$_2$ group, the former at $1350-1290$ cm^{-1} and the latter at $1160-1120$ cm^{-1} (Fig. 43). Both bands are strong in the infrared, while only the symmetric vibration results in a strong Raman line (Fig. 44).
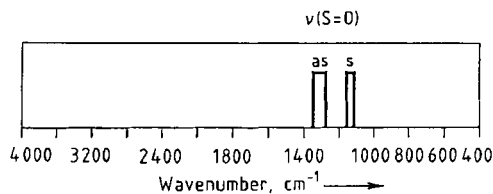
Figure 43. Characteristic group frequencies of sulfones

17.4.10. Computer-Aided Spectral Interpretation

After his first experiences an interpreting vibrational spectra the spectroscopist soon realizes the limited capacity of the human brain to store and selectively retrieve all required spectral data. Powerful micro- and personal computers are now available together with fast and efficient software systems to help the spectroscopist to identify unknown compounds. IR Mentor [41] is a program that resembles an interactive book or chart of functional group frequencies. Although the final interpretation remains in human hands, this program saves the spectroscopist time by making tabular correlation information available in computer form, and, moreover, it is also an ideal teaching tool. To facilitate the automated identification of unknown compounds by spectral comparison numerous systems based on library search, e.g., SPECTACLE [42], GRAMS/32 [43], or those offered by spectrometer manufacturers, are used. These systems employ various algorithms for spectral search [44], and the one that uses full spectra according to the criteria of LOWRY and HUPPLER [45] is the most popular. The central hypothesis behind library search is that if spectra are similar then chemical structures are similar [46]. In principle, library search is separated into identity and similarity search systems. The identity search is expected to identify the sample with only one of the reference compounds in the library. If the sample is not identical to one of the reference compounds the similarity search presents a set of model compounds similar to the unknown one and an estimate of structural similarity. The size and contents of the library are crucial for a successful library search system, especially a similarity search system. A smaller library with carefully chosen spectra of high quality is more useful than a comprehensive library containing spectra of all known chemical compounds or as many as are available because, in similarity search, the retrieval of an excessive
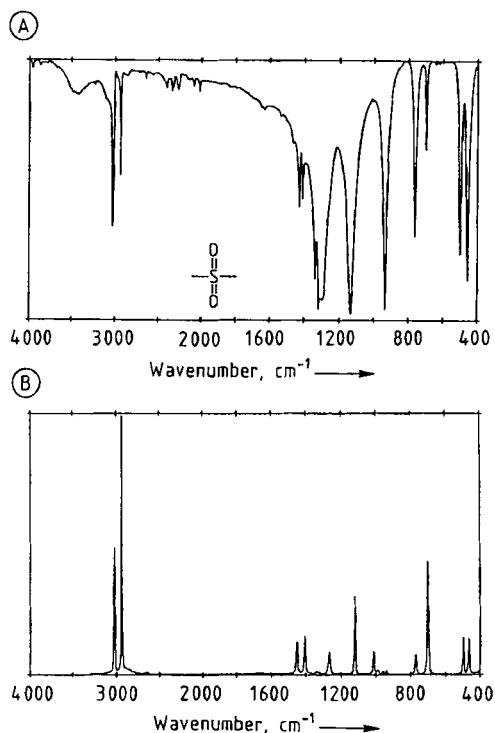
Figure 44. FT-infrared (A) and FT-Raman (B) spectra of dimethyl sulfone

number of closely similar references for a particular sample only increases output volume without providing additional information [46]. A critical discussion of the performance of library search systems is presented by CLERC [46]. While library search systems are well-established, invaluable tools in daily analytical work, expert systems (i.e., computer programs that can interpret spectral data) based on artificial neural networks (ANNs) are more and more emerging [47], [48].

17.5. Applications of Vibrational Spectroscopy

17.5.1. Infrared Spectroscopy

17.5.1.1. Transmission Spectroscopy

Transmission spectroscopy is the simplest sampling technique in infrared spectroscopy, and is generally used for routine spectral measure-

Objective mirror

Field mirror

Cell windows
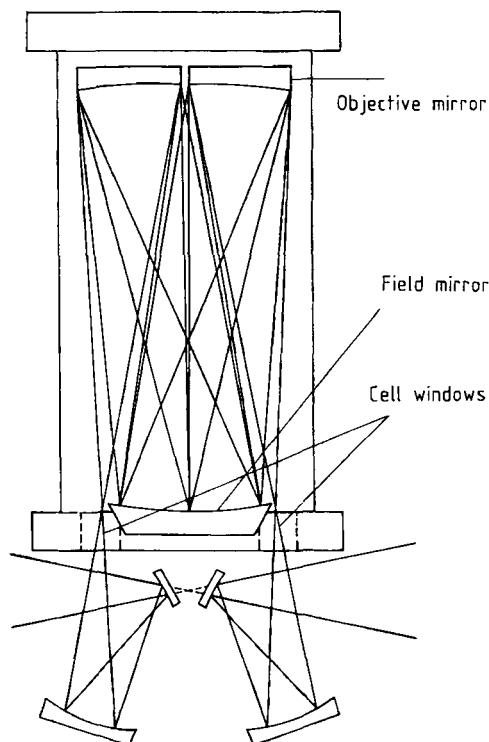
Figure 45. Optical diagram of a White gas cell (A mode) showing eight passes, to change the path length only one objective mirror is rotated, the rotation is nonlinear and is very small at long path lengths. For B mode operation, the light beam enters and exits on one and the same side of the field mirror, the objective mirrors are coupled and are rotated together linearly over the entire operational range (Reproduced by permission of Harrick Scientific Corporation, Ossining, NY 10562)

ments on all kinds of samples. It requires only simple accessories, such as infrared-transparent windows [e.g., potassium bromide, sodium chloride, or thallium bromoiodide (KRS-5)] for gases and liquids, or a sample holder for solids. The sample is placed in the light beam of an infrared spectrometer, and the intensity of the incident beam is compared with that transmitted by the sample. According to the fundamental relation governing the absorption of radiation as a function of transmittance (see Section 17.3.2), parameters that can be determined in transmission spectroscopy are the thickness or the concentration of the sample. The thickness can range from micrometers for solids and liquids to even kilometers for gas samples.

Apart from general laboratory gas analysis, gas-phase transmission spectroscopy is used for the characterization of air pollutants as well as the monitoring of stack gas and of air quality in work places. As air quality has become a major concern for environmentalists, industrial manufacturers and many others, spectroscopic measurement in the gas phase is receiving increasing attention. However, two major problems must be overcome: absorptions in the gas phase are many times weaker than those in the liquid or solid phase, and the species to be measured usually makes up only a small fraction of the total gas volume. The most popular way of surmounting these problems is to increase the path length by reflecting the beam several times from precisely controlled mirrors. Such multiple-pass gas cells are all based on the White design [49] or variations thereof. Figure 45 shows an optical diagram of a White cell aligned for eight passes. The length of the gas cell and the number of passes employed determines the optical path length, which may vary from centimeters to kilometers. The number of passes is adjusted in four pass increments by rotating one of the objective mirrors. In the A mode of operation, only one objective mirror is rotated, while in the B mode the objective mirrors are coupled and rotated together linearly over the entire operational range to change the path length. To prepare a sample, the gas cell is evacuated, the sample is bled into the cell, and the cell is sealed.

Liquid samples are easily studied with the aid of a wide variety of liquid cells, including heatable, flow through, and variable path length cells. These cells are constructed of two infrared transparent windows with a spacer between them, thereby forming a cavity for the sample. As the samples are often strong infrared absorbers, for this spectral range the liquid transmission cell must be constructed with short optical path lengths (0.025 – 1 mm).

Polycrystalline or powdered samples can be prepared as a suspension in mineral oil (Nujol mull), as a potassium bromide disk (pellet), or as thin films deposited on infrared-transparent substrates. The potassium bromide pellet is the most common way of preparing powder samples; in this method a small amount, usually 1 mg, of finely-ground solid sample is mixed with powdered potassium bromide, usually 300 mg, and then pressed in an evacuated die under high pressure. The resulting disks are transparent and yield excellent spectra. The only infrared absorption in the potassium bromide matrix is due to small amounts of adsorbed water, which can, however, be confused with OH-containing impurities in the sam-

Figure 46. Near-normal (θ_1) and grazing angle (θ_2) incidence

Figure 47. Types of specular reflection
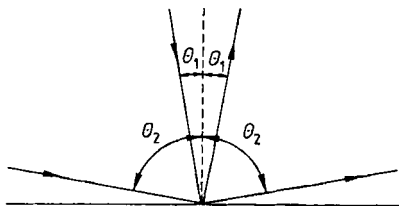
ple. To minimize adsorbed water the potassium bromide is ground as little as possible. The sample is ground separately and then mixed with potassium bromide, after which the mixture is ground only so far as is necessary to achieve good mixing.

17.5.1.2. External Reflection Spectroscopy

In external reflection spectroscopy, electromagnetic radiation is reflected at the interface of two materials differing in refractive index or from a smooth surface, a conductor such as metal, a semiconductor, a dielectric, or a liquid. The intensity of light reflected is measured to obtain information about the surface chemistry and structure. The angle of collection is the same as the angle of incidence and is measured from the normal to the surface (Fig. 46). Small-angle incidence (θ_1, usually 10°) is called near-normal incidence while that at large angles (θ_2, usually 80°) is known as grazing angle incidence. A nondestructive method requiring no sample preparation, external reflection spectroscopy comprises two different types of measurements, both being termed specular reflectance. In the first type, the reflection from a smooth mirrorlike surface, e.g., free-standing film, single crystal face, or surface of an organic material, is referred to as true specular reflection or simply specular reflection (Fig. 47). In the second, a film of sample is deposited on a highly reflecting substrate, usually a metal, from which the infrared beam is reflected, the radiation is transmitted through the sample layer, reflected from the substrate surface, and finally transmitted back through the sample. This technique is called reflection – absorption (RA), infrared reflection-absorption spectroscopy (IRRAS), or transflectance.

A material's object properties are characterized by its complex refractive index, $n' = n + ik$, where n is the refractive index and k is the absorption index, both of which are dependent on the frequency of incident light. Unlike normal trans-

mission spectra, true specular reflection spectra often have distorted, derivative bands because specular reflectance from a surface frequently has a large refractive index component. However, by using a mathematical operation known as the Kramers – Kronig transformation [50] a transmission-like spectrum showing the expected absorptivity information can be calculated from a measured, distorted spectrum. As true specular reflectance is often measured at or near normal incidence, the reflected energy is small, usually between 0 and 12 % for most organic materials. However, when this low signal-to-noise ratio is overcome by using more sensitive cooled MCT detectors the true specular reflection technique has the important advantage of permitting the non-contact analysis of bulk solids and liquids. Depending on the thickness of the film of sample, two types of reflection – absorption measurement are used: near-normal and grazing angle. In near-normal reflection – absorption, the thickness of the sample usually ranges between 0.5 and 20 μm and the angle of incidence varies between 10 and 30°. Hence, the amount of energy reflected from the substrate is much greater than the amount reflected from the front surface of the sample film, and a completely transmission-like reflection – absorption spectrum is obtained. Grazing angle reflection – absorption measurements are more sensitive than near-normal reflectance and are used to analyze sample films less than 1 μm thick, and even monolayers, on highly reflective materials. The improved sensitivity is due to the enhanced electromagnetic field that can be produced at the reflection surface by using polarized light at large angles of incidence [51], [52]. As the plane of incidence contains the incident and reflected rays, as well as the normal to the reflection surface, the electrical field of p-polarized light is parallel to it whereas that of s-polarized light is perpendicular to it. For s-polarized light, the phase of the reflected beam is shifted 180° from the incident
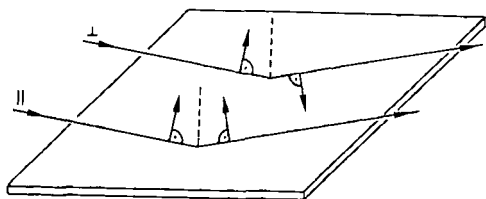
Figure 48. Phase shift of the reflected beam occurring at grazing angle incidence with perpendicular (180° shift) and parallel (90° shift) polarized light

beam at all angles of incidence (Fig. 48), and the two beams interfere at the surface to produce a node. Hence little interaction occurs between the infrared radiation and films thinner than one-quarter of the wavelength. However, for p-polarized light, the phase shift is approximately 90° at large angles of incidence, and the combination of incident and reflected beams at the surface produces a standing wave, giving an intense electrical field oriented normal to the surface. For metal substrates the greatest enhancement of the electrical field is produced by using p-polarization at large angles of incidence (e.g., 80°).

Reflection – absorption measurements are used in the analysis of a wide range of industrial products, including polymer coatings on metals, lubricant films on hard magnetic disks (Fig. 49), semiconductor surfaces, or the inner coating of food and beverage containers. In addition, grazing angle reflectance is especially suited to the study of molecular orientations at surfaces [53]. External reflectance at different angles of incidence can be employed to determine both the thickness and refractive index of a film of sample by measuring the interference fringe separation [54]; Figure 50 shows one of the various external reflection attachments which are available.

17.5.1.3. Internal Reflection Spectroscopy

Internal reflection spectroscopy, also known as attenuated total reflectance (ATR) or multiple internal reflectance (MIR), is a versatile, non-destructive technique for obtaining the infrared spectrum of the surface of a material or the spectrum of materials either too thick or too strongly absorbing to be analyzed by standard transmission spectroscopy.

The technique goes back to NEWTON [55] who, in studies of the total reflection of light at the interface between two media of different refractive
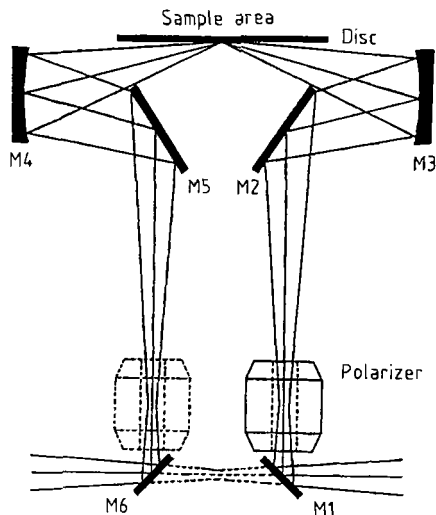
Figure 49. Ray diagram of the wafer/disk checker attachment for recording spectra of thin films on large samples, e.g., lubricants on hard magnetic disks;
$M_1 - M_6 =$ mirrors
(Reproduced by permission of Harrick Scientific Corporation, Ossining, NY 10562)

indices, discovered that an evanescent wave in the less dense medium extends beyond the reflecting interface. Internal reflection spectroscopy has been developed since 1959, when it was reported that optical absorption spectra could conveniently be obtained by measuring the interaction of the evanescent wave with the external less dense medium [56], [57]. In this technique, the sample is placed in contact with the internal reflection element (IRE), the light is totally reflected, generally several times, and the sample interacts with the evanescent wave (Fig. 51) resulting in the absorption of radiation by the sample at each point of reflection. The internal reflection element is made from a material with a high refractive index; zinc selenide (ZnSe), thallium iodide – thallium bromide (KRS-5), and germanium (Ge) are the most commonly used. To obtain total internal reflection the angle of the incident radiation θ must exceed the critical angle θ_c [58]. The critical angle is defined as:

$$\theta_c = \sin^{-1} \frac{n_2}{n_1} \tag{17}$$

where n_1 is the refractive index of the internal reflection element and n_2 is the refractive index of the sample. What makes ATR a powerful technique is the fact that the intensity of the evanescent
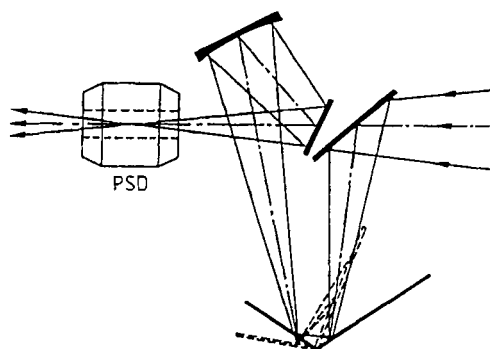
Figure 50. Ray diagram of the versatile reflection attachment (VRA) with single diamond polarizer (PSD)
(Reproduced by permission of Harrick Scientific Corporation, Ossining, NY 10562)

Figure 51. Schematic representation of total internal reflection with: A) Single reflection; B) Multiple reflection
IRE (internal reflection element)
n_1 = Refractive index of the internal reflection element; n_2 = Refractive index of the sample with $n_2 < n_1$; θ = Angle of incidence; d_p = Depth of penetration

wave decays exponentially with the distance from the surface of the internal reflection element. As the effective penetration depth is usually a fraction of a wavelength, total internal reflectance is generally insensitive to sample thickness and so permits thick or strongly absorbing samples to be analyzed. The depth of penetration d_p, defined as the distance required for the electrical field amplitude to fall to e^{-1} of its value at the interface, is given by:

$$d_p = \frac{\lambda_1}{2\pi\left(\sin^2\theta - n_{21}^2\right)^{1/2}} \tag{18}$$

where $\lambda_1 = \lambda/n_1$ is the wavelength in the denser medium, and $n_{21} = n_2/n_1$ is the ratio of the refractive index of the less dense medium divided by that of the denser [56].

Although ATR and transmission spectra of the same sample closely resemble each other, differences are observed because of the dependency of the penetration depth on wavelength: longer wavelength radiation penetrates further into the sample, so that in an ATR spectrum bands at longer wavelengths are more intense than those at shorter ones.

The depth of penetration also depends on the angle of incidence; hence, an angle of 45°, which allows a large penetration depth, is generally used to analyze organic substances, rather than an angle of 60°, which results in a substantially weaker spectrum due to the decreased depth of penetration.

The degree of physical contact between sample and internal reflection element determines the sensitivity of an ATR spectrum. To achieve this, a

horizontal ATR accessory such as FastIR [58], in which the top plate is the sampling surface, is used [59]; reproducible contact is ensured by a special sample clamp or powder press. Good quality spectra are thus obtained for many materials that present problems of analysis with routine transmission methods, e.g., powders, pastes, adhesives, coatings, rubbers, fibers, thick films, textiles, papers, greases, foams, and viscous liquids. Possible methods of obtaining a spectrum from a variety of samples are discussed in [60]; in situ ATR spectroscopy of membranes is described in [61], [62]. Liquid samples are also well suited to ATR analysis. Most liquids require a very short path length; aqueous samples, for instance, are measured at path lengths of no more than ca. 15 µm, which makes the design of transmission cells difficult because flow of liquids is hindered; they also exhibit interference fringes because of the small spacing between the high refractive index infrared windows. These problems are eliminated by using liquid ATR cells, a variant of solid ATR, in which the internal reflection element is surrounded by a vessel into which the liquid is poured. Various such liquid cells are available, e.g., the Circle [63], the Prism Liquid Cell (PLC) [58], the Squarecol [64] and the Tunnel Cell [65]. With optimized optical design and fixed internal reflection elements these cells provide a highly reproducible path length which permits the quantitative analysis of liquids and aqueous solutions [66]. Liquid ATR
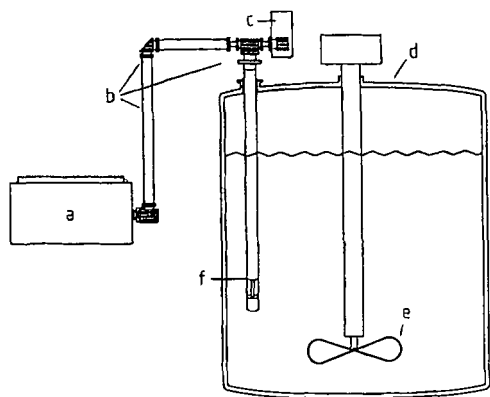
Figure 52. Schematic drawing of an FT-IR measurement system utilizing the Deep Immersion Probe Model DPR-124 mounted in a batch reaction vessel [68]
a) FT-IR spectrometer; b) Optical transfer elements; c) Detector assembly; d) Reaction vessel; e) Mixing blade; f) ATR sensing head
(Reproduced by permission of Axiom Analytical, Inc., Irvine, CA 92614)

cells are uniquely suited to fully automatable, on-line process-monitoring of liquids and viscous fluids. In addition, with suitable optical transfer modules such as the Axiot system [67], it is possible to operate liquid cells outside the sample compartment of the FT-IR spectrometer. This is important for on-line applications in which it is not practicable to pipe the sample to the spectrometer because the analysis can be carried out at the most desirable location, even in a fume hood or on a process line. With an ATR sensing head, immersion probes [68] offer the possibility of in situ FT-IR spectroscopy, for monitoring batch process reactions (Fig. 52), laboratory process development, the identification of hazardous waste, the verification of the contents of storage drums, and the inspection of incoming raw materials.

17.5.1.4. Diffuse Reflection Spectroscopy

Diffuse reflection spectroscopy, also known as diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy, enables the analysis of many samples for which traditional techniques fail, to be made with little or no sample preparation. Many substances in their natural state, especially powders but also any solid with a rough surface, such as dyed textiles and printed papers, exhibit diffuse reflection, i.e., incident light is

scattered in all directions, unlike specular reflection (see Section 17.5.1.2) where the angle of incidence equals the angle of reflection. Diffuse reflectance spectra are in practice complex, because of spectral distortions caused by the mixing of both absorbance and reflectance features owing to contributions from transmission, internal, and specular components in the measured radiation. Hence, DRIFT spectra are affected by particle size, packing, and dilution of the sample.

In diffuse reflectance spectra large glossy particle faces or crystal surfaces cause specular reflection which produces inverted bands, known as reststrahlen bands; this effect is reduced by fine grinding of the sample. Reststrahlen bands may also occur in neat samples with a rough surface, such as dyed textiles, printed papers, and agricultural and botanical specimens [69], measured by direct nondestructive, DRIFT spectroscopy. In this case, roughening the surface with emery paper reduces the disturbing bands because with smaller, thinner particles on the roughened surface the transmission component of the collected light increases, thus decreasing the reststrahlen effect. In strongly absorbing samples, the slight penetration of the incident beam produces mainly specular reflection at the surface, which also results in strong reststrahlen bands. By diluting the sample with a nonabsorbing powder such as KBr the effect is minimized or even eliminated, as this ensures deeper penetration of the incident beam and, therefore, an increased contribution of transmission and internal reflection components to the DRIFT spectrum. The diffuse reflectance spectrum is usually calculated as the ratio of sample and reference reflectance, with the pure diluent being taken as reference. Typically, a mixture of 90–95 % diluent and 10–5 % sample is used. Although the diluted spectra and the bulk reflectance spectra of weakly absorbing compounds have a similar appearance, they show enhanced intensity at lower wavenumbers compared to transmission spectra.

The interpretation of diffuse reflectance is based on a theory developed by KUBELKA and MUNK [70] and extended by KORTÜM, BRAUN, and HERZOG [71] for diluted samples in nonabsorbing matrices. The diffuse reflectance R_∞ of a diluted sample of infinite thickness (i.e., a sample for which an increase in thickness does not appreciably change the spectrum) is linearly related to the concentration c of the sample, as given by the Kubelka–Munk (K–M) equation:

$$f(R_\infty) = \frac{(1 - R_\infty)^2}{2R_\infty} = \frac{2.303ac}{s} \qquad (19)$$

where a is the absorptivity and s is the scattering coefficient. Because the scattering coefficient depends on both particle size and degree of sample packing, the linearity of the K–M function can only be assured if particle size and packing method are strictly controlled. If these conditions are fulfilled, accurate quantitative diffuse reflectance measurements can be obtained [72], [73].

Initiated by the work of GRIFFITHS et al. [74], diffuse reflectance accessories have been developed that pay special attention to the optimal reduction of the disturbing specular component. For example, the Praying Mantis diffuse reflectance attachment (Fig. 53) uses an off-line collection angle incorporating two 6 : 1 90° off-axis ellipsoidal mirrors. One ellipsoidal mirror focuses the incident beam on the sample, while the radiation diffusely reflected by it is collected by the other. Both ellipsoidal mirrors are tilted forward; therefore, the specular component is deflected behind the collecting ellipsoid, permitting the diffusely reflected component to be collected. As diffuse reflectance is a very sensitive technique for the analysis of trace quantities of materials, all available attachments allow for microsampling [75]. The analysis of thin-layer chromatograms by DRIFT spectroscopy, in situ or after transfer, is described in [76], [77]. Catalytic reactions can be monitored by using a controlled reaction chamber attached to a DRIFT accessory [78], [79]. Intractable samples, such as paint on automobile panels, can be analyzed by using silicon carbide abrasives [80]: a silicon carbide disk is rubbed against the sample, a small amount of which adheres to the disk. The DRIFT spectrum of the sample is then quickly obtained by placing the disk in the diffuse reflectance attachment.

17.5.1.5. Emission Spectroscopy

Infrared emission spectroscopy is a useful and effective technique for studying the surface of liquid and solid organic, inorganic, and polymeric materials, and for the in situ process-monitoring of high-temperature reactions and thermal transformations [81], [82]. In contrast to the conventional transmission technique, in emission spectroscopy the sample itself is the infrared source. The temperature of the sample is raised, which increases the Boltzmann population of the vibrational energy states. Infrared radiation is emitted on return to the ground state. In theory, emission spectra can always be measured provided the sample is at a different temperature to the detector. Like transmission spectra, emission spectra are obtained by dividing the emission of the sample by that of the reference, which is often a blackbody source at the same temperature as the sample. The main reason for the lack of any extensive work in this technique before 1970 was the low intensity of infrared emission, which results in spectra with poor signal-to-noise ratio. However, the availability of FT-IR spectrometers with the potential for low-energy sampling and for the development of highly efficient FT-IR emission spectroscopy accessories [83], [84] has led to a resurgence of interest in infrared emission spectroscopy.

Because of the phenomenon of self-absorption the ideal sample for conventional emission studies is a thin layer (e.g., a polymer film), on both metal and semiconductor surfaces [81]. A sample is usually heated from below the emitting surface, the lower surface thus having a higher temperature than the upper one. Therefore, radiation emitted from below the upper surface is absorbed before it reaches the surface, and this self-absorption of previously emitted light severely truncates and alters features in the emission spectra of optically thick samples. This problem is overcome by using a laser for controlled heat generation within a thin surface layer of the sample, self-absorption of radiation thus being minimized. These methods, known as laser-induced thermal emission (LITE) spectroscopy [85], [86] and transient infrared emission spectroscopy (TIRES) [87], [88] can produce analytically useful emission spectra from optically thick samples. Quantitative applications of infrared emission spectroscopy are described in [89]–[91].

17.5.1.6. Photoacoustic Spectroscopy

Photoacoustic spectroscopy (PAS) [92]–[95] is a fast, nondestructive method for analyzing various materials in the gas, liquid, or solid state with virtually no sample preparation. In cases where samples are insoluble, difficult to grind into a powder, or of irregular shape—such as coal, carbon-filled and conducting polymers, pharmaceutical preparations, spin coatings on fibers, and coatings on irregular surfaces—photoacoustic spectroscopy complements other established techniques such as attenuated total reflectance (ATR) and diffuse re-
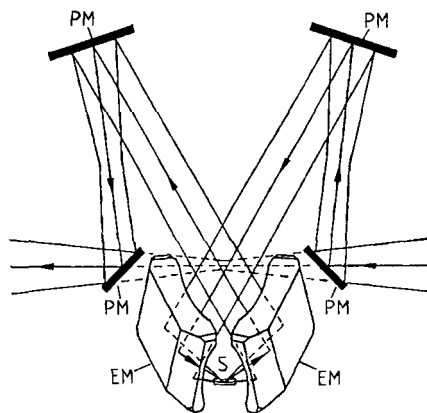
Figure 53. Ray diagram of the Praying Mantis diffuse reflectance attachment
EM = Ellipsoidal mirror; PM = Planar mirror; S = Sample
(Reproduced by permission of Harrick Scientific Corporation, Ossining, NY 10562)

Figure 54. Schematic of a photoacoustic cell

flectance (DRIFT). Moreover, measurements on samples that show the Christiansen effect [96] when prepared as KBr pellets for standard transmission spectroscopy can be better performed by the photoacoustic method. As most photoacoustic measurements are made with solid or semisolid samples, the sample is sealed in a small-volume cell containing a nonabsorbing gas such as helium and having windows for optical transmission (Fig. 54). The sample is illuminated with modulated radiation from the monochromator of a dispersive instrument or the interferometer of a Fourier transform infrared spectrometer. At wavelengths where the sample absorbs a fraction of the incident radiation, a modulated temperature fluctuation at the same frequency, but not necessarily with the same phase, as that of the incident radiation is generated in the sample, and the surrounding inert gas produces periodic pressure waves in the sealed cell. These fluctuations in pressure can be detected by a microphone because the modulation frequency of the incident beam usually lies in the acoustic range, e.g., between 100 and 1600 Hz. The low power available from dispersive instruments results in spectra with a rather poor signal-to-noise ratio; hence, photoacoustic mid-infrared spectra of condensed substances are preferably obtained by using an FT-IR spectrometer with a photoacoustic cell accessory. In a rapid-scanning Michelson interferometer with a mirror velocity of v (in cm/s), radiation of wavenumber $\tilde{v}$ (in cm^{-1}) is modulated with a frequency of $2\,v\,\tilde{v}$ Hz [8]. Therefore, FT-IR spectrometers are ideal for photoacoustic measurements in the mid-

infrared, where mirror velocities on the order of $0.05-0.2$ cm s^{-1} provide modulation frequencies in the acoustic range. As carbon black absorbs all incident radiation, the sample single-beam spectrum is usually ratioed against a carbon black single-beam spectrum to obtain an absorbance-like spectrum.

Before the first Fourier transform infrared photoacoustic spectrum was reported in 1978 [97], most infrared photoacoustic experiments were performed by using high-intensity tunable laser sources. However, as the source is monochromatic, this type of experiment is only useful for monitoring a single line in the spectrum. Highly-developed photoacoustic accessories are now available from FT-IR manufacturers or other companies [98], which have a high sensitivity also for microsamples. As the signal is inversely proportional to the cell volume, cells optimized for the measurement of condensed samples have a small volume, typically well under 1 cm^3, to enhance the signal amplitude [99]. Water vapor or carbon dioxide generate good photoacoustic signals, so even traces of these impurities cause serious interference in the photoacoustic spectroscopy signal; therefore, the purity of the gas in the cell is essential for good-quality photoacoustic spectra.

Important applications of photoacoustic spectroscopy are near-surface analysis [100] and depth profiling of polymers and polymer layers [101]–[103]. The theory of depth profiling is described in detail in [92]. It is based on the fact that thermal diffusion length (the distance a thermal wave travels in the sample before its intensity is reduced by 1/e in magnitude) and hence intensity of the photoacoustic signal are increased by lowering the modulation frequency. In a rapid-scanning interferometer this is achieved by decreasing the mirror velocity. To date, rapid-scanning interferometers that provide modulation frequencies

which continuously change with wavenumber have generally been used. However, step-scan interferometers accommodating any desired modulation frequency, constant over the entire wavelength range, produce improved photoacoustic spectra and are thus more effective in depth profiling [104] – [107].

Saturation effects, which can occur for stronger absorption bands, are the main limitation of photoacoustic spectroscopy. These distortions can be overcome by using thin samples [93] or by analyzing the phase of the photoacoustic signal [108]. The latter method is particularly useful for quantitative analysis.

17.5.1.7. Chromatography/Fourier Transform Infrared Spectroscopy

The combination of chromatographic separation with Fourier transform infrared spectroscopy has significantly improved the analysis of complex mixtures [109]. In chromatography/FT-IR systems, an infrared detector providing information on the structure of separated species is used instead of standard bulk chromatography detectors such as thermal conductivity, flame ionization, ultraviolet, or fluorescence; these detectors can be used for quantitative analysis when the identities of the mixture components are known.

In these systems the following factors must be taken into consideration: the physical state of the matter to be analyzed, the optimization of the interface between chromatograph and FT-IR, and the specific requirements and constraints of all three parts. In gas chromatography (GC)/FT-IR spectroscopy [110], one type of interface is composed of temperature-controlled, long, thin flow-through gas cells or "light pipes" coated with gold for high reflectivity. As the light-pipe techniques have a detection limit in the nanogram range, sub-nanogram limits can be routinely attained with the matrix-isolation technique (GC/MI/FT-IR) [109], and even better detection limits are achieved with the subambient trapping GC/FT-IR technique [109], [111]. Rapid-scanning interferometers able to record the entire spectrum in less than one second at low resolution (usually 4 or 8 cm^{-1}), together with highly sensitive detectors (e.g., cooled MCT detectors), are essential in all column chromatography/FT-IR spectroscopic techniques, with capillary GC/FT-IR requiring the most rapid data acquisition. In comparison to vapor phase species, rotations of matrix-isolated species are hindered, which results in sharper and more intense bands. Hence, for accurate spectra representation, GC/MI/FT-IR spectra are usually measured at 1 cm^{-1} resolution.

The Gram – Schmidt vector orthogonalization method [112] is the most sensitive and computationally most efficient and, hence, the most popular procedure for the reconstruction of the total chromatographic trace from interferometric infrared data. Selective detection is obtained by functional-group-specific chromatograms (chemigrams [113]), which are produced by computing absorbance spectra during chromatographic separation and plotting integrated absorbance for prespecified wavelength windows as a function of separation time, representing the quantity of a specific functional group that elutes as a function time. In chromatography/FT-IR spectral evaluation, in order to classify unknowns as to compound type, library searching (see Section 17.4.10) has proved to be very useful.

The addition of a complementary detector significantly increases the power of GC/FT-IR analysis. Mass spectrometry (MS) is an ideal choice [114] because the disadvantages of each method — infrared spectroscopy often cannot distinguish between long-chain homologues, and mass spectrometry frequently fails to distinguish isomeric species — are offset by the other.

The crucial point in liquid chromatography (LC)/FT-IR methods, however, is the interferences caused by the liquid phase. Actually, two viable LC-IR systems are commercially available: the LC-Transform [115] and the InfraRed Chromatograph (IRC) [116], [117]. While the LC-Transform is a solvent-elimination interface, the IRC provides full FT-IR spectra of the eluent stream from an LC column. The LC-Transform system utilizes a high-performance ultrasonic nebulizer or a pneumatic linear capillary nozzle to remove the mobile phase from samples as they elute from the chromatograph. The capillary is surrounded by hot, flowing sheath gas which provides sufficient thermal energy to evaporate the mobile phase and to contain the nebulized spray in a tight, focused cone as the spray emerges from the nozzle. The sample is deposited on a germanium sample-collection disk as spots with diameters between 0.5 and 1 mm. In a consecutive, uncoupled step, good-quality infrared spectra can be obtained from these spots, which contain microgram to sub-microgram amounts, by the use of a specially designed optics module, comprising a beam condenser. While this optics module can be used in any FT-IR spectrometer uncoupled from the LC-Transform interface,

the IRC includes an FT-IR spectrometer, a data station for driving the spectrometer and for data-processing, and a vacuum interface for depositing and scanning the sample. Being a real-time direct deposition interface, the IRC uses an ultrasonic nebulizer, a desolvation tube and a vacuum chamber to evaporate the solvent from the LC stream. The sample residue is then deposited onto a moving zinc selenide window. The data station collects a continuous array of spectra as the dry sample track on the window moves through the focused beam of the spectrometer.

LC-IR provides powerful identification possibilities, based on molecular structure. A high degree of confidence can be placed on substance identification that simultaneously matches infrared spectral characteristics and chromatographic elution time. This is especially so when molecules that have various conformational isomers are to be distinguished.

Supercritical fluid chromatography (SFC)/FT-IR spectroscopy, generally with carbon dioxide as the mobile phase, bridges the gap between GC/FT-IR and LC/FT-IR, and is particularly useful for separating nonvolatile or thermally labile materials not amenable to gas chromatographic separation [109]. Flow cells, mobile phase elimination and matrix-isolation techniques are used as SFC/FT-IR interfaces.

Applications of chromatography/FT-IR spectroscopy involve toxins and carcinogens, wastewater constituents, sediment extracts, airborne species, pesticides and their degradation products, fuels and fuel feedstocks, flavors and fragrances, natural products, pharmaceuticals, biomedicine, and polymers.

17.5.1.8. Thermogravimetry/Fourier Transform Infrared Spectroscopy

Thermogravimetry (TG), also called thermogravimetric analysis (TGA), measures changes in sample mass as a function of increasing temperature and has become an important tool for the analysis and characterization of materials [118], [119]. To correlate weight changes with chemical changes, it is necessary to identify gases evolved from the thermogravimetric analyzer furnace. This is made possible by coupling thermogravimetry with FT-IR spectroscopy, which provides the necessary speed and sensitivity.

A system has been described in which the infrared beam of the FT-IR spectrometer is led directly into the TG equipment [120], but in commercially available systems [121], [122] the evolved gases are conducted from the TG system to the spectrometer by carrier gas flow via a transfer line. The interface is heated to prevent recondensation and incorporates the transfer line and a flow-through gas cell with nitrogen or helium as carrier gas.

After infrared spectra at low resolution (usually 4 cm^{-1}) have been recorded as a function of time and temperature, the further procedure is similar to GC/FT-IR (see Section 17.5.1.7) and, in fact, GC/FT-IR software, such as Gram–Schmidt thermogram reconstruction and library search with a vapor-phase spectra database is commonly used. Helpful for the identification in the vapor phase is the virtual absence of interaction between molecules, and therefore the nondestructive IR detection enables in many cases an easy and fast interpretation of single and overlapping absorption bands.

Providing structural information for sample components down to nanogram levels, TG/FT-IR applications include decomposition studies of polymers and laminates [123], the analysis of coal, oil shales, and petroleum source rocks [124], [125], and the determination of activation energies [126] and thermal decomposition of energetic materials [127].

The IR detection of evolved gases opens new possibilities in the quantification of thermogravimetric analysis. Besides measuring the mass loss of a TG step, the total amount can also be calculated after calibration of a single component and integration over time [128]. A further improvement in quantitative TG/FT-IR is achieved by PulseTA [129]. Here, a certain amount of gas or liquid is injected into the system and used as internal calibration standard for the quantification of the unknown amount of evolved gas.

17.5.1.9. Vibrational Circular Dichroism

Vibrational circular dichroism (VCD) extends the functionality of electronic circular dichroism (CD) into the infrared spectral region, with VCD bands of enantiomeric pairs of chiral molecules having opposite signs. Early VCD spectrometers were dispersive instruments equipped with a CaF$_2$ photoelastic modulator for creating left and right circularly polarized radiation and a liquid-nitrogen-cooled InSb detector. FT-IR-VCD measurements are carried out by using the so-called double modulation technique [130] in which the VCD interferogram is carried by the polarization

modulation frequency in the region of 50 kHz, and the IR transmission interferogram occurs in the usual range of Fourier frequencies. typically below 5 kHz.

Combining the structural specificity of FT-IR spectroscopy with the stereosensitivity of circular dichroism, VCD allows the determination of optical purity without enantiomeric separation or derivatization, as well as of absolute configuration without crystallization. Simultaneous monitoring of the optical purity of multiple chiral species, such as reactant and product molecules in a reaction process is also possible, as is the determination of conformations of biological molecules such as proteins, peptides, and DNA [131], [132].

17.5.2. Raman Spectroscopy

In mid-infrared spectroscopy, Fourier transform instruments are used almost exclusively. However, in Raman spectroscopy both conventional dispersive and Fourier transform techniques have their applications, the choice being governed by several factors [133], [134]. Consequently, a modern Raman laboratory is equipped with both Fourier transform and CCD-based dispersive instruments. For a routine fingerprint analysis, the FT system is generally used, because it requires less operator skill and is quicker to set up; the FT system is also be tried first if samples are highly fluorescent or light sensitive. However, if the utmost sensitivity is required, or if Raman lines with a shift smaller than 100 cm^{-1} are to be recorded, conventional spectrometers are usually preferred.

Raman spectroscopy offers a high degree of versatility, and sampling is easy because no special preparation is required. As it is a scattering technique, samples are simply placed in the laser beam and the backscattered radiation is analyzed.

Glass gives only a weak Raman signal, so that FT-Raman spectra from liquids are recorded by using a capillary cell [34], a sapphire sphere [34], or glass cuvettes with the back surface silvered for the favored 180° backscattering arrangement (see Section 17.2.2). Moreover, the study of aqueous samples in glass or quartz vessels, which is difficult in the mid-infrared region owing to strong infrared absorption bands of water and quartz, is possible in Raman spectroscopy because there are no significant bands causing interference in the solution spectrum. Powders are either measured by using microspherical cells [33] or are filled into cells consisting of a brass cylinder with a 3 mm

hole drilled through, into which a reflective metal rod is placed to compress the powder [14]. The Raman intensity is related to the sample thickness, but once the thickness exceeds 2 mm the intensity does not increase. Optical fibers are used for remote sampling, and for FT-Raman spectroscopy this is especially advantageous as the transmission by the fibers, excited by a near-infrared Nd : YAG laser [33], [135], can be at its highest just in the range of the Raman spectrum. Guidelines for optimizing FT-Raman recording conditions are given in [34]. Conventional dispersive Raman spectroscopy has not been widely used for quantitative measurements, mainly because of difficulties with band reproducibility. However, by using an FT-Raman spectrometer, spectra with excellent reproducibility, both in band position and intensity, are obtained, so that this technique has great potential for quantitative analysis [136]–[138].

Along with general reviews [139]–[142] and surveys of major areas of Raman spectroscopy [16], [17], various reviews covering special applications of Raman spectroscopy, such as polarization measurements [143], polymers [144], proteins and nucleic acids [145], pharmaceuticals and biomaterials [146]–[150], paint systems [151], [152], inorganic and organometallic materials [153], and thin-layer chromatography [154], have been published. For further applications and special Raman techniques, such as resonance Raman (RR), surface-enhanced Raman spectroscopy (SERS), coherent anti-Stokes Raman scattering (CARS), and stimulated Raman scattering (SRS), see [14], [155].

17.5.3. Fourier Transform Infrared and Raman Microspectroscopy

In conventional infrared microsampling, beam size is reduced four to six times by using beam condensers. These accessories, which permit submicrogram samples of between 2 and 0.5 mm in diameter to be measured, have, however, drawbacks such as heating effects, since the entire power of the infrared beam is condensed on the sample, and the fact that the sample cannot be visibly located, so that a skilled operator is needed for this time-consuming work. These disadvantages are largely overcome by coupling an optical microscope to an FT-IR spectrometer. As infrared lenses are poor in performance, infrared microscopes use reflecting optics such as Cassegrain objectives instead of the glass or quartz lenses
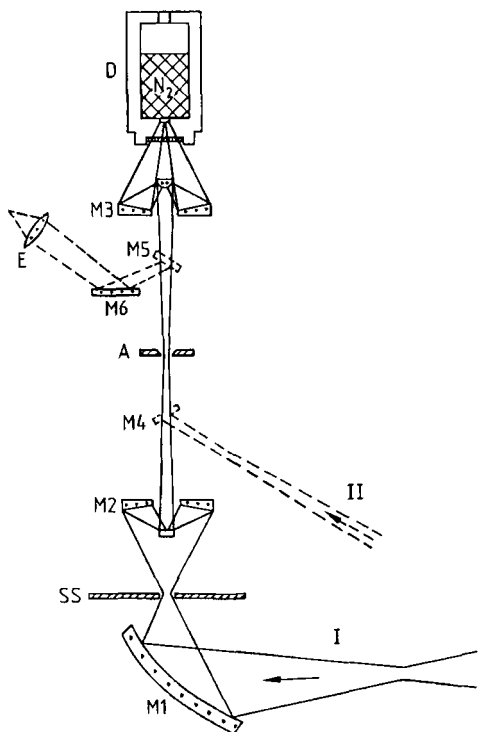
Figure 55. Beam path of a typical infrared microscope
I = Infrared transmittance beam; II = Infrared reflectance
beam; M1 = Condensing mirror; M2 and M3 = Cassegrain
objectives; M4 = Movable, semi-transparent mirror;
M5 = Movable mirror; M6 = Mirror; SS = sample stage;
A = Aperture; E = Eye-piece; D = MCT-Detector, liquid N_2
cooled
(Reproduced by permission of Bruker Optik GmbH,
D-76275 Ettlingen, Germany)

employed in conventional optical microscopes.
Figure 55 shows the beam path of a typical in-
frared microscope with both transmission and re-
flection modes. The infrared beam from the FT-IR
instrument is focussed onto the sample placed in a
standard microscope, and the beam that passes
through the sample is collected by a Cassegrain
objective with 15- or 36-fold magnification that
produces an image of the sample within the barrel
of the microscope. The radiation passes through
the variable aperture placed in the image plane and
is focussed by another Cassegrain objective on a
small area (250×250 μm or less) of the specially
matched MCT detector. As the visible optical train
is collinear with the infrared light path, it is pos-
sible to position the sample and to isolate and
aperture the area for analysis visually. This is

the main advantage of an FT-IR microscope over
a beam condenser, and it enables the measurement
of, for example, microcontaminants that cannot be
removed from the sample, or the individual layers
of a polymer laminate. FT-IR microscopy requires
virtually no sample preparation. A further advan-
tage of the technique is its unsurpassed sensitivity,
which permits samples in the picogram range to be
analyzed, with the diffraction limit of infrared
radiation (10 – 20 μm) as the limiting condition
for the size of measurement spot. To minimize
stray light due to diffraction being collected by
the microscope, particularly when the size of the
area investigated is approximately equal to the
wavelength employed, a second aperture is intro-
duced into the optics that focus the radiation onto
the sample [156].

Uses of FT-IR microspectroscopy include gen-
eral characterization of particulate matter, dichroic
measurements with polarized light, polymer char-
acterization, semiconductor measurements, the
identification of contaminants, as well as forensic,
biological, and pharmaceutical applications
[157] – [159].

By using diamond cells with the beam con-
denser or microscope, the thickness of a sample
(e.g., paint chips) can be adjusted by squeezing.
An alternative to this is an ultrasmall sample
analyzer [160], which allows nondestructive inter-
nal reflectance studies of microgram and nano-
gram samples to be performed.

In principle, Raman spectroscopy is a
microtechnique [161] since, for a given light flux
of a laser source, the flux of Raman radiation is
inversely proportional to the diameter of the laser-
beam focus at the sample, i.e., an optimized Ra-
man sample is a microsample. However, Raman
microspectroscopy able to obtain spatially re-
solved vibrational spectra to ca. 1 μm spatial res-
olution and using a conventional optical micro-
scope system has only recently been more widely
appreciated. For Raman microspectroscopy both
conventional [162] and FT-Raman spectrometers
[163],[164] are employed, the latter being coupled
by near-infrared fiber optics to the microscope.

Applications of Raman microspectroscopy in-
clude the analysis of a wide variety of organic and
inorganic materials, e.g., semiconductors,
polymers, single fibers, molecular crystals, and
minerals [165] – [168].

17.5.4. Time-Resolved FT-IR and FT-Raman Spectroscopy

Time-resolved FT-IR and FT-Raman techniques allow studies of molecular reaction mechanisms with time resolutions up to nanoseconds [169], which is a prerequisite for the understanding of biological processes at the molecular level. The major problem in measuring biological reactions consists in selecting small absorption bands of the molecular groups undergoing reactions from the large background absorption of water and the entire biological molecule. This difficulty is met by obtaining difference spectra of two reaction states of the molecules under investigation. The intrinsic advantages of FT spectroscopy, i.e., the multiplex, the throughput, and the accuracy (see Section 17.2.1) enable such small absorption changes to be reliably detected. With operation of the FT spectrometer in the step-scan mode [169], a time resolution of a few nanoseconds is achieved. Biological applications of time-resolved FT-IR and FT-Raman spectroscopy are described in [169].

17.5.5. Vibrational Spectroscopic Imaging

Although vibrational spectroscopic imaging methods are relative newcomers to the panoply of vibrational techniques, they inherit the versatility of the traditional single-point infrared and Raman approaches and, in combination with digital imaging technology, allow for new perspectives in the interpretation of the spectra. Imaging advantages are derived, in part, from the ability of an investigator to process and discern effectively a complex two-dimensional spatial representation of a sample in terms of a variety of spectrally related molecular parameters. In particular, a single spectroscopic image, or chemical map, is capable of summarizing and conveying a wealth of spatial, chemical, and molecular data of benefit to both specialists and nonspecialists. Regardless of the specific method used to collect data during a vibrational spectroscopic imaging experiment, the final representation is typically expressed as a three-dimensional image cube, or hypercube, consisting of two spatial dimensions and one spectral dimension. The image cube concept, shown in Figure 56, is interpreted in one of two ways, either (1) a series of images collected as a function of wavenumber (Figure 56 A), or (2) a vibrational spectrum corresponding to each spatial location,

or pixel, within the image plane (Figure 56 B). The ability to merge and recall, within a single analytical technique, information corresponding to the spatial and spectral axes of the image cube provides extraordinary flexibility in probing and extracting structural and compositional information from samples. Thus, a data set may be either summarized as a single image or expressed as a series of images derived from one or more spectroscopic parameters, such as spectral peak intensity, band frequency, or linewidth. The chemical and morphological interpretations of the resulting images are then based on the spatial variations of the intrinsic molecular property being highlighted and their relevance to the biological questions being pursued. In contrast to images generated by conventional point mapping, which generally suffer from low spatial resolution due to relatively few sampling points, direct imaging approaches using the advantages inherent in two-dimensional array detectors derive an enormous benefit from the highly parallel nature of the data collection. That is, a single data set which typically contains many tens of thousands of independent spectral channels may be analyzed and interpreted by using a variety of statistical algorithms, allowing subtle spatial and spectral variations that are often overlooked or misinterpreted in traditional spectroscopic studies to be tested and revealed.

Vibrational spectroscopic imaging has been widely applied to diverse materials, including polymers, semiconductors, pharmaceuticals, cosmetics, consumer products, and biological materials. The strengths and adaptability of vibrational spectroscopic imaging rest not only on the ability to determine chemical compositions and component distribution within a sample, but also on being able to extract localized molecular information relevant to the sample architecture. For example, vibrational spectra are sensitive to alterations in protein secondary and tertiary structures. Spectral shifts could therefore be used to image specifically the distributions of specific structural moieties within a biological sample as opposed to measuring simply an overall distribution of a general class of molecules. Multivariate approaches may also be used to generate a composite metric indicative of either disease progression or biological function. Although these metrics can not be readily interpreted in terms of biochemical changes, they may often be statistically correlated with disease. In addition, by simultaneously recording data on a two-dimensional focal-plane array detector from a myriad of spatial locations
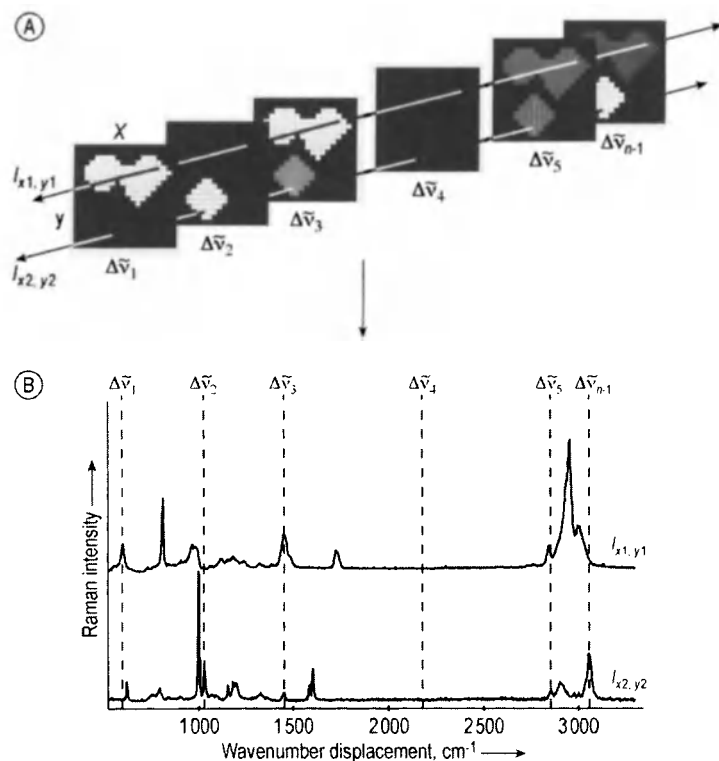
Figure 56. Conceptualization of a three-dimensional hyperspectral image data cube
(A) Images as a function of wavenumber: B) Selected spectra associated with pixels (i, j) in the image
(Reproduced by permission of Marcel Dekker, Inc., 270 Madison Avenue, New York, NY 10016-0602)

within a biological matrix, pooled spectra may be treated statistically. In this manner, analysis of variance can be employed to more robustly test the significance of observed spectral changes appearing, for example, because of a perturbed or diseased state.

Raman and infrared imaging techniques as well as their biological applications are described in detail in [170].

17.5.6. Infrared and Raman Spectroscopy of Polymers

As polymer analysis is a discipline of its own beyond the scope of this article, the reader is referred to the comprehensive works by HUMMEL [171], KOENIG [172], and ZERBI et al. [173].

17.6. Near-Infrared Spectroscopy

17.6.1. Comparison of Mid-Infrared and Near-Infrared Spectroscopy

Near-infrared spectroscopy as a valuable tool for qualitative and quantitative analysis [174], [175] has experienced a revival through the development of chemometric methods and the introduction of fiber optics during the 1980s, and high-quality near-infrared spectra are provided by improved conventional and by Fourier transform instruments.

Various aspects differentiate a mid-infrared spectrum from the corresponding near-infrared spectrum (Table 6). In the mid-infrared region fundamental molecular vibrations always occur between 4000 and 200 cm^{-1}, while in the near-infrared region overtones and combination bands of these fundamentals are observed between 12 800 and 4000 cm^{-1}. Near-infrared bands are due primarily to hydrogenic stretches of C–H, N–H, and

Table 6. Comparison of mid-infrared and near-infrared spectroscopy

Mid-Infrared	Near-Infrared
4000 – 200 cm^{-1}	12 800 – 4000 cm^{-1}
Fundamentals	overtones and combinations
High intensity	low intensity
High sensitivity (trace analysis possible)	low sensitivity (not suited for trace analysis)
Sharp, separated bands	strongly overlapping bands
Easy quantitation with isolated bands	quantitation complex, chemometrics necessary
FT-instruments	grating, filter, FT and acousto-optical tunable scanning instruments

O–H bonds whose fundamental vibrations give rise to strong bands between 4000 and 2000 cm^{-1}, so that their overtones and combinations occur above 4000 cm^{-1} as the most intense absorption bands in the near-infrared spectrum. Since the near-infrared absorptions of polyatomic molecules thus mainly reflect vibrational contributions from very few functional groups, near-infrared spectroscopy is less suitable for detailed qualitative analysis than mid-infrared spectroscopy, which shows all (active) fundamentals and the overtones and combinations of low-energy vibrations. The near-infrared overtone and combination bands depend more on their environment than does the fundamental of the same vibration; slight perturbation in the bonding scheme causes only small changes in the fundamental but significant frequency shifts and amplitude changes in the near-infrared. In going from the fundamental to the first overtone the intensity of an absorption band decreases by a factor of about 10 – 100, so that the sensitivity of near-infrared spectroscopy is reduced in comparison with the mid-infrared. This is a disadvantage when gases are to be measured, but for liquids, it is a considerable advantage as regards sample handling because cells with convenient path lengths between 1 mm and 10 cm can be used. Glass or quartz is used as window material. Near-infrared bands of liquids and solids have relatively large bandwidths between 30 and 60 cm^{-1}, so that they strongly overlap and direct assignment of bands is generally not possible for larger, complex molecules. This is why, in the near-infrared, easy quantitative analysis using isolated bands, as in the mid-infrared, is not possible. Chemometrics [37], [176]–[178], however, by the application of mathematical and statistical procedures, generates correlations between experimental data and the chemical composition or physical properties of the sample, and can be used in a general manner to solve quantitative and qualitative analytical problems.

17.6.2. Applications of Near-Infrared Spectroscopy

Applications of near-infrared spectroscopy include chemistry [179], [180], the oil industry [181], clinical analysis [182], biological and medical analysis [183], and the pharmaceutical industry [184], [185].

Since the intensities of characteristic near-infrared bands are independent of or only slightly dependent on the state of the system, near-infrared spectroscopy is widely applicable for the quantitative study of liquid and compressed gaseous systems, including fluid systems, up to high pressures and temperatures.

The application of near-infrared data to routine quantitative analysis was initiated by NORRIS [186], who used diffuse reflectance measurements to quantitatively determine major components, such as moisture, protein and fat, in agricultural commodities. In comparison with mid-infrared, near-infrared diffuse reflectance [187] is able to measure powdered samples with minimal sample preparation, and lends itself extremely well to quantitative analysis because the smaller absorptivities and larger scattering coefficients at shorter wavelengths result in a larger scattering/absorption ratio. Although near-infrared reflectance analysis was initially developed for agricultural and food applications [188], in combination with chemometrics it is now applied to many other areas, e.g., polymers [189], pharmaceuticals [190], organic substances [191], geological samples [192], and thin-layer chromatography [193], [194].

The excellent transmittance of quartz in the near-infrared range has led to a further enhancement of the potential of near-infrared spectroscopy by the introduction of fiber optics. Fiber-optic waveguides are used to transfer light from the spectrometer to the sample and, after transmission or reflection, back to the spectrometer. Most fiber optic cables consist of three concentric compo-

nents: the inner part, through which the light propagates, is called the core, the middle layer is the cladding, and the outer protective layer is the jacket. Crucial to the light throughput of the fiber optic cable are the optical properties of the core and cladding. The difference in refractive index of the two materials defines the highest angle at which the core/cladding interface exhibits total internal reflection, and hence allows the light to propagate through the cable. This angle is related to the numerical aperture (N.A.) of the fiber by the following equation:

$$\sin\alpha = \text{N.A.} = \left(n_{core}^2 - n_{cladding}^2\right)^{1/2} \qquad (20)$$

where α is the half-angle of acceptance and n is the refractive index of the material. The greater the difference in refractive index the greater the half-angle of acceptance. Any light entering at an angle greater than the half-angle of acceptance is lost to the cladding through absorption. The core and cladding are usually of quartz (of different refractive index), while the outer protective layer is a polymer; other fiber optic materials are zirconium fluoride (ZrF) and chalcogenide (AsGeSe) [195]. The obvious advantage of the optical fiber technique is that the location of measurement can be separated from the spectrometer by between two and several hundred meters, depending on the length of the waveguide. Since almost all practically relevant substances show characteristic near-infrared absorption bands, quantitative analysis by near-infrared spectroscopy is generally applicable to on-line concentration measurements in connection with chemical reactions, chemical equilibria, and phase equilibria. Various probes can be integrated into different reactors or bypasses and offer numerous remote-sensing and on-line process control applications [196]. In the analysis of hazardous or toxic materials this has proved of outstanding importance. In the refining, petrochemical, polymer, plastics, and chemicals industries, typical applications of near-infrared remote spectroscopy are the determination of the octane number and oxygenates in gasoline, aromatics in fuels, the composition of solvent mixtures [197], [198], the hydroxyl number of polyols, low contents of water in solvents, and for polymer analysis.

17.7. References

[1] W. W. Coblentz: *Investigations of Infrared Spectra*. Carnegie Institution, Washington 1905, re-published. The Coblentz Society, Norwalk 1962.

[2] G. Herzberg: *Molecular Spectra and Molecular Structure I*, Van Nostrand, Princeton 1953.

[3] G. Herzberg: *Molecular Spectra and Molecular Structure II*, Van Nostrand, Princeton 1960.

[4] E. B. Wilson, Jr., J. C. Decius, P. C. Cross: *Molecular Vibrations*. McGraw-Hill, New York 1955, re-published Dover, Mineola 1980.

[5] D. Steele: *Theory of Vibrational Spectroscopy*. Saunders, Philadelphia 1971.

[6] J. A. Koningstein: *Introduction to the Theory of the Raman Effect*, Reidel, Dordrecht 1972.

[7] D. A. Long: *Raman Spectroscopy*, McGraw-Hill, New York 1977.

[8] P. R. Griffiths, J. A. de Haseth: *Fourier Transform Infrared Spectroscopy*, J. Wiley & Sons, New York 1986.

[9] M. W. Mackenzie: *Advances in Applied Fourier Transform Infrared Spectroscopy*, J. Wiley & Sons, New York 1988.

[10] N. B. Colthup, L. H. Daly, S. E. Wiberley: *Introduction to Infrared and Raman Spectroscopy*, 3rd ed., Academic Press, San Diego 1990.

[11] B. Schrader (ed.): *Infrared and Raman Spectroscopy*, VCH Verlagsgesellschaft, Weinheim 1995.

[12] K. W. F. Kohlrausch: *Ramanspektren*, Akademische Verlagsgesellschaft, Leipzig 1943, re-published Heyden & Son, London 1973.

[13] H. Baranska, A. Labudzinska, T. Terpinski: *Laser Raman Spectroscopy*, J. Wiley & Sons, New York 1983.

[14] P. Hendra, C. Jones, G. Warnes: *Fourier Transform Raman Spectroscopy*, Ellis Horwood, Chichester 1991.

[15] J. R. Ferraro, K. Krishnan (eds.): *Practical Fourier Transform Infrared Spectroscopy*, Academic Press, San Diego 1990.

[16] M. J. Pelletier (ed.): *Analytical Applications of Raman Spectroscopy*. Blackwell Science, Oxford 1999.

[17] J. G. Grasselli, B. J. Bulkin (eds.): *Analytical Raman Spectroscopy*, J. Wiley & Sons, New York 1991.

[18] G. Socrates: *Infrared Characteristic Group Frequencies*, J. Wiley & Sons, New York 1980.

[19] L. J. Bellamy: *The Infrared Spectra of Complex Molecules*, 3rd ed., vol. 1, Chapman & Hall, London 1975.

[20] L. J. Bellamy: *The Infrared Spectra of Complex Molecules*, 2nd ed., vol. 2, Chapman & Hall, London 1980.

[21] F. R. Dollish, W. G. Fateley, F. F. Bentley: *Characteristic Raman Frequencies of Organic Compounds*, Wiley Interscience, New York 1974.

[22] D. Lin-Vien, N. B. Colthup, W. G. Fateley, J. G. Grasselli: *The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*, Academic Press, Boston 1991.

[23] K. Nakamoto: *Infrared and Raman Spectroscopy of Inorganic and Coordination Compounds*, 5th ed., J. Wiley & Sons, Chichester 1997.

[24] H.-U. Gremlich, B. Yan (eds.): *Infrared and Raman Spectroscopy of Biological Materials*, Marcel Dekker, New York 2000.

[25] B. Schrader: *Raman/Infrared Atlas of Organic Compounds*, 2nd ed., VCH Verlagsgesellschaft, Weinheim 1989.

[26] Sadtler IR Digital Spectra Libraries, Sadtler Division, Bio-Rad Laboratories, Inc., 3316 Spring Garden Street, Philadelphia, PA 19104.

[27] F. M. Mirabella (ed.): *Modern Techniques in Applied Molecular Spectroscopy,* J. Wiley & Sons, New York 1998.

[28] A. A. Michelson, *Philos. Mag.* **31** (1891) 256–259.

[29] W. Herres, J. Gronholz, *Comput. Appl. Lab.* **4** (1984) 216–220.

[30] J. W. Cooley, J. W. Tukey, *Math. Comput.* **19** (1965) 297–301.

[31] J. Gronholz, W. Herres, *Instrum. Computers* **3** (1985) 10–19.

[32] B. Schrader, A. Hoffmann, A. Simon, J. Sawatzki, *Vib. Spectrosc.* **1** (1991) 239–250.

[33] B. Schrader in J. R. Ferraro, K. Krishnan (eds.): *Practical Fourier Transform Infrared Spectroscopy,* Academic Press, San Diego 1990, p. 167.

[34] B. Schrader, A. Hoffmann, S. Keller, *Spectrochim. Acta* **47 A** (1991) 1135–1148.

[35] J. Coates in W. O. George, H. A. Willis (eds.): *Computer Methods in UV, Visible and IR Spectroscopy,* The Royal Society of Chemistry, Cambridge 1990, p. 95.

[36] P. S. Wilson in W. O. George, H. A. Willis (eds.): *Computer Methods in UV, Visible and IR Spectroscopy.* The Royal Society of Chemistry, Cambridge 1990, p. 143.

[37] F. Despagne, D. L. Massart, *Analyst* **123** (1998) 157R–178R.

[38] G. M. Barrow: *Introduction to Molecular Spectroscopy.* McGraw-Hill, New York 1962.

[39] S. R. La Paglia: *Introductory Quantum Chemistry.* Harper & Row, New York 1971.

[40] *Ullmann,* 4th ed., **5,** 303.

[41] Available from Sadtler Division, Bio-Rad Laboratories, Inc., 3316 Spring Garden Street, Philadelphia, PA 19104.

[42] Available from LabControl GmbH, D-50858 Köln 40, Max-Planck-Str. 17 a, Germany.

[43] Available from Galactic Industries Corporation, 395 Main Street, Salem NH 03079.

[44] J. P. Coates, R. W. Hannah in T. Theophanides (ed.): *Fourier Transform Infrared Spectroscopy.* D. Reidel Publ., Dordrecht 1984, p. 167.

[45] S. R. Lowry, D. A. Hupplar, *Anal. Chem.* **53** (1981) 889–893.

[46] J. T. Clerc in W. O. George, H. A. Willis (eds.): *Computer Methods in UV, Visible and IR Spectroscopy.* The Royal Society of Chemistry, Cambridge 1990, p. 13.

[47] D. A. Cirovic, *Trends Anal. Chem.* **16** (1997) 148–155.

[48] H. Yang, P. R. Griffiths, *Anal. Chem.* **71** (1999) 751–761.

[49] J. U. White, *J. Opt. Soc. Am.* **32** (1942) 285–288.

[50] R. T. Graf, J. L. Koenig, H. Ishida, *Appl. Spectrosc.* **39** (1985) 405–412.

[51] R. G. Greenler, *J. Chem. Phys.* **44** (1966) 310–315.

[52] R. G. Greenler, *J. Chem. Phys.* **50** (1969) 1963–1968.

[53] D. L. Allara, J. D. Swalen, *J. Phys. Chem.* **86** (1982) 2700–2704.

[54] N. J. Harrick, *Appl. Opt.* **10** (1971) 2344–2349.

[55] I. Newton: *Opticks,* Dover, New York 1952.

[56] N. J. Harrick: *Internal Reflection Spectroscopy.* Harrick Scientific Corporation, Ossining, New York 1979.

[57] J. Fahrenfort in A. Mangini (ed.): *Adv. Mol. Spectrosc. Proc. Int. Meet. 4th 1959,* vol. **2,** 1962, p. 701.

[58] Available from Harrick Scientific Corporation, 88 Broadway, Ossining, NY 10562.

[59] N. J. Harrick, M. Milosevic, S. L. Berets, *Appl. Spectrosc.* **45** (1991) 944–948.

[60] S. Compton, P. Stout, FTS 7/IR Notes no. 1, Bio-Rad, Digilab Division, Cambridge, MA 02139, 1991.

[61] U. P. Fringeli in F. M. Mirabella (ed.): *Internal Reflection Spectroscopy: Theory and Application,* Marcel Dekker, New York 1992, p. 255.

[62] U. P. Fringeli, *Chimia* **46** (1992) 200–214.

[63] Available from Spectra-Tech Inc., Stamford, CT 06906.

[64] Available from GRASEBY SPECAC Ltd., Lagoon Road, St. Mary Cray, Orpington, Kent BR5 3QX, U.K.

[65] W. M. Doyle, *Appl. Spectrosc.* **44** (1990) 50–69.

[66] B. E. Miller, N. D. Danielson, J. E. Katon, *Appl. Spectrosc.* **42** (1988) 401–405.

[67] Available from Axiom Analytical, Inc., 17751-C Sky Park Circle, Irvine, CA 92614.

[68] W. M. Doyle, *Process Control Quality* **2** (1992) 11–41.

[69] N. L. Owen, D. W. Thomas, *Appl. Spectrosc.* **43** (1989) 451–455.

[70] P. Kubelka, F. Munk, *Z. Tech. Phys.* **12** (1931) 593–601.

[71] P. Kortüm, W. Braun, G. Herzog, *Angew. Chem., Int. Ed. Engl.* **2** (1963) 333–341.

[72] P. H. Turner, W. Herres, BRUKER FT-IR Application Note 23, Bruker Analytische Meßtechnik GmbH, D-76153 Karlsruhe, Germany.

[73] D. Reinecke, A. Jansen, F. Fister, U. Schernau, *Anal. Chem.* **60** (1988) 1221–1224.

[74] M. P. Fuller, P. R. Griffiths, *Anal. Chem.* **50** (1978) 1906–1910.

[75] M. Fuller in *Nicolet FT-IR Spectral Lines,* Nicolet Analytical Instruments, Madison, WI 53711, Spring/Summer 1990, p. 15.

[76] G. Glauninger, K.-A. Kovar, V. Hoffmann, *Fresenius J. Anal. Chem.* **338** (1990) 710–716.

[77] H. Yamamoto, K. Wada, T. Tajima, K. Ichimura, *Appl. Spectrosc.* **45** (1991) 253–259.

[78] S. A. Johnson, R. M. Rinkus, T. C. Diebold, V. A. Maroni, *Appl. Spectrosc.* **42** (1988) 1369–1375.

[79] K. A. Martin, R. F. Zabransky, *Appl. Spectrosc.* **45** (1991) 68–72.

[80] Si-Carb Sampler, available from Spectra-Tech Inc., Stamford, CT 06906.

[81] F. J. DeBlase, S. Compton, *Appl. Spectrosc.* **45** (1991) 611–618.

[82] A. M. Vassallo, P. A. Cole-Clarke, L. S. K. Pang, A. J. Palmisano, *Appl. Spectrosc.* **46** (1992) 73–78.

[83] M. Handke, N. J. Harrick, *Appl. Spectrosc.* **40** (1986) 401–405.

[84] R. T. Rewick, R. G. Messerschmidt, *Appl. Spectrosc.* **45** (1991) 297–301.

[85] L. T. Lin, D. D. Archibald, D. E. Honigs, *Appl. Spectrosc.* **42** (1988) 477–483.

[86] A. Tsuge, Y. Uwamino, T. Ishizuka, *Appl. Spectrosc.* **43** (1989) 1145–1149.

[87] R. W. Jones, J. F. McClelland, *Anal. Chem.* **61** (1989) 650–656.

[88] R. W. Jones, J. F. McClelland, *Anal. Chem.* **61** (1989) 1810–1815.

[89] R. J. Pell, B. C. Erickson, R. W. Hannah, J. B. Callis, B. R. Kowalski, *Anal. Chem.* **60** (1988) 2824–2827.

[90] J. A. McGuire, B. Wangmaneerat, T. M. Niemczyk, D. M. Haaland, *Appl. Spectrosc.* **46** (1992) 178–180.

[91] B. Wangmaneerat et al., *Appl. Spectrosc.* **46** (1992) 340–344.

[92] A. Rosencwaig: *Photoacoustics and Photoacoustic Spectroscopy*. J. Wiley & Sons, New York 1980.

[93] J. F. McClelland, *Anal. Chem.* **55** (1983) 89 A – 105 A.

[94] J. A. Graham, W. M. Grim III, W. G. Fateley in J. R. Ferraro, L. J. Basile (eds.): *Fourier Transform Infrared Spectroscopy*, **vol. 4,** Academic Press, Orlando 1985, p. 345.

[95] H. Coufal, J. F. McClelland, *J. Mol. Struct.* **173** (1988) 129 – 140.

[96] G. Laufer, J. T. Huneke, B. S. H. Royce, Y. C. Teng, *Appl. Phys. Lett.* **37** (1980) 517 – 519.

[97] G. Busse, B. Bullemer, *Infrared Phys.* **18** (1978) 631 – 634.

[98] MTEC Photoacoustics, Inc., Ames, IA 50010.

[99] R. O. Carter III, S. L. Wright, *Appl. Spectrosc.* **45** (1991) 1101 – 1103.

[100] C. Q. Yang, *Appl. Spectrosc.* **45** (1991) 102 – 108.

[101] J. C. Donini, K. H. Michaelian, *Infrared Phys.* **24** (1984) 157 – 163.

[102] M. W. Urban, J. L. Koenig, *Appl. Spectrosc.* **40** (1986) 994 – 998.

[103] C. Q. Yang, R. R. Bresee, W. G. Fateley, *Appl. Spectrosc.* **41** (1987) 889 – 896.

[104] R. Rubinovitz, J. Seebode, A. Simon, *Bruker Report* **1** (1988) 35 – 38.

[105] M. J. Smith, C. J. Manning, R. A. Palmer, J. L. Chao, *Appl. Spectrosc.* **42** (1988) 546 – 555.

[106] R. A. Crocombe, S. V. Compton, FTS/IR Notes no. 82, Bio-Rad, Digilab Division, Cambridge, MA 02139, 1991.

[107] K. H. Michaelian, *Appl. Spectrosc.* **45** (1991) 302 – 304.

[108] L. Bertrand, *Appl. Spectrosc.* **42** (1988) 134 – 138.

[109] R. White: "Chromatography/Fourier Transform Infrared Spectroscopy and Its Applications," *Practical Spectroscopy Series*, vol. 10, Marcel Dekker, New York 1990.

[110] W. Herres: *HRGC-FTIR: Capillary Gas Chromatography-Fourier Transform Infrared Spectroscopy*, Hüthig Verlag, Heidelberg 1987.

[111] S. Bourne et al., FTS/IR Notes no. 75, Bio-Rad, Digilab Division, Cambridge, MA 02139, 1990.

[112] R. L. White, G. N. Giss, G. M. Brissey, C. L. Wilkins, *Anal. Chem.* **53** (1981) 1778 – 1782.

[113] P. Coffey, D. R. Mattson, J. C. Wright, *Am. Lab.* **10** (1978) 126 – 132.

[114] K. S. Chiu, K. Biemann, K. Krishnan, S. L. Hill, *Anal. Chem.* **56** (1984) 1610 – 1615.

[115] Available from Lab Connections, Inc., 201 Forest Street, Marlborough, MA 01752.

[116] Available from Bourne Scientific, Inc., 65 Bridge Street, Lexington, MA 02421-7927.

[117] S. Bourne, *Am. Lab.* **16** (1998) 17F – 17J.

[118] C. M. Earnest, *Anal. Chem.* **56** (1984) 1471 A – 1486 A.

[119] D. Dollimore, *Anal. Chem.* **64** (1992) 147 R – 153 R.

[120] J. A. J. Jansen, J. H. van der Maas, A. Posthuma de Boer, *Appl. Spectrosc.* **46** (1992) 88 – 92.

[121] K. Ichimura, H. Ohta, T. Tajima, T. Okino, *Mikrochim. Acta* **1** (1988) 157 – 161.

[122] R. C. Wieboldt, S. R. Lowry, R. J. Rosenthal, *Mikrochim. Acta* **1** (1988) 179 – 182.

[123] J. Mullens et al., *Bull. Soc. Chim. Belg.* **101** (1992) 267 – 277.

[124] R. M. Carangelo, P. R. Solomon, D. J. Gerson, *Fuel* **66** (1987) 960 – 967.

[125] J. K. Whelan, *Energy Fuels* **2** (1988) 65 – 73.

[126] L. A. Sanchez, M. Y. Keating, *Appl. Spectrosc.* **42** (1988) 1253 – 1258.

[127] M. D. Timken, J. K. Chen, T. B. Brill, *Appl. Spectrosc.* **44** (1990) 701 – 706.

[128] E. Post, S. Rahner, H. Möhler, A. Rager, *Thermochim. Acta* **263** (1995) 1 – 6.

[129] M. Maciejewski, F. Eigenmann, A. Baiker, E. Dreser, A. Rager, *Thermochim. Acta* (2000) in press.

[130] L. A. Nafie in M. W. Mackenzie (ed.): *Advances in Applied FTIR Spectroscopy*, John Wiley and Sons, New York 1988, p. 67.

[131] L. A. Nafie, T. B. Freedman in H.-U. Gremlich, B. Yang (eds.): *Infrared and Raman Spectroscopy of Biological Materials*, Marcel Dekker, New York 2000, p. 15.

[132] T. A. Keiderling in H.-U. Gremlich, B. Yang (eds.): *Infrared and Raman Spectroscopy of Biological Materials*, Marcel Dekker, New York 2000, p. 55.

[133] B. Schrader, S. Keller, *Proc. SPIE Int. Soc. Opt. Eng. 1575 (Int. Conf. Fourier Transform Spectrosc., 8th, 1991)* (1992) 30 – 39.

[134] N. J. Everall, J. Lumsdon, *Spectrosc. Eur.* **4** (1992) 10 – 21.

[135] C. G. Zimba, J. F. Rabolt, *Appl. Spectrosc.* **45** (1991) 162 – 165.

[136] T. Jawhari, P. J. Hedra, H. A. Willis, M. Judkins, *Spectrochim. Acta* **46 A** (1990) 161 – 170.

[137] F. T. Walder, M. J. Smith, *Spectrochim. Acta* **47 A** (1991) 1201 – 1216.

[138] H. Sadeghi-Jorabchi, R. H. Wilson, P. S. Belton, J. D. Edwards-Webb, D. T. Coxon, *Spectrochim. Acta* **47 A** (1991) 1449 – 1458.

[139] L. A Lyon et al., *Anal. Chem.* **70** (1998) 341R – 361R.

[140] B. Chase in A. M. C. Davies, C. S. Creaser (eds.): *Anal. Appl. Spectrosc. 2*. The Royal Society of Chemistry, Cambridge 1991, p. 13.

[141] K. P. J. Williams, S. M. Mason, *TRAC* **9** (1990) 119 – 127.

[142] P. J. Hendra, *J. Mol. Struct.* **266** (1992) 97 – 114.

[143] A. Hoffmann et al., *J. Raman Spectrosc.* **22** (1991) 497 – 503.

[144] P. J. Hendra, C. H. Jones, *Makromol. Chem. Macromol. Symp.* **52** (1991) 41 – 56.

[145] G. J. Thomas, *Ann. Rev. Biophys. Biomol. Struct.* **28** (1999) 1 – 27.

[146] A. M. Tudor et al., *J. Pharm. Biomed. Anal.* **8** (1990) 717 – 720.

[147] P. R. Carey, *J. Biol. Chem.* **274** (1999) 26 625 – 26 628.

[148] J. Góral, V. Zichy, *Spectrochim. Acta* **46 A** (1990) 253 – 275.

[149] I. W. Levin, E. N. Lewis, *Anal. Chem.* **62** (1990) 1101 A – 1111 A.

[150] S. Nie et al., *Spectroscopy* **5** (1990) 24 – 32.

[151] G. Ellis, M. Claybourn, S. E. Richards, *Spectrochim. Acta* **46 A** (1990) 227 – 241.

[152] D. Bourgeois, S. P. Church, *Spectrochim. Acta* **46 A** (1990) 295 – 301.

[153] T. N. Day, P. J. Hendra, A. J. Rest, A. J. Rowlands, *Spectrochim. Acta* **47 A** (1991) 1251 – 1262.

[154] N. J. Everall, J. M. Chalmers, I. D. Newton, *Appl. Spectrosc.* **46** (1992) 597 – 601.

[155] D. L. Gerrard, J. Birnie, *Anal. Chem.* **64** (1992) 502 R – 513 R.

[156] B. Yan, *Acc. Chem. Res.* **31** (1998) 621 – 630.

[157] R. G. Messerschmidt, M. A. Harthcock (eds.): *Infrared Microspectroscopy: Theory and Applications*, Marcel Dekker, New York 1988.

[158] K. Krishnan, S. L. Hill in J. R. Ferraro, K. Krishnan (eds.): *Practical Fourier Transform Spectroscopy*, Academic Press, San Diego 1990, p. 103.

[159] S. Compton, J. Powell, FTS 7/IR Notes no. 2, Bio-Rad, Digilab Division, Cambridge, MA 02139, 1991.

[160] N. J. Harrick, M. Milosevic, S. L. Berets, *Appl. Spectrosc.* **45** (1991) 944–948.

[161] B. Schrader, *Fresenius J. Anal. Chem.* **337** (1990) 824–829.

[162] J. D. Louden in D. J. Gardiner, P. R. Graves (eds.): *Practical Raman Spectroscopy*, Springer Verlag, Berlin 1989, p. 119.

[163] J. Sawatzki, *Fresenius J. Anal. Chem.* **339** (1991) 267–270.

[164] A. Hoffmann, B. Schrader, R. Podschadlowski, A. Simon, *Proc. SPIE-Int. Soc. Opt. Eng. 1145 (Int. Conf. Fourier Transform Spectrosc., 7th, 1989)* (1989) 372–373.

[165] F. Bergin, *Proc. Annu. Conf. Microbeam Anal. Soc.* **25** (1990) 235–239.

[166] B. W. Cook in A. M. C. Davies, C. S. Creaser (eds.): *Anal. Appl. Spectrosc.* **2**, The Royal Society of Chemistry, Cambridge 1991, p. 61.

[167] A. J. Sommer, J. E. Katon, *Appl. Spectrosc.* **45** (1991) 527–534.

[168] D. Gernet, W. Kiefer, H. Schmidt, *Appl. Spectrosc.* **46** (1992) 571–576.

[169] K. Gerwert in H.-U. Gremlich, B. Yang (eds.): *Infrared and Raman Spectroscopy of Biological Materials*, Marcel Dekker, New York 2000, p. 193.

[170] M. D. Schaeberle, I. W. Levin, E. N. Lewis in H.-U. Gremlich, B. Yang (eds.): *Infrared and Raman Spectroscopy of Biological Materials*, Marcel Dekker, New York 2000, p. 231.

[171] D. O. Hummel: *Atlas of Polymer and Plastics Analysis*, 3rd ed., VCH Verlagsgesellschaft, Weinheim 1991.

[172] J. L. Koenig: *Spectroscopy of Polmers*, American Chemical Society, Washington, D.C. 1992.

[173] G. Zerbi et al.: *Modern Polymer Spectroscopy*, Wiley-VCH, Weinheim 1998.

[174] J. Workman, Jr., *J. Near Infrared Spectrosc.* **1** (1993) 221.

[175] J. Murray, I. A. Cowe (eds.): *Making Light Work: Advances in Near Infrared Spectroscopy*, VCH Verlagsgesellschaft, Weinheim 1992.

[176] P. Geladi, E. Dabbak, *J. Near Infrared Spectrosc.* **3** (1995) 119.

[177] D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. de Jong, P. J. Lewi, J. Smeyers-Verbecke: *Handbook of Chemometrics and Qualimetrics, Part A*, Elsevier, Amsterdam 1997.

[178] D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. de Jong, P. J. Lewi, J. Smeyers-Verbecke: *Handbook of Chemometrics and Qualimetrics, Part B*, Elsevier, Amsterdam 1998.

[179] S. M. Donahue, C. W. Brown, B. Caputo, M. D. Modell, *Anal. Chem.* **60** (1988) 1873–1878.

[180] J. Lin, C. W. Brown, *Vibr. Spectrosc.* **7** (1994) 117–123.

[181] S. R. Westbrook, *SAE Tech. Paper Ser.* **930734** (1993) 1–10.

[182] H. M. Heise in H.-U. Gremlich, B. Yan (eds.): *Infrared and Raman Spectroscopy of Biological Materials*, Marcel Dekker, New York 2000, p. 259.

[183] R. J. Dempsey, D. G. Davis, R. G. Buice, Jr., R. A. Lodder, *Appl. Spectrosc.* **50** (1996) 18A–34A.

[184] J. D. Kirsch, J. K. Drennen, *Appl. Spectrosc. Rev.* **30** (1995) 139–174.

[185] W. Plugge, C. van der Vlies, *J. Pharm. Biomed. Anal.* **14** (1996) 891–898.

[186] I. Ben-Gera, K. H. Norris, *Israel J. Agric. Res.* **18** (1968) 125–132.

[187] D. L. Wetzel, *Anal. Chem.* **55** (1983) 1165 A–1176 A.

[188] P. Williams, K. Norris (eds.): *Near-Infrared Technology in the Agricultural and Food Industries*, AACC, St. Paul 1987.

[189] C. E. Miller, *Appl. Spectrosc. Rev.* **26** (1991) 277–339.

[190] B. F. MacDonald, K. A. Prebble, *J. Pharm. Biomed. Anal.* **11** (1993) 1077–1085.

[191] L. G. Weyer, *Appl. Spectrosc. Rev.* **21** (1985) 1–43.

[192] D. E. Honigs, T. B. Hirschfeld, G. M. Hieftje, *Appl. Spectrosc.* **39** (1985) 1062–1065.

[193] E. W. Ciurczak, L. J. Cline-Love, D. M. Mustillo, *Spectrosc. Int.* **3** (1991) no. 5, 39–42.

[194] E. W. Ciurczak, W. R. Murphy, D. M. Mustillo, *Spectrosc. Int.* **3** (1991) no. 7, 39–44.

[195] M. J. Smith, T. E. May, *Int. Lab.* **22** (1992) 18–24.

[196] L. P. McDermott, *Ad. Instrum. Control.* **45** (1990) 669–677.

[197] Guided Wave, Inc., Application Note no. A3-987, El Dorado Hills, CA 95630.

[198] Guided Wave, Inc., Application Note no. A4-188, El Dorado Hills, CA 95630.

18. Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy

REINHARD MEUSINGER, Institute of Organic Chemsitry, University of Mainz, Mainz, Federal Republic of Germany

A. MARGARET CHIPPENDALE, ZENECA Specialties, Manchester, United Kingdom

SHIRLEY A. FAIRHURST, AFRC Institute for Plant Science Research, Brighton, United Kingdom

Symbols

a	ESR isotropic hyperfine coupling constant; ESR total hyperfine interaction
AT	acquisition time
B	magnetic field
DW	dwell time
g	electronic or nuclear g factor
h	Planck constant
I	nuclear spin quantum number
J	spin–spin coupling constant
k	Boltzmann constant
K	intensity of ESR response
M	magnetization vector
m_I, m_S	magnetic quantum numbers
N	Avogadro number
s	shielding constant
S	electronic spin quantum number
SW	spectrum width
T	ESR anisotropic hyperfine coupling
T_1	spin–lattice relaxation time
T_2	spin–spin relaxation time
T_d	relaxation delay
β	electronic Bohr or nuclear magneton
γ	magnetogyric ratio
δ	chemical shift
μ	magnetic dipole moment
τ	rotational correlation time, pulse width
v	resonance frequency
ω_0	Larmor frequency

Units

ppm	parts per million (unit of the dimensionless δ scale of the chemical shift)
T	tesla (unit of magnetic field strength, $1\ \mathrm{T} = 10^4$ gauss)

18.1. Introduction

It is now about 50 years since the phenomena of electron spin resonance (ESR) and nuclear magnetic resonance (NMR) were first discovered. During the last thirty years NMR spectroscopy has developed into perhaps the most important instrumental measuring technique within chemistry. This is due to a dramatic increase in both the sensitivity and the resolution of NMR and ESR instruments. NMR and ESR spectroscopy are used today within practically all branches of chemistry and in related sciences, such as physics, biology,

and medicine, both at universities and in industrial laboratories. NMR has its most important applications in the determination of molecular structures. It can today be applied to a wide variety of chemical systems, from small molecules to proteins and nucleic acids. ESR has played a key role in the study of free radicals, many transition metal compounds, and other species containing unpaired electrons.

The continuing development of each technique has resulted from a series of step changes. Both methods were first developed using continuous wave methodology. By the late 1960s the pulsed Fourier transform techniques were developed and additionally the NMR spectroscopists began to use magnets based on superconducting material. On grounds of these developments the quality of spectra, expressed both in terms of sensitivity and resolution, improved quickly during the 1970s. The increase in sensitivity has made it possible to study small amounts of material as well as chemically interesting isotopes of low natural occurrence, e.g., carbon-13 (^{13}C NMR), routinely. Pulsed ESR techniques have since followed.

Since the 1970s two-dimensional NMR methods and more recently three- and multi-dimensional varieties have allowed ever more complex molecules to be studied. In the last two decades an additional increase of sensitivity was obtained by inverse detection techniques simultaneously combined with a drastic shortening of measuring time by gradient enhanced spectroscopy. In the area of solid states NMR techniques have been developed that obtain solution-like spectra from solids. Superconducting magnet systems for NMR frequencies up to 900 MHz (21.13 T) allow now the elucidation of the structure of large proteins and make NMR spectroscopy invaluable for the development of new drugs.

Since the mid-1970s NMR and ESR imaging has also been developed. The construction of whole-body magnets with magnetic field strength of up to 8.0 T and with magnet bores of up to 90 cm has paved the way for investigations on humans and NMR imaging today is an indispensable aid for medical diagnosis.

Table 1. Nuclear properties

Mass number	Atomic number	Nuclear spin I
Odd	even or odd	$1/2, 3/2, 5/2 \ldots$
Even	even	0
Even	odd	$1, 2, 3, \ldots$

18.2. Principles of Magnetic Resonance [1], [92]–[96]

Many of the principles of magnetic resonance are common to both ESR and NMR spectroscopy. To exhibit magnetic resonance the system concerned must have a magnetic moment μ. For NMR this is provided by a nucleus with nonzero nuclear spin, and for ESR by an unpaired electron in an atom, radical, or compound. The two phenomena differ in the signs and magnitudes of the magnetic interactions involved. When matter is placed in a strong magnetic field, these systems (nuclei or electrons) behave like microscopic compass needles. According to the laws of quantum mechanics these tiny compass needles can orient themselves with respect to the magnetic field only in a few ways. These orientations are characterized by different energy levels. If the sample is exposed to electromagnetic waves of an exactly specified frequency the nuclear spins or the unpaired electrons can be forced to jump between different energy levels. In magnetic fields with a field strength of a few teslas, the resonance frequencies for the NMR and ESR are in the radio and microwave frequency regions, respectively. During the experiment the frequency is varied. When it exactly matches the characteristic frequency of the nuclei or the electrons an electric signal is induced in the detector. The strength of the signal is plotted as a function of this resonance frequency in a diagram called the NMR or ESR spectrum, respectively.

18.2.1. Nuclear and Electronic Properties

The nuclear spin I of an atomic nucleus has values of $0, 1/2, 1, 3/2$, etc. in units of $h/2\pi$, where h is the Planck constant. The value for a given nucleus is dependent on the mass and atomic numbers of the nucleus (Table 1). Thus some important, common nuclei such as ^{12}C and ^{16}O which have both even mass and atomic numbers have zero spin. An electron has a spin S of $1/2$. Nuclear

(μ_N) and electronic (μ_e) magnetic moments are proportional to the magnitude of the spin:

$$\mu_N = \gamma_N \frac{hI}{2\pi}$$

where γ_N is the magnetogyric ratio of the nucleus. In SI units μ_N is in amperes meter squared and γ_N in radian per tesla per second. Alternatively the magnetic moment can be expressed as:

$$\mu_N = g_N \beta_N I$$

where g_N is a dimensionless constant (nuclear g factor) and β_N the nuclear magneton:

$$\beta_N = \frac{eh}{4\pi Mc}$$

where e and M are the charge and mass of the proton, and c the velocity of light. The quantities g_N and I distinguish nuclei from one another.

The magnetic moment of an electron is given by an analogous expression:

$$\mu_e = -g\beta S$$

where g is the dimensionless electron g factor, and β the electronic Bohr magneton:

$$\beta = -\frac{eh}{4\pi mc}$$

where $-e$ and m are the charge and mass of the electron. Because the electron is negatively charged, its moment is negative.

18.2.2. Nuclei and Electrons in a Stationary Magnetic Field

When a nucleus with nonzero spin or an electron is placed in a strong magnetic field B the orientation of the spin becomes quantized, with each possible orientation having a different energy. A nucleus of spin I has $2I+1$ possible orientations, with associated magnetic quantum numbers m_I of $-I, -I+1, \ldots I-1, I$. Hydrogen nuclei (protons) have a spin quantum number $I = 1/2$. Consequently the proton can exist in only two spin states, which are characterized by their magnetic quantum numbers $m_I = 1/2$ and $m_I = -1/2$. The proton can therefore be pictured as a magnetic dipole, the z component μ_z of which can have a parallel or antiparallel orientation with respect to the z direction of the coordinate system
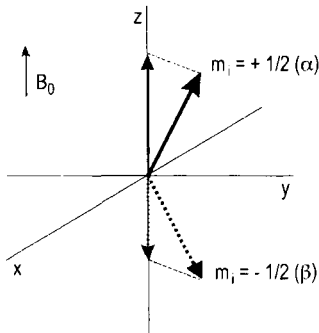
Figure 1. Orientation of the nuclear magnetic moments in the magnetic field B_0.

(Fig. 1). In the following discussion the denotations α and β are used for these two orientation states, where α denotes the lower, more stable energy level and β the higher one. For a single electron with $S = 1/2$, allowed values for the magnetic quantum number m_S are also $+1/2$ and $-1/2$.

In the absence of a magnetic field the α and β states have the same energy. Only in a static magnetic field B_0 is this degeneracy destroyed as a result of the interaction of the nuclear or electronic moment μ with B_0 and an energy difference for the two orientation states results. The energy of interaction is proportional to the nuclear or electronic moment and the applied field. For a nucleus and electron, respectively, this can be written as:

$$E_1 = -g_N\beta_N m_I B_0 = \frac{-\gamma_N h m_I B_0}{2\pi}$$

$$E_S = g\beta m_S B_0$$

Energy level diagrams for nuclei with spins of 1/2 and 3/2 are given in Figure 2.

The energy levels diagram for an electron looks exactly like that of the nucleus in Figure 2 A), but with exchanged signs of the magnetic quantum numbers m_S. The selection rule for NMR and ESR allows transitions for which m_I and m_S change by one unit. Thus:

$$\Delta E_1 = g_N\beta_N B_0 = \frac{\gamma_N h B_0}{2\pi}$$

$$\Delta E_S = g\beta B_0$$

18.2.3. Basic Principles of the NMR and ESR Experiments

To observe transitions, radiation given by $\Delta E = h\nu$ must be applied. The fundamental resonance conditions for NMR and ESR experiments respectively are:

$$h\nu = g_N\beta_N B_0 = \frac{\gamma_N h B_0}{2\pi} \tag{1}$$

and

$$h\nu = g\beta B_0$$

where ν is the resonance frequency. For a proton in a magnetic field of 2.349 T, ν is 100 MHz. The resonance frequency of a free electron in the same field is 65.827 MHz.

Resonance absorption can only be detected if the spin levels involved differ in population. For a very large number N of protons in thermal equilibrium, the distribution of the spin 1/2 nuclei between the ground (α) state and the excited (β) state is given by the Boltzmann relation:

$$\frac{N_\beta}{N_\alpha} = \exp(-h\nu/kT)$$

The excess of spins in the lower state is only ca. 1 in 10^5 for NMR at a magnetic field strength of 9 T and at room temperature. This is the fundamental reason for the low sensitivity of NMR compared with IR and UV spectroscopy and even ESR.

The magnetic resonance experiment itself is best explained and visualized in classical as opposed to quantum mechanical terms. How is the energy transferred to nuclei or electrons aligned in a magnetic field, and how is the energy absorbed measured? The following description is for nuclei of spin 1/2 but is equally applicable to electrons. In a magnetic field B_0, applied in the z-direction, the nuclei have an orientation which is either aligned or opposed to the direction of B_0. Additionally, the nuclei precess randomly around the z-axis with a frequency ω_0, known as the Larmor frequency, where $\omega_0 = 2\pi\nu$. A macroscopic magnetization vector M along the z-axis but not in the x-y plane results from the individual nuclear magnetic moments μ_z defined in Figure 1. The slight excess of nuclei aligned with the magnetic field as shown above, gives rise to the alignment of this magnetization vector M_z in direction of B_0 (Fig. 3).
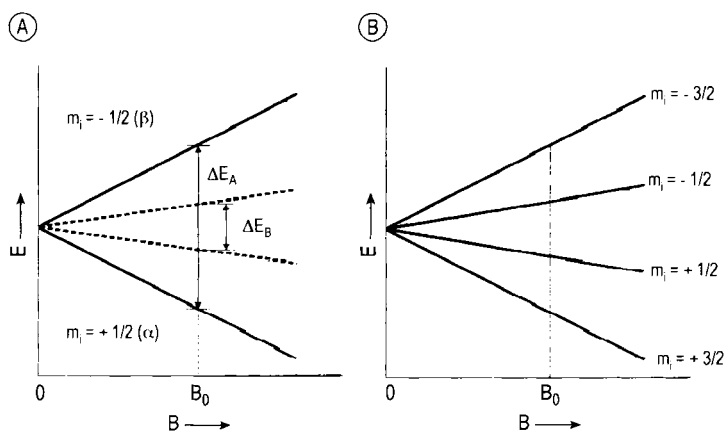
Figure 2. Energy levels for a nucleus with a spin of A) $I = 1/2$, and B) $I = 3/2$. On the left-hand side the levels for two different isotopes A and B with different magnetogyric ratios are represented ($\gamma_A > \gamma_B$). For the nuclei of the isotope B an obviously smaller energy difference ΔE results at the same magnetic field strength B_0. Only the value and not the direction of the magnetic field B_0 is plotted on the abscissas.

The quantum description of the lower and higher energy levels can be equated with the classical description of precessing nuclei aligned with and against the applied field. Up to here the description of the nuclear motions was referred to in the stationary or laboratory frame of reference. If the laboratory could be rotated at the Larmor frequency ω_0, the nuclei would no longer appear to precess but would become stationary and coincident with the magnetic field axis B_0 as shown in Figure 4; x' and y' are the rotating x- and y-axes. The magnetic behavior is now completely described by a magnetization vector M_z acting along B_0. Use of the rotating frame system greatly simplifies the description of pulsed NMR experiments (see Section 18.3.1.2).

In the magnetic resonance experiment the longitudinal net magnetization M_z is tipped toward the $x-y$ plane. Radio-frequency electromagnetic energy is applied such that its magnetic component B_1 is at right angles to B_0 and is rotating with the precessing nuclei. The RF irradiation is applied along the x-axis and, at the resonance frequency, tilts the net magnetization towards the $x-y$ plane. This creates an x-y component, the so-called transverse magnetization, which induces a signal in the detector coil. Conventionally the receiver is located on the y-axis of the Cartesian coordinates (Fig. 4).

To obtain a spectrum, either the RF frequency or the magnetic field B_0 can be scanned over a narrow range. Historically, this mode, known as continuous wave (CW), was employed in both NMR and ESR instruments. For NMR it has been almost completely superceded by the more sensitive and sophisticated pulsed Fourier transform (FT) method. Pulsed methods are now becoming more widely used in ESR.

18.2.4. Relaxation

In the absence of any mechanism enabling the nuclear or electronic spins to return to the ground state, all the excess population in the lower energy state will eventually be raised to the higher energy state, and no more energy will be absorbed. However, such mechanisms do exist, the so-called relaxation phenomena. That is, the magnitudes of both longitudinal and transversal magnetization are time dependent . One is the spin–lattice or longitudinal relaxation process, which occurs by loss of energy from excited nuclear spins to the surrounding molecules. By this process the component of magnetization along the z-axis relaxes back to its original value M_z, with an exponential decay characterized by a time constant T_1, the spin–lattice relaxation time.

After application of the RF energy the magnetization has components in the $x-y$ plane, and these also decay at least as fast as the spin–lattice relaxation returns the magnetization to the z-axis. However, $x-y$ magnetization can also be lost by other processes, which cause the $x-y$ magnetism to dephase to give a net $x-y$ magnetization of zero. The rate constant in the $x-y$ plane for this process is T_2, the spin–spin or transverse relaxation time. Such relaxation involves transfer of
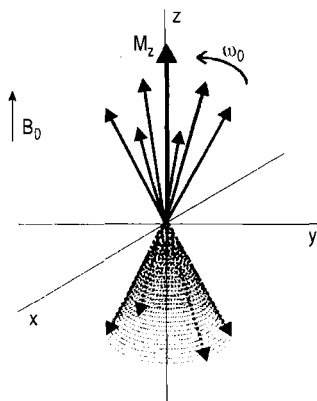
Figure 3. Motion of spin 1/2 nuclei in a magnetic field B_0. Note that more nuclei are arranged parallel than antiparallel to the direction of the static magnetic field B_0.

Figure 4. Motion of spin 1/2 nuclei in the rotating frame and generation of transverse magnetization M_y by rotation of the vector of the macroscopic magnetization M_z. The B_1 field caused by the transmitter coil along the x-axis deflects M_z into the $x'-y'$-plane.

energy from one nucleus to another. An important feature of spin–spin relaxation is that it determines the natural linewidths of the signals in a spectrum, which is defined in terms of the half-height linewidth Δ. In hertz one has:

$$\Delta = \frac{1}{\pi T_2}$$

if T_2 is measured in seconds. However, the observed linewidth is always larger because of a contribution from the magnetic field inhomogeneity.

topes of important spin 1/2 nuclei such as ^{13}C and ^{15}N require FT techniques for observation in natural abundance. Nuclei with spin $I \geq 1$ have a quadrupole moment in addition to their magnetic moment. The interaction of the nuclear quadrupole moment with the electric field gradient at the nucleus provides a very efficient process for nuclear relaxation. Often, the quadrupolar mechanism is dominant, the resulting relaxation times are very short, and the signals broad. In extreme cases (e.g., covalently bonded ^{35}Cl or ^{37}Cl), the signals are too broad to detect. However, nuclei with very small quadrupole moments such as ^{2}H and ^{11}B can be observed in the usual manner as the broadening is much less severe.

18.3. High-Resolution Solution NMR Spectroscopy [2], [3], [4]–[6], [97]–[109].

Knowing the gyromagnetic ratio γ of a nucleus enables the resonance frequency from Equation (1) to be calculated for any magnetic field strength. Table 2 lists useful, commonly studied nuclei and their magnetic properties. The frequencies are for a 2.35 T applied magnetic field, i.e., relative to ^{1}H at 100 MHz. Table 2 gives the relative sensitivity for equal numbers of each nucleus and their relative receptivities, i.e., the product of the relative sensitivity and natural abundance.

The abundant nuclei, ^{1}H, ^{19}F, and ^{31}P, with $I = 1/2$ were the first to be studied by CW NMR. Owing to sensitivity problems, less abundant iso-

18.3.1. The NMR Experiment

The first successful NMR experiments were reported in 1945 by two independent groups in the USA [7], [8]. The discoveries by BLOCH at Stanford and by PURCELL at Harvard were awarded a Nobel Prize in Physics in 1952. Already in September 1952 the first commercial NMR spectrometer was installed. Early instruments used permanent magnets or electromagnets with fields from 0.94 up to 2.35 T, corresponding to 40 up to 100 MHz for proton resonance. Permanent magnets give very stable magnetic fields and are often used without further stabilization. Electromagnets need additional control mechanisms to provide the necessary stability. The commonest

Table 2. Magnetic properties of some nuclei

Isotope	Spin	Natural abundance, %	NMR frequency,* MHz	Relative sensitivity	Relative receptivity
^{1}H	1/2	99.985	100.000	1.00	1.00
^{2}H	1	0.015	15.351	9.65×10^{-3}	1.45×10^{-6}
^{11}B	3/2	80.42	32.084	0.17	0.13
^{13}C	1/2	1.108	25.144	1.59×10^{-2}	1.76×10^{-4}
^{14}N	1	99.63	7.224	1.01×10^{-3}	1.01×10^{-3}
^{15}N	1/2	0.37	10.133	1.04×10^{-3}	3.85×10^{-6}
^{17}O	5/2	0.037	13.557	2.91×10^{-2}	1.08×10^{-5}
^{19}F	1/2	100	94.077	0.83	0.83
^{23}Na	3/2	100	26.451	9.25×10^{-2}	9.25×10^{-2}
^{29}Si	1/2	4.70	19.865	7.84×10^{-3}	3.69×10^{-4}
^{31}P	1/2	100	40.481	6.63×10^{-2}	6.63×10^{-2}

* At $B = 2.35$ T.
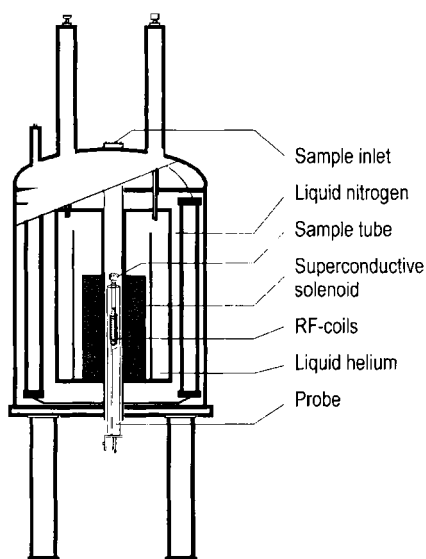
Figure 5. Schematic diagram of a cryomagnet. The superconducting coil is immersed in liquid helium. The sample tube is placed from above into the middle of the solenoid, where the field is strongest and has the best homogeneity. The lines of the static magnetic field are parallel to the long axis of the sample tube.

way to do this is to use the NMR signal of another nucleus in the sample to provide a field-frequency locking mechanism. The other nucleus may be the reference nucleus (homo-lock) or a different nucleus (hetero-lock), e.g., the deuterium resonance of a deuterated solvent. The need for instruments with higher resolution and sensitivity has led to the development of commercially available systems operating at up to 900 MHz for protons. All instruments operating above 100 MHz use helium-cooled superconducting solenoids to provide the magnetic field. High-resolution NMR requires both the magnetic field and the RF source to be

homogeneous and stable to better than 1 part in 10^8. To achieve this performance, which allows to obtain the maximum information possible from an NMR experiment, the magnetic field profile has to be made extremely uniform in the region of the sample. The two parameters which quantify these performance characteristics are magnetic field "homogeneity" and field "drift". The magnets are designed to minimize any magnetic field variations in time. Inevitable imperfections are introduced during manufacture of the wire of the solenoid and also caused by different samples, temperature variations, etc. They are eliminated using correction coils and so-called shim coils to remove residual field gradients. Also the magnetic field stability is of major importance. In conventional magnets the superconducting wire is bathed in liquid helium under atmospheric pressure where it has a boiling point of 4.2 K. By reducing the pressure the boiling point of helium can be lowered to 2.3 K. By doing this the performance of the superconducting wire is enhanced and higher magnetic fields can be achieved (cryostabilization). In Figure 5 a schematic diagram of a superconducting magnet is shown.

The sample, normally a solution in a deuterated solvent in a 5 or 10 mm glass tube (smaller tubes have been employed for specific applications for some years as well), is placed in the instrument probe between the poles of an electromagnet/permanent magnet or inside the solenoid of a superconducting magnet. The probe contains the RF transmitter and receiver coils and a spinner to spin the tube about its vertical axis. Sample spinning is used to average out magnetic field inhomogeneity across the sample. Multidimensional NMR spectra are measured in modern instruments without rotating.
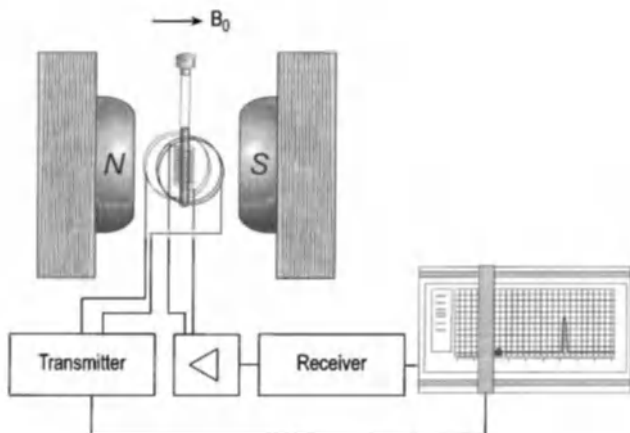
Figure 6. Block diagram of a CW NMR spectrometer.

The ideal NMR solvent should dissolve the sample, be inert, nonpolar, liquid over the temperature range of interest, inexpensive, and not give rise to large interfering peaks in the spectrum being recorded. Deuterated solvents such as deuterochloroform and deuterodimethyl sulfoxide are most commonly used. They offer the advantage of being "transparent" in proton spectra and provide a deuterium signal which can be used to stabilize and lock the spectrometer system. A clear, mobile solution is required for a high-quality spectrum. Solid particles or viscous solutions degrade the resolution. Traces of paramagnetic impurities in a solution can dramatically reduce relaxation times and cause peak broadening. With increasing field strength the demands both on the sample and on the used NMR tubes are usually growing.

Spectra are recorded either by the CW scan or pulsed FT method. The amount of sample required depends on the method of detection (CW or FT) and the receptivity of the nucleus being studied. For most routine CW proton spectrometers 10 – 20 mg of compound of M_r 200 – 300, dissolved in 0.4 mL of solvent should give a reasonable spectrum. On a 500 MHz FT instrument it should be possible to obtain a proton spectrum from less than one milligram of such a compound. Even smaller amounts can be measured with specific techniques, e.g., special NMR tube inserts. For specially designed probes amounts of few nanograms are sufficient [110], [144].

To measure nuclei with low receptivity, such as ^{13}C, a more concentrated solution must be used, in order to obtain a spectrum in as short a time as possible. Larger diameter sample tubes (10 or 15 mm) and, consequently, larger sample volumes

can be used to increase the amount of sample. Recently the so-called inverse detection techniques were used for observation of nuclei with small sensitivities (Section 18.3.6.4).

18.3.1.1. Continuous Wave Methodology

Until the late 1960s all NMR instruments were of the CW type. Figure 6 shows a block diagram of such a spectrometer. To obtain a spectrum from a sample, either the magnetic field or the RF frequency is slowly varied. When the resonance condition (Eq. 1) is met for the nuclei being observed, the sample absorbs RF energy and the resulting signal is detected by the receiver coils, amplified, and recorded. Spectra are recorded on precalibrated charts relative to a reference compound.

The CW method is suitable for recording spectra of abundant nuclei from relatively large amounts of sample. As a result, early NMR work was essentially limited to the ^1H, ^{19}F, and ^{31}P nuclei. Nowadays the use of CW instruments, which are relatively inexpensive, is mainly limited to ^1H observation. Such instrumentation is usually to be found in university teaching or synthetic organic chemical laboratories rather than specialist NMR facilities.

One of the main limitations of NMR spectroscopy is its inherent lack of sensitivity. As a result, the CW method is unsuitable for recording spectra from nuclei of low natural abundance or even from abundant nuclei in solutions of very low concentration. One way of improving the signal-to-noise (S/N) ratio is to record several spectra from a sample and add them together. The noise averages
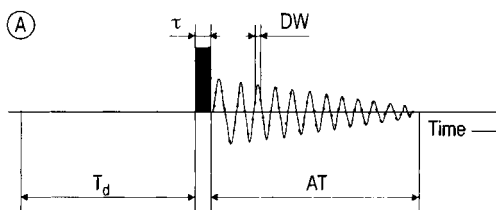
Figure 7. Schematic presentation of a pulse or FT NMR experiment (A); the free induction decay (FID) (B); and the appendant Fourier-transformed signal (NMR spectrum) (C). In A the FID is shown in a simplified manner also in the digital form, which will be generated by an analog-to-digital converter (ADC).

out, and the S/N ratio increases by a factor equal to the square root of the number of spectra accumulated. Using a computer to store and add the spectra allows many thousands of spectra to be accumulated in this way. The main drawback to this method on a CW instrument is the time to acquire each spectrum, typically ca. 4 min, resulting in excessive and often prohibitive total experiment times. A technique was required which could simultaneously excite the whole spectral region of interest. The pulsed FT technique, which first became available on commercial instruments in the late 1960s, provided the answer.

18.3.1.2. Fourier Transform Methodology

In 1966 ERNST and ANDERSON discovered that the sensitivity of NMR spectra could be increased dramatically if the sample was not exposed to the slow radio frequency sweep but to short and intense radio frequency pulses. The intensity of the signal was measured as a function of time after this pulse. The next pulse and signal acquisition were started after a few seconds, and the signals after each pulse were summed in a computer. If the RF frequency v is turned on and off very rapidly to obtain a pulse τ seconds long, this is equivalent to irradiating the sample with a range of frequencies centered about v over a frequency range $v \pm 1/\tau$. Thus if $\tau = 50 \, \mu s$ a range of $v \pm 20\,000$ Hz is covered.

If an RF pulse, at the resonant frequency v_0, is applied along the x-axis in the laboratory frame, this is equivalent to applying a static field $\boldsymbol{B}_1$ along the x'-axis of the rotating frame. This drives the sample magnetization about the x'-axis by an angle θ, as shown in Figure 4. Spectrometers are normally designed to detect signals along the y'-axis. The component of magnetization along this direc-

tion is given by $M_0 \sin \theta$, and the maximum signal is obtained when $\theta = 90°$. A pulse producing this effect is known as a 90° pulse, and the time for which the pulse must be applied to achieve this is known as the 90° pulse time. On modern FT spectrometers 90° pulse times are usually significantly less than 50 μs. Since the pulse applied is very short it must be powerful enough to excite nuclei in the time available across the whole spectral region of interest. To do this it must satisfy the condition:

$$\gamma \boldsymbol{B}_1 \geq 2\pi (\mathrm{SW})$$

where SW is the spectral width required.

At the end of the RF pulse, the spin system begins to relax back towards its equilibrium condition by means of the two relaxation mechanisms described in Section 18.2.4. After a time $5\,T_1$ the magnetization along y' — the signal detection axis — has decayed essentially to zero. For the remainder of this article the rotating frame is assumed and the axes are referred to simply as x, y, and z.

In the CW experiment, descibed in Section 18.3.1.1, the intensity of the NMR signal is recorded as function of frequency. However, in the pulsed experiment, the intensity of the decaying signal, following the RF pulse, is detected and recorded as a function of time. The signal is known as the free induction decay (FID) and it is not amenable to a simple interpretation. In Figure 7 a pulse experiment is represented as diagram in the usual notation. After the delay period T_d the excitation occurs through a pulse with the pulse width (length) τ and the subsequent acquisition. The resultant FID is also illustrated in Figure 7 together with the conventional NMR spectrum, for a sample which gives a spectrum with a single peak. For a spectrum containing more than one peak, the FID is an interference pattern resulting from the frequencies corresponding to all the individual peaks in the spectrum. The FID is often referred to as the time domain signal, whereas the corresponding NMR spectrum provides a frequency domain signal.

Time and frequency domain data can be interconverted by the mathematical process of Fourier transformation (FT), which can be performed rapidly by the computer. The intensity of the FID is measured at n intervals equally spaced by the so-called dwell time DW and stored in digital form. In other words, the dwell time is the time used to produce a particular data point. The total time to

acquire the data is referred to as the acquisition time AT. In order to obtain a spectral width SW the following equation must be satisfied:

$$AT = \frac{n}{2\,(SW)} = n \cdot DW$$

Thus, typical values for obtaining a 100 MHz ^{13}C NMR spectrum would be:

SW	$= 20\,000$ Hz (equivalent to 200 ppm)
N	$= 16\,384$ (stored in 16 K computer words)
DW	$= 25\ \mu s$
AT	$= 0.41$ s

Fourier transformation of the 16 K data points gives a spectrum with 8 K points since the transform contains both real and imaginary components, the real component being identical to the conventional CW spectrum.

The principal advantage of the FT method is that full spectral data can be obtained in a few seconds acquisition time as compared with minutes for the CW method. Signal averaging, by adding together the FIDs obtained after each of a series of RF pulses is therefore a much more efficient process than its CW analogue. The introduction of FT instruments made the routine study of less abundant isotopes, and in particular ^{13}C, possible.

FT instruments are available today with magnetic fields from 1.9 to 21.1 T, i.e., 80 to 900 MHz for ^{1}H observation. Most such instruments are available with a range of probeheads for studying different nuclei. Some probeheads are selectively tuned for observation of a single nucleus at maximum sensitivity. Dual- and trial-tuned probeheads for the convenient observation of ^{1}H and other fixed nuclei (e.g., ^{13}C, ^{15}N, ^{31}P) on a single probehead are now widely used, as are tunable probeheads which enable a wide range of nuclei to be studied. However, the most extensive development occurred in the case of the pulse sequences. Several hundred pulse sequences were described in the last twenty years and it is impossible to know all. A very good overview of the 150 most important ones is given in [109]. Furthermore, numerous information about basics and new developments are to be found in the Internet. Some of the most impressive web pages are listed in [111], [112].

The relatively high sensitivity and computing facilities associated with modern FT instruments have led to the development of fully automated systems with robotic sample changers. This ena-

bles unattended, round-the-clock use of expensive high-field instrumentation. A typical sample changer can accommodate from 6 up to 100 samples, and data is automatically acquired, processed, and plotted.

18.3.2. Spectral Parameters

Around 1950, it was discovered that nuclear resonance frequencies depend not only on the nature of the atomic nuclei, but also on their chemical environment. The possibility of using NMR as a tool for chemical analysis soon became apparent. All applications of solution NMR spectroscopy make use of one or more of the spectral parameters chemical shift, spin–spin coupling, signal intensity, and relaxation time. In the following sections examples from ^{1}H and ^{13}C NMR are used but the concepts described apply to magnetic nuclei in general.

18.3.2.1. Chemical Shift

Equation 1 might suggest that only a single peak would be obtainable from each magnetic isotope. However, when the molecule under observation is placed in the magnetic field, the electrons in the molecule shield the nuclei from the applied magnetic field to a small extent. For a particular nucleus within a molecule the degree of shielding depends on the density of the electrons circulating about that nucleus. Protons in different chemical environments therefore have different shieldings. Figure 8 shows diagramatically what happens to spherically symmetrical s electrons around a nucleus under the influence of an external magnetic field B_0. They circulate around the nucleus, producing their own magnetic field which opposes the applied field. The effective field B at the nucleus is then given by

$$B = B_0(1 - \sigma)$$

Thus, to obtain the resonant condition (Eq. 1) it is necessary to increase the applied field above that for an isolated nucleus. The shielding constant σ is a small fraction, usually of the order of parts per million. For electrons in p-orbitals where there is no spherical symmetry the situation is more complex since shielding and deshielding effects depend on the orientation of the molecule with respect to the applied field. Such effects are considered in more detail in Section 18.3.3.1.
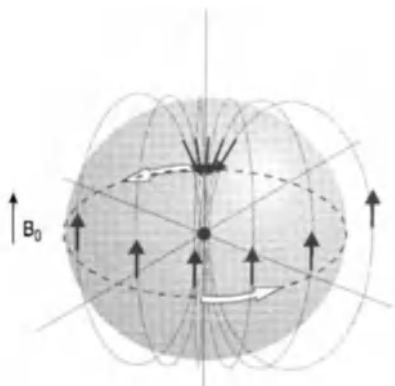
Figure 8. Electronic shielding of a nucleus (●) by circulating *s*-electrons (.......) in a magnetic field B_0

Figure 9. 250 MHz ^{1}H NMR spectrum of methyl acetate

The magnitude of the nuclear shielding is proportional to the applied magnetic field. The difference in the peak position of a particular nucleus from the peak position of a reference nucleus is called the chemical shift δ. In ^{1}H NMR if two peaks are separated by 80 Hz in a magnetic field corresponding to 100 MHz they will be separated by 320 Hz in a field corresponding to 400 MHz. Chemical shifts δ can be expressed in dimensionless units as parts per million (ppm) of the applied frequency in MHz as follows:

$$\delta = \frac{\left(\nu_{sample} - \nu_{reference}\right)}{\text{Applied frequency}}$$

Thus for the example given above

At 100 MHz $\delta = \frac{80\,\text{Hz}}{100 \times 10^6\,\text{Hz}} = 0.8\,\text{ppm}$

At 400 MHz $\delta = \frac{320\,\text{Hz}}{400 \times 10^6\,\text{Hz}} = 0.8\,\text{ppm}$

The chemical shift δ is therefore a molecular parameter that does not depend on the spectrometer frequency. ^{1}H and ^{13}C nuclei in organic molecules give peaks which cover a range of approximately 20 ppm and 220 ppm, respectively.

Tetramethylsilane (TMS) is the most commonly used chemical shift reference compound for ^{1}H and ^{13}C NMR. In both types of spectra TMS gives a single sharp resonance at lower frequency than most other proton or carbon resonances. The TMS is usually added to the sample solution as an internal reference. The accepted sign convention is for chemical shifts to high and low frequency of the reference peak to be positive and negative, respectively. Figure 9 shows the 250 MHz ^{1}H spectrum of methyl acetate

referenced to TMS; the peaks at $2.0\,\delta$ and $3.6\,\delta$ arise from the acetyl and methyl protons, respectively.

18.3.2.2. Spin – Spin Coupling

Chemical shift is one source of information on fine structure in NMR spectra. The second is spin – spin coupling. Figure 10 shows the aromatic region of the proton spectrum of 3-methoxy-6-nitro-*o*-xylene. One might expect two peaks from the two different aromatic protons in the molecule. In fact four peaks are observed. This results from the indirect coupling of proton spins through the intervening bonding electrons. In Figure 11 this is shown schematically for the two diastereotopic protons in a methylene group. The magnetic B_0 field experienced by nucleus A has a small component arising from nucleus B. Nucleus B has two equally probable orientations in the magnetic field, $m_B = +1/2$ or α and $m_B = -1/2$ or β. As a result nucleus A gives two equally intense peaks corresponding to the two orientations of B. By the same reasoning B itself also gives two equally intense peaks with the same separation as those for A. This splitting is called the spin – spin coupling J, where J_{AB} is the coupling between nuclei A and B. The magnitude of the coupling, which is measured in hertz, is independent of the applied magnetic field. Couplings between protons may be observed from nuclei separated by up to five bonds. Their magnitude rarely exceeds 20 Hz.

If more than two nuclei are interacting with each other all possible orientations of the coupling nuclei must be considered. When three nuclei have different chemical shifts, as for the olefinic protons of vinyl acetate, then each nuclear signal is split by coupling to both of the other nuclei. Thus
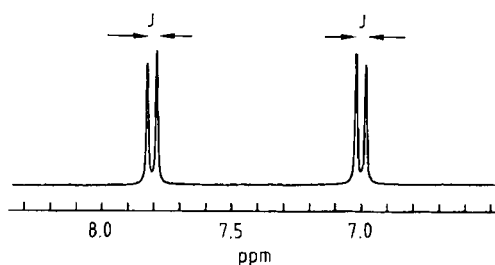
Figure 10. Aromatic region of the 250 MHz ^{1}H NMR spectrum of 3-methoxy-6-nitro-o-xylene

each resonance can be envisaged as being split into a doublet by the first coupling, each line of which is further split into a doublet to give a quartet pattern for each nucleus. When two of the three nuclei are equivalent, the total number of lines in the spectrum is reduced because some of the couplings are identical. Chemically equivalent protons (e.g., in a methyl group) do not exhibit splitting from coupling among each other.

Simple splitting patterns, for nuclei of spin 1/2, obey the following general rules:

1) If a nucleus interacts with n other nuclei with a different coupling to each then the signal will consist of 2^n lines of equal intensity
2) If a nucleus A couples to n equivalent nuclei B then the pattern for A comprises $(n + 1)$ lines with relative intensities given by the coefficients of the binomial expansion of $(1 + x)^n$

These simple rules are only strictly applicable to so-called first-order spectra. For spectra to be first order a single criterion must be obeyed. The chemical shift difference in hertz Δv for the coupled nuclei must be much larger than the coupling between them $(\Delta v/J \geq 10)$. If this condition is not met the spectra may show more lines than the rules predict, and their intensities are distorted, with a build up of intensity towards the middle of the coupling pattern. It is beyond the scope of this article to descibe these second-order spectra.

Because the magnitude of the spin – spin coupling interaction is independent of the applied magnetic field, whereas the chemical shift separation is proportional to it, a complex NMR spectrum can be simplified by using higher applied fields.

All the above examples of spin – spin coupling involve homonuclear coupling in proton spectra. In natural abundance ^{13}C spectra homonuclear coupling is not normally observed since the probability of two adjacent ^{13}C atoms is extremely low
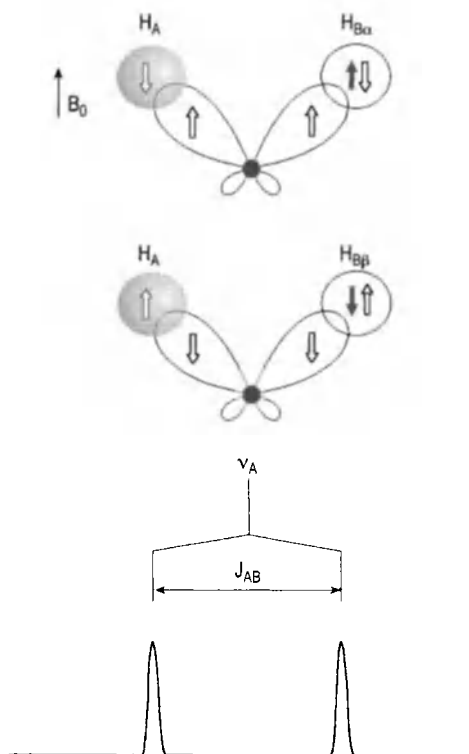
Figure 11. Nuclear spin-spin-interaction through the bonding electrons in a CH_2-group. The low-energy state corresponds to the antiparallel orientation of the magnetic moments of the nuclei (black arrows) and electrons (white arrows). The two possible states of the magnetic moments of hydrogen nucleus B are shown. As a result the resonance signal of nucleus A shows a splitting into a doublet. The effect is only observable in a methylene group in the case of diastereotopic, i.e., chemically inequivalent protons.

(10^{-4}). For organic molecules the most commonly encountered couplings are those to protons. One bond couplings range from about 110 to 250 Hz. Most longer range couplings are < 10 Hz. Proton-coupled ^{13}C spectra often show complex overlapping multiplets. To simplify the spectra and obtain them in a shorter time, routine ^{13}C measurements are usually proton noise decoupled (see Section 18.3.4.2). Thus, in the absence of other coupling nuclei such as ^{31}P or ^{19}F, a single sharp peak is observed for each chemically inequivalent ^{13}C atom. Figure 12 shows the proton-coupled and -decoupled ^{13}C NMR spectra of ethylbenzene.

18.3.2.3. Signal Intensity [113]

NMR is an inherently quantitative technique. For any nucleus the NMR signal obtained, at least
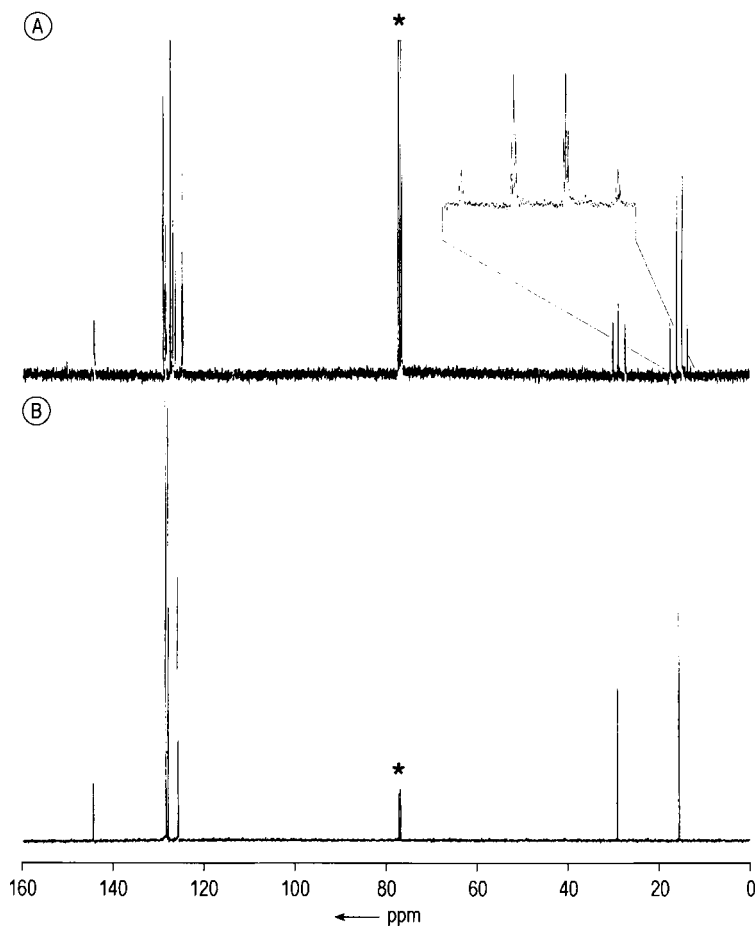
Figure 12. Proton-coupled (A) and proton broadband decoupled (B) ^{13}C NMR spectra of 10 % ethylbenzene in CDCl$_3$ (*).
Applied frequency 100.6 MHz: puls width 13.5 μs: repetition time 5 s; number of accumulations 1024 and 128; measuring time
85 min and 11 min, respectively.

for a single-scan CW experiment or a single-pulse
FT experiment, is directly proportional to the
number of nuclei producing it. For FT experiments
in general to be quantitative the spin system must
be given sufficient time to relax completely be-
tween successive pulses. Most proton spectra are
recorded under essentially quantitative conditions
and are integrated to give the relative number of
protons contributing to each resonance. Figure 13
shows an integrated proton spectrum of camphor
with relative peak areas of approximately
1 : 1 : 1 : 1 : 1 : 2 : 3 : 3 : 3 for the CH-protons, the
diastereotopic methylene, and the methyl protons,
respectively.

^{13}C spectra and those of other low natural
abundance nuclei are not normally recorded under
quantitative conditions. For structural identifica-

tion work with such nuclei, conditions are usually
selected for optimum sensitivity, not quantifica-
tion. There is an additional problem in spectra
such as ^{13}C where broadband proton decoupling
is usually used. Apart from collapsing the multi-
plet structure, the decoupling irradiation causes an
increase in signal intensity due to nuclear Over-
hauser enhancement (NOE, see Section 18.3.4.2).
The enhancement factor is often different for the
various ^{13}C resonances in the spectrum. Quantita-
tive ^{13}C work is possible but the experimental
conditions must be carefully selected [113].

18.3.2.4. Relaxation Times

Spin – lattice and spin – spin relaxation, the two
processes by which a perturbed spin system re-
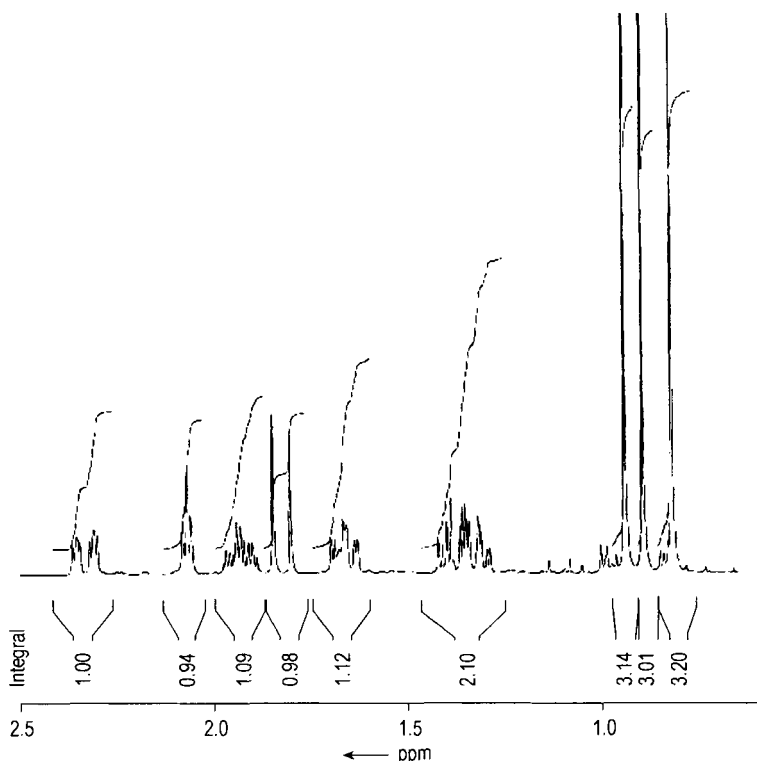
Figure 13. Integrated 400 MHz ^{1}H NMR spectrum of camphor. The reason for the differences between the experimental and the theoretical integral values are impurities and spectrum processing effects.

turns to equilibrium, are outlined in Section 18.2.4. The various mechanisms that contribute to these processes are discussed in [3]. This section is concerned with the effect variations of T_1 and T_2 have on the appearance of a spectrum. Proton spectra are usually little affected by differences in relaxation times for different protons in a molecule because most protons have short relaxation times (a few seconds). For ^{13}C and many other nuclei this may not be the case. ^{13}C relaxation times in organic materials range from a few milliseconds to > 100 s, depending largely on the molecular mass of the molecule and the environment of a particular ^{13}C atom in the molecule.

For spin 1/2 nuclei the principle T_1 mechanism is often dipole – dipole relaxation, which results from interactions between two nuclei with magnetic spins. Its efficiency is inversely proportional to the sixth power of the internuclear distance. In ^{13}C spectra of organic molecules, protons provide the main source of such relaxation. Where a ^{13}C

nucleus is directly bonded to a proton, T_1 is usually quite short, whereas a quaternary carbon atom in the same molecule will probably have a significantly longer value. In a FT ^{13}C spectrum this will result in a relatively low intensity for a quaternary resonance unless the interpulse delay is sufficiently long to allow complete relaxation of all nuclei. Such effects can be used as an aid in spectral assignment.

For polymers and other high molecular mass molecules, T_1 can be very short, in the range 10^{-3} to 1 s. For organic molecules with molecular masses < 1000, values for proton-bonded carbons lie typically in the range $0.1 - 10$ s, and for non-proton-bonded carbons, $10 - 300$ s.

The most important effect of the spin – spin relaxation time T_2 is that it determines the natural width of the lines in the spectrum. However, for most ^{1}H and ^{13}C spectra the observed linewidth is determined by the homogeneity of the applied magnetic field.

Table 3. ^{1}H chemical shifts of CH_3X compounds

X	δH, ppm
SiMe$_3$	0.0
H	0.13
Me	0.88
CN	1.97
COCH$_3$	2.08
NH$_2$	2.36
I	2.16
Br	2.68
Cl	3.05
OH	3.38
F	4.26

18.3.3. NMR and Structure

In this section the relationship between chemical structure and the NMR parameters of chemical shift and spin – spin coupling are discussed for some of the more commonly studied nuclei.

18.3.3.1. Hydrogen (^{1}H and ^{2}H)

The two naturally occurring isotopes of hydrogen are the proton and the deuteron with natural abundances of 99.985 % and 0.015 %, respectively. Whilst the former with $I = 1/2$ is the most commonly studied of all magnetic nuclei, the latter with $I = 1$ is rarely studied in natural abundance and then only for a few specialist applications. One such application, the determination of deuterium levels in samples, is discussed in Section 18.3.8.2. If small isotope effects are disregarded, corresponding ^{1}H and ^{2}H chemical shifts are the same. ^{2}H spectra are less well dispersed than the corresponding ^{1}H spectra by a factor of 6.51, the ratio of the magnetogyric ratios. The smaller magnetogyric ratio also results in correspondingly smaller coupling constants. Whilst the following discussion is exemplified by reference to the proton, most of the content is equally applicable to the deuteron. However, since deuterium homonuclear coupling constants are very small, ^{2}H spectra with broadband ^{1}H decoupling give a single peak from each chemically inequivalent deuteron even for deuterium enriched samples.

Chemical Shifts. The shielding effect of spherically symmetrical s electrons is discussed in Section 18.3.2.1. This diamagnetic upfield shift affects all nuclei since every molecule has s electrons. For electrons in p-orbitals there is no spherical symmetry and the phenomenon of diamagnetic anisotropy is used to explain some otherwise

anomalous chemical shifts. Diamagnetic anisotropy means that shielding and deshielding depend on the orientation of the molecule with respect to the applied magnetic field.

The hydrogen atom is a special case since it has no p electrons. A consequence of this is that in proton NMR the direct influence of the diamagnetic term can be seen. For example in substituted methanes CH_3X, as X becomes more electronegative, the electron density round the protons decreases and they resonate at lower fields (Table 3). For the same reason acidic protons in carboxylic acid groups resonate at very low fields. Figure 14 shows the chemical shift ranges for a selection of organic groups.

Whilst the electronegativity concept can be used to explain chemical shifts for saturated aliphatic compounds, the effects of diamagnetic anisotropy must also be considered for some other classes of compounds. For example, aromatic compounds such as benzene are strongly deshielded, the protons of acetylene are much more strongly shielded than those of ethylene (acetylene 1.48 ppm compared with ethylene 5.31 ppm), and aldehydic protons resonate at very low fields. In the case of benzene the deshielding results from the so-called ring current effect. When a benzene molecule is orientated perpendicular to the applied magnetic field B_0, there is a molecular circulation of π electrons about the field, as shown in Figure 15. The resulting ring current produces an additional magnetic field which opposes the applied magnetic field along the sixfold axis of the benzene ring but adds to it at the protons on the benzene ring, giving a low-field shift. The ring current is only induced when the applied magnetic field is perpendicular to the benzene ring. In practice only a proportion of the rapidly tumbling molecules are perpendicular to the magnetic field, but the overall averaged shift is affected by them.

Similar arguments can be invoked to explain the apparently anomalous shifts of acetylenic and aldehydic protons. The free circulation of electrons which gives rise to the diamagnetic effects in spherically symmetric atoms and the benzene ring can also occur around the axis of any linear molecule when the axis is parallel to the applied field. In the case of acetylene the π electrons of the bond can circulate at right angles to the applied field, inducing a magnetic field that opposes the applied field. Since the protons lie along the magnetic axis they are shielded and an upfield shift results. In the case of the aldehydic proton, the effect of the applied magnetic field is greatest
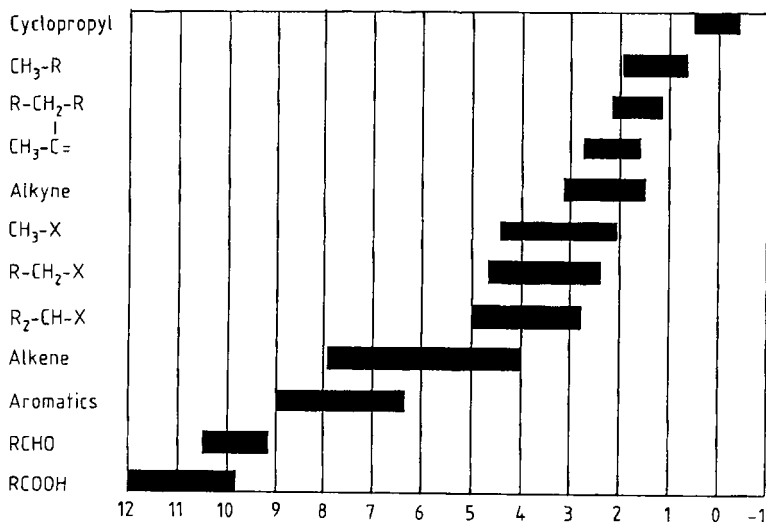
Figure 14. ^{1}H NMR chemical shift ranges
X = Halogen, –N⟨, –S–, –O–; R=Alkyl
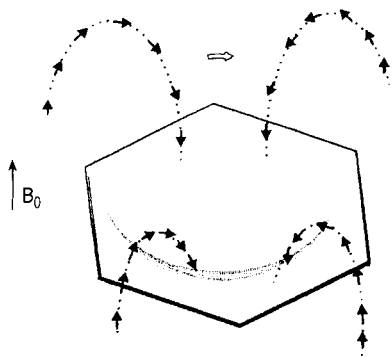
Figure 15. Ring current effects in benzene. The overlapping of the $2p_z$ orbitals leads to two current loops, above and below the *s*-bond plane of the benzene molecule. Only one loop is represented here. A magnetic field (black arrows) is induced by the ring current (white arrows), changing the shielding constant.

along the transverse axis of the C=O bond. This anisotropy causes deshielding of protons lying in a cone whose axis is along the C=O bond, and shielding outside this cone. The aldehydic proton, which lies within the cone, thus experiences a low-field shift and consequently resonates at low fields (9.5 – 10.0 ppm).

The effect of substituents on the proton chemical shifts of hydrocarbons has been extensively investigated. Schoolery's rules [2], [137] enable the prediction of the chemical shift of any CH_2XY and $CHXYZ$ proton. The chemical shift is the shift for

methane plus the sum of substituent contributions, which are largely determined by the electronegativity of the substituent. In aromatic compounds strongly electron-withdrawing groups such as nitro and carboxyl deshield all the protons, but the effect is largest at the *ortho* and *para* positions. The opposite is true for strongly electron-donating groups such as NH_2 and OH, while the halogens show smaller effects. For the various types of compounds the substituent increments can be used in an additive manner for compounds with more than one substituent. However, the accuracy of prediction decreases as the number of substituents increases.

Spin – Spin Coupling. The origin of spin – spin coupling and the resulting multiplet structure in proton NMR spectra is introduced in Section 18.3.2.2. The magnitude of couplings plays an important role in chemical structure determination, particularly in conformational analysis, stereochemistry, and in differentiating between isomeric structures. In proton NMR couplings are observed between protons and to other magnetic nuclei in the molecule such as ^{19}F and ^{31}P. Couplings to rare nuclei are only observed in accordance with their small abundance as so-called "satellites" in ^{1}H NMR spectra. A vicinal proton – proton coupling over three bonds is denoted by $^3J_{HH}$, whereas $^1J_{PH}$ represents a one bond coupling between a proton and a phosphorus atom. Depending
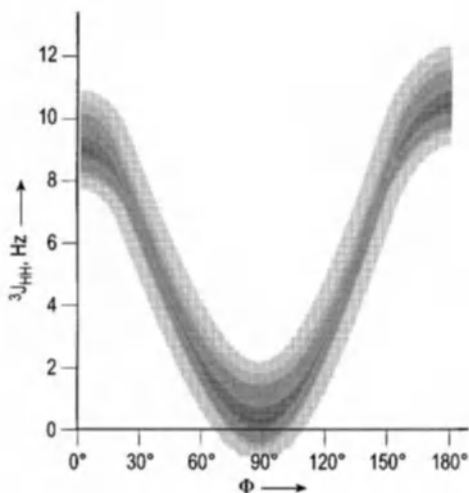
Figure 16. The Karplus curve shows the dependence of vicinal proton-proton coupling $^3J_{HH}$ on the dihedral angle Φ. The shaded area symbolizes the range of the empirical results.

on whether the coupling occurs over two, three, ore more bonds it is designated geminal, vicinal, or long-range coupling.

The magnitude of geminal couplings (^2J) depends to a first approximation on the H–C–H angle between the coupled protons. For tetrahedrally substituted carbon atoms a coupling in the range 12–15 Hz is typical. In aliphatic systems where there is free rotation about bonds geminal couplings are only observed where methylene protons are diastereotopic, that is, when the molecule contains a chiral center (C*):

$$\begin{array}{ccc} H_a & X & \\ | & | & \\ R-C-C*-Y & \\ | & | & \\ H_b & Z & \end{array}$$
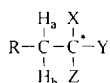
In such compounds H_a and H_b cannot be interchanged by symmetry operations or by rapid rotation and therefore may have different chemical shifts and exhibit coupling to each other. The most commonly encountered coupling in NMR is the vicinal proton-proton coupling. For protons on vicinal carbon atoms in saturated aliphatic groups the coupling constant depends primarily on the dihedral angle Φ between the CH bonds:

$$^3J_{(H-C-C-H)} = A + B\cos\Phi + C\cos^2\Phi$$

where A, B and C are constants with the values 4.22, −0.5, and 4.5, respectively. The graph of this function-the so-called Karplus curve-is shown in

Figure 16. A series of important regularities is explained by the Karplus curve, e.g., those for couplings in six-membered rings of fixed geometry. An axial-axial coupling J_{aa} in a cyclohexane ring where $\Phi = 180°$ would be ca. 10 Hz according to the Karplus equation. Experimentally, J_{aa} varies from 8 to 13 Hz, J_{ae} from 2 to 4 Hz, and J_{ee} from 1 to 3 Hz, depending on the nature of the substituents present. In acyclic systems the same angular dependence exists, with *gauche* and *trans* couplings typically in the ranges of 2–4 and 8–12 Hz, respectively. However, in systems with free rotation around the intervening C-C bond, an average coupling of ca. 7 Hz is observed.

In olefinic compounds the $^3J_{HH}$ coupling constant is used to differentiate *cis* and *trans* isomers since $J_{trans} > J_{cis}$. Ranges for *trans* and *cis* couplings are 12–18 and 6–12 Hz, respectively. Geminal couplings in olefins lie in the range 0–3 Hz. In aromatic compounds the observed couplings are also characteristic of the disposition of the hydrogen atoms. For benzene derivatives the following ranges are typical:

$$^3J_{ortho} = 6 - 10 \text{ Hz}$$
$$^4J_{meta} = 1 - 3 \text{ Hz}$$
$$^5J_{para} = 0 - 1 \text{ Hz}$$

Coupling of protons to other nuclei are discussed in the sections concerned with those nuclei.

18.3.3.2. Carbon (^{13}C) [9]–[11], [114]–[116]

Although present in natural abundance at only 1.1 %, the ^{13}C nucleus has, since the advent of FT methodology, been extensively studied. Only the approximately 5700 times more sensitive ^{1}H NMR is more widely used. In organic chemistry the information obtained from the two nuclei is often complementary for solving structural problems. The large ^{13}C chemical shift range and the small line width of the ^{13}C signals increases spectral resolution effectively. So, ^{13}C NMR is the method of choice for structural investigations of complex organic molecules, of complex mixtures of compounds, as well as biological oligomers and macromolecules.

Chemical Shifts. For organic chemicals most shifts are spread over a range of 0–200 ppm — about twenty times that for the proton. Figure 17 shows chemical shift ranges for some common organic groups. By comparison with Figure 14 it can be seen that there is an overall similarity in

Figure 17. ^{13}C NMR chemical shift ranges
R = H. alkyl; Y = All substituents

13.7 22- 34.- 22.- 13.7
 .6   5   6
$CH_3-CH_2-CH_2-CH_2-CH_3$
13.9 22- 28.- 32.- 62.0
 .9   5   7
$CH_3-CH_2-CH_2-CH_2-CH_2-OH$

Substituent chemical shifts in benzene derivatives have been extensively studied [13]. Such shifts can be used with caution in an additive manner to predict shifts and hence probable assignments for multi-substituted aromatics.

Spin – Spin Coupling. Carbon – proton couplings are usually sacrificed by broadband proton decoupling in the interests of spectral simplification and obtaining an acceptable spectrum more quickly. However, such couplings can provide valuable structural information.

One-bond carbon – proton couplings J_{CH} are of particular interest. At the simplest level the resulting multiplet provides information about the type of carbon atom; for example, a methyl group gives a quartet resulting from coupling to three equivalent hydrogen atoms. However, such information can be more conveniently obtained from editing experiments (see Section 18.3.5.3). $^1J_{CH}$ values vary from ca. 110 to 320 Hz, increasing with increasing s character of the C–H bond. Thus the couplings for methane (sp^3), ethylene (sp^2), and acetylene (sp) are 125, 156, and 248 Hz, respectively. Polar substituents also have significant effects on the magnitudes of couplings.

$^2J_{CH}$ values in aliphatic hydrocarbons range from 5 to 60 Hz. Couplings are often characteristic of substitution and/or bond hybridization. Thus whilst $^2J_{CH}$ for ethane is 4.5 Hz, the coupling to the methyl carbon of acetaldehyde is 26.7 Hz. Acetylene has a very large $^2J_{CH}$ of 49.3 Hz.

$^3J_{CH}$ couplings show orientational effects analogous to those observed in the corresponding H–H couplings. Thus $^3J_{cis}$ is always less than $^3J_{trans}$ in olefins, and the same Karplus type of dihedral angle dependence exists for saturated aliphatic groups. In aromatic rings $^3J_{CH}$ values are characteristically larger than $^2J_{CH}$ values. For benzene itself, $^3J_{CH} = 7.4$ Hz and $^2J_{CH} = 1.0$ Hz.

ordering on going downfield from TMS to that observed for proton chemical shifts. However, in ^{13}C NMR the effects of substitution are often significant several bonds away from the point of substitution. In simple alkanes α- and β-carbons result in deshielding effects of approximately 9 ppm each whilst a γ-carbon shields by about 2.5 ppm. As a result the C-1 shifts for methane, ethane, and propane are –2.3, 5.7, and 15.4 ppm, respectively. For alkanes and most alkenes the effects are additive and a set of rules have been derived [12]. Similarly, substitution leads to typical stereochemical effects in cycloalkanes as also in open-chain compounds. This is of importance for conformational analysis, where the stereochemistry of different conformers can be assessed through their ^{13}C data.

Polar substituents have large effects as can be seen by comparing the shifts for n-pentane and 1-pentanol.

Table 4. Representative $^{13}C - X$ coupling constants (Hz)

Compound	X	$^{1}J_{CX}$	$^{2}J_{CX}$	$^{3}J_{CX}$	$^{4}J_{CX}$
C_6H_5F	F	245.3	21.0	7.7	3.3
$n\text{-}C_6H_{13}F$	F	166.6	19.9	5.3	
$(C_6H_5)_3P$	P	12.4	20.0	6.7	
$(n\text{-}C_4H_9)_3P$	P	10.9	11.7	12.5	
C_6H_6	H	157.5	1.0	7.4	
C_6D_6	D	25.5			
$CHCl_3$	H	209.0			
$CDCl_3$	D	31.5			

Table 5. Representative ^{19}F chemical shifts*

Compound	δ, ppm	Compound	δ, ppm
CH_3F	−271.9	CF_2Cl_2	−6.9
C_2H_5F	−211.5	CF_3CO_2H	−78.5
CF_3CF_3	−88.2	$C_6H_5CF_3$	−63.9
$FCH=CH_2$	−114.0	F_2	422.9
$F_2C=CH_2$	−81.3	AsF_3	−43.5
$F_2C=CF_2$	−135.2	AsF_5	−67.2
C_6F_6	−162.9	SF_6	48.5
C_6H_5F	−113.2	PF_3	−36.2
$C_6H_5SO_2F$	65.5	PF_5	−77.8

* Referenced to $CFCl_3$.

Note that heteronuclear couplings to magnetic nuclei other than ^1H (e.g., ^{19}F, ^{31}P, and ^2H) are observed in proton decoupled ^{13}C NMR spectra. Thus coupling to deuterium is seen in signals from deuterated solvents. Some representative C–X couplings constants are given in Table 4.

18.3.3.3. Fluorine (^{19}F) [14], [16], [17], [106], [127]

With 100 % natural abundance and $I = 1/2$, ^{19}F is ideally suited to NMR investigations. Since the earliest days of NMR spectroscopy ^{19}F applications in organic, inorganic, and organometallic chemistry have been widespread. As in the case of the ^{13}C nucleus broadband proton decoupling is often used to simplify spectra.

Chemical Shifts. ^{19}F NMR spectroscopy has the advantage of a very large chemical shift range. Organofluorine compounds alone cover over 400 ppm, whilst the range is extended to some 700 ppm if inorganic compounds are included. $CFCl_3$ is now the generally adopted internal reference. Most organofluorine resonances are high field shifted to the $CFCl_3$ reference peak. In NMR spectroscopy the accepted convention is for shifts

to high field of the reference to be assigned negative values. Table 5 gives some representative ^{19}F chemical shifts. The shifts are sensitive not only to directly attached substituents but also to long range effects and thus are very good probes of changes in molecular structure. The chemical shifts given for the following compound provide an idea of the wide chemical shift range observed for some fluorocarbons.

$$\begin{array}{c} -72.5 \\ -81.6 \quad -123.0 \quad -117.1 \quad CF_3 \; O \\ F_3C-CF_2-CF_2-CF_2-CF-C-F \\ -126.8 \quad -120.4 \quad -180.4 \quad +32.3 \end{array}$$
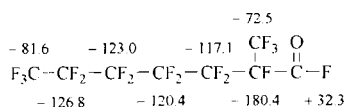
Spin–Spin Coupling. Large compilations of data are available for both fluorine–fluorine and fluorine–hydrogen coupling constants. The presence of several relatively large couplings can lead to complicated spectra even from relatively simple molecules if they are highly fluorinated. However, the spectra are often amenable to first-order analysis, and in general the coupling constants are very sensitive to structure and substituents.

Three-bond fluorine–hydrogen couplings show orientational effects similar to those observed in the corresponding H–H couplings, i.e., the dihedral angle dependence in saturated aliphat-

Table 6. Selected F–F and F–H coupling constant ranges

Group	Coupling	J_{FX} coupling constant range, Hz	
		X = F	X = H
\C—F / \C\X (^2J)	^2J	40–370	40–80
—C—C— F X (^3J)	^3J	40–80	0–40
\C=C F /\X (^2J)	^2J	0–110	70–90
F\C=C/\X (^3J)	^3J	100–150	10–50
F\X C=C (^3J)	^3J	0–60	0–20
Fluorobenzenes:			
ortho-X	^3J	18–35	7–12
meta-X	^4J	0–15	4–8
para-X	^5J	4–16	0–3

ics and that $^3J_{cis}$ is less than $^3J_{trans}$ in olefins. Typical coupling constant ranges are given in Table 6.

18.3.3.4. Phosphorus (^{31}P) [18], [19], [117]

The ^{31}P nucleus has 100% natural abundance and $I = 1/2$. However, a relatively small magnetogyric ratio results in the sensitivity of the ^{31}P nucleus being only 0.066 that of the proton. Nevertheless, with Fourier transform methodology, the nucleus is easy to detect even at low levels. There is much data in the literature covering applications involving inorganic compounds and complexes, organophosphorus chemistry, and biological and medical studies. In addition to these, even the correlation between ^{31}P and ^{13}C is observable with modern spectrometers in 2D correlation experiments [118].

Chemical Shifts. The accepted standard for ^{31}P NMR work is 85% H_3PO_4, which is normally used as an external reference. Phosphorus compounds are often reactive and no suitable internal reference standard has been found for general use. Relative to H_3PO_4, ^{31}P shifts cover the range from ca. 250 ppm to −470 ppm. Phosphorus(III) compounds cover the complete range, showing that the chemical shifts are very substituent dependent,

whilst phosphorus(V) compounds cover a much smaller range of −50 ppm to 100 ppm. To a first approximation the substituents directly attached to the phosphorus atom determine the chemical shift. Three variables are important: the bond angles, the electronegativity of the substituent, and the π-bonding character of the substituent. Some representative ^{31}P shifts are given in Table 7.

The phosphorus nuclei are well suited for biological and medical investigations. Relatively simple high resolution ^{31}P NMR spectra are obtained for the chemically different phosphorus atoms of cell metabolism products. The quality of the spectra which can be obtained from living systems (in vivo ^{31}P NMR spectroscopy) allows to detect metabolic changes. Furthermore, the position of the ^{31}P resonance of inorganic phosphate in the cell is a sensitive indicator for the intracellular pH value [119].

Spin–Spin Coupling. Coupling constants have been studied between ^{31}P and a wide variety of other nuclei. Couplings to protons, fluorine, and other phosphorus nuclei have been extensively used for structure elucidation. One-bond phosphorus–hydrogen couplings depend on the oxidation state of the phosphorus atom. Thus for P^{III} compounds $^1J(^{31}P,^1H)$ values are on the order of 200 Hz, whilst for P^V compounds they are usually in the range 400–1100 Hz. By contrast, one-bond phosphorus–fluorine coupling constants for both P^{III} and P^V compounds have a common range of ca. 800 to 1500 Hz. Phosphines are widely used as ligands in inorganic complexes and therefore phosphorus couplings to some transition metals are important aids to structure determination.

18.3.3.5. Nitrogen (^{14}N and ^{15}N) [20], [21], [106], [127]

There are two magnetic isotopes of nitrogen, ^{14}N and ^{15}N, with natural abundances of 99.63% and 0.36%, respectively. In a magnetic field of 9.4 T the ^{14}N resonance frequency is 28.9 MHz, while that for ^{15}N is 40.5 MHz. There are problems associated with observing both nuclei. The ^{14}N nucleus possesses a quadrupole moment since $I = 1$. As a result, although ^{14}N signals can often be detected easily, the lines can be very broad with widths often between 100 and 1000 Hz. With a 9.4 T applied magnetic field this is equivalent to 3.5–34.6 ppm. This can cause resolution problems when a molecule contains more than one type of nitrogen atom.

Table 7. Representative ^{31}P shifts*

P^{III} compounds		P^V compounds	
Compound	δ, ppm	Compound	δ, ppm
PF$_3$	97	PF$_5$	−35
PCl$_3$	219	PCl$_5$	−80
PBr$_3$	227	PBr$_5$	−101
P(OMe)$_3$	141	P(OEt)$_5$	−71
P(OPh)$_3$	127	P(OPh)$_5$	−86
PH$_3$	−240	(MeO)$_3$P=O	2
MePH$_2$	−164	(PhO)$_3$P=O	−18
Me$_2$PH	−99	Me$_3$P=S	59
Me$_3$P	−62	Me$_3$P=O	36
Ph$_3$P	−6	Ph$_3$P=O	25

* Referenced to 85 % H$_3$PO$_4$.
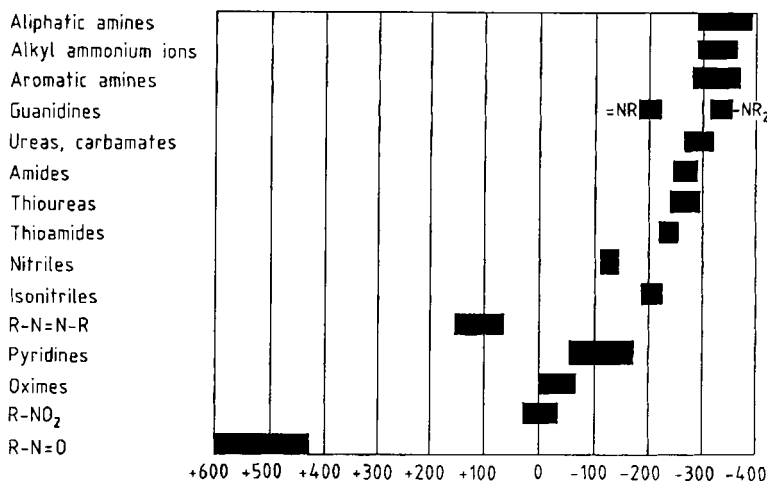
Figure 18. ^{14}N/^{15}N NMR chemical shift ranges
R = alkyl, aryl

Although $I = 1/2$ for the ^{15}N nucleus its very low natural abundance results in low receptivity. It also has a negative magnetogyric ratio which leads to a negative NOE when broadband proton decoupling is used. This reduces the intensity of peaks and can result in negative and even null resonances. It has therefore been necessary to use experimental procedures which minimize or eliminate the NOE. The two principal ones are inverse-gated proton decoupling (see Section 18.3.4.2) and the use of a paramagnetic relaxation reagent such as [Cr(acac)$_3$], which reduces T_1 and consequently the NOE. With modern FT equipment it is now possible to obtain ^{15}N spectra from most small and medium-sized nitrogen-containing molecules at natural abundance level. However, for work with large biological macromolecules ^{15}N enrichment is commonly used.

Chemical Shifts. In principle the use of two isotopes raises the possibility of a primary isotope shift, but in practice ^{14}N and ^{15}N shieldings differ by no more than a few tenths of a ppm and are interchangeable for practical purposes. Nitromethane is used as reference standard for calibration purposes and the convention of assigning negative values to signals occurring high field of the reference is used here also. However, the chemical shift of nitromethane is somewhat solvent and concentration dependent. The total nitrogen chemical shift range covers about 1000 ppm, and ranges for a selection of chemical classes are shown in Figure 18.

Spin–Spin Coupling. It is unusual to observe coupling in ^{14}N spectra where resonances are very broad as a result of quadrupolar relaxation. Cou-
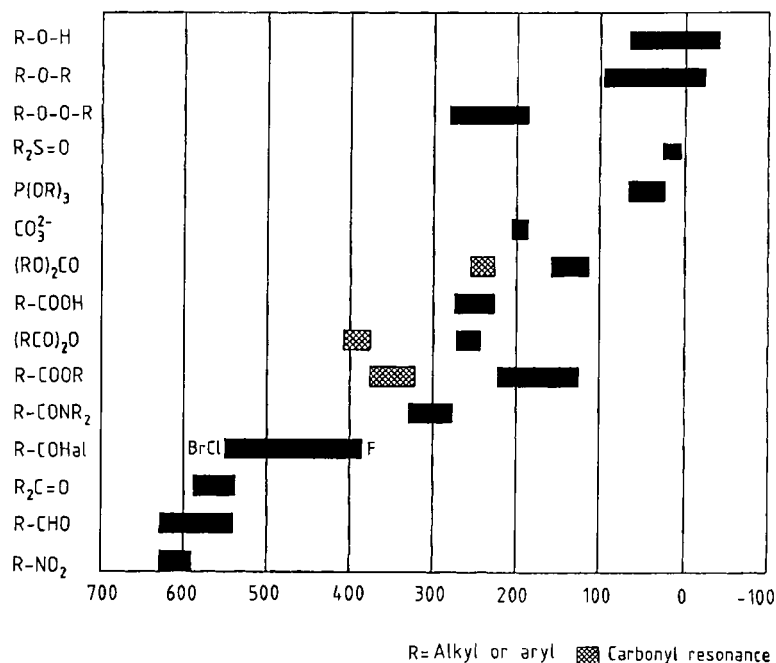
Figure 19. ^{17}O NMR chemical shift ranges
R = Alkyl, aryl

pling is, however, often resolved for ammonium salts. The relative magnitudes of the corresponding ^{14}N and ^{15}N coupling constants are given by the ratio of their magnetogyric ratios. The discussion here is restricted to ^{15}N coupling. The observation of coupling to low-abundance nuclei such as ^{13}C or another ^{15}N nucleus normally requires isotopic enrichment. Couplings to abundant nuclei such as ^1H, ^{19}F, and ^{31}P are more readily observed, as in the case of ^{13}C spectroscopy.

One-bond couplings to hydrogen depend on hybridization and electronegativity. Thus the $^1J(^{15}N, ^1H)$ values for NH_4^+, $O=C(NH_2)_2$, and $HC\equiv NH^+$ are 73.3, 89.0, and 134.0 Hz, respectively. One-bond couplings to ^{13}C also show a reasonable correlation with hybridization. The magnitudes of one- and two-bond couplings to ^{13}C are often comparable.

18.3.3.6. Oxygen (^{17}O) [120], [121]

The only magnetic isotope of oxygen is ^{17}O. With a natural abundance of only 0.037 %, a receptivity that is 0.61 that of ^{13}C, and a quadrupole moment ($I = 5/2$), ^{17}O is difficult to study. Nevertheless, because of the importance of oxygen in both inorganic and organic chemistry as well as biology, the ^{17}O nucleus has been quite widely studied both in natural abundance and in isotopically enriched samples. For small molecules linewidths are typically in the range of several tens to several hundreds of hertz. With large molecules or viscous solutions much larger linewidths may be encountered.

Chemical Shifts. Water is the usual chemical shift reference, being readily available in enriched form. A chemical shift range of over 1000 ppm affords wide spectral disperison. Figure 19 shows representative chemical shift ranges for some organic groups. There is a clear distinction between single and double carbon–oxygen bonds. However, a single resonance is observed from a carboxylic group at the average position expected for the carbonyl and OH groups as a result of rapid proton exchange facilitated by dimeric species. Primary alcohols and ethers resonate close to the water position, whilst lower field shifts are observed for secondary and tertiary species.

Spin–Spin Coupling. Since ^{17}O resonances are usually broad, the majority of reported values are

Table 8. ^{29}Si chemical shifts for SiX_nMe_{4-n} (ppm)

X	$n = 1$	$n = 2$	$n = 3$	$n = 4$
H	−15.5	−37.3	−65.2	−93.1
Et	1.6	4.6	6.5	8.4
Ph	− 5.1	− 9.4	−11.9	−14.0
OMe	17.2	− 2.5	−41.4	−79.2
NMe$_2$	5.9	− 1.7	−17.5	−28.1
Cl	30.2	32.3	12.5	−18.5

for couplings between directly bonded nuclei. $^1J(^{17}O, H)$ in water and alcohols are ca. 80 Hz. The $^1J(P, ^{17}O)$ couplings involving P=O groups cover the range 145–210 Hz. In contrast, couplings involving P–O groups are usually smaller (90–100 Hz). Some two-bond couplings have been reported. Examples of two-bond hydrogen–oxygen couplings are 38, 7.5, and 10.5 Hz for the O–CH, O–CH$_3$, and HC=O interactions in methyl formate.

18.3.3.7. Silicon (^{29}Si) [22], [122]

^{29}Si, with an abundance of 4.7 %, is the only naturally occurring isotope of silicon with nonzero spin ($I = 1/2$). Whilst its receptivity is 2.1 times that of the ^{13}C nucleus there are problems associated with its detection. In common with the ^{15}N nucleus a negative magnetogyric ratio results in negative NOEs when broadband proton decoupling is employed, as is usually the case. Inverse-gated decoupling and relaxation reagents are used to circumvent the problem. Broad signals from the glass in the NMR probehead or tube can also cause problems. The glass can be replaced by alternative materials such as teflon or the spectrum obtained from a blank run (i.e., without sample) can be subtracted from the spectrum of the sample. ^{29}Si is very important for solid state NMR, e.g., for analysis of glasses and zeolites [128].

Chemical Shifts. The ^{29}Si resonance of TMS is the usual chemical shift reference. ^{29}Si chemical shifts cover a range of ca. 400 ppm. The large dispersion arising from structural effects makes ^{29}Si NMR a valuable tool for molecular structure determination. The nature of the atoms directly bonded to silicon is of particular importance. Table 8 lists chemical shifts for a series of compounds SiX_nMe_{4-n}. ^{29}Si NMR has proved to be an ideal tool for structure determination of polysiloxane macromolecules.

Spin–Spin Coupling. Couplings to abundant nuclei such as the proton and fluorine can be measured either in a coupled ^{29}Si spectrum or by observing the ^{29}Si satellites in the spectrum of the abundant nucleus. The latter are relatively weak compared with peaks from the main ^{28}Si isotopomer. Nevertheless many determinations were made in this way during the 1960s by using CW spectrometers. In a well-resolved ^{1}H NMR spectrum the satellites due to a $^2J_{SiH}$ coupling of 6.8 Hz are readily observed at the base of the TMS reference peak. One-bond proton–silicon coupling constants in silanes are usually in the range 150–380 Hz, the magnitude of J increasing with substituent electronegativity. The literature contains data about coupling to a range of other nuclei, including ^{31}P, ^{13}C, ^{15}N, and ^{29}Si itself.

18.3.4. Double Resonance Techniques

In Section 18.3.2.2 reference is made to the use of proton noise (or broadband) decoupling to simplify ^{13}C spectra and improve the signal-to-noise ratio. This is just one example of a large number of double resonance techniques which are available to either simplify spectra or as spectral interpretation tools to aid structure elucidation.

18.3.4.1. Homonuclear Spin Decoupling

The principles of homonuclear double resonance can be illustrated by considering two coupled protons, A and B, as shown in Figure 11. In the double resonance experiment nucleus A is observed in the usual way in a CW or FT experiment whilst simultaneously selectively irradiating nucleus B with a second much stronger RF field at the "decoupling" frequency. This irradiation induces transitions between the two spin states of nucleus B. If sufficient irradiating power is applied, B flips between the α and β spin states so rapidly that nucleus A can no longer distinguish
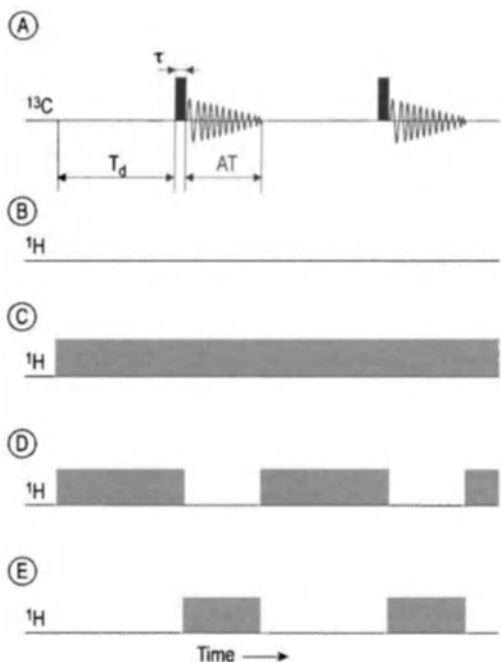
Figure 20. Schematics of an ^{13}C NMR experiment
(A) without (B) and with different ^1H decoupling techniques:
C) broadband decoupling; D) gated decoupling and
E) inverse gated (igated) decoupling

between the two orientations of B and perceives an average orientation. The coupling J_{AB} disappears and the A signal doublet collapses to give a single line. In the same way irradiation of an arbitrary signal in a complex NMR spectrum results in the obvious collapsing of all multiplets which are coupled with the irradiated proton. At time before 2D NMR was feasible, this method was often used to prove the connectivities between different protons in a molecule.

18.3.4.2. Heteronuclear Spin Decoupling

In the case of heteronuclear spin decoupling the decoupled and observed nuclei are of different chemical types. The experiments are illustrated with examples in which protons are decoupled and the effect observed on the ^{13}C spectrum. However the methodologies are equally applicable to other heteronuclear spin systems.

Selective spin decoupling is analogous to the homonuclear decoupling experiment described above. In general higher RF power is required than for homonuclear decoupling but lower than is gen-

erally used for broadband heteronuclear decoupling. An individual proton signal is selected and irradiated to remove all couplings to that specific proton in the ^{13}C spectrum.

Broadband Spin Decoupling. If the proton decoupler frequency is set to the center of the proton spectrum and modulated by using a noise generator with a bandwidth wide enough to cover the complete proton region, then this is equivalent to simultaneously irradiating every proton frequency. This effectively decouples all the protons in the molecule producing the effect shown in Figure 12 B. Two problems are, however, associated with this type of broadband decoupling. The high RF power required heats the sample and, secondly, decoupling is ineffective if the bandwidth to be irradiated is large. The latter is more of a problem on higher field spectrometers and where broadband decoupling of other nuclei such as ^{19}F or ^{31}P, with their large spectral widths, is required. In recent years alternative methods such as WALTZ-16 [23] based on composite pulse decoupling have largely overcome these problems.

Broadband decoupling also produces a change in intensity, known as nuclear Overhauser enhancement, of the observed signals. For ^{13}C the increase in intensity can be as much as 200 %.

Gated Decoupling. When decoupling is turned on or off, coupling effects appear and disappear instantaneously. The NOE effects however, grow and decay at rates related to some relaxation process. As a result it is possible by computer control of the decoupler to completely separate the two effects.

Sometimes it is necessary to record broadband proton decoupled spectra without any NOE. The NOE may be undesirable either because quantitative information is needed about the relative number of atoms contributing to each peak in, for example, a ^{13}C spectrum, or because the nucleus being studied has a negative NOE (e.g., ^{15}N or ^{29}Si). Such spectra are also used as controls when measuring NOE values. In other instances coupled spectra with NOE are used as a means of improving sensitivity.

Figure 20 shows a FT ^{13}C experiment with three possible broadband decoupling schemes. The repetitive pulse experiment consists of three parts: a relaxation delay T_d, a ^{13}C RF pulse, and the acquisition time AT during which time the FID is sampled. AT and T_d are on the order of seconds, whilst the pulse duration is a few microseconds.

For conventional broadband decoupling (Fig. 20 C) the decoupler is on throughout the experiment. If the decoupler is gated off just before the RF pulse as in Figure 20 D, NOE effects can be observed without any decoupling. Conversely, if the decoupler is switched on when the pulse is applied and off as soon as the acquisition is complete, as shown in (E), the resultant decoupled spectrum does not exhibit an NOE. The latter experiment is referred to as inverse gated decoupling.

Off-Resonance Decoupling. Whilst the broadband proton decoupling described above produces considerable simplification of ^{13}C and other spectra, it removes all the coupling information. As a result the spectra may be difficult to assign. In the early days of ^{13}C NMR a technique known as off-resonance decoupling was used to circumvent this problem. Off-resonance decoupling is achieved by setting the central frequency of the broadband proton decoupler 1000 – 2000 Hz outside the proton spectrum. This results in residual couplings from protons directly bonded to ^{13}C atoms, whereas longer range couplings are lost. The magnitude of the residual coupling, which is smaller than the true coupling J, is determined by the amount of the offset and the power of the decoupler. In such spectra it is usually possible to determine the multiplicities of all the individual carbon atoms, i.e., methyl carbons appear as quartets, methylenes as triplets, etc. Nowadays the use of gated decoupling for multiplicity determination has largely been replaced by spectral editing experiments (Section 18.3.5.3).

18.3.4.3. NOE Difference Spectroscopy [24], [25]

The phenomenon of nuclear Overhauser enhancement was introduced in Section 18.3.2.3 to explain the increase in intensity of ^{13}C peaks when broadband proton decoupling is used. In general terms the NOE is defined as the change in signal intensity of a nucleus when a second nucleus is irradiated. It is beyond the scope of this article to give a theoretical treatment of the origin of the effect, which is caused by through space nuclear relaxation by dipolar interaction. Among other factors the magnitude of the enhancement has a $1/r^6$ dependence on the distance between the two nuclei. Therefore, homonuclear NOEs between protons have been used extensively to obtain in-

formation about internuclear distances and thereby to distinguish between possible structures.

The most convenient way of observing such enhancements, which rarely exceed 20 % in the most favorable cases and are sometimes < 1 %, is by a technique known as NOE difference spectroscopy. In such an experiment two spectra are recorded. The first is a conventional ^1H spectrum, while in the second a chosen resonance is selectively irradiated during the interpulse delay of an FT experiment. The second spectrum is then subtracted from the first to give a "difference" spectrum.

1
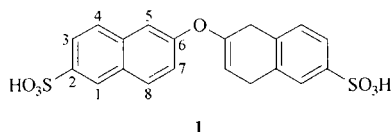
Figure 21 A shows the ^1H NMR spectrum for **1**. From this spectrum alone it is not obvious whether the ether linkage is at C-6 or C-7. The peak at 8.2 ppm is clearly due to H-1 whilst that at ca. 8.0 ppm is from H-5 or H-8, depending on which isomer is present. Figure 21 B shows the difference spectrum obtained when the resonance at 8.0 ppm is selectively irradiated. The irradiated resonance, which is partially or fully saturated, gives a negative signal in the difference spectrum. Positive signals are obtained from those resonances which exhibit an enhancement. Since a positive enhancement is observed for H-1 this is a good indication that H-1 and H-8 are close together. Measurable effects can be observed between atoms up to about 0.4 nm apart. Since the above description is rather simplistic it should be stressed that because dipolar relaxation of a particular nucleus may involve interaction with more than one other nucleus a basic understanding of the processes involved is a prerequisite to applying the technique successfully. Also spin-spin coupling may complicate NOE measurements. If irradiation is applied to spin multiplets, NOE measurements may fail if selective population transfer is present and special precautions have to be taken in such cases. Furthermore, the NOE effect depends on the molecular correlation time τ. During slow molecular motions frequencies dominate the relaxation process. For macromolecules or small molecules in viscous solvents the NOE may completely vanish or become negative. A positive NOE effect is observable, therefore, only under condition $\omega \tau \ll 1$, which is always true for small
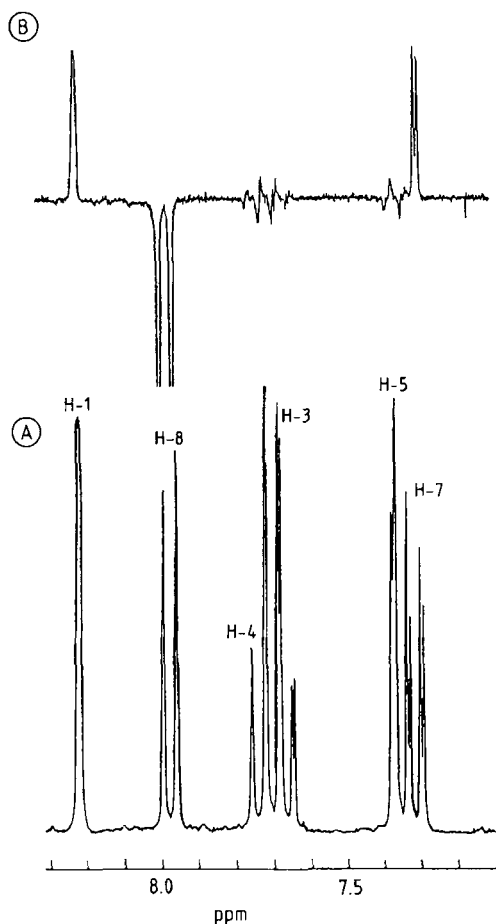
Figure 21. (A) 250 MHz ^{1}H NMR spectrum of **1**; (B) NOE difference spectrum on irradiation of H-8

and isotropic molecules of low molecular weight compounds ($M_r < 500$) in solutions of low viscosity. In the last years the two-dimensional NMR technique has been used more frequently for measurement of the NOE effect (NOESY spectroscopy, see Section 18.3.6.5).

18.3.5. One-Dimensional Multi-Pulse FT Experiments

The FT experiment described in Section 18.3.1.2. and depicted in Figure 7 is a single-pulse experiment. This means that only one RF pulse is applied prior to recording the FID response. A large number of experiments have been developed, for a variety of purposes, in which more

than one pulse is applied prior to recording the FID. One of these, the inversion recovery method of determining T_1, is described below in some detail. Others will be mentioned only in so far as their application is concerned.

18.3.5.1. T_1 Measurement

Spin–lattice relaxation is described briefly in Section 18.3.2.4. T_1 values can provide valuable information about a molecule. Each of the magnetic nuclei in a different environment in a molecule has a different T_1. Thus each type of hydrogen or carbon atom will have a different value. A knowledge of these values and their relative sizes can give an insight into such things as the type of atom (e.g., whether it is a quaternary or proton-bonded carbon atom), the mechanism responsible for the relaxation, and the molecular motion of that part of the molecule. A knowledge of T_1 values is also required for using NMR as an accurate quantitative tool (see Section 18.3.8.2.).

The most common method of measuring T_1 is the inversion recovery experiment, which uses the pulse sequence: $[T_d - 180° - \tau - 90° \, (\text{FID})]_n$. The principle is pictured in Figure 22. Whereas a 90° pulse rotates the magnetization vector, which at equilibrium is M_{z0} and lies along the z-axis (Fig. 22 B), to the y-axis of the rotating frame, a 180° pulse which is twice as long inverts the magnetization to the $-z$-axis. Immediately after the pulse the magnetization vector M_z begins to relax back to its equilibrium value M_{z0} according to the following equation:

$$M_z = M_{z0}[1 - 2\exp(-\tau/T_1)] \tag{2}$$

With time, M_z becomes less negative, passes through zero, and eventually relaxes back to M_{z0}. This process proceeds at a different rate for each peak in a spectrum. At a time τ after the 180° pulse, a 90° pulse is applied which tips the magnetization vector onto the $-y$- or $+y$-axis, depending on whether M_z was negative or positive. The delay T_d required before the sequence can be repeated should be at least five times the longest T_1 being measured to enable the system to relax completely. As in the case of the single pulse experiment the sequence is repeated n times and the FID added together. The resulting FID can then be transformed in the usual way. The sign and the amplitude of the resulting peaks depends on the length of the interpulse delay τ and T_1 for each nucleus (Fig. 22 C).
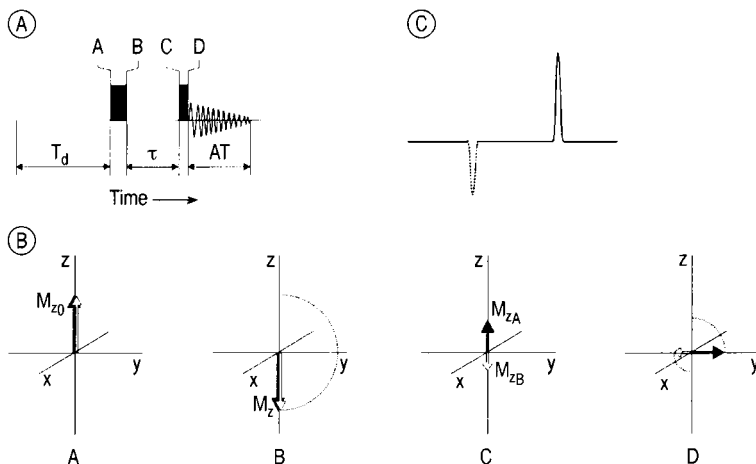
Figure 22. The principle of the inversion-recovery experiment for T_1 measurement: Puls sequence (A) and position of the macroscopic magnetization M_z (B) during the experiment at four selected points (A-D), shown for two nuclei A and B with different T_1 values, resulting in the spectrum (C)
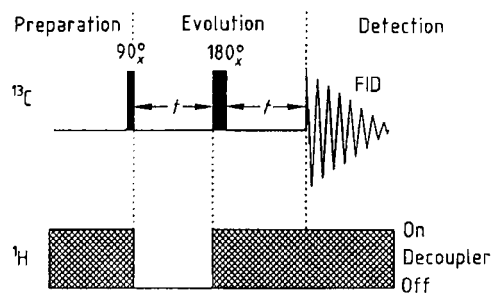
Figure 23. J-modulated spin-echo pulse sequence

Rearranging Equation (2) and taking logarithms gives:

$$\ln(M_z - M_{z0}) = -\ln(2M_{z0}) - \tau/T_1$$

Thus a plot of $\ln(M_z - M_{z0})$ against τ for each peak gives a straight line with a gradient of $-1/T_1$. In practice a series of experiments is carried out with increasing values of τ. The value of M_{z0} is determined by performing an experiment in which τ is very long.

18.3.5.2. T_2 Measurement

It is much more difficult to measure T_2 than T_1. One of the main reasons is that imhomogeneities in the magnetic field can make a significant con-

tribution to the apparent spin–spin relaxation time. Complex, so-called spin-echo pulse sequences [27], [28] have been developed for such measurements.

18.3.5.3. Spectral Editing Experiments [0], [26]

The most frequently used multi-pulse experiments are probably those for multiplicity determination in ^{13}C NMR. Both the J-modulated spin-echo [29] and the distortionless enhancement by polarization transfer (DEPT) [30] experiments are extensively used for this purpose.

Figure 23 shows the pulse sequence used for the *J-modulated spin-echo* experiment which includes the three periods characteristic of all modern pulse experiments, namely preparation, evolution, and detection. The preparation period is a relaxation delay ending with a $90°_x$ pulse along the x-axis of the rotating frame. The decoupler is switched off, and during the first half of the evolution period the magnetization in the $x-y$ plane evolves under the influence of the proton–carbon couplings. At the beginning of the second period the decoupler is switched on again and a $180°_x$ pulse refocusses all effects other than J-modulation. In the detection period a proton-decoupled FID signal with NOE is observed. If $t = 1/J$, signals from quaternary and CH_2 carbons are in anti-phase to those from CH and CH_3 carbons. Figure 24
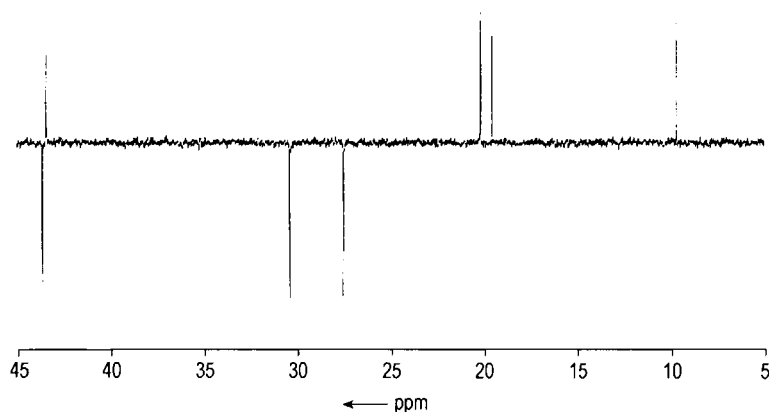
Figure 24. Part of the J-modulated ^{13}C NMR spin-echo spectra of camphor. Note the distinction between the C-3 methylene and the C-4 methine carbons with a shift difference of only 0.3 ppm.

shows such a J-modulated spin-echo ^{13}C spectrum for camphor (**2**).

2
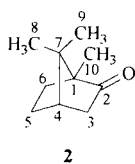
When $t = 1/2J$ only signals from quaternary carbons are observed. Whilst widely used, this experiment has two shortcomings for multiplicity assignment. Firstly, signals from CH and CH_3 carbons cannot be distinguished, and secondly it is sensitive to variations in J values. One-bond carbon–proton couplings are typically ca. 125 and ca. 160 Hz for saturated aliphatic groups and aromatic/olefinic carbon atoms, respectively. For such compounds use of a compromise value for t works well. However, for carbon atoms with much larger couplings such as alkynyl carbons and some unsaturated heterocyclic carbons it is less successful.

In such instances new methods for ^{13}C assignment are more suitable. Compared with the J-modulated spin-echo experiment, the *DEPT* (Distortionless Enhancement by Polarization Transfer) pulse sequence is insensitive to variations in $^1J_{CH}$. The method is based on the use of polarization transfer, whereby magnetization is transferred from protons to ^{13}C nuclei, and as a result quaternary carbon atoms are not detected in such experiments. This experiment involves pulsing both the ^1H and ^{13}C channels. The final proton pulse is along the y-axis and by performing two

experiments with angles for this pulse of 90° and 135°, full multiplicity data can be obtained. The 135° experiment gives negative CH_2, but positive CH and CH_3 carbon signals whilst in the 90° experiment only CH signals are observed.

18.3.6. Multi-Dimensional NMR [123]

A wide range of two-dimensional (2 D) experiments are now available as routine tools, and more recently three- and four-dimensional methods have been developed. In general the experiments are used to extract information from complex spectra (e.g., about which nuclei are J coupled to each other) and to measure J couplings. In this article it is only possible to give a very brief introduction to two-dimensional methodology together with a few examples of its application in ^1H and ^{13}C NMR. Further information about this important field is available in [31], [32], [99], [103], [108], [109].

18.3.6.1. Basic Principles

The common feature of the 1 D multipulse experiments described above was the time sequence preparation – evolution – detection, whereby the detected signal is only a function of the detection time t_2. The important difference in 2 D NMR is that the evolution time t_1 is now a variable. In a 2 D experiment n separate experiments are performed with incremented values of t_1. For each experiment a free induction decay $S(t_2)$ is measured. In this way a matrix $S(t_1, t_2)$ is built up.
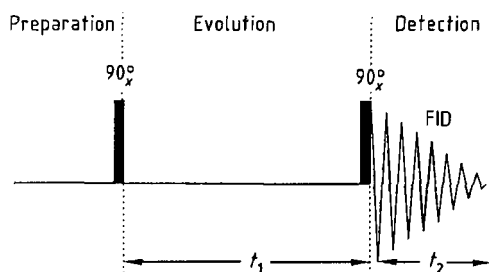
Figure 25. Basic COSY pulse sequence

Fourier transformation of this matrix with respect to t_2 gives a series of spectra $S(t_1, f_2)$. Further transformation with respect to t_1 gives a matrix $S(f_1, f_2)$ which is a spectrum in two independent frequency dimensions. The dispersion of the signals as a function of f_1 depends on the effect of the pulse sequence used on the spin system as t_1 is varied. In general there are two distinct types of 2 D experiments, referred to as J-resolved and correlated spectroscopy. In a *J-resolved spectrum* one frequency axis (f_1) contains spin–spin coupling information and the other (f_2) chemical shift information. In *correlated spectroscopy* both axes contain chemical shift information, the connection between the two being via spin–spin coupling, NOE, or exchange effects.

18.3.6.2. *J*-Resolved Spectra

J-Resolved spectra provide the possibility to observe the resonance frequencies δ and the coupling constants J on two distinct frequency axes separately. This is achieved by a pulse sequence $[T_d - 90° - t_1/2 - 180° - t_1/2 - \text{FID}(t_2)]_n$. After excitation by a 90° pulse the evolution time t_1 is divided by a 180° pulse. The signal (FID) is then detected in t_2. This experiment, known as 2 D J-resolved or J,δ-spectroscopy is mainly used in analysis of crowded spectra. The overlapping absorptions in a 1D proton spectrum can sometimes be resolved by this experiment, in which chemical shifts are presented on one axis and coupling constants on the other. The projection of the 2 D spectrum onto the δ-axis is effectively a "^{1}H broad band decoupled" proton spectrum. However, the relatively long measuring time, which are usually larger than that of other 2 D methods and a number of artifacts, shown in strongly coupled spin systems, caused a decrease of the application frequency.

In heteronuclear *J*-resolved spectra chemical shifts of an arbitrary nucleus X which couples with protons (this is mostly ^{13}C) are presented on one axis and proton-X J couplings on the other. The information content is equivalent to that in a proton-coupled ^{13}C spectrum (Fig. 12) but without the severe overlap of multiplets which is usually encountered in the latter. In common with off-resonance proton decoupling, J-modulated spin echo, and DEPT experiments, it facilitates multiplicity determination. In addition, it enables proton-X coupling constants to be measured.

18.3.6.3. Homonuclear Chemical Shift Correlation (COSY)

The most frequently applied 2 D technique is proton homonuclear correlation using one of the many variants of the COSY (COrrelated SpectroscopY) experiment. The basic COSY experiment consists of two $90°_x$ pulses separated by a time t_1, as shown in Figure 25. Figure 26 shows the aliphatic part of the COSY spectrum obtained from camphor 2.

The spectrum is presented as a contour plot in which intensities above a chosen threshold level are plotted like height contours on a map. Since both frequency domains are proton chemical shifts the matrix is square. Contours along the diagonal of the square correspond to the peaks in a one-dimensional spectrum. Homonuclear couplings give rise to off-diagonal contours or cross peaks. The one-dimensional spectrum is also plotted as a projection along both axes. The cross-peaks provide the same information about proton–proton connectivities as can be obtained from a series of homonuclear decoupling experiments, but all in a single spectrum.

The connectivities for a particular proton can be extracted by drawing a horizontal line from the relevant diagonal peak that intercepts cross peaks corresponding to correlations. In Figure 26 the multiplet at 2.35 ppm is due to H-3$_{\text{exo}}$, and from the cross-peaks it shows correlations to three other multiplets. By drawing a vertical line from each cross-peak back to the diagonal it can be determined with which contour the first diagonal contour is correlated (i.e., coupled). In this case correlations are observed at 2.1, and 1.84 ppm from H-4 and the attached H-3$_{\text{endo}}$, respectively, and a weak one at 1.9 ppm from the H-5$_{\text{exo}}$, four bonds away from the H-3$_{\text{exo}}$. Note that the cross-peaks are found symmetrically on both sides of the diagonal.
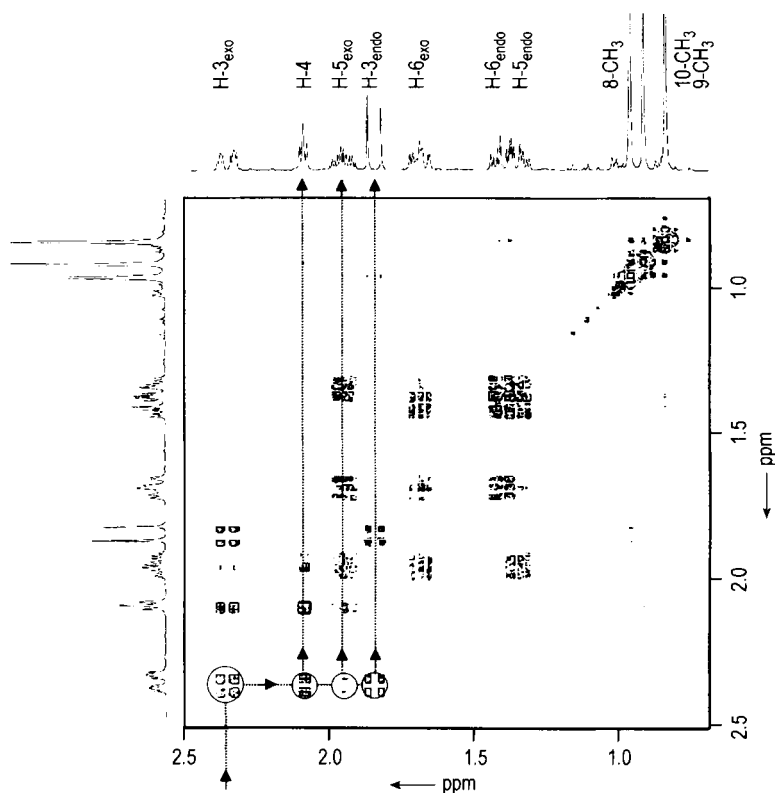
Figure 26. ^{1}H COSY spectrum of camphor

2 D methodology is inherently much less sensitive than 1 D, principally because of the requirement to perform a large number of individual experiments. However, because of the relatively high sensitivity of the proton, a few milligrams of sample are sufficient to yield a good COSY spectrum on a modern high-field instrument. Even shorter measuring time have become possible by the application of gradient accelerated or enhanced COSY spectroscopy, a new technique where the traditional phase cycling techniques will be replaced by the use of pulsed field gradients. Specially designed probe-heads and electronics are necessary to employ the gradients. Depending on the manufacture of the spectrometer, either one or three additional coils are required to generate the z- or the x-, y-, z-field gradients within the sample. There exists, however, no real enhancement by field gradients, and the abbreviation gs-$COSY$ (gradient selected) become increasingly common. Indeed, gradients select desired and undesired co-

herences. Simplified, a coherence describes all possible mechanism for the exchange of spin population between different energy levels in a NMR diagram. This task was previously performed by time-consuming phase cycling. The selection of the desired coherence already happens in the probe, and only one single transient is sufficient, provided that enough substance is available. Thus, a typical gradient enhanced COSY experiment with 256 time increments can be recorded in a few minutes.

In addition, several other modifications of the COSY pulse sequence exist: COSY-45 for the reduction of the diagonal signals, long-range COSY to emphasize small couplings, relayed-COSY for the observation of remote protons, and the double quantum filter DQF-COSY for elimination of singlet signals. The last one is very important for suppression of water signals in biological samples [124].
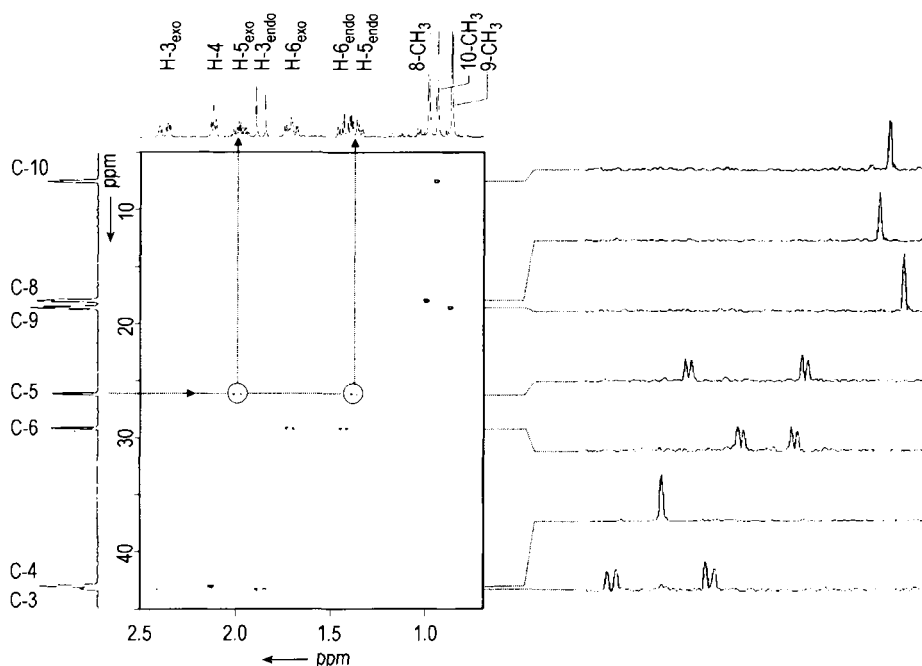
Figure 27. Aliphatic part of the heteronuclear chemical shift correlated (HETCOR) spectrum of camphor. Contour diagram (left) and the traces in the ^{1}H dimension (right) of all C-H-systems.

18.3.6.4. Heteronuclear Chemical Shift Correlation (HETCOR, HMQC)

In the *HETCOR* experiment the peaks of an insensitive nucleus (^{13}C, ^{15}N) are correlated with those of a sensitive nucleus (^{1}H, ^{19}F, ^{31}P). In Figure 27 the aliphatic part of the HETCOR spectrum of camphor (2) shows the specific resonances of the protons which are attached to each ^{13}C nucleus. The relevant parts of the corresponding 1 D spectra are plotted along the axes. A correlation is observed as a cross-peak at the intersection of two lines drawn from a proton resonance and from a carbon peak, respectively. The three pairs of diastereotopic methylene protons H-3$_{endo/exo}$, H-5$_{endo/exo}$ and H-6$_{endo/exo}$ give individual cross peaks at the same carbon resonance, respectively. Correlations are not observed for quaternary carbon atoms. The technique is an important tool for chemical shift assignment and thus structure elucidation.

The cross peaks in the basic HETCOR experiment are generated by a magnetization transfer from the sensitive (^{1}H) to the adjacent insensitive nucleus (^{13}C). A number of variants have been developed to increase the sensitivity and to enable

observation of long range proton-carbon correlations.

A dramatic increase in sensitivity has been obtained by introduction of the reverse or inverse shift correlation, where the sensitive nucleus (^{1}H) is used for signal detection [125]. This method is known by the acronym *HMQC* (heteronuclear multiple quantum coherence). In theory, a ^{1}H-excited and ^{1}H-detected H,C-correlation experiment is 32 times more sensitive (S/N) compared to a normal ^{13}C detection. In practice, an average S/N-enhancement of 2 – 3 for a HMQC compared to a proton decoupled 1 D ^{13}C spectrum will be found. For detection of long-range connectivities between carbons and protons the *COLOC sequence* (correlation via long -range couplings) and an inverse technique with the highest possible intensity, the *HMBC* (heteronuclear shift correlations via multiple bond connectivities) experiment, were developed [126]. In Figure 28 the gradient selected HMBC of camphor are shown. The corresponding 1D ^{1}H and ^{13}C spectra are plotted on the axis analogous to the HMQC spectrum, but in contrast cross peaks are shown here only for H,C-couplings over two (^{1}H-C-^{13}C) or three bonds (^{1}H-C-C-^{13}C). That is illustrated
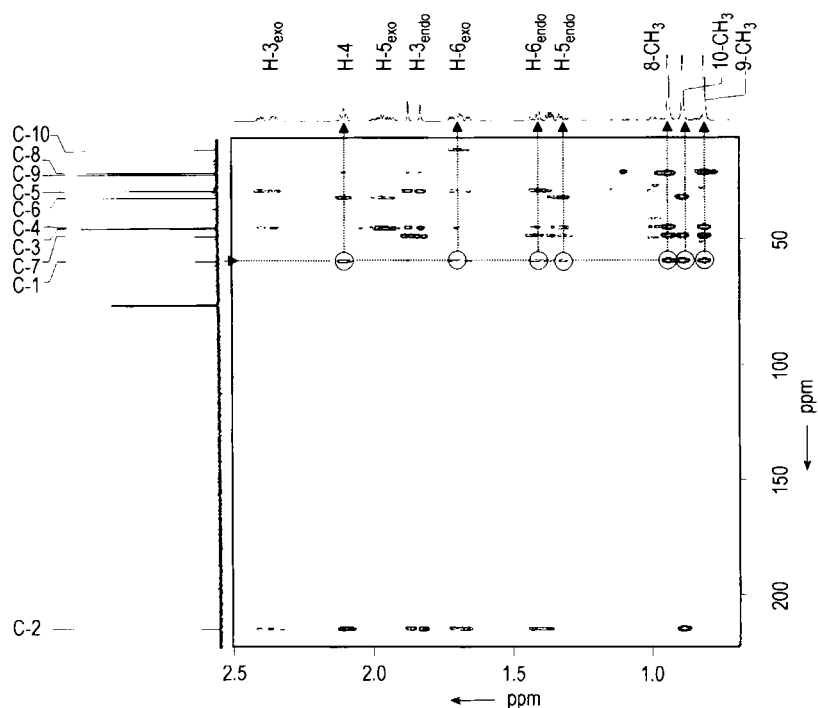
Figure 28. HMBC (*h*eteronuclear shift correlations via *m*ultiple *b*ond connectivities) spectrum of camphor

clearly for the carbons C-1 and C-7 in Figure 28. Both quaternary carbons show cross peaks to all three methyl groups whereas the couplings over one bond are suppressed, as shown for the C-10 methyl carbon. Therefore, the interpretation of a HMBC spectrum requires the previous analysis of the HETCOR or HMQC spectrum.

However, there are some problems connected with inverse correlation experiments that should be mentioned. In contrast to the HETCOR, in the inverse experiment the ^{13}C resolution does not depend on the acquisition time AT, but on the number of time increments, because of the interchange of the f_1 and f_2 frequency axes (see Section 18.3.6.1). Another large problem is the suppression of the undesired signals of protons bound to ^{12}C. In theory, phase cycling should remove all non-^{13}C bonded protons. But a perfect cancellation can never be achieved. Thus, a severe so-called t_1 noise ridge is usually shown at all ^1H chemical shift frequencies. In this article it is only possible to name the two methods that help to reduce these problems dramatically: The BIRD (bilinear rotational decoupling) sequence and the pulsed field gradients (PFG, see Section 18.3.6.3). With PFG the homogeneity of the magnetic field

can be destroyed in a controlled way. This yields artifact-free correlation spectra in a fraction of the time need previously (e.g., gs-HMQC). Today with modern spectrometers, the 2 D gs-HMQC is a good alternative to the normal 1 D ^{13}C NMR spectrum and there is no reason to run conventional HETCOR experiments anymore.

18.3.6.5. Homonuclear NOE Correlation

The 2 D NOESY (nuclear Overhauser and exchange spectroscopy) technique in principal enables all proton – proton NOE effects to be assembled in a single spectrum. Such spectra are comparable to the respective COSY spectrum. The cross peaks arise from through-space proton-proton interactions between nonbonded protons that are nearby in space. The technique is extensively used in the determination of three-dimensional structures and conformations of molecules. This is shown in Figure 29 with a part of the NOESY spectrum of camphor (**2**). The ^1H NMR signals at 0.83 ppm (^{13}C 19.8 ppm) and 0.97 ppm (^{13}C 19.2 ppm) can be unambiguously assigned to the methyl groups 8 and 9 by use of the NOE effects. The high-field proton methyl signal shows an ob-
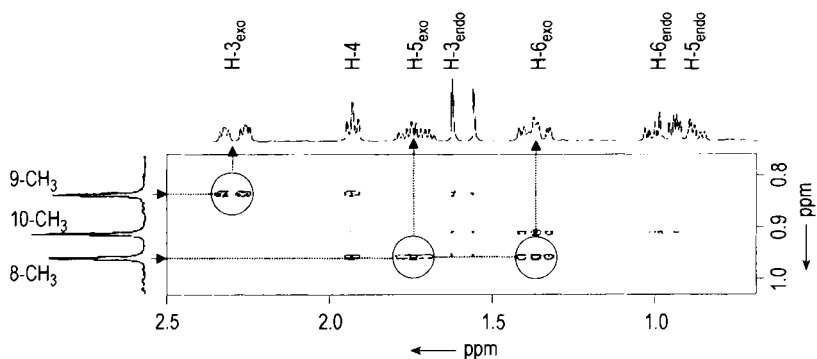
Figure 29. Part of the NOESY spectrum of camphor. Only the NOE effects of the three methylgroups (left) to all other protons (top) are shown here.

vious NOE to the H-3_{exo} proton what is only possible from the 9-position whereas the low-field signal shows effects to the H-5_{exo} and H-6_{exo} protons.

18.3.7. NMR Spectral Collections, Databases, and Expert Systems

NMR is a powerful tool for the determination of structures from first principles and the chemical shift is the most important NMR parameter in structural analysis. For estimating the relationship between chemical structures and chemical shifts three possibilities exist: the calculation of the chemical shift values by empirical methods [137], the computation by quantum chemical procedures, e.g., with the IGLO-method (Individual Gauge for Localized Orbitals [129]), or the use of large compilations of NMR spectra and the associated chemical structures. The access to relevant reference data for identical or similar compounds can facilitate the assignment process enormously. Reference data may assist by reducing the amount of experimental and/or interpretive effort required or increase confidence in the suggested structure.

Many data compilations exist, containing significant amounts of reference data, especially those dedicated to a single nucleus. In addition there are many spectral libraries, some available in hard-copy format only. One of the first ^{1}H NMR libraries was published by Varian Associates always in the early 1960s [33]. Such older hard-copy libraries contain only low-field data and are normally searchable by parameters such as name, molecular formula, chemical shift value, or compound type [34]–[38]. [137]. However, state of

the art are search softwares in computer-readable form. Computer-searchable libraries, especially those which allow substructure searching of compound data in conjunction with spectral data searching, are much more powerful than hardcopy libraries. In 1973 BREMSER published a procedure for the description of the chemical environment of nuclei by their hierarchically ordered spheres of environment (HOSE code) [130]. Today several databases exist containing hundreds of thousands of chemical shift values in particular for ^{13}C nuclei and the corresponding information about the chemical environment of the individual carbons. Only some examples of electronically stored databases should be mentioned here: SpecInfo [39], CSEARCH [132], and CNMR [133]. Their current contents are 149 000 (500), 265 000 (0) and >67 000 (>82 000) spectra for ^{13}C NMR (^{1}H NMR) spectroscopy, respectively. All three systems listed above contain assigned NMR data which can be searched by molecular formula, name, peak positions/-spectrum similarity, and structure/substructure. Users are also able to add their own spectral data to the libraries. All databases can be used to identify known substances contained within the database, and they are an efficient tool for elucidating the structure of unknown substances.

The SpecInfo system which was originally developed in-house by BREMSER et al. at BASF and was also marketed in a searchable microfiche version [40], is the premier spectroscopic archive and interpretation system for ^{13}C, ^{1}H, hetero-NMR, IR, UV/VIS, Raman, and mass spectra. In a more automated way the structure assignment can be achieved by using a structure generator, which assembles all possible isomers for a given

Table 9. ^{13}C chemical shift estimation of the camphor carbons with different databases

C	CH_n	exp	SpecTools [a] [153]	C_shift [a] [134]	SBSD [135]	ACD [133]
1	C	57.7	64.7	60.3	57.6	57.2
2	C	219.7	213.9	216.5	219.3	217.3
3	CH_2	43.3	33.1	42.9	43.3	43.7
4	CH	43.1	36.3	46.6	43.1	43.3
5	CH_2	27.1	22.1	26.5	27.1	27.3
6	CH_2	30.0	26.3	29.0	29.95	30.1
7	C	46.8	31.8	47.6	46.8	46.6
8	CH_3	19.2	21.4[b]	20.2[b]	19.15	18.7
9	CH_3	19.8	21.4[b]	20.2[b]	19.8	20.7
10	CH_3	9.3	12.7	11.8	9.25	9.5

[a] In contrast to the databases which fall back on the experimental spectra, the shift values were calculated with this PC programs by use of additional increments (SpecTools) or neural networks (C_shift).
[b] Exchangeable since stereochemistry is not considered here.

molecular formula and performs an automatic ranking of the candidate structures (SPECTACLE [131]). In 2000 a PC program was developed on basis of the SpecInfo database for the calculation of ^{13}C NMR spectra of any proposed organic molecular structure [134]. The spherically encoded chemical environments of more than 500 000 carbons were used here to train artificial neural networks which allow the fast determination of ^{13}C chemical shift values.

CSEARCH spectrum estimation is based on scripts written by R. Bobrovsky, G. Löffler ,and W. Robien and can be accessed via an e-mail based server system at the University of Vienna [132]. Four different versions are available. Optionally the stereochemistry of the molecules can be considered and/or neural networks can be used for the fast estimation of chemical shifts.

The Advanced Chemistry Development (ACD) presents a toolset of PC and web-based software for NMR prediction, processing, and database management also for ^{19}F (> 11 500 spectra), ^{31}P NMR (> 18 500 spectra) and for MS, IR, UV-Vis and chromatographic databases. The operation of an actual NMR spectrometer can be simulated here, allowing to choose among different spectra modes (off-resonance, J-modulation, DEPT) just as the operating frequency, the solvent and the concentration of the solute.

Another web-based integrated spectral data base system for organic compounds (SDBS) was developed by the National Institute of Materials and Chemical Research of Tsukuba (Japan) [135]. This system includes six different types of spectra: ^{13}C NMR (ca. 11 000 spectra), ^1H NMR (ca. 13 500 spectra), MS, IR, laser-Raman, and ESR (ca. 2000 spectra). The studies on the SDBS started already in early 1970. This system only

allows the search for available structures and not the determination of spectra from unknown compounds.

However, spectra databases are statistical tools to establish the relationships between NMR spectral parameters and chemical environment of individual atoms. So, the results can only be offered with a statistical probability, depending on the quantity and quality of the available database entries. In other words, the accuracy of the predicted data cannot be more precise than the stored data. Usually, ^{13}C NMR spectra can be calculated for almost any drawn organic structure to an accuracy of ± 3 ppm or better, apart from stereochemical problems which can not be considered by some databases. Table 9 summarizes the results of the determination of the ^{13}C NMR chemical shifts for the ten carbon atoms in camphor (2) obtained with different methods.

Many publications on a variety of expert systems and computer-based structure determinations will be found in the recent literature [136]. In general they are not associated with the large commercially available databases but they are useful for the elucidation of a completely unknown structure. However, in most practical cases the more common type of structure determination is structure verification. Here, the structure information achieved via the chemical shift is usually sufficient.

18.3.8. Applications

Solution NMR spectra usually contain a wealth of information. Previous sections have provided an overview of experimental techniques, the parameters which can be determined from NMR spectra

and techniques which can be used for the assignment of signals in spectra. Here some of the applications of solution NMR spectroscopy are briefly summarized.

18.3.8.1. Chemical Structure Determination

NMR spectroscopy is the most powerful tool available for the unambiguous determination of molecular structures of compounds in solution. In a modern chemical research laboratory the principal tools for structure determination are usually mass spectrometry, to obtain molecular mass data, and NMR spectroscopy which provides detailed information about the groups present and how they are assembled. The abundant nuclei ^1H, ^{19}F, and ^{31}P have been used for over 40 years to study the structures of organic and inorganic chemicals. In the early 1970s the advent of FT methodology enabled a much wider range of nuclei to be studied and in particular resulted in an explosive growth in the use of ^{13}C NMR as a complementary technique to ^1H NMR in the identification of organic compounds. The most useful NMR parameters are, in general, chemical shifts, coupling constants, and integrals from which structural information can be deduced.

In addition to the study of monomeric species, NMR is a long established tool for studying the structures of synthetic polymers [41], [138]. In the case of copolymers the various groups present can be identified, their relative concentrations determined, and often information about sequence distributions obtained. The latter provides a measure of the random/block character of the polymer. Whilst ^1H and ^{13}C are the most commonly studied nuclei for such work several others such as ^{19}F and ^{29}Si have been used where appropriate. For vinyl polymers ^{13}C and ^1H NMR spectra provide information on tacticity. Chain branching and end groups can also be identified from NMR spectra of polymers.

During the last decade NMR has developed as a powerful tool for biochemical structure determination. The most impressive applications are in the determination of protein structures with molecular masses up to the 25 000 – 30 000 range [42], [139]. Such work has been made possible by the development of multidimensional NMR techniques. 2 D and 3 D techniques are widely used for such work and 4 D experiments are being developed. 3 D and 4 D methodologies are used in conjunction with ^{15}N and/or ^{13}C labelling of the amino acid residues to reduce resonance overlap.

To date NMR protein structures have usually been calculated on the basis of interproton distances and dihedral angles derived from NOE and J coupling measurements, respectively. Each structure determination necessitates a large number of sophisticated experiments, so that data acquisition and subsequent analysis may take several months. The value of the solution structure lies in the fact that it provides a starting point for the study of protein/substrate interactions, a knowledge of which may then aid drug design.

18.3.8.2. Quantiative Chemical Analysis by NMR [43]

NMR spectroscopy offers several important advantages for quantitative analysis over other techniques, including LC and GC chromatography. Firstly a single technique can be used to unambiguously confirm the identity of the components and quantify them. For chromatographic analysis spectroscopic techniques such as MS and NMR are generally required initially to identify components. Secondly and most importantly pure samples of the compounds of interest are not required to calibrate the response of the instrument in an NMR experiment. This is a result of the fact that, given due attention to the experimental NMR conditions, the integrated resonance intensity is directly proportional to the concentration of nuclei giving rise to the resonance. Relative concentrations can be obtained directly from relative resonance intensities and absolute concentrations by adding a known amount of another compound as an internal intensity standard. The nondestructive nature of the NMR experiment offers an additional advantage. On the negative side, the NMR experiment is inherently much less sensitive than other spectroscopic and chromatographic methods. Therefore the technique is rarely suitable for quantifying components present at very low concentrations even when the more sensitive nuclei such as ^1H, ^{19}F, or ^{31}P are used.

In a modern NMR laboratory, pulsed FT techniques are normally used for quantitative analysis. For such work careful selection of the experimental parameters is required to obtain accurate intensity relationships between the resonances in a spectrum. Whilst ^1H NMR spectra are normally integrated for structural identification work, the accuracy required to identify the relative numbers of different types of protons in a molecule is much less than for quantitative analysis.

For proton NMR the main consideration is ensuring complete relaxation between successive pulses for all the different types of hydrogen atoms present. This requires the interpulse delay to be at least 5 times the longest T_1. In addition, for nuclei such as ^{13}C, where broadband decoupling is usually required, the inverse gated technique (Section 18.3.4.2) should be used to prevent the occurrence of NOE effects. A further consideration in the case of spectra from nuclei such as ^{13}C and ^{19}F, which may have very wide spectral widths, is whether the RF pulse has sufficient power to irradiate all the nuclei equally effectively. The digital resolution and data processing requirements for a particular application also require careful selection.

The wide range of chemical, biochemical, and clinical applications include strength determinations, mixture analyses, polymer analyses, and reaction monitoring. Choice of nucleus obviously depends on the individual application. For organic chemical applications, the proton would normally be the preferred nucleus because of its relatively high sensitivity. However, proton spectra of mixtures are often highly congested and the greater dispersion afforded by ^{13}C NMR may make it the nucleus of choice. Two specialist applications are described briefly below.

Isotope Content Determination. A highly developed example of isotope content determination is the site-specific natural isotope fractionation (SNIF) NMR of deuterium [44]. The natural abundance of deuterium at different sites in a natural product can vary significantly depending on its biochemical origin. This pattern of natural abundances of deuterium is often quite characteristic of the source of the material, as is the case for alcohol in wine. By using a standard with a known deuterium content, ^{2}H NMR can be used to determine the deuteration levels in the alcohol. This data provides information about the origin of the wine and is a means of detecting watering down or artificial enrichment.

Enantiomeric Purity Determination. NMR has become an increasingly important tool for the determination of enantiomeric purity [45]. Many pharmaceuticals and agrochemicals are chiral, with only one enantiomer having the required effect. At best the other is inactive, at worst it may have undesirable properties. This situation has led to a surge in enantioselective synthesis, the products of which require analysis. Chromatographic methods which use chiral columns are usually used for quality control purposes with an established process, but at the research/development stage NMR is often the method of choice. In an achiral medium, enantiomers cannot be distinguished by NMR because their spectra are identical. Diastereoisomers, however, which have different physical properties, can be distinguished by NMR. Therefore the determination of enantiomeric purity by NMR requires the use of a chiral auxiliary that converts a mixture enantiomers into a mixture of diastereoisomers. Chiral lanthanide shift reagents [46] and chiral solvating agents form diastereomeric complexes in situ and can be used directly for NMR analysis. A third method involves reaction with chiral derivatising agents prior to NMR examination.

Lanthanide shift reagents (LSR) are *tris* (β-diketonate) complexes. Further complexation with organic compounds may result in large shifts in the NMR resonances of the latter. This is caused by the magnetic properties of the lanthanide ions. Compounds containing a wide range of functional groups, including alcohols, amines, and ketones, form complexes with LSRs. Prior to the widespread introduction of high-field instruments such materials were widely used to increase chemical shift dispersion in spectra. If an LSR chelate containing an optically pure ligand interacts with a pair of optical isomers, two diastereoisomers result which can be distinguished by NMR. Many of the ligands used in such reagents are derivatives of camphor.

Chiral solvating agents (CSA) form diastereoisomeric solvation complexes with solute enantiomers via rapidly reversible equilibria in competition with the bulk solvent.

Where quantitative chiral purity determinations are required, the minimum level of detection for the minor component and accuracy achieved depend on factors which include the resolution between the resonances from the diastereoisomers formed and the signal-to-noise ratio in the spectrum.

18.3.8.3. Rate Processes and NMR Spectra [47]

NMR is an important tool for the study of certain types of rate processes. In deciding whether a process is amenable to study by NMR the rate must be compared with the "NMR timescale," which refers to lifetimes of the order of 1 s to 10^{-6} s. An example of a process which can be studied in this way is rotation about the C–N bond in an amide such as N,N-dimethylacetamide (3).
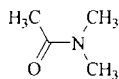
3

Rotation causes the two *N*-methyl groups to exchange positions. In the room-temperature ^{1}H NMR spectrum separate peaks are observed for the two *N*-methyl groups. If the temperature is raised, the rate of rotation increases and the *N*-methyl resonances broaden until a temperature is reached when the two peaks coalesce into a single broad resonance. This peak sharpens as the temperature is raised further until finally the fast exchange limit is reached, above which no further change is detected. The temperatures at which slow exchange, coalescence, and fast exchange are observed for a particular process depend on the rate of the process and the frequency difference between the exchanging lines. Thus at higher field coalescence will take place at a higher temperature. Analysis of such spectra enables exchange rates to be determined over a range of temperatures. Examples of other processes which have been studied in this way include keto–enol tautomerism, ring inversion, and proton-exchange equilibria.

18.3.8.4. NMR Methods Utilized in Combinatorial Chemistry and Biochemistry
[110], [140]

Combinatorial chemistry provides a powerful means of rapidly generating the large numbers of structurally diverse compounds necessary for the biological screening required by drug discovery. Thus, combinatorial chemistry has become an indispensable tool in pharmaceutical research. The diversity of techniques being employed has caused the adaptation of a number of NMR methods for combinatorial chemistry purposes. These methods include the NMR analysis of reaction intermediates and products in solution or in the gel state and the analysis of ligands interacting with their receptors. High-throughput screening strategies were developed which involve the acquisition of 2 D spectra of small organic molecules in a few minutes. Libraries of more than 200 000 compound can be tested in less than one month. There are many advantages of high-throughput NMR-based screening compared to conventional assays, such as the ability to identify high-affinity ligands for protein targets. This suggests that the method will

be extremely useful for screening the large number of targets derived from genomics research [145].

SHUKER et al. described a NMR method in which small organic molecules that bind to proximal subsites of a protein and produce high-affinity ligands are identified. This approach was called "SAR by NMR" because structure-activity relationships (SAR) are obtained here from NMR. This technique involves a series of 2 D spectra of a labeled receptor protein in the presence and absence of potential ligands. The method reduces the amount of chemical syntheses and time required for the discovery of high-affinity ligands and appears particularly useful in target-directed drug research [141].

NMR has become established also as a valuable technique in clinical biochemistry where ^{1}H NMR in particular is used to study complex biochemical fluids such as plasma, urine, and bile. The major problem associated with such work is the removal or reduction of the very large water signal. One method involves freeze-drying the sample and redissolving it in deuterated water. NMR techniques are also available for selective suppression of the water signal. These methods, such as presaturation of the water, are required to reduce the water signal such that acceptable suppression can be attained from phase cycling. The same effect can be accomplished with magnetic field gradients alone because the removal of the undesirable resonances is accomplished in a single scan.

Under the conditions used, only low molecular mass constituents are observable, the signals from proteins and other large molecules are not resolved. Applications include screening and monitoring for metabolic diseases, the study of biochemical mechanisms associated with disease processes, and the identification of drug metabolites.

Translational diffusion measurements are a useful tool for studying supramolecular complexes and for characterizing the association state of molecules that aggregate at NMR concentrations. NMR spectra quality can provide a direct measure of the mobility of organic polymeric compounds and pulsed field gradients can be used to measure diffusion coefficients. Therefore, diffusion measurements can play an important role in the study of the properties and dynamics of the resins used in combinatorial chemistry. *Diffusion-ordered NMR* (DOSY) is another powerful tool for the analysis of complex mixtures. It allows to resolve the NMR signals of discrete compounds in a *mixture on*

basis of variance of their molecular diffusion coefficients [142].

One of the most powerful methods for performing analysis of complex mixtures represents the redevelopment of HPLC-NMR coupling techniques [143]. HPLC-NMR has been widely applied in drug metabolism and also used for combinatorial chemistry. Data collection usually occurs with the system operating in either on-flow or stopped-flow mode. In the latter, up to three dozen fractions may be collected within storage loops. If the differences in the retention time of individual compounds are smaller than the required NMR measuring time, the storage loop acts like a sample changer.

18.4. NMR of Solids and Heterogeneous Systems [48]

Solids give very broad NMR signals devoid of the fine structure which provides detailed information about chemical structures. The restriction of NMR spectroscopy to the liquid, solution, or gas phase, however, imposes considerable limitations on its applicability since some compounds are insoluble or have very low solubility. Moreover, a compound may experience a change in molecular structure on dissolution (e.g., by tautomerism). Studying samples as solids should provide some insight about how molecules pack as well as about molecular structure. Since the 1970s developments in experimental design and spectrometer hardware have made it possible in some situations to obtain spectra from solids in which the resolution approaches that obtainable in a solution experiment.

This section provides a brief explanation of the origins of the broadening effects, followed by an overview of the techniques employed to obtain high-resolution NMR spectra from solids and examples of their applications. Only spin 1/2 nuclei are considered. There are two principal factors which cause broadening in the NMR spectra of spin 1/2 nuclei obtained from solid samples. The first is chemical shift anisotropy: the chemical shifts are orientation dependent and the different crystallites in a sample have a range of orientations to the applied magnetic field. Broadening resulting from dipole–dipole interactions is the second factor. Both of these interactions are present in solution but are usually averaged out by molecular tumbling. A further problem results

from the fact that nuclei in solids generally have very long relaxation times. This means that they take a long time to relax back to the equilibrium magnetization, which in turn affects the amount of time required to obtain a spectrum with a reasonable signal-to-noise ratio.

18.4.1. High-Resolution NMR of Solids

The study of "dilute" nuclear spin systems such as ^{13}C has been more amenable to obtaining high-resolution solid-state spectra than that of abundant nuclei such as the proton. The reason for this is that the problems associated with homonuclear dipolar interactions are essentially eliminated in a dilute spin system. Dilution may be the result of low natural abundance, as in the case of ^{13}C, or of abundant isotopes present in low concentration (e.g., ^{31}P in organophosphates). In either case homonuclear dipolar interactions are negligible, since dipolar interaction has a $1/r^3$ dependence. Heteronuclear dipolar interactions, usually to protons, remain but these can be removed effectively by the use of high-power proton decoupling. A technique known as magic angle spinning (MAS) is used to remove the effects of chemical shift anisotropy. This involves spinning the sample at an angle of $54°44'$ to the applied magnetic field. To be effective the rate of rotation must be comparable to the frequency range being observed (i.e., several kilohertz). To overcome the problems associated with low sensitivity a technique known as cross-polarization (CP) is used. This involves the transfer of magnetization from highly polarizable abundant spins such as protons to dilute spins such as ^{13}C.

The combined application of CP, MAS, and high-power proton decoupling can give high-resolution spectra from powders and amorphous solids, comparable in quality to those obtained in solution. The ^{13}C, ^{29}Si, ^{15}N, and ^{31}P nuclei have been widely studied in this way in the last 20 years [128], [146], [147] . There are two principal reasons for obtaining a high-resolution NMR spectrum from a solid. Either it is insoluble or information is being sought about the solid-state structure. In the latter case crystallographic information is sought either because a single-crystal X-ray structure is not available or to enable a comparison of solid and solution state structures/conformations to be made when spectra are available in both phases. Crystallographic effects frequently give rise to splitting of lines in spectra of solids
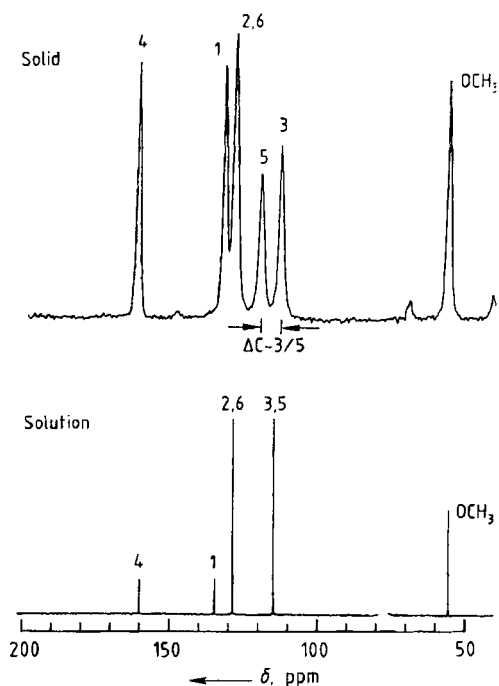
Figure 30. ^{13}C NMR spectra of 4,4'-dimethoxybiphenyl (**4**)

when compared with those obtained in solution. For a particular type of atom to yield a single line in a spectrum obtained from a solid, all the carbons of that type must be related by symmetry in the crystal. Nonequivalence is not an uncommon occurrence and may be intramolecular or intermolecular. Figure 30 shows the solid-state and solution spectra of 4,4'-dimethoxybiphenyl (**4**). In solution a single peak is observed for C-3 and C-5, whereas for the solid two peaks separated by 6.4 ppm are present. This is attributed [49] to the fact that in the solid the compound exists in the locked conformation (**4**) whereas in solution rotation about the C-4–O bond causes an average shift to be observed.

4
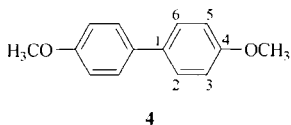
Solid-state NMR spectroscopy, particularly ^{13}C, has been widely used to study polymorphism since polymorphs invariably give spectra which are discernably different.

18.4.2. Low Resolution ^{1}H NMR of Heterogeneous Systems

Thus far the discussion has assumed a requirement for the highest resolution NMR spectra. However, there are important analytical applications where it is not necessary to resolve the signals from all the different hydrogen atoms in a sample. As in the case of high-resolution NMR, early instruments used CW methodology. Pulsed NMR, which is much better suited to and more versatile for this type of measurement, is now used. After excitation the bulk magnetization of a sample is studied. The initial amplitude of the signal is proportional to the total number of hydrogen nuclei in the sample. Each component of the signal decays at a different rate, so, for example, signals from hydrogen atoms in solid phases decay far more rapidly than those from liquids. There are three types of measurements on such instruments. For many years the methodology has been used to measure solid/liquid ratios (e.g., to determine the solid fat content in margarine). An absolute measurement method is also available where the absolute quantity of protons in a sample can be measured and expressed as a percentage by weight. The third more sophisticated method involves the measurement of bulk relaxation times, which can be correlated with a property of the sample. For instance the decrease in the spin – spin relaxation time T_2 has been used to follow the styrene polymerization reaction [50].

18.5. NMR Imaging [148] – [151]

Since 1970 a technique known as magnetic resonance imaging (MRI) has been developed which has revolutionized diagnostic medicine. The technique produces an NMR picture or image. In addition to the NMR methodology described above, MRI also uses magnetic field gradients in the x, y, and z directions to make the resonance frequency a function of the spatial origin of the signal. First, a slice at height z and thickness dz of the object is selected. This is done by means of a selective RF pulse, combined with a field gradient in z direction. Next, the spatial coding for the x direction is obtained by means of an x-gradient. Finally, the spatial coding for the second direction, i.e., the y direction, is obtained by means of a y-gradient. In medical imaging the protons in water are usually detected. While the water con-

tent in different types of tissue may show little variation, the T_1 and T_2 relaxation times are different. Pulse sequences are therefore used which produce contrast that reflects the different relaxation times. The technique has become particularly important in the diagnosis of cancer, since cancerous tissue has a longer relaxation time than healthy tissue. In recent years this method acquired special importance in neurochemistry [152].

Equipment is available for a range of applications from whole human body imagers operating at field strengths of up to ca. 8 T to equipment for imaging small samples up to 11.7 T. The achievable resolution depends on the magnitude and duration of the field gradients applied. For small objects the best resolution which has been achieved is of the order of 5 – 10 µm. For whole body in-vivo applications it is much lower.

The techniques developed for biomedical applications to study the liquid state have also been used in materials science. Such applications include the study of solvents ingressed into polymers, liquids absorbed onto porous media, and polymerization reactions [51], all of which can be studied as a function of space and time.

18.6. ESR Spectroscopy

Electron spin resonance (ESR) spectroscopy is also known as electron paramagnetic resonance (EPR) spectroscopy or electron magnetic resonance (EMR) spectroscopy. The main requirement for observation of an ESR response is the presence of unpaired electrons. Organic and inorganic free radicals and many transition metal compounds fulfil this condition, as do electronic triplet state molecules and biradicals, semicon-ductor impurities, electrons in unfilled conduction bands, and electrons trapped in radiation-damaged sites and crystal defect sites.

The principles of ESR and general applications are covered in various monographs [52] – [57]. Reviews of technique developments and applications have been published [58], [59]; industrial applications have been reviewed [60]; and dosimetry [61] and biological applications described [62] – [66].

18.6.1. The ESR Experiment

The resonance condition for ESR, for a system having a spin value of 1/2, is:

$$\Delta E = h\nu = g\beta B \qquad (3)$$

where ΔE is the separation of energy levels produced by the application of an external magnetic field B, and β is the Bohr magneton $(9.274 \times 10^{24}$ J/K$)$. Most ESR spectrometers operate at a fixed frequency ν and record an ESR spectrum by sweeping the external field B. The most convenient frequency is ca. 9 GHz, which is in the microwave region. It corresponds to a wavelength of ca. 3 cm and is known as X-band. Spectrometers have been built which operate at other microwave frequencies: L-band (1.0 – 1.2 GHz), S-band (3.8 – 4.2 GHz), K-band (24 GHz), and Q-band (34 GHz); and some even operate at a few megahertz in the earth's field. The lower frequencies are used when "lossy" samples, that is those with high dielectric constants such as aqueous or biological samples, are examined. The higher frequencies are used when greater dispersion is required, e.g., when improved separation of anisotropic g components in solid samples is required to reduce overlap of lines. Unfortunately, at the higher frequencies the sample size is restricted and noisier spectra often result.

18.6.1.1. Continuous Wave ESR

The basic components of a continuous wave (CW) ESR spectrometer are a frequency source (klystron or Gunn diode), an electromagnet (having a field of ca. 350 mT for X-band) and a resonant cavity (or loop gap resonator) in which the sample is placed (Fig. 31). To improve the signal-to-noise ratio, it is customary to modulate the microwave frequency at 100 kHz (other frequencies are also used) and to detect at this frequency by using a phase sensitive detector. This results in the usual recorder tracing from the spectrometer being the first-derivative instead of absorption as found in NMR spectrometers. A simple ESR spectrum is shown in Figure 32.

18.6.1.2. ENDOR and Triple Resonance

Electron – nuclear double resonance (ENDOR) offers enhanced resolution compared with conventional ESR. This is achieved mainly because the technique is ESR-detected NMR spectroscopy.
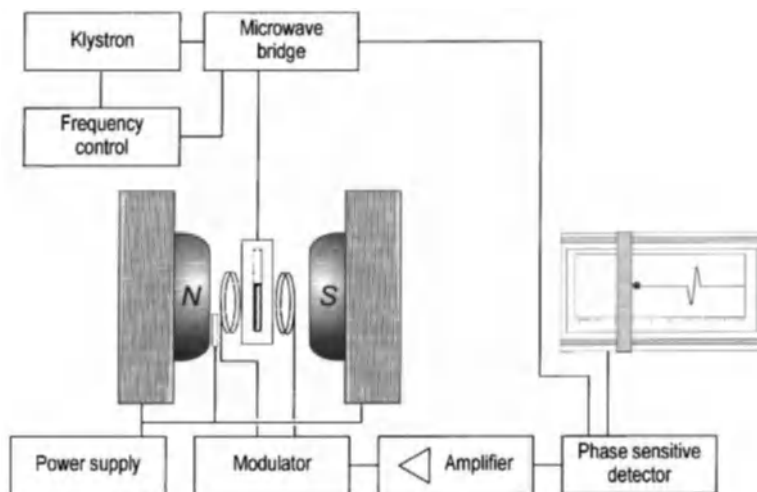
Figure 31. Block diagram of a CW ESR spectrometer
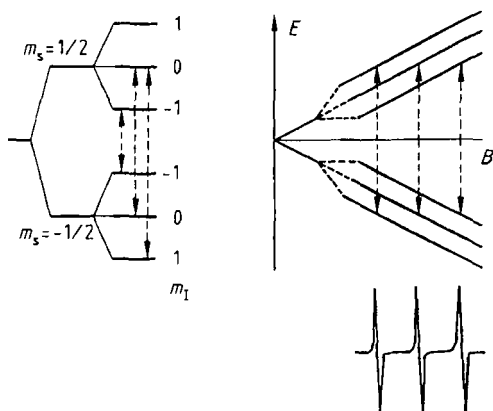
Figure 32. Simple ESR spectrum
P. L. Nordio in [64] (with permission)

Another advantage is that only two ENDOR lines arise from each group of equivalent nuclei, resulting in a simplification in the spectra when compared with ESR. ENDOR has been applied to the study of organic [67] and inorganic radicals and transition metal complexes [62], [68] in both the solid (powdered and crystalline) and liquid states.

The ENDOR experiment involves setting the spectrometer on a chosen line in the ESR spectrum and then increasing the microwave power to saturate the transition while sweeping the selected NMR frequency.

Two NMR transitions are observed at the frequencies:

$$v = ||v_n \pm a/2|| \tag{4}$$

where v_n is the NMR frequency for the applied magnetic field of the ESR spectrometer and a is the ESR hyperfine coupling constant. Equation (4) applies when $v_n > |a/2|$; however, when $v_n < |a/2|$, the two lines are centred at $|a/2|$ and are separated by $2 v_n$.

Other experiments performed with an ENDOR spectrometer include special and general electron-nuclear-electron triple resonance, where three frequencies are employed. In the special triple experiment, two NMR transitions belonging to one set of equivalent nuclei are irradiated. This results in improved spectral resolution compared with ordinary ENDOR. In the general triple experiment, NMR transitions belonging to different sets of nuclei are irradiated. This allows the relative signs of the hyperfine coupling constants to be determined. As an example the ENDOR spectrum, and the corresponding ESR, general triple and special triple spectra, of bacteriochlorophyll a (**5**) are shown in Figure 33.

Apart from the improvement in resolution already described, ENDOR spectroscopy has several additional advantages over CW ESR, namely:
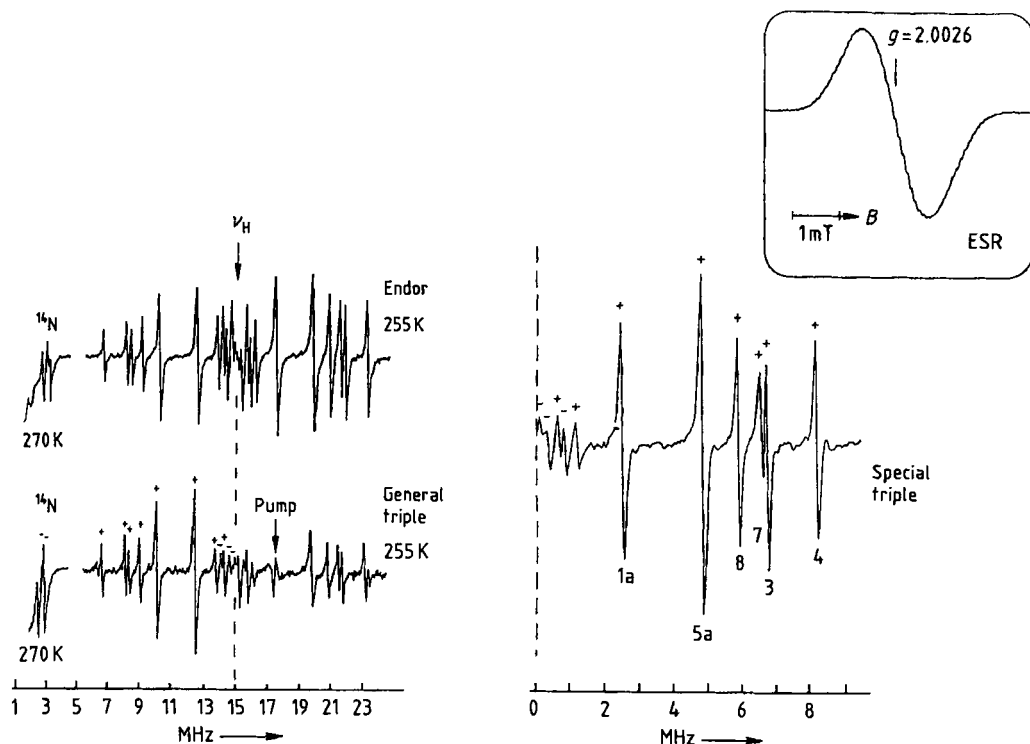
Figure 33. Typical ESR, ENDOR, general triple and special triple spectra of bacteriochlorophyll a (5) cation radical (iodine oxidation in 6 : 1 $CH_2Cl_2 : CH_3OH$)
K. Mobius, W. Lubitz and M. Plato in [66] (with permission)
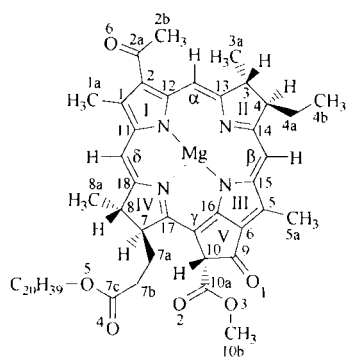
5

1) Simplification of the spectrum when several equivalent nuclei are present
2) Direct measurement of hyperfine splittings
3) Unambiguous assignment of the hyperfine splittings of magnetic nuclei since the spectrometer is set at the appropriate NMR frequency for that particular nucleus

4) Direct determination of quadrupolar splittings
5) Determination of relative signs of hyperfine and quadrupolar interactions
6) Spectra of individual species in a mixture can be determined
7) Very small hyperfine coupling constants can be measured

18.6.1.3. Pulse ESR [69]–[73]

Although the basic principles of ESR and NMR are similar, practical difficulties mean that pulse methods are less useful in ESR. This is because the pulse power required to produce the frequency span of a typical ESR spectrum would be several kilowatts and the pulse very short (nanoseconds). The pulse or FT (Fourier transform) ESR spectrometer is usually based on a standard CW instrument because it is often useful to record a standard ESR spectrum before carrying out pulse experiments. The pulse microwave source is usually a travelling wave tube; other essentials

include a travelling wave tube amplifier, a programmable pulse generator, and fast digitizer.

SCHWEIGER [72] has reviewed pulse ESR and pulse ENDOR methods and HOFF [66] includes many recent applications. Echo methods are similar to those used in NMR. The electron spin echo (ESE) method involves the application of two or three microwave pulses to a spin system in a constant magnetic field. The ESE signal decay due to relaxation is accompanied by periodic rises and falls in the echo amplitude. These modulation effects depend on electron–nuclear and electron–electron couplings in the spin system, and are termed electron spin echo envelope modulation (ESEEM) [70], [71]. ESEEM is used to investigate hyperfine interactions in magnetically diluted solids, and most experiments are performed at temperatures between 4 and 77 K. ESEEM has been applied to the study of surface complexes, coordination of water in metmyoglobin, structures of ligands in copper proteins, and magnetic properties of electronic triplets and solitons in polyacetylene. Pulses and detection need to be on a very short time scale since a typical spin–lattice relaxation time is about 1 μs.

Spin dynamic studies, including saturation and inversion recovery, Hahn echo and stimulated echo decay, as well as Carr–Purcell–Meiboom–Gill sequences, can be performed, yielding relaxation times. When rapid reaction kinetics are being measured, information is also gained on transient phenomena such as chemically induced electron polarization (CIDEP).

Pulsed ENDOR has several advantages over conventional ENDOR for solids [72], [73]. This means that pulsed ENDOR and ESE–ENDOR can be used at any temperature, provided that a spin echo can be detected.

18.6.1.4. ESR Imaging

ESR imaging [74] is used for both biological and materials research and, as with the more commonly used NMR imaging, it is based on the application of field gradients in the x, y, and z directions, to allow volume elements to be selectively studied. Large field gradients are required because of the wide field range of the ESR spectrum, typically ca. 3 mT.

Imaging has been performed at various irradiation frequencies. X-band (9 GHz) imaging is limited to samples with a maximum diameter of 10 mm and is unsuitable for most biological samples because of their high water content. Larger or

biological samples are usually examined at lower frequency. Small-animal studies have been performed at L-band (1.0–1.2 GHz), and the use of even lower frequencies permits samples as large as 100 mm diameter to be examined.

ESR microscopy (at X-band) has been applied to a variety of materials including in situ studies of the oxidation of coals; defects in diamonds; radiometric dosimeters; polymer swelling and to the diffusion of oxygen in model biological systems. The resolution of ESR imaging depends on the field gradient and line width, and the best resolution to date from an X-band imager is 1 μm.

18.6.2. Spectral Parameters

18.6.2.1. g-Factor

The ESR g-factor is also known as the Landé g-factor or spectroscopic splitting factor and depends on the resonance condition for ESR (Eq. 3) and is independent of both applied field and frequency. The g-factor of a free electron g_e is 2.002322, while the g-factors of organic free radicals, defect centers, transition metals, etc. depend on their electronic structure. The g-factors for free radicals are close to the free electron value but may vary from 0 to 9 for transition metal compounds. The most comprehensive compilations of g-factors are those published in [75], [76]. The magnetic moments and hence g-factors of nuclei in crystalline and molecular environments are anisotropic, that is the g-factor (and hyperfine interactions) depend on the orientation of the sample. In general, three principal g-factors are encountered whose orientation dependence is given by:

$$g^2 = (g_{xx})^2(l_{xx})^2 + (g_{yy})^2(l_{yy})^2 + (g_{zz})^2(l_{zz})^2 \qquad (5)$$

where l_{xx}, l_{yy}, and l_{zz} are the direction cosines between the direction of B and the principal g-factors, g_{xx}, g_{yy}, and g_{zz}. When the site shows threefold or fourfold symmetry only two g-factors, $g_{\parallel}$ (g-parallel) and $g_{\perp}$ (g-perpendicular) are observed, where $g_{\perp} = (g_{xx} + g_{yy})/2$. For solutions, free molecular tumbling leads to averaging and the isotropic g-factor g_{iso} is observed:

$$g_{iso} = 1/3 (g_{xx} + g_{yy} + g_{zz}) \qquad (6)$$

or

$$g_{iso} = 1/3 \left(g_{\parallel} + 2g_{\perp}\right) \qquad (7)$$

9.5 GHz   35 GHz

B ‖ x

B ‖ y

B ‖ z

Rigid glass

Solution

5.0 mT
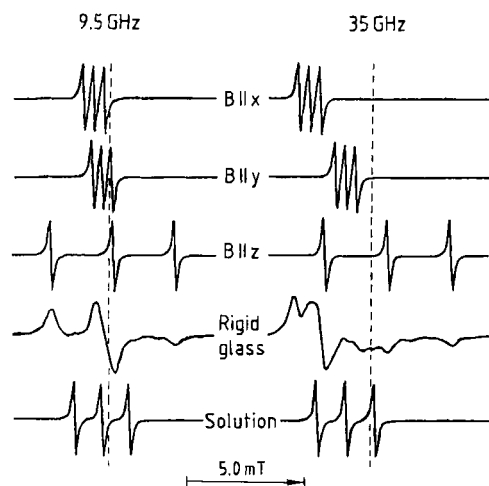
Figure 34. Simple crystal, powder and solution spectrum of a typical aminoxyl radical such as TEMPO (2,2,6,6-tetramethylpiperidine-N-oxyl), (**6**).
O. Hayes Griffith, P. Jost in [64] (with permission)

Figure 34 shows the single crystal, powder and solution spectra expected for an aminoxyl free radical (**6**): this sample has anisotropic g and hyperfine interactions (see Section 18.6.2.2).

6
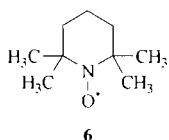
18.6.2.2. Nuclear Hyperfine Interaction

If the interaction of an electron with a magnetic field were the only effect operative, then all ESR spectra of free radicals would consist of one line. When the nuclear spin quantum number I is nonzero a nuclear hyperfine interaction A is observed. When several equivalent nuclei are present (e.g., $\cdot CH_3$, $\cdot C_2H_5$), the number of lines in the spectrum is given by $\Pi_i(2n_iI_i+1)$, where n is the number of magnetically equivalent nuclei i.

The isotropic, or Fermi, hyperfine interaction a arises from the presence of nonzero electron density at the nucleus. Interaction between electrons and nuclear dipoles gives rise to the anisotropic, or dipolar, hyperfine coupling T. This interaction is orientation dependent, but averages to zero in freely tumbling solution. The total hyperfine interaction A is the sum of a and T, where A and T are tensors.

The orientation dependence is given by:

$$A^2 g^2 = (A_{xx})^2 (g_{xx})^2 (l_{xx})^2$$
$$+ (A_{yy})^2 (g_{yy})^2 (l_{yy})^2 + (A_{zz})^2 (g_{zz})^2 (l_{zz})^2 \qquad (8)$$

where A_{xx}, A_{yy}, and A_{zz} are the principal hyperfine constants. As with the g-factor, an axially symmetric spectrum with only $A_{\parallel}$ and $A_\perp$ is often observed.

18.6.2.3. Quantitative Measurements

g-**Factor and Hyperfine Splitting.** The determination of g-factor and hyperfine splitting depends on accurate determination of the external magnetic field at the sample. This can be done directly by using a NMR magnetometer. The microwave frequency is measured directly with a frequency counter, and the g-factor can then be found from Equation (3). Alternatively, the g-factor and magnetic field sweep of the spectrometer can be determined by comparison with secondary standards. Several authors list suitable materials [52], [77].

Spin Concentration. To determine the electron spin concentration it is essential to obtain, as accurately as possible, the intensity K of the ESR response. Preferably, this is done by double integration of the first derivative spectrum using ESR data acquired in a computer; however, if this is not available or if there is considerable overlap with lines from a second species then the relation:

$$K = hw^2 \qquad (9)$$

is used, where h is the peak-to-peak amplitude and w is the peak-to-peak line width. The latter relationship works satisfactorily if relative intensities are required for signals with similar line shapes. Generally, concentration determinations are made relative to a concentration standard. Some of the points which must be considered in concentration determinations including the choice of concentration standard are noted in [52], [77], [78]. Examples of ESR concentration standards include solutions of α,α'-diphenylpicrylhydrazyl (DPPH), potassium peroxylamine disulfonate ($K_2NO(SO_3)_2$), nitroxide radicals, copper sulfate pentahydrate ($CuSO_4 \cdot 5\,H_2O$), and manganese sulfate monohydrate ($MnSO_4 \cdot H_2O$). Solid standards include single crystals of ruby and $CuSO_4 \cdot 5\,H_2O$, or powders of weak and strong pitch in potassium
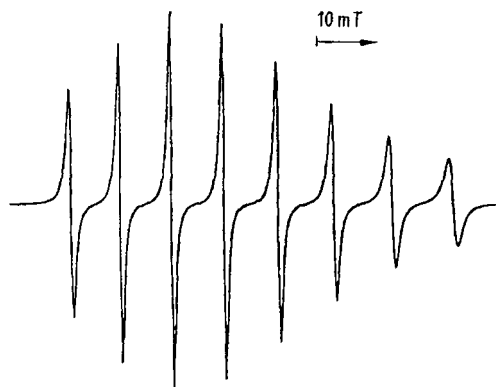
Figure 35. ESR spectrum of vanadyl acetylacetonate (7) in 1 : 1 toluene : chloroform, showing asymmetric line broadening

chloride; DPPH; $MnSO_4 \cdot H_2O$; F-centers in alkyl halides and defects in magnesium oxide. Some of the latter are suitable as g-factor standards also.

For magnetically dilute samples the concentration of the unknown C_X is given by:

$$C_x = C_{std} K_x (g_{std})^2 [S(S + 1)]_{std} / K_{std} (g_x)^2 [S(S + 1)]_x \tag{10}$$

where S is the total spin (e.g., $S = 1/2$ for organic free radicals, 5/2 for manganese(II) which has five d electrons), and the unknown (X) and standard samples (std) have been examined under identical conditions (temperature, spectrometer gain, modulation and scan range), without microwave power saturation. For neat materials the following applies:

$$M_p = g^2 \beta^2 N_0 S(S + 1) B_r / 3k(T - \theta) \tag{11}$$

where M_p is the magnetization of a nondilute paramagnetic material, β is the Bohr magneton, N_0 is the Avogadro number, B_r is the applied field at resonance, k is the Boltzmann constant, T is the absolute temperature, and θ is the Curie temperature.

18.6.3. ESR in the Liquid State

In solution, molecular motion leads to a simplification of the spectra due to the averaging of anisotropic interactions to zero, resulting in the observation of isotropic spectra.

There have been many studies on organic and inorganic free radicals in solution, including their

role as reaction intermediates, mechanistic studies, and studies of molecular dynamics.

18.6.3.1. Slow Molecular Tumbling

When molecular tumbling leads to incomplete averaging, perhaps because of high solution viscosity, variations are observed in the line width of the hyperfine lines. This variation is shown in Figure 35 for vanadyl acetylacetonate (7) dissolved in 1 : 1 toluene : chloroform at room temperature.

7
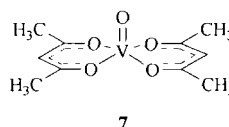
Asymmetric line broadening is commonly observed for aminoxyl radicals, which are therefore used in a wide variety of spin label and spin probe studies (see Section 18.6.5.2).

The variation of line width $1/T_2$ with m_I has been studied extensively [79], [80]. For isotopic tumbling with correlation times of 10^{-11} to 10^{-9} s the following expression is generally used:

$$[T_2(m_I)]^{-1} = A + B m_I + C m_I^2 + D m_I^3 \tag{12}$$

The parameters A, B, C, and D depend on the sign and magnitudes of the g and hyperfine anisotropies and upon the rate of molecular reorientations in the liquid. D is usually small compared with B and C and only terms up to m_I^2 need normally be retained.

In Equation (12) the A term includes broadening contributions arising from instrumental factors and mechanisms which are independent of the magnetic anisotropies. The term in m_I progressively broadens the line across the spectrum, transitions with the largest negative m_I value being broadened the least.

The term in m_I^2 broadens the lines at the extremities of the spectrum more than those in the middle, but its effect is symmetrical. The B and C parameters are given by:

$$B = 4/15 (b \Delta \gamma B_0 \tau_c) \tag{13}$$

$$C = b^2 \tau_c / 8 \tag{14}$$

where B_0 is the operating field of the spectrometer,

$$b = 4\pi[A_{zz} - (A_{xx} + A_{yy})/2]/3 \qquad (15)$$

$$\Delta\gamma = -\beta_e h^{-1} [g_{zz} - (g_{xx} + g_{yy})/2] \qquad (16)$$

and τ_c is the rotational correlation time.

According to the Debye diffusion model, a measure of the length of time a molecule remains in a given orientation τ_c can be calculated from the expression:

$$\tau_c = 4\pi\eta r^3/3kT \qquad (17)$$

where r is the spherical radius of the molecule and η is the viscosity of the solvent. This method has been used to estimate radical radii and the effects of hindered environments [81].

Correlation times with time scales of 10^{-11} to 10^{-9} s are determined by the above method and are said to be in the fast tumbling region. In the slow tumbling region (10^{-9} to 10^{-7} s) more sophisticated theoretical approaches [79], [80] are used. The very slow region (10^{-6} to 10^{-3} s) cannot be studied by conventional ESR and requires special techniques such as saturation transfer spectroscopy (see Section 18.6.5.4).

18.6.3.2. Exchange Processes

Generally, it is assumed that the electrons in free radicals and paramagnetic materials are completely independent and noninteracting. Several effects become important when electron spins interact magnetically and chemically with each other and with their environment.

Electron spin exchange is a bimolecular reaction in which the unpaired electrons of two free radicals exchange their spin states. At low concentration (ca. 10^{-3} mol/L) the usual isotropic spectrum is observed. As the radical concentration is increased the lines gradually broaden until at high concentration ($> 10^{-1}$ mol/L) the lines coalesce to a single line. This single line sharpens further as the concentration is increased.

The spectrum is said to be exchange narrowed since the electron spins are exchanging so fast that the time average of the hyperfine field is close to zero. Electron spin exchange must be avoided if narrow lines are desired in solution and this is readily achieved if the concentration is maintained below ca. 10^{-3} mol/L.

Exchange narrowed spectra (where hyperfine splittings are coalesced) are observed for many pure solid free radicals, such as DPPH and Cu-$SO_4 \cdot 5\, H_2O$. For this reason, single crystal spectra are obtained by incorporating about 1 % of the paramagnetic compound into an isomorphous diamagnetic host crystal.

18.6.4. Computer Simulation of Spectra

Isotropic Spectra. Numerous computer programs have been written for the simulation of solution state (isotropic) ESR spectra. Often these include the facility to simulate spectra from mixtures of radicals and to allow for contributions from satellite lines arising from the presence of low-abundance magnetic nuclei such as ^{13}C, ^{15}N, and ^{33}S.

Time-dependent phenomena, including the alternating line width effect [52], [82] and asymmetric line broadening [79], [80], require simulated spectra to obtain the ESR parameters and rates.

Anisotropic Spectra. Programs are available for the simulation of single crystal, partially oriented, and powder simulation (sometimes in the same program) and in several cases allowance for noncoincidence of the axes of the **g** and **A** tensors can be included. Various programs are described in detail in [53], [56], [57].

18.6.5. Specialist Techniques

18.6.5.1. Spin Trapping

A limitation to the application of ESR spectroscopy to the study of short-lived transient radicals is the difficulty of producing sufficient free radicals for direct observation. Spin trapping provides a simple and efficient means of overcoming this problem [83], [84].

A diamagnetic compound (spin trap) is introduced into the radical producing system to give a relatively stable ESR-observable free radical (spin adduct). Typical spin traps are nitrone compounds such as phenyl-*N-tert*-butylnitrone (PBN) (**8**) and 5,5-dimethyl-1-pyrroline-*N*-oxide (DMPO) (**9**) and nitroso compounds such as 2-methyl-2-nitrosopropane (MNP) (**10**) and 2,4,6-tri-*tert*-butylnitrosobenzene (TBN) (**11**)

PBN (8)

DMPO (9)

$$H_3C-\underset{\underset{CH_3}{|}}{\overset{\overset{CH_3}{|}}{C}}-N=O$$
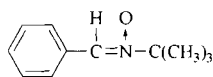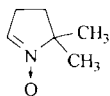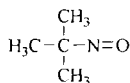
MNP (10)
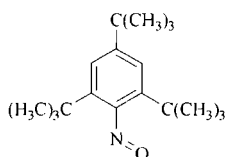
TBN (11)

Spin trapping of a radical X with nitrone and nitroso compounds occurs as follows:

$$R-CH=\overset{\overset{O}{\uparrow}}{N}-R' + X\cdot \longrightarrow R-\underset{\underset{X}{|}}{\overset{\overset{H \quad O\cdot}{|\quad |}}{C}}-N-R'$$

$$R-N=O + X\cdot \longrightarrow \underset{\underset{X}{\diagdown}}{\overset{\overset{R}{\diagdown}}{N}}-O\cdot$$
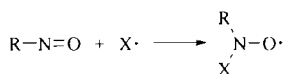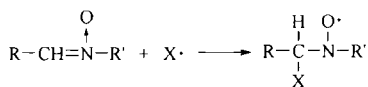
In favorable cases, the original free radical can be identified from the g-factor and hyperfine coupling constants of the spin adduct.

When a methyl radical is trapped by MNP (10) the resulting ESR spectrum is a $1:1:1$ triplet of $1:3:3:1$ quartets. However, if PBN (8) is used, the hyperfine structure is not indicative of the attached radical and its identity must be deduced from hyperfine splittings originating from the nitrogen and lone proton already present in PBN.

The main use of spin trapping is in identifying radical intermediates in organic and inorganic reactions. The reactions can be thermally, electrochemically, or radiation induced. Probably the most important application of spin trapping is the study of radicals in biological systems.

18.6.5.2. Spin Labeling [64], [65], [85]

Spin labeling is a powerful means of studying biochemical and biophysical characteristics of lipids, proteins, and amino acids. Many investigations have been devoted to the synthesis of radicals having a suitable reactive functional group for coupling selectively to the biological system whilst causing minimum perturbation of biological activity.

In spin labeling, a stable free radical is chemically attached to a macromolecule, and the increase in molecular mass of the radical reduces its rate of tumbling, which can be easily monitored by ESR spectroscopy, as described in Section 18.6.3.1. The most suitable class of free radicals for the purpose are aminoxyls (also referred to as nitroxides and nitroxyls) because of their stability and because their anisotropic g and A tensors are sufficiently different in the x, y, and z directions to provide high sensitivity to a change in the rate of molecular tumbling. Rates of tumbling can be measured by the change from a completely isotropic spectrum (rates of about 10^{10} s^{-1}) to a near-powder spectrum (rates of about 10^6 s^{-1}).

Thousands of aminoxyl spin labels have been synthesized and some examples are given below (compounds $12-15$) [85].

A technique related to spin labeling is that of using stable free radicals as spin probes to monitor molecular motions. In this technique a radical, usually an aminoxyl, is synthesized that mimics as closely as possible the structure of the molecule under investigation. The major difference compared with spin labeling is that the radical is not chemically attached to the molecule under investigation.

12

13
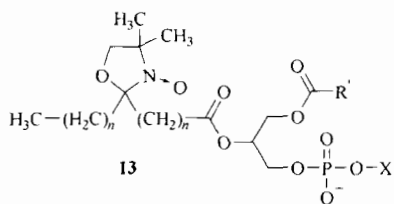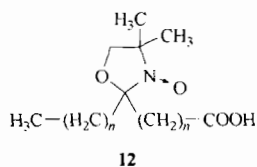
X = (CH$_2$)$_2$N$^+$(CH$_3$)$_3$ (phosphatidylcholine)
X = (CH$_2$)$_2$NH$_3^+$ (phosphatidylethanolamine)
X = CH$_2$CH(OH)CH$_2$OH   (phosphatidylglycerol)
X = CH$_2$CH(NH$_3$)COO$^-$ (phosphatidylserine)
X = H   (phosphatidic acid)

14
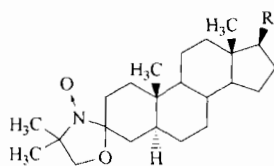
R =   (cholestane derivative)

R = —OH   (androstanol derivative)

15

X =

X =

X =
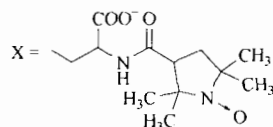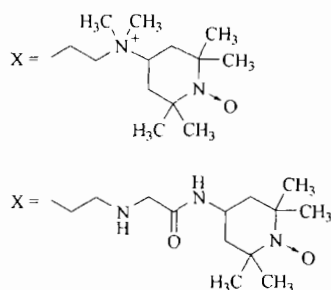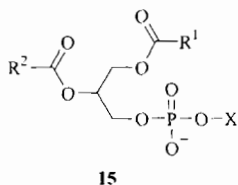
18.6.5.3. Oximetry

Molecular oxygen is paramagnetic and this property is used to provide an indirect method for the detection and measurement of oxygen by oximetry [86]. Bimolecular collisions of oxygen with free radicals results in broadening of the ESR spectrum. By using radicals with sharp lines and measuring the extent of broadening produced by oxygen it is possible to determine the oxygen concentration. In principle, dissolved oxygen concentrations as low as 10^{-7} mol/L can be rapidly measured.

Probably the most important application of this technique is the measurement of oxygen concentrations in cells and animals, both in vivo and in vitro [87]. Aminoxyl radicals can be used, the most commonly used radical being 3-carboxamido-2,2,5,5-tetramethylpyrrolidine-N-oxyl, CTPO (**16**) and its deuterated analogue 3-carbamido-2,2,5,5-tetra-[^{2}H$_3$]methylpyrrolidine-N-oxyl:

16
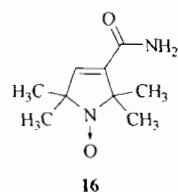
Another approach is to use a solid which is pervious to oxygen and whose ESR spectrum has a single very sharp line. This can be achieved with a small crystal of lithium phthalocyanine, which, in the absence of oxygen has a peak-to-peak line width of less than 0.002 mT [88]. A less good but cheaper material is a type of coal called fusinite.

18.6.5.4. Saturation Transfer

Saturation transfer is a technique that is sensitive to very slow motions with correlation times from 10^{-6} to 10^{-3} s — a region where conventional ESR is not sensitive [89]. It has been used extensively in spin label studies of biological systems [90], [91]. Spectra are recorded under saturation conditions and high modulation amplitudes, as second harmonic in-phase and quadrature spectra. The method measures the transfer of saturation throughout a spectrum by reorientational molecular motions. The line shape is sensitive to molecular motions and the saturation properties of the spin label.

18.7. References

[1] A. Abragam: *The Principles of Nuclear Magnetism,* Clarendon Press, Oxford 1961.
[2] R. J. Abraham, J. Fisher, P. Loftus: *Introduction to NMR Spectroscopy,* J. Wiley & Sons, Chichester 1988.
[3] R. K. Harris: *Nuclear Magnetic Resonance Spectroscopy,* Longman Scientific & Technical, Harlow 1986.
[4] J. Mason: *Multinuclear NMR,* Plenum Publishing, New York 1987.
[5] D. Shaw: *Fourier Transform NMR Spectroscopy,* 2nd ed., Elsevier, Amsterdam 1984.
[6] T. C. Farrar: *An Introduction to Pulse NMR Spectroscopy,* Farragut Press, Chicago 1987.
[7] F. Bloch, W. W. Hansen, M. E. Packard, *Phys. Rev.* **69** (1946) 127.
[8] E. M. Purcell, H. C. Torrey, R. V. Pound, *Phys. Rev.* **69** (1946) 37.
[9] G. C. Levy, R. L. Lichter, G. L. Nelson: *Carbon-13 Nuclear Magnetic Resonance for Organic Chemists,* 2nd ed., Wiley, New York 1980.
[10] F. W. Wehrli, A. P. Marchand, S. Wehrli: *Interpretation of Carbon-13 NMR Spectra,* 2nd ed., Wiley, New York 1988.
[11] E. Breitmaier, W. Voelter: *Carbon-13 NMR Spectroscopy,* 3rd ed., VCH Publishers, New York 1987.
[12] L. P. Lindeman, J. Q. Adams, *Anal. Chem.* **43** (1971) 1245.
[13] D. E. Ewing, *Org. Magn. Reson.* **12** (1979) 499.
[14] E. F. Mooney: *An Introduction to ^{19}F NMR Spectroscopy,* Heyden-Sadtler, London 1970.
[15] J. W. Emsley, L. Phillips, *Prog. Nucl. Magn. Reson. Spectrosc.* **7** (1971) 1–526.

[16] V. Wray, *Annu. Rep. NMR Spectrosc.* **10 B** (1980) 1.
[17] V. Wray, *Annu. Rep. NMR Spectrosc.* **14** (1983) 1.
[18] D. G. Gorenstein: *Phosphorus-31 NMR: Principles and Applications,* Academic Press, New York 1984.
[19] J. Emsley, D. Hall: *The Chemistry of Phosphorus,* Harper & Row, New York 1976.
[20] G. C. Levy, R. L. Lichter: *Nitrogen-15 NMR Spectroscopy,* J. Wiley & Sons, New York 1979.
[21] M. Witanowski, L. Stefaniak, G. A. Webb, *Annu. Rep. NMR Spectrosc.* **18** (1986) 1.
[22] E. A. Williams, J. D. Cargioli, *Annu. Rep. NMR Spectrosc.* **9** (1979) 221.
[23] A. J. Shaka, J. Keeler, R. Freeman, *J. Magn. Reson.* **53** (1983) 313.
[24] J. H. Noggle, R. E. Schirmer: *The Nuclear Overhauser Effect – Chemical Applications,* Academic Press, New York 1971.
[25] D. Neuhaus, M. P. Williamson: *The Nuclear Overhauser Effect in Structural and Conformation Analysis,* VCH Publishers, New York 1989.
[26] J. K. M. Sanders, B. K. Hunter: *Modern NMR Spectroscopy,* Oxford University Press, Oxford 1987.
[27] H. Y. Carr, E. M. Purcell, *Phys. Rev.* **94** (1954) 630.
[28] S. Meiboom, D. Gill, *Rev. Sci. Instrum.* **29** (1958) 688.
[29] C. LeCocq, J. Y. Lallemand, *J. Chem. Soc. Chem. Commun.* **1981,** 150.
[30] D. M. Doddrell, D. T. Pegg, M. R. Bendall, *J. Magn. Reson.* **48** (1982) 323.
[31] G. A. Morris, *Magn. Reson. Chem.* **24** (1986) 371.
[32] R. Benn, H. Günther, *Angew. Chem. Int. Ed. Engl.* **22** (1983) 350.
[33] Varian Associates, Palo Alto: *High Resolution NMR Spectra Catalogue.* **vol. 1, 1962; vol. 2, 1963.**
[34] The Sadtler Collection of High Resolution Spectra, Sadtler Research Laboratories, Philadelphia.
[35] C. J. Pouchert: *Aldrich Library of NMR Spectra,* vols. 1 and 2, Aldrich, Milwaukee, WI, 1983.
[36] W. Brugel: *Handbook of NMR Spectral Parameters,* vols. 1–3, Heyden, Philadelphia 1979.
[37] L. F. Johnson, W. C. Jankowski: *Carbon-13 NMR Spectra, a Collection of Assigned, Coded, and Indexed Spectra.* Wiley, New York 1972.
[38] E. Breitmaier, G. Haas, W. Voelter: *Atlas of C-13 NMR Data,* vols. 1–3, Heyden, Philadelphia 1979.
[39] SPECINFO, Chemical Concepts, Weinheim.
[40] W. Bremser et al.: *Carbon-13 NMR Spectral Data (microfiche),* 4th ed., VCH Publishers, New York 1987.
[41] J. L. Koenig: *Spectroscopy of Polymers,* American Chemical Society, Washington 1992.
[42] A. Bax, S. Grzesiek, *Acc. Chem. Res.* **26** (1993) 131.
[43] D. L. Rabenstein, D. A. Keire, *Pract. Spectrosc.* **11** (1991) 323.
[44] G. J. Martin, X. Y. Sun, C. Guillou, M. L. Martin, *Tetrahedron* **41** (1985) 3285.
[45] D. Parker, *Chem. Rev.* **91** (1991) 1441.
[46] T. C. Morrill (ed.): *Lanthanide Shift Reagents in Stereochemical Analysis,* VCH Publishers, New York 1986.
[47] J. I. Kaplan, G. Fraenkel: *NMR of Chemically Exchanging Systems,* Academic Press, New York 1980.
[48] C. A. Fyfe: *Solid State NMR for Chemists.* CFC Press, Guelph 1983.
[49] A. M. Chippendale et al., *Magn. Reson. Chem.* **24** (1986) 81.
[50] Bruker Minispec Application Note 30, Bruker Analytische Messtechnik GmbH, Karlsruhe.
[51] P. Jackson et al., *Poly. Int.* **24** (1991) 139.

[52] J. E. Wertz, J. R. Bolton: *Electron Spin Resonance; Elementary Theory and Practical Applications,* Chapman & Hall, New York 1986.

[53] J. A. Weil, J. R. Bolton: *Electron Paramagnetic Resonance Spectroscopy, Theory and Examples,* Wiley, New York 1993.

[54] N. M. Atherton: *Principles of Electron Spin Resonance,* Ellis Horwood, Chichester 1993.

[55] A. Abragam, B. Bleaney: *Electron Paramagnetic Resonance of Transition Ions,* Clarendon Press, Oxford 1970.

[56] J. R. Pilbrow: *Transition Ion Electron Paramagnetic Resonance,* Clarendon Press, Oxford 1990.

[57] F. E. Mabbs, D. Collison: *Electron Paramagnetic Resonance of d Transition Metal Compounds.* Elsevier, Amsterdam 1992.

[58] M. C. R. Symons (ed.): "Specialist Periodical Reports," *Electron Spin Resonance,* vols. 1–13, Royal Society of Chemistry, London 1965–1993.

[59] J. A. Weil (ed.): *Electron Magnetic Resonance of the Solid State,* Canadian Society for Chemistry, Ottawa 1987.

[60] S. A. Fairhurst: "Specialist Periodical Reports," in M. C. R. Symons (ed.): *Electron Spin Resonance,* vol. 13 A, Royal Society of Chemistry, London 1992.

[61] M. I. Ikeya: *New Applications of Electron Spin Resonance Dating, Dosimetry and Microscopy,* World Scientific Publ., New York 1993.

[62] H. Sigel (ed.): "ENDOR, EPR and Electron Spin Echo for Probing Co-ordination Spheres," *Metal Ions in Biological Systems,* vol. 22, Marcel Dekker, New York 1987.

[63] L. J. Berliner, J. Reuben (eds.): "Biological Magnetic Resonance," *EMR of Paramagnetic Molecules,* vol. 13, Plenum Press, New York 1993.

[64] L. J. Berliner (ed.): *Spin Labeling, Theory and Applications,* Academic Press, New York 1976.

[65] L. J. Berliner, J. Reuben (eds.): "Biological Magnetic Resonance," *Spin Labeling Theory and Applications,* vol. 8, Plenum Press, New York 1989.

[66] A. J. Hoff (ed.): *Advanced EPR: Applications in Biology and Biochemistry,* Elsevier, Amsterdam 1989.

[67] H. Kurreck, B. Kirste, W. Lubitz: *Electron Nuclear Double Resonance Spectroscopy of Radicals in Solution; Applications to Organic and Biological Chemistry,* VCH Publishers, New York 1988.

[68] L. Kevan, L. D. Kispert: *Electron Spin Double Resonance Spectroscopy,* Wiley, New York 1976.

[69] C. P. Keijzers, E. J. Reijerse, J. Schmidt (eds.): *Pulsed EPR: A New Field of Applications,* North Holland, Amsterdam 1989.

[70] L. Kevan, M. K. Bowman (eds.): *Modern Pulsed and Continuous-Wave Electron Spin Resonance,* Wiley, New York 1990.

[71] S. A. Dikanov, Y. D. Tsvetkov: *Electron Spin Echo Envelope Modulation (ESEEM)* Spectroscopy, CRC Press, Boca Raton 1992.

[72] A. Schweiger, *Angew. Chem. Int. Ed. Engl.* **30** (1991) 265.

[73] A. Schweiger, *Pure Appl. Chem.* **64** (1992) 809.

[74] G. R. Eaton, S. S. Eaton, K. Ohno: *EPR Imaging and In vivo EPR,* CRC Press, Boca Raton 1991.

[75] *Landolt-Börnstein,* **New Series Group II, vols. 1, 9 a, b, c1, c2, d1, d2.**

[76] *Landolt-Börnstein,* **New Series Group II, vols. 2, 8, 10, 11.**

[77] I. B. Goldberg, A. J. Bard: *Treatise on Analytical Chemistry,* part 1, vol. 10, Wiley Interscience, New York 1983.

[78] R. G. Kooser, E. Kirchmann, T. Matkov, *Concepts Magn. Reson.* **4** (1992) 145.

[79] J. H. Freed in L. J. Berliner (ed.): *Spin Labeling Theory and Applications,* Academic Press, New York 1976.

[80] J. H. Freed in L. J. Berliner, J. Reuben (eds.): *Spin Labeling Theory and Applications,* vol. 8: "Biological Magnetic Resonance," Plenum Press, New York 1989.

[81] R. F. Boyer, S. E. Keinath: *Molecular Motions in Polymers by ESR,* Harwood Academic Press, Chur 1980.

[82] P. D. Sullivan, J. R. Bolton in J. S. Waugh (ed.): *Advances in Magnetic Resonance,* vol. 4, Academic Press, New York 1970.

[83] M. J. Perkins in V. Gold, D. Bethell (eds.): *Advances in Physical Organic Chemistry,* vol. 17, Academic Press, London 1980.

[84] E. G. Janzen, D. L. Haire in D. D. Tanner (ed.): *Advances in Free Radical Chemistry,* vol. 1, JAI Press, London 1990.

[85] D. Marsh, *Tech. Life Sci. B4/II* **B 426** (1982) 1–44.

[86] J. S. Hyde, W. K. Subczynski in L. J. Berliner, J. Reuben (eds.): "Biological Magnetic Resonance," *Spin Labeling Theory and Applications,* **vol. 8,** Plenum Press, New York 1989.

[87] R. K. Woods et al., *J. Magn. Reson.* **85** (1989) 50.

[88] X.-S. Tang, M. Mousavi, G. C. Dismukes, *J. Am. Chem. Soc.* **113** (1991) 5914.

[89] J. S. Hyde, L. R. Dalton in L. J. Berliner (ed.): *Spin Labeling Theory and Applications,* Academic Press, New York 1979.

[90] M. A. Hemminga, P. A. Jager: "Biological Magnetic Resonance," in L. J. Berliner, J. Reuben (eds.): *Spin Labeling Theory and Applications,* vol. 8, Plenum Press, New York 1989.

[91] P. F. Knowles, D. Marsh, H. W. E. Rattle: *Magnetic Resonance of Biomolecules; An Introduction to the Theory and Practice of NMR and ESR in Biological Systems,* Wiley-Interscience, New York 1976.

[92] D. M. Grant, R. K. Harris (eds.): *Encyclopedia of Nuclear Magnetic Resonance,* J. Wiley, Chichester 1996.

[93] C. P. Slichter: *Principles of Magnetic Resonance,* 3rd ed., Springer Verlag, Berlin 1996.

[94] R. S. Macomber: *A Complete Introduction to Modern NMR Spectroscopy,* Wiley, New York 1998.

[95] R. Freeman: *A Handbook of Nuclear Magnetic Resonance,* 2nd ed., Addison Wesley Longman, Essex 1997.

[96] S. W. Homans: *Dictionary of Concepts in NMR,* Oxford University Press, Oxford 1992.

[97] A. E. Derome: *Modern NMR Techniques for Chemistry Research,* Pergamon Press, Oxford 1987.

[98] J. K. M. Sanders, B. K. Hunter: *Modern NMR Spectroscopy,* Oxford University Press, Oxford 1987.

[99] R. R. Ernst, G. Bodenhausen, A. Wokaun: *Principles of Nuclear Magnetic Resonance in One and Two Dimensions,* Clarendon Press, Oxford 1990.

[100] D. M. S. Bagguley: *Pulsed Magnetic Resonance,* Oxford Univ. Press, Oxford 1992.

[101] J. W. Akitt: *NMR and Chemistry: An Introduction to the Multinuclear Fourier Transform Era,* 3rd ed., Chapman & Hall, London 1992.

[102] P. J. Hore: *Nuclear Magnetic Resonance,* Oxford University Press, Oxford 1995.

[103] H. Günther: *NMR Spectroscopy: Basic Principles, Concepts and Applications in Chemistry*. J. Wiley & Sons, New York 1995.

[104] R. Freeman: *Spin Choreography. Basic Steps in High Resolution NMR*, Freeman, Basingstoke 1996.

[105] D. Canet: *Nuclear Magnetic Resonance: Concepts and Methods*, Wiley, Chichester 1996.

[106] S. Berger, S. Braun, H. O. Kalinowski: *NMR-Spectroscopy of the Non-metallic Elements*, John Wiley, Chichester 1996.

[107] G. Batta, K. Kovir, L. Jossuth, C. Szankey: *Methods of Structure Elucidation by High-Resolution NMR*, Elsevier, London 1997.

[108] H. Friebolin: *Basic One and Two-Dimensional NMR Spectroscopy*, Wiley-VCH, Weinheim 1998.

[109] S. Braun, H.-O. Kalinowski, S. Berger: *150 and More Basic NMR Experiments*, Wiley-VCH, Weinheim 1998.

[110] P. A. Keifer, *Drugs of the Future* **23** (1998) 301.

[111] J. P. Hornak, http://www.cis.rit.edu/htbooks/nmr/inside.htm, Rochester Institute of Technology, Rochester, NY.

[112] P. A. Petillo, http://www-chem430.scs.uiuc.edu/lessons/lessons.htm, University of Illinois, Urbana, IL.

[113] D. J. Cookson, B. E. Smith, *J. Magn. Reson.* **57** (1984) 355.

[114] H.-O. Kalinowski, S. Berger, S. Braun: *Carbon-13 NMR Spectroscopy*, Wiley, Chichester 1988.

[115] K. Pihlaja, E. Kleinpeter: *Carbon-13 NMR Chemical Shifts in Structural and Stereochemical Analysis*, VCH, Weinheim 1994.

[116] Attar-ur-Rahman, V. U. Ahmad: *13C NMR of Natural Products*, **Vol. 1 and 2**, Plenum Press, New York 1992.

[117] L. D. Quin, J. G. Verkade: *Phosphorus-31 NMR Spectral Properties in Compound Characterization and Structural Analysis*, VCH Publishers, New York 1994.

[118] S. Berger, T. Faecke, R. Wagner, *Magn. Reson. Chem.* **34** (1996) 4.

[119] D. G. Gadian: *Nuclear Magnetic Resonance and its Application to Living Systems*, Clarendon Press, Oxford 1982.

[120] D. W. Boykin: *17O NMR Spectroscopy*, CRC Press, Boca Raton 1990.

[121] S. Kuroki: *O-17 NMR in I. Ando, T. Asakura (ed.): Solid State NMR of Polymers*, Elsevier Science Publ, Amsterdam 1998, p. 236.

[122] Y. Takeuchi, T. Takayama: "Si-29 NMR spectroscopy of organosilicon compounds" in Z. Rappoport, Y. Apeloig (eds.): *Chemistry of Organic Silicon Compounds*, **vol. 2**, J. Wiley & Sons, Chichester 1998, p. 267.

[123] F. J. M. van de Veen: *Multidimensional NMR in Liquids: Basic Principles and Experimental Methods*. VCH, Weinheim 1995.

[124] A. E. Derome, M. P. Williamson, *J. Magn. Reson.* **88** (1990) 177.

[125] A. A. Maudsely, L. Müller, R. R. Ernst, *J. Magn. Reson.* **28** (1977) 463.

[126] W. Bermel, C. Griesinger, H. Kessler, K. Wagner, *J. Magn. Reson. Chem.* **25** (1987) 325.

[127] Landolt Bernstein: *III/35B Nuclear Magnetic Resonance (NMR)Data: Chemical Shifts and Coupling Constants for 19F and 15N*. Springer Verlag, Berlin 1998.

[128] G. Engelhardt, D. Michel: *High Resolution Solid State NMR of Silicates and Zeolites*. Wiley, Chichester 1987.

[129] M. Schindler, W. Kuttzelnigg, *J. Chem. Phys.* **76** (1982) 1919.

[130] W. Bremser, *Anal. Chim. Acta* **103** (1973) 355.

[131] LabControl Scientific Consulting and Software Development GmbH, *Spectroscopic Knowledge Management System — SPECTACLE*, Köln, Germany.

[132] V. Purtuc, V. Schütz, S. Felsinger, W. Robien, Institute for Organic Chemistry, University of Vienna, Vienna, Austria, http://felix.orc.univie.ac.at/

[133] Advanced Chemistry Development Inc., *CNMR data base*, Toronto, Canada.

[134] J. Meiler, R. Meusinger, M. Will, *J. Chem. Inf. Comput. Sci.* **40** (2000) 1169.

[135] K. Hayamizu, M. Yanagisawa, O. Yamamoto, *Integrated Spectral Data Base System for Organic Compounds (SDBS)*, National Institute of Materials and Chemical Research, Tsukuba, Japan, http://www.aist.go.jp/RIODB/SDBS/menu-e.html

[136] E. M. Munk, *J. Chem. Inf. Comput. Sci.* **38** (1998) 997.

[137] E. Pretsch, J. T. Clerc: *Spectra Interpretation of Organic Compounds*, VCH, Weinheim 1997.

[138] F. A. Bovey, P. A. Mireau: *NMR of Polymers*, Academic Press, San Diego 1996.

[139] D. G. Reid: *Protein NMR Techniques*, Humana Press, Totowa, N.J. 1997.

[140] M. J. Shapiro, J. S. Gounarides, *Prog. NMR Spectr.* **35** (1999) 153.

[141] S. B. Shuker, P. J. Hajduk, R. P. Meadows, S. W. Fesik, *Science* **274** (1996) 1531.

[142] H. Barjat, G. A. Morris, S. Smart, A. G. Swanson, S. C. R. Williams, *J. Magn. Reson. Ser. B.* **108** (1995) 170.

[143] M. Spraul, M. Hoffman, P. Dvortsak, J. K. Nicholson, I. D. Wilson, *Anal. Chem.* **65** (1993) 327.

[144] D. L. Olson, T. L. Peck, A. G. Webb, R. L. Magin, J. V. Sweedler, *Science* **270** (1995) 1967.

[145] P. J. Hajduk, T. Gerfin, J. M. Boehlen, M. Häberli, D. Marek, S. W. Fesik, *J. Med. Chem.* **42** (1999) 2315.

[146] I. Ando, J. Austin: *Solid State NMR Spectroscopy*, Elsevier, Amsterdam 1995.

[147] E. O. Stejskal, J. D. Memory: *High Resolution NMR in the Solid State*. Oxford University Press, New York 1994.

[148] B. Hills: *Magnetic Resonance Imaging in Food Science*, Wiley-VCH, Weinheim 1998.

[149] R. Kimmich (ed.): *NMR Tomography, Diffusometry, Relaxometry*. Springer Verlag, Berlin 1997.

[150] R. R. Edelmann, J. R. Hesselink, M. B. Zlatkin (eds.): *Clinical Magnetic Resonance Imaging*, W. B. Saunders Comp. 1996.

[151] M. T. Vlaardingerbroek, J. A. den Boer: *Magnetic Resonance Imaging*, Springer Verlag, Berlin 1996.

[152] H. Bachelard (ed.): *Magnetic Resonance Spectroscopy and Imaging in Neurochemistry*. Plenum Press, New York 1997.

[153] A. Gloor, M. Cadisch, R. Bürgin-Schaller, M. Farkas, T. Kocsis, J. T. Clerc, E. Pretsch, R. Aeschimann, M. Badertscher, T. Brodmeier, A. Fürst, H.-J. Hediger, M. Junghans, H. Kubinyi, M. E. Munk, H. Schriber, D. Wegmann: *SpecTool 1.1, A Hypermedia Book for Structure Elucidation of Organic Compounds with Spectroscopic Methods*, Chemical Concepts GmbH, Weinheim 1995.

19. Mössbauer Spectroscopy

P. GÜTLICH, Institut für Anorganische Chemie und Analytische Chemie, Johannes Gutenberg-Universität, Mainz, Germany

J. ENSLING, Institut für Anorganische Chemie und Analytische Chemie, Johannes Gutenberg-Universität, Mainz, Germany

19.1. Introduction

Since the discovery of the Mössbauer effect by RUDOLF MÖSSBAUER [1]–[3] in 1958 this nuclear spectroscopic method has found a wide variety of applications in materials science, solid state physics, chemistry, metallurgy, and earth sciences.

Up to the present time, the Mössbauer effect has been observed for almost 100 nuclear transitions in more than 40 different elements. Not all of these transitions are suitable for actual studies, for reasons that will be discussed below. But ca. 15–20 elements remain that are appropriate for applications, of which the best known Mössbauer nuclide is iron, because on the one hand the Mössbauer experiment is easy to perform with the 14.4 keV transition in ^{57}Fe, and on the other hand the element iron is very abundant in nature.

It is the purpose of this article to introduce the reader unfamiliar with the field to the principles of Mössbauer spectroscopy, and to the various types of chemical information that can be extracted from the electric and magnetic hyperfine interactions reflected in the Mössbauer spectrum. For pedagogical reasons we shall focus on the principles of ^{57}Fe Mössbauer spectroscopy in the introductory part, and shall then present examples of typical applications. The more seriously interested reader is advised to consult the various text books, reviews, and special compilations of original communications [4]–[13].

19.2. Principle and Experimental Conditions of Recoil-free Nuclear Resonance Fluorescence

19.2.1. Nuclear Resonance

The resonance absoption of electro-magnetic radiation is a phenomenon well known in many branches of physics. The scattering of sodium light by sodium vapor, the excitation of a tuning fork by
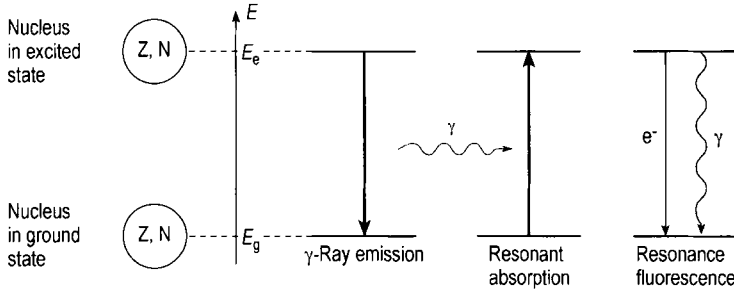
Figure 1. Schematic representation of nuclear resonance absorption of γ-rays (Mössbauer effect) and nuclear resonance fluorescence

sound, and of a dipole by radiofrequency radiation are some familiar examples. The possible observation of the phenomenon in nuclear transitions involving γ-radiation was suggested in 1929 by KUHN [14] but for many years unsuccessful attempts were reported. It was the great merit of RUDOLF MÖSSBAUER who established the experimental conditions as well as the theoretical description of the occurrence of the nuclear resonant absorption (fluorescence) of γ-rays [1]–[3]. The following conditions have to be fulfilled for the observation of the Mössbauer effect.

Consider a nucleus in an excited state of energy E_e with the mean life time τ. Transition to the ground state of energy E_g occurs by emitting a γ-quantum with the energy

$$E_0 = \_hw = E_e - E_g \tag{1}$$

This γ-quantum may be absorbed by a nucleus of the same kind in its ground state, thereby taking up the energy difference E_0. This phenomenon is called nuclear resonant absorption of γ-rays, and is schematically sketched in Figure 1.

The fact that the lifetime τ of an excited nuclear state is finite means that its energy has a certain distribution of width Γ, which, due to the Heisenberg uncertainty relationship, is connected to τ by

$$\tau\Gamma = \_h \tag{2}$$

Consequently, the probability for the emission of a quantum is a function of the energy, or, in other words, the energy spectrum exhibits an intensity distribution $I(E)$ about a maximum at E_0 with a width at half maximum, Γ, which corresponds to the uncertainty in energy of the ex-

cited nuclear state. The explicit shape of this distribution is Lorentzian according to

$$I(E) \quad \frac{\Gamma/2\pi}{(E - E_0)^2 + (\Gamma/2\pi)^2} \tag{3}$$

The mean lifetime τ of the excited states of the Mössbauer isotopes are in the range of 10^{-6} to 10^{-10} s; the corresponding uncertainties of the transition energies are between 10^{-9} and 10^{-5} eV. Nuclear resonance fluorescence is known to possess an amazingly high energy resolution given by the ratio Γ/E_0, which is of the order of 10^{-12}. As an example, the first excited state of ^{57}Fe has a mean lifetime $\tau = 141$ ns, which leads to a natural linewidth $\Gamma = \hbar/\tau = 4.7 \cdot 10^{-9}$ eV and hence to a relative energy resolution of $\Gamma/E_0 = 3.3 \cdot 10^{-13}$.

The absorption line is also described by a Lorentzian curve. Nuclear resonant absorption can only be observed when the difference of the transition energies for emission and absorption are not much greater than the sum of the linewidths, i.e., when the emission and absorption lines overlap sufficiently.

After resonance absorption, the excited nucleus will decay by either emitting isotropically a γ-quantum (as in the primary γ-ray emission of Figure 1) or a conversion electron e^-, preferentially from the K-shell. This phenomenon is termed *nuclear resonance fluorescence* and may be used in Mössbauer scattering experiments (surface investigations).

19.2.2. Recoil Effect

Consider the emission of a γ-quantum with energy E_γ from an excited nucleus of mass M in a free atom or molecule (in the gas or liquid phase). The energy of the excited state is E_0 above

$E_R = \dfrac{E_0^2}{2Mc^2}$ Nucleus of mass M $E_\gamma = E_0 - E_R$
(recoil energy)   and mean energy E_0 (γ-ray energy)
   (at rest before γ-ray
   emission)
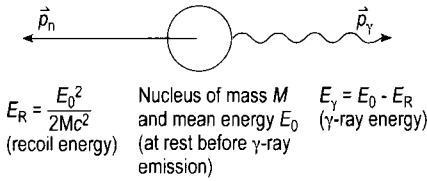
Figure 2. Recoil of momemtum $\to p_n$ and energy E_R imparted to an isolated nucleus upon γ-ray emission

that of the ground state. If the excited nucleus is supposed to be at rest before the decay, a recoil effect will be imparted to the nucleus upon γ-emission, causing the nucleus to move with velocity v in opposite direction to that of the emitted γ-ray (see Fig. 2).

The resulting recoil energy is

$$E_R = \frac{1}{2}Mv^2 = \frac{p_n^2}{2M} \tag{4}$$

Momentum conservation requires that

$$p_\gamma = E_\gamma/c = -p_n \tag{5}$$

where p_γ and p_n are the linear momenta of the γ-quantum and the recoiled nucleus, respectively, and c is the velocity of light. Thus the energy of the emitted γ-quantum is

$$E_\gamma = E_0 - E_R \approx E_0 \tag{6}$$

Since E_R is very small compared to E_0, we may write for the recoil energy (in a nonrelativistic approximation)

$$E_R = \frac{p_n^2}{2M} = \frac{E_\gamma^2}{2Mc^2} \approx \frac{E_0^2}{2Mc^2} \tag{7}$$

For many nuclear transitions $E_R \approx 10^{-2} \dots 10^{-3}$ eV. The recoil effect causes the energy of the emitted photon to be decreased by E_R and the emission line appears at an energy $E_0 - E_R$ rather than at E_0. In the absorption process, the γ-ray to be absorbed by a nucleus requires the total energy $E_0 + E_R$ to make up for the transition from the ground to the excited state and the recoil effect, which causes the absorbing nucleus to move in the same direction as the incoming γ-ray photon. As shown schematically in Figure 3 the transition lines for emission and absorption are separated by $2E_R$, which is about 10^6 times larger than the natural line width Γ.
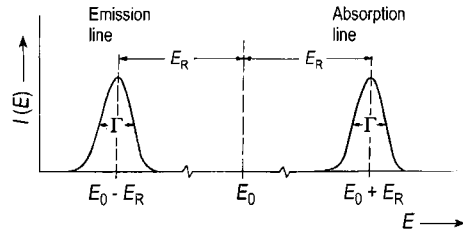
Figure 3. Consequences of the recoil effect caused by γ-ray emission and absorption in isolated nuclei. The transition lines for emission are separated by $2E_R \approx 10^6 \, \Gamma$. There is no overlap between emission and absorption line and hence no resonant absorption is possible

The condition for overlap of emission and absorption line, viz. $2E_R \leq \Gamma$ is not fulfilled at all. Thus, the cross section of nuclear resonant absorption in isolated atoms and molecules in the gaseous and liquid phases is extremely small.

In optical spectroscopy, however, the recoil energy is negligible compared with the linewidth and the resonance condition is certainly fulfilled for freely emitting and absorbing particles.

Before the discovery of the Mössbauer effect some successful experiments on nuclear resonance absorption in the gas phase had been carried out utilizing the Doppler effect. High velocities were required ($\approx 10^2 - 10^3$ ms^{-1}). This could be achieved either by mounting a source of emitting nuclei on a high speed rotor or by heating the source and the absorber, so that some overlap of the lines could take place.

Rudolf Mössbauer, while working on his Ph.D. thesis, carried out experiments of this kind with ^{191}Ir. But surprisingly, on lowering the temperature he found an increase in the absorption effect rather than a decrease. Mössbauer was able to explain this unexpected phenomenon, and was awarded the Nobel Prize in 1961 for the observation and correct interpretation of the *recoilless nuclear resonance absorption (fluorescence)* which is now known as the *Mössbauer Effect* and which provided the basis for a powerful technique in solid state research.

19.2.3. Mössbauer Effect

As explained in Section 19.2.2, nuclear resonance absorption and fluorescence of γ-rays is not possible for nuclei in freely emitting or absorbing particles. This is different in the solid state. The nuclei are more or less rigidly bound to the

lattice. If a γ-ray photon is emitted from the excited nucleus, the concomitant recoil energy may now be assumed to consist of two parts,

$$E_R = E_{tr} + E_{vib} \tag{8}$$

E_{tr} refers to the translational energy transferred through linear momentum to the crystallite of mass M. As M is $\approx 10^{20}$ times larger than the mass of a single atom, the corresponding recoil energy E_{tr} imparted to the crystallite becomes negligibly in comparison with γ [Equation (7)].

Most of the recoil energy E_R will be dissipated into processes other than linear momentum, viz. to the lattice vibrational system. E_{vib}, the vibrational energy part, still being of the order of 10^{-3} eV and thus several orders of magnitude smaller than the atom displacement energy (ca. 25 eV) and also much smaller than the characteristic lattice phonon energies (ca. $10^{-1} - 10^{-2}$ eV), is dissipated by heating the nearby lattice surroundings. As a consequence of this phonon creation with frequencies up to w_i, the resulting energy of the emitted γ-ray will be

$$E_\gamma = E_0 - \sum_i n\,\hbar w_i \tag{9}$$

This again would destroy the resonance phenomenon by shifting the emission and absorption lines too far apart from each other on the energy scale. Fortunately, however, the lattice vibrations are quantized and E_{vib} changes the vibrational energy of the lattice oscillators by integral multiples of the phonon energy, $n\hbar w_i$ ($n = 0, 1, 2, ...$). This means that there is a certain probability f that no lattice excitation (energy transfer of $0 \cdot \hbar w_i$, called the zero-phonon-process) occurs during the γ-ray emission or absorption process. Mössbauer has shown that only to the extent of this probability f of unchanged lattice vibrational states during γ-ray emission and absorption does nuclear resonance absorption become possible. This has been termed the *Mössbauer effect*. The probability f is called the recoil-free fraction and denotes the fraction of nuclear transitions which occur without recoil.

The recoil-free fraction in Mössbauer spectroscopy is equivalent to the fraction of X-ray scattering processes without lattice excitation; this fraction of elastic processes in X-ray and neutron diffraction is described by the Debye – Waller factor:

$$f = e^{-k^2 \langle x^2 \rangle} \tag{10}$$

where $\langle x^2 \rangle$ is the mean square amplitude of vibration of the resonant nucleus in the photon propagation direction and $k = 2\pi/\lambda = E_\gamma/(\hbar c)$ is the wavenumber of the γ-quantum. The value of f depends on the lattice properties, described by $\langle x^2 \rangle$, and on the photon properties, represented by k^2. The ratio of the mean square displacement and wavelength λ of the photon must be small ($\ll 1$), in order to have a sufficiently high Debye – Waller factor f. It is also obvious that the Mössbauer effect cannot take place in liquids, where the molecular motion is characterized by unbound $\langle x^2 \rangle$, causing the recoil-free fraction to vanish. The following consequences emerge.

1) f decreases exponentially with k^2, i.e., E_γ^2. The Mössbauer effect can hardly be detected with γ-ray energies above 150 keV.
2) The mean square displacement $\langle x^2 \rangle$ should be small, in other words, the Mössbauer atom should be bound tightly to the crystal lattice. However, crystallinity of the material is not a necessary condition for observation of the Mössbauer effect. The effect is also detectable in amorphous materials, glasses, and in frozen solutions.
3) The recoil-free fraction f is temperature dependent, because $\langle x^2 \rangle$ decreases with decreasing temperature and hence forces f to increase. However, f does not reach unity even at $T = 0$ due to the fact that $\langle x^2 \rangle \neq 0$ at $T = 0$ because of the quantum-mechanical zero-point motion of the atoms (nuclei). In order to express f in terms of the usual experimental variables, the mean square amplitude $\langle x^2 \rangle$ is calculated using lattice dynamic models (e.g., Einstein, Debye) [6], [15].
4) $\langle x^2 \rangle$ and hence f can be anisotropic. As a consequence, a dependence of the intensity of absorption lines on the observation direction in single crystal experiment may be observed.

19.3. Mössbauer Experiment

19.3.1. Doppler Effect

The main components of a Mössbauer spectrometer are the γ-ray source containing the *Mössbauer active* nuclides, the absorber (or scatterer), and a detector for low-energy γ-radiation (see Fig. 4) plus electronics for automatic recording of the spectrum. The source and the absorber are

Figure 4. Schematic illustration of the experimental arrangement for recoilless nuclear resonance absorption, and relative transmission of recoilless γ-quanta as a function of Doppler velocity

where E_0 is the energy of the γ-quantum emitted from the same source at rest and is the velocity of light. This relationship is valid for a moving source, the velocity being positive for the source moving towards the absorber, and negative for moving away from the absorber. Owing to the very narrow linewidth Γ of the resonance line, it is generally sufficient to produce a small Doppler energy change E_D of the order of the line width Γ by Doppler shifting the source in order to sweep over the resonance. The Doppler velocities needed in the case of ^{57}Fe spectroscopy are in the range 0 to ± 10 mm s^{-1}, and for most of the other Mössbauer nuclides velocities of less than ± 100 mm s^{-1} are sufficient.

The Mössbauer spectrum, a plot of the relative transmission as a function of Doppler velocity, shows maximum resonance and therefore minimum relative transmission at relative velocities where emission and absorption lines overlap ideally (cf. Fig. 4). At high positive or negative velocities the overlap of emission and absorption lines is negligible, the resonance effect being practically zero, i.e., the relative transmission yields the *base line*.

19.3.2. Mössbauer Spectrometer

The Mössbauer spectrometers which are in use nowadays generate the spectrum by the velocity-sweep method. The drive system moves the source (or absorber) repeatedly over a range of velocities, while simultaneously counting the γ-quanta behind the absorber into synchronized channels.

The essential components of a modern Mössbauer spectrometer as illustrated in the block-diagram of Figure 5 are: the velocity transducer, the wave form generator and synchronizer, the multi-channel analyzer, γ-ray detection system, a cryostat or oven for low and temperature dependent measurements, a velocity calibration device, the source and the absorber, and a read-out unit.

The source (and in a few cases the absorber) is mounted on the vibrating axis of an electromagnetic transducer (loudspeaker system) [16] which is moved according to a voltage waveform applied to the driving coil of the system. The usual velocity functions are of the triangular, sawtooth, or sinusoidal form. In special cases the source may be moved with a constant velocity.

A function generator produces the desired waveform of the motion. A feedback system (feedback amplifier) operates in such a manner that the

moved relative to each other (either by moving the source and keeping the absorber fixed or vice versa). The transmitted (scattered) γ-quanta are registered as a function of the relative velocity. In this way, it is possible to trace, stepwise, the absorption line by the emission line utilizing the Doppler effect. A γ-quantum which is emitted from the source moving at a velocity v receives a Doppler energy E_D modulation such that

$$E_\gamma = E_0 + E_D = E_0\left(1 + \frac{v}{c}\right) \qquad (11)$$

Figure 5. Block diagram illustrating the principle of a Mössbauer spectrometer

actual velocity of the source follows the reference waveform of the function generator with an accuracy better than 0.1 %. Using the triangular waveform one obtains an undistorted spectrum, because the velocity is a linear function of time. In extreme cases, where high velocities are required or the mass to be moved is large, it is preferable to use a sinusoidal waveform. For convenience, this type of spectrum is converted after the measurement into the linear form.

The counts detected behind the absorber are stored in a memory (mostly in the memory of a mutlichannel analyzer containing typically 512 or 1024 channels). Synchronization of the channel number in the memory and the instantaneous velocity of the source (absorber) is achieved by advancing the address of the memory one by one through an external clock which subdivides the period of the waveform into the number of available channels. A start pulse, which coincides with the beginning of the waveform, triggers the multichannel analyzer to start with channel No. 1, which is subsequently advanced by the clock pulses. In this way synchronization is achieved, i.e., during each period of motion a certain channel of the analyzer always corresponds to the same instantaneous velocity of the source. The whole velocity range of interest is scanned many (typically 10–50) times per second. As each instantaneous velocity occurs twice during one scan period, the spectrum is registered twice in the memory, where both halves of the memory appear as

mirror images, symmetrically about the center of the channel series.

It is essential to have a standard reference line position which other signals can be referred to, particularly in Mössbauer spectroscopy, where different host metals for a particular Mössbauer isotope are often utilized, e.g., ^{57}Co/Cu, ^{57}Co/Pt, ^{57}Co/Rh in case of ^{57}Fe measurements with emitting γ-rays of slightly different energy. The most widely used reference material for velocity calibration in Mössbauer spectroscopy is a natural iron foil of which the six line positions of the magnetic hyperfine split spectrum are accurately known. One uses the distance between the two outermost lines of the iron spectrum as a calibration basis for the measurements of the hyperfine interactions with strengths corresponding to a Doppler velocity region of $0-10\ \mathrm{mm\,s^{-1}}$. This region is suitable for Mössbauer effect measurements in, e.g., ^{57}Fe, ^{119}Sn, ^{61}Ni, ^{197}Au, and others.

The usual detectors for low-energy γ-radiation are thin NaI(Tl) scintillation crystals coupled to a photomultiplier, and gas-filled proportional counters. Considerable improvements in energy resolution can be achieved with solid-state detectors such as the Li drifted Ge counter.

It is only with low-energy γ-emitters, such as the excited state of ^{57}Fe, that the Mössbauer effect can be observed at room temperature and even above. With most Mössbauer isotopes it is necessary to cool the absorber and sometimes also the source to liquid-nitrogen or liquid-helium temper-

Table 1. Some important isotopes in Mössbauer spectroscopy

Mössbauer isotope	Abundance of stable element, %	γ-Ray energy, keV	Half-life of Mössbauer transition, 10^{-9} s	Parent isotope	Half-life of parent isotope
^{57}Fe	2.14	14.4	97.8	^{57}Co	270 d
^{119}Sn	8.58	23.83	17.8	^{119m}Sn	245 d
^{121}Sb	57.25	37.15	3.5	^{121m}Sn	76 y
^{129}I	0 (β^-)	27.77	16.8	^{129}Te	33 d
^{151}Eu	47.82	21.54	9.5	^{151}Sm	90 y
^{197}Au	100	77.35	1.89	^{197}Pt	18 h

$$
\begin{array}{ll}
7/2 - & \underset{27}{^{57}}\text{Co} \quad 270\ \text{d} \\
 & \quad\ \ EC\ (99.84\%) \\
5/2 - & \underline{\qquad}\ 136.32 \\
3/2 - & \underline{\qquad}\ 14.41\ \text{keV} \\
1/2 - & \underline{\qquad}\ 0 \\
 & \underset{26}{^{57}}\text{Fe}
\end{array}
$$
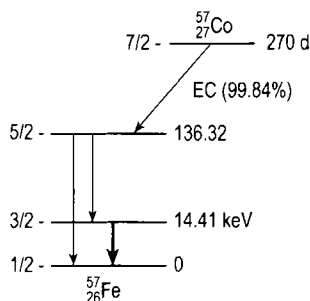
Figure 6. Decay scheme of ^{57}Co

atures in order to obtain a measurable effect. Apart from these practical requirements it is often desirable to vary the temperature of the absorber in order, e.g., to obtain information about phase transitions or the electronic state and the molecular symmetry of the compound of interest. Many types of cryostats are nowadays commercially available. The temperature range varies from 1.2 to 300 K in He-bath cryostats and from 6 to 300 K in He-flow cryostats.

A He-bath cryostat equipped with a superconducting magnet can be utilized for Mössbauer investigations of materials in an applied magnetic field. Magnetic fields up to 10 T are very common today.

Above room temperature electrical furnaces are used. In order to keep the risk of oxidation of the material low, a moderate vacuum of about $10^{-2} - 10^{-3}$ is appropriate.

19.4. Preparation of Mössbauer Source and Absorber

The source for a Mössbauer experiment is a radioactive isotope of reasonable half-life and appropriate nuclear transition energy and excited state lifetime. By radioactive disintegration, the isotope populates a nuclear excited state which decays to the ground state by emitting low-energy γ-radiation. Table 1 lists a selection of isotopes which are suitable for practical Mössbauer effect measurements.

As a typical example, Figure 6 shows the decay scheme of ^{57}Co which populates the 14.4 keV Mössbauer level of ^{57}Fe with a lifetime of $\tau = 140$ ns. The isotope ^{57}Co can be produced in a cyclotron by the nuclear reaction ^{56}Fe(d,n)^{57}Co. The decay of ^{57}Co occurs essentially by electron capture (99.8 %) from the K-shell leaving a hole in this shell which is filled from higher shells under emission of a 6.4 keV X ray. Sources of ^{57}Co are usually prepared by electrochemically depositing the carrier-free isotope on metallic supports and then diffusing it into the metal at high temperatures.

A Mössbauer source should meet the following requirements.

The emission line should be as narrow and intense as possible, unsplit and unbroadened

The recoil-free fraction should be as high as possible

The source material should be chemically inert during the lifetime of the source and resistant against autoradiolysis

The host material should not give rise to interfering X rays, and Compton scattering and photoelectric processes should be insignificant

A uniform absorber with randomly oriented crystallites can be prepared easily by sandwiching the finely ground material between the thin windows of a Perspex holder. In order to obtain undistorted line shapes it is desirable that the absorber is "thin". The natural line width Γ_{nat} of the Mössbauer line having Lorentzian shape [see Section 19.2.1, Eq. (3)] is determined, as already discussed, by the half-life of the excited nuclear state and the Heisenberg uncertainty principle. The ex-

perimental line (Γ_{exp}), however, is broadened for several reasons. The hyperfine fields may have a distribution which is typically observed in amorphous materials and small particles (nano particles). The narrow hyperfine field distribution caused by imperfections and strains in molecular crystals is typically not observed by a line broadening. Dynamic behavior of hyperfine fields lead to line broadenings if the eneries $\hbar\tau_f$ corresponding to the fluctuation time τ_f of the fields at the nucleus are of the same order of magnitude as the difference of the hyperfine energies belonging to the field values. The sample thickness gives rise to the so-called thickness broadening which, however, can be taken into account evaluating the spectra by the fitting routines.

Heavy elements in the compound under investigation will scatter and absorb the Mössbauer γ-rays, and/or emit photons of similar energy which are detected along with the Mössbauer photons leading to a decrease in the signal to noise ratio. Mössbauer lines of appropriate intensity can therefore be obtained by decreasing the absorber thickness and increasing the recording time. It is therefore preferable to estimate an optimum absorber thickness for which the measuring time is as short as possible in order to achieve the maximum information from the experiment [17], [18].

19.5. Hyperfine Interactions

Owing to the narrow linewidth of the Mössbauer nuclear transition (of the order of 10^{-8} eV), the resonance spectrum is extremely sensitive to energy variations of the γ-radiation. Interactions between the nucleus and the surrounding electrons with energies comparable to the width of the transition lines manifest themselves in the Mössbauer spectrum. It is the influence of the electronic environment on the emission and absorption lines which determines the hyperfine structure of the spectrum. This interaction of the positively charged nucleus with the electric and magnetic fields caused by the orbital electrons in the region of the nucleus is called *hyperfine interaction*. This perturbation shifts or splits degenerate nuclear levels.

The Hamiltonian for the nucleus in a solid usually is expressed in a multipole expansion up to the second order

$$H = H_{E_0} + H_{M_1} + H_{E_2} + \dots \tag{12}$$

Higher order terms are negligibly small and cannot be detected by Mössbauer effect measurements. The first term, the so-called monopole interaction, E_0, gives rise to the *isomer shift* (δ) and shifts the energy levels of both the ground state and the excited state. It originates from the Coulombic interaction between the nucleus and electrons at the nuclear site. The magnetic dipole (M_1) interaction removes the degeneracy of the nuclear levels if the nucleas has a magnetic moment. It is commonly known as the magnetic hyperfine interaction and describes the influence of a magnetic field on the nuclear spin (*magnetic splitting*). The third term in Equation (12) represents the electric quadrupole interaction, i.e., the interaction between the nuclear quadrupole moment and an electric field gradient at the nuclear site. The resulting Mössbauer parameter is called *quadrupole splitting*.

The different interactions and the kind of information one can obtain from the respective Mössbauer parameters are summarized in Table 2. In the following we shall discuss the three hyperfine interactions separately.

19.5.1. Isomer Shift

The isomer shift (other names occasionally used are chemical isomer shift or center shift) arises from the fact that the nucleus of an atom possesses a finite volume, and s-electrons have the ability of penetrating the nucleus and spending a fraction of time inside the nuclear region. p-, d-, and f-electrons do not have this ability.

The electric monopole energy of a uniform spherical nucleus with radius and charge in a constant electron charge density $-e|\psi(0)|^2$ is given by

$$H_{E_0} = \frac{2}{5}\pi Ze^2|\psi(0)|^2 R^2 \tag{13}$$

The nuclear radii of the ground and excited states, in general, differ by a small amount δR. This results in a gain or loss of Coulombic energy δE of the nucleus, when it emits a γ-quantum. This energy is transferred to the γ-ray and therefore produces a shift. The same argument holds for the absorption process, so that the total shift, δ, is expressed by

$$\delta = \frac{4}{5}\pi Ze^2 R^2 (\delta R/R)\left\{\left|\psi_a(0)\right|^2 - \left|\psi_s(0)\right|^2\right\} \tag{14}$$

Table 2. The interactions between the nucleus and surrounding electrons, the respective Mössbauer parameters, and the type of information

Type of interaction	Mössbauer parameter	Information
Electric monopole interaction between nucleus and electrons at the nuclear site	Isomer shift δ	a. Oxidation state (nominal valency) of the Mössbauer atom b. Bonding properties in coordination compounds (covalency effects between central atom/ion and ligands. Delocalization of d-electrons due to back-bonding effects, shielding of s-electrons by p- and d-electrons) c. Electronegativity of ligands
Electric quadrupole interaction between electric quadrupole moment of the nucleus and electric field gradient at the nuclear site	Quadrupole splitting ΔE_Q	a. Molecular symmetry b. Oxidation state (nominal valency) c. Spin state d. Bonding properties
Magnetic dipole interaction between magnetic dipole moment of the nucleus and a magnetic field at the nucleus	Magnetic splitting ΔE_M	Magnetic properties (e.g., ferro-, antiferro-, para-, diamagnetism, absolute value and direction of local magnetic fields)
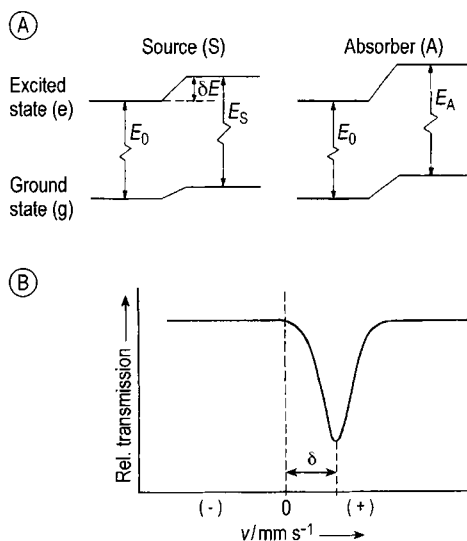
Figure 7. Origin of the isomer shift.
A) Electric monopole interaction in source and absorber shifts the nuclear energy levels without affecting the degeneracy; B) resultant Mössbauer spectrum (schematic)

where $\delta R = R_e - R_g$ is the change in nuclear radius on going from the excited (e) to the ground state (g). The subscripts a and s refer to absorber and source, respectively. Figure 7 shows schematically the origin of the isomer shift. From Equation (14) it is obvious that the isomer shift is always a relative shift, between absorber and source.

Therefore, the measured isomer shift is always given with respect to a standard material; this can be the Mössbauer source used in the particular experiment or any conventional absorber material.

Thus, the isomer shift is expressed by a nuclear term $\delta R/R$ and an electronic term $\left\{ |\psi_a(0)|^2 - |\psi_s(0)|^2 \right\}$, which measures the electron density at the nucleus of an absorber *relative to a given source*. For a given Mössbauer atom the nuclear factor is constant; the isomer shift is thus exclusively dependent on the difference of the electron densities between absorber and source. For [57]Fe $\delta R/R$ is negative, i.e., the nuclear radius in the excited state is smaller than in the ground state; in the case of [119]Sn the reverse is true.

The isomer shift can reflect minute changes in the electron density at the nucleus. In a non-relativistic approximation, the electron density at the nucleus is large only for electrons with zero angular momentum (s-electrons) and can be approximated by $|\psi(0)|^2$. Thus, the isomer shift of the Mössbauer spectra is a relative measure of the total s-electron density at the nuclear probe in a compound.

The total s-electron density of an atom in a chemical compound is the sum of all contributions from the inner (filled) and outer (valence electron) shells. The contribution from the outer electron shells is sensitive to changes in the chemical environment of the element (e.g., change of charge state, change of bonding properties by electron delocalization) and will exert a *direct* influence on the isomer shift. The valence p-, d-, and f-electrons have only an *indirect* influence on isomer shift by their shielding effect on the outer s-electron shells against the positive nuclear charge.

The measurement of the isomer shift gives information on
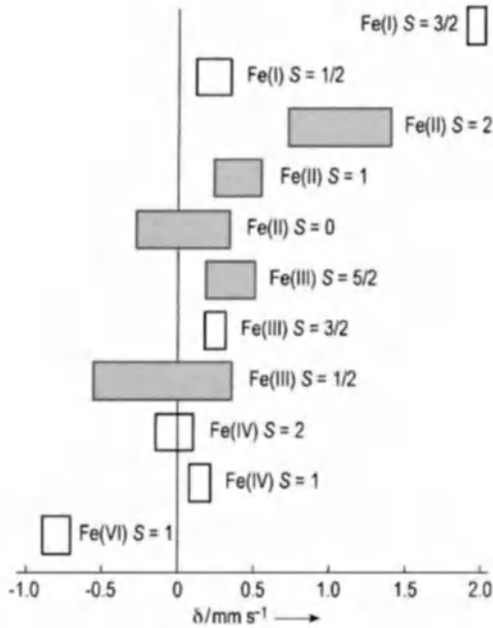
Figure 8. Approximate ranges of isomer shifts observed in iron compounds (relative to metallic iron). S refers to the electronic spin quantum number. Note that $(\delta R/R) < 0$ for ^{57}Fe

the oxidation state (e.g., Fe(II), Fe(III), Fe(IV)),
the spin state (high spin, low spin),
the bond properties (covalency, ionicity),
the electronegativity of the ligands

of a Mössbauer atom in solid material (coordination compounds, alloys, intermetallic phases, amorphous materials, etc.).

Figure 8 shows approximate ranges of the isomer shifts determined in iron compounds with different oxidation and spin states of the central metal ion.

It was shown [19] that the electronic density difference on going from the configuration 3d^6 (Fe^{2+}) to 3d^5 (Fe^{3+}) essentially originates from changes in the 3s shell. The filled 1s and 2s orbitals remain practically constant for both configurations. The removal of the 3d electron leads to an increase in the electron density at the nucleus due to the less effective shielding of the 3s electrons from the nuclear charge by the 3d electrons in Fe^{3+} as compared with Fe^{2+}. As $\delta R/R$ is negative for ^{57}Fe the isomer shift becomes more negative for Fe^{3+} than for Fe^{2+} (with respect to the same standard). The more or less wide-spread ranges of the δ values for each oxidation state is a direct con-

sequence of the nature of the chemical bond. The electron distribution in the molecular orbitals is the result of the variable abilities of the ligands to donate electrons to the metal ion via σ-bonding and to accept electrons from the metal ion via π-bonding.

Similar correlation diagrams as shown in Figure 8 are available for numerous Mössbauer atoms [20].

19.5.2. Magnetic Hyperfine Interaction

An atomic nucleus, either in the ground or excited state, with nuclear spin quantum number $I > 0$ possesses a magnetic dipole moment $\rightarrow\mu$ and interacts with a magnetic field $\rightarrow B$ at the nuclear site. The magnetic field $\rightarrow B$ may arise from the electronic environment or from an external magnet. This interaction is called *magnetic dipole interaction* and is described by the Hamiltonian

$$H_{M_1} = -\rightarrow\mu\rightarrow B = -g_N\mu_N\hat{I}\rightarrow\rightarrow B \tag{15}$$

where g_N is the nuclear Landé factor and $\mu_N = e\hbar/2Mc$ (M is the mass of the nucleus) is the nuclear magneton. The magnetic dipole interaction splits a nuclear state $|I\rangle$ into $2I+1$ equally spaced substates $|I, m_I\rangle$, each of these being characterized by the nuclear magnetic spin quantum number $m_I = -I, -I+1, ..., I$ (nuclear Zeeman effect). The Zeeman energies of the substates are $|I, m_I\rangle$

$$E_M(m_I) = -\mu B M_I/I = -g_N\mu_n B m_I \tag{16}$$

Since the Mössbauer transition in ^{57}Fe is of the magnetic dipole type (M_1), there are only transitions between nuclear sublevels with $\Delta m_1 = 0, \pm 1$ and $\Delta I = \pm 1$. This selection rule yields only six transitions between the two ground state sublevels $(I = 1/2)$ and the four excited state sublevels $(I = 3/2)$. Figure 9 illustrates the splitting of the nuclear levels, the allowed transitions, and the resulting ^{57}Fe Mössbauer spectrum, a sextet due to the magnetic hyperfine interaction.

The magnetic hyperfine field $\rightarrow B$ at the nucleus can originate from the own electron shell $(\rightarrow B_{ion})$ or can be generated by paramagnetic ions in the nearby lattice $(\rightarrow B_{lat})$, or can be due to an applied magnetic field $(\rightarrow B_{app})$:

$$\rightarrow B = \rightarrow B_{ion} + \rightarrow B_{lat} + \rightarrow B_{app} \tag{17}$$
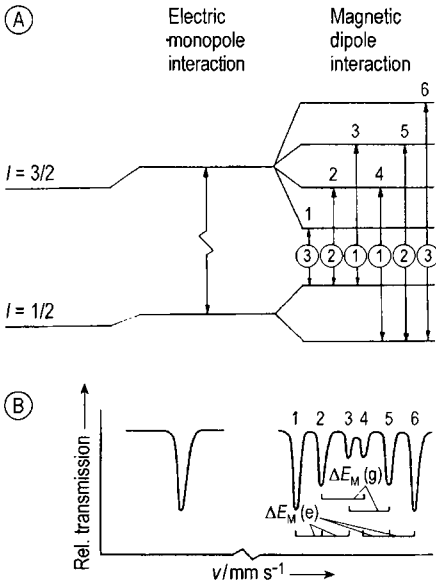
Figure 9. Origin of the magnetic hyperfine splitting
A) Magnetic dipole interaction in ^{57}Fe in absorber;
B) resultant Mössbauer spectrum (schematic)

The dipolar field of the moments of the surrounding paramagnetic ions $\rightarrow B_{lat}$ is a very small contribution.

$\rightarrow B_{FC}$ is the so-called Fermi contact field [21] and represents the most important contribution to the field $\rightarrow B_{ion}$. This field is the direct coupling between the nucleus and the unpaired s-electron density at the nucleus [22], [23],

$$\rightarrow B_{FC} = -\frac{16\pi}{3}\mu_B\left(\sum(\uparrow \psi_s^2(0) - \downarrow \psi_s^2(0))\right) \quad (18)$$

where $\uparrow \psi_s^2(0)$ is the spin density at the nucleus of s-electrons with spin parallel to the total spins of the valence d-electrons spin and $\downarrow \psi_s^2(0)$ that with spin antiparallel to the d-electrons spins.

Owing to the fact that the exchange interaction between the spin-up polarized d-shell and the spin-up s-electron is attractive, while that between the spin-up d-shell and the spin-down s-electron is repulsive, the radial parts of the two s-electrons distort, one being closer to the nucleus and the other being more distant. As a consequence, the spin density with spin-down is enhanced at the nucleus. The direction of the Fermi contact field is antiparallel to this spin, and consequently the Fermi contact field is negative by defintion, which means anti-parallel to the magnetic moments of the d-electrons.

For high spin ferric compounds with electron configuration $3d^5$, ^6S, the dipolar field due to its own electrons and the contribution of the orbital momentum to the total field are zero. The Fermi contact term, estimated to be -11 T per unpaired 3d electron, dominates over the contributions of the neigboring dipoles and the applied field. In high spin ferrous compounds, however, the negative Fermi contact term is opposed by a positive orbital contribution of almost the same order of magnitude. The positive dipolar contribution is in general quite small. In this case it is impossible to predict the sign of the internal hyperfine field, B_{ion}. This problem can be solved by applying an external magnetic field that aligns the 3d moments. Sign and absolute value of the internal field can then be obtained from the vector sum of the applied and the internal field.

19.5.3. Quadrupole Splitting

The third term in the multipole expansion [cf. Equation (12)] describes the *electric quadrupole interaction*. This type of nuclear–electronic interaction is also electrostatic in nature and occurs if the following conditions are fulfilled.

1) The Mössbauer nucleus must possess a measurable electric quadrupole moment eQ, which arises from non-spherical nuclear charge distribution leading to spin quantum numbers $I > 1/2$. A measure of the deviation from spherical symmetry is given by the electric quadrupole moment, eQ, which can be calculated by the spatial integral

$$eQ = \int^2 (3\cos^2\theta - 1)\, dV \quad (19)$$

where e is the proton charge and p is the charge density at the spherical coordinates r and θ. The sign of the nuclear quadrupole moment refers to the shape of the distorted nucleus: Q is negative for a flattened and positive for an elongated nucleus.

2) The electric field at the Mössbauer nucleus must be inhomogeneous, measured by the electric field gradient (EFG) $V_{ij} = \partial^2 V/\partial i\partial j$ ($i, j = x, y, z$). This arises from a non-cubic charge distribution of the electrons and/or of the neighboring ions around the nucleus.

With respect to the principal axes, the EFG tensor is described by only two independent pa-
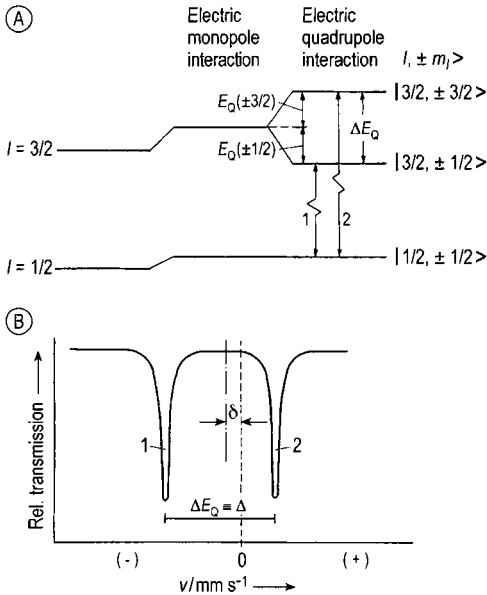
Figure 10. Origin of the quadrupole splitting in ^{57}Fe
A) electric quadrupole interaction in case of $eQV_{zz} > 0$;
B) resultant Mössbauer spectrum (schematic) with equal intensities as for a powder sample; the inevitable shift of the nuclear levels due to electric monopole interaction giving rise to the isomer shift is also shown

rameters, usuallly chosen as $V_{zz} = eq$ and $0 \leq \eta = (V_{xx} - V_{yy})/V_{zz} \leq 1$.

The spin Hamiltonian of the nucleus in the principal axes system of the EFG is given by:

$$H_{E_2} = \frac{eQV_{zz}}{4I(2I-1)} \left\{ 3\hat{I}_z^2 - I(I+1) + \frac{\eta}{2}\left(\hat{I}_+^2 + \hat{I}_-^2\right) \right\}$$

(20)

where $\hat{I}_{z,+,-}$ are the usual spin operators.

As an example, the effect of the electric quadrupole interaction in a Mössbauer nucleus with $I = 3/2$ in the excited state and $I = 1/2$ in the ground state, as is the case in ^{57}Fe and ^{119}Sn, is given in Figure 10. The quadrupole interaction gives rise to a doublet with the splitting

$$\Delta Eq = \frac{1}{2}e\left|QV_{zz}\right|\left(1 + \frac{\eta^2}{3}\right)^{\frac{1}{2}}$$

(21)

The quadrupole splitting in a Mössbauer spectrum is a highly sensitive probe for the chemical environment that determines the EFG at the nucleus of the Mössbauer atom. In general, there are two

fundamental sources which can contribute to the total EFG.

1) Charges on ions surrounding the Mössbauer probe atom in noncubic symmetry give rise to the so-called *lattice contribution*, (EFG)$_{lat}$ or V_{zz}^{lat} (for axial symmetry).

2) Noncubic distribution of the electrons in the partially filled valence orbitals of the Mössbauer atom generate the *valence contribution*, (EFG)$_{val}$ or V_{zz}^{val} (for axial symmetry).

According to the different origins (valence/lattice contributions) it is expected that observed quadrupole splittings may reflect information about the electronic structure (oxidation state, spin state), bond properties, and molecular symmetry. As an example the spectra of three iron coordination compounds are shown in Figure 11.

a) The observed large quadrupole splitting of $\Delta E_Q = 3.4$ mm s^{-1} for FeSO$_4 \cdot$7H$_2$O essentially arises from the valence contribution V_{zz}^{val}. This compound is a typical high spin compound where the Fe(II) ion is octahedrally surrounded by six water molecules. This type of ions with a $^5T_{2g}$ ground state are subject to Jahn–Teller distortion, which lifts the degeneracy of the t_{2g} orbitals partly or completely and leads then to a noncubic distribution of the six electrons in the 3d shell of the [Fe(H$_2$O)$_6$]$^{2+}$ ions. The doubly occupied lowest d_{xy} orbital gives rise to the relatively large quadrupole splitting in the tetragonally compressed [Fe(H$_2$O)$_6$]$^{2+}$ octahedron. Further details can be found in reference [12].

b) K$_4$[Fe(CN)$_6$]$\cdot$3H$_2$O is a typical low spin Fe(II) compound. The single line Mössbauer spectrum results from the absence of a quadrupole interaction due to the cubic charge distribution of the electron configuration t_{2g}^6 which consequently leads to zero valence contribution V_{zz}^{val} [12]. The lattice contribution V_{zz}^{lat} also vanishes since the [Fe(CN)$_6$]$^{4-}$ complex ion forms a regular octahedron.

c) The relatively large quadrupole splitting in the case of Na$_2$[Fe(CN)$_5$NO]$\cdot$2H$_2$O is solely due to the fact that the presence of different ligand molecules can no longer form a regular octahedron as in the case of [Fe(CN)$_6$]$^{4-}$. The magnitude of the quadrupole splitting is mainly determined by the lattice contribution.
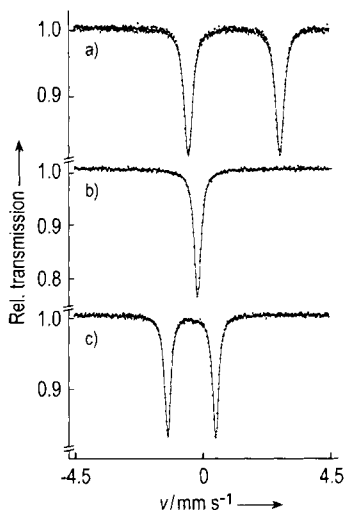
Figure 11. Mössbauer spectra of: a) $FeSO_4 \cdot 7H_2O$; b) $K_4[Fe(CN)_6] \cdot 3H_2O$; and c) $Na_2[Fe(CN)_5NO] \cdot 2H_2O$

19.5.4. Combined Quadrupole and Magnetic Hyperfine Interactions

Quite frequently electric quadrupole interactions and magnetic dipole interactions may be present in addition to the electric monopole interactions which are always active. The total Hamiltonian H_{HFS} for the combined electric quadrupole and magnetic dipole interaction cannot be solved analytically in the general case of comparable interaction strengths and arbitrary direction of the magnetic hyperfine field in the PAS of the EFG. In case of a small EFG and large magnetic field, a situation often met, a perturbation treatment may be sufficient. For an axially symmetric EFG with the polar angle ϑ of the direction of the magnetic field in the PAS of the EFG the eigenvalues of the nuclear sublevels of the $I = 3/2$ state are given by

$$E_{M,Q}(I, M_I) = -g_N \mu_N B m_I$$
$$+ (-1)^{|m_I| + \frac{1}{2}} (eQV_{zz}/8)(3\cos^2\vartheta - 1)$$

$$(22)$$

We find in the case of ^{57}Fe, which is depicted in Figure 12, that the sublevels $|3/2, \pm 3/2\rangle$ are shifted by an amount $E_Q(\pm m_I)$ to a higher energy and the sublevels $|3/2, \pm 1/2\rangle$ are shifted by E_Q to a lower energy, if V_{zz} is positive. These energy shifts by E_Q are reversed if V_{zz} is negative. As a result, the sublevels of the excited state $I = 3/2$ are no longer equally spaced, unless $\cos\vartheta = \sqrt{1/3}$. Combined

electric quadrupole and magnetic dipole interactions generally manifest themselves in the Mössbauer spectrum as an asymmetrically split sextet (for $I = 3/2 \longleftrightarrow I = 1/2$ transitions as in ^{57}Fe) as described in Figure 12. As the sublevel spacings and thus the asymmetry of the spectrum is directly correlated with the sign of V_{zz}, one can determine the sign of the EFG of a polycrystalline material from the magnetically split hyperfine spectrum. The effect of a magnetic dipole interaction (applied magnetic field) as a perturbation of the electric quadrupole splitting is described in detail in reference [24].

These three types of hyperfine interactions with the relevant Mössbauer parameters are most important in solid state research. In addition, one often extracts further helpful information from the temperature and pressure dependence of the Mössbauer parameters, the shape and width of the resonance lines (relaxation phenomena), and the second-order Doppler shift (lattice dynamics).

19.5.5. Relative Intensities of Resonance Lines

The relative intensities of the lines in a Mössbauer spectrum provide additional information about the orientation of the hyperfine fields with respect to the crystal system in the case of single crystals or partially oriented (textured) samples. For pure magnetic or electric hyperfine fields the intensities are analytical expressions dependent on the direction of the fields with respect to the direction of γ-radiation. For powder samples the intensity ratios of pure transitions do not depend on the strength of the fields.

For the case of a magnetically split spectrum of a $3/2 \rightarrow 1/2$ transition the intensity ratios are $3:2:1:1:2:3$ for the hyperfine components of the Zeeman pattern. The corresponding quadrupole spectrum has equal intensities as already mentioned above. For more details the reader is referred to the references [25], [26].

19.6. Evaluation of Mössbauer Spectra

The evaluation of a Mössbauer spectrum implies the determination of the physical parameters, the isomer shift, Debye–Waller factor, and electric and magnetic hyperfine fields. When fluc-
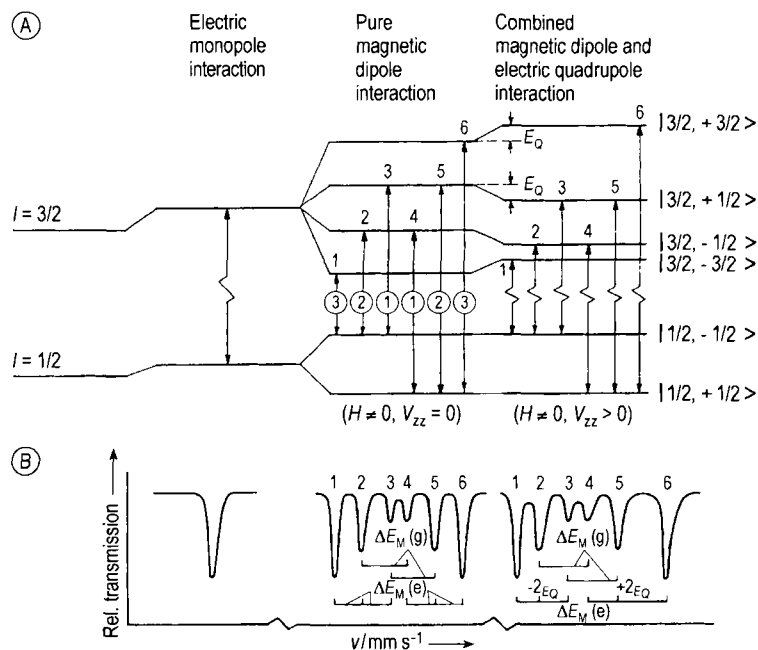
Figure 12. Combined quadrupole and magnetic hyperfine interaction: (a) splitting and shifts, (b) resultant Mössbauer spectrum (schematic)

tuation times of the hyperfine fields of the order influencing the shape of the spectra are present the fluctuation times themselves may be a parameter of interest.

Fitting of Mössbauer spectra requires the existence of a model for the hyperfine interaction at the Mössbauer nuclei in the material under study. Computer programs are available for a variety of models. The interested reader is referred to the homepage of the Mössbauer community: http://www.kfki.hu/mixhp/.

19.7. Selected Applications

Since the discovery more than 40 years ago, Mössbauer spectroscopy has become an extremely powerful analytical tool for the investigation of various types of materials. In most cases, only two parameters are needed, viz. the isomer shift and the quadrupole splitting, to identify a specific sample. In case of magnetically ordered materials, the magnetic dipole interaction is a further helpful parameter for characterization. In the following

we shall discuss some applications selected from different research areas.

19.7.1. Magnetic Materials

As an example we will consider the magnetic ordering in magnetite, Fe_3O_4.

Magnetite is an inverse spinel, with iron atoms occupying interstitial sites in the close-packed arrangement of oxygen atoms. The tetrahedral sites (A) are occupied by ferric cations, and the octahedral sites (B) half by ferric and half by ferrous cations. In the unit cell there are twice as many B-sites as A-sites. The structural formula can therefore be written: $(Fe(III))_{tet}[Fe(II)Fe(III)]_{oc-tO_4}$. There is little distortion in the cubic close packing of the oxygens, so that the structure is close to ideal.

Magnetite has a Curie temperature of 840 K. In the magnetically ordered phase the iron atoms on A-sites are antiferromagnetically coupled to those on B-sites resulting in a complete cancelling of the moments of the Fe(III) ions. Thus, a ferromagnetic moment, originating in the Fe(II) ions on A-sites, remains. On the whole, magnetite shows ferrimagnetic behavior.
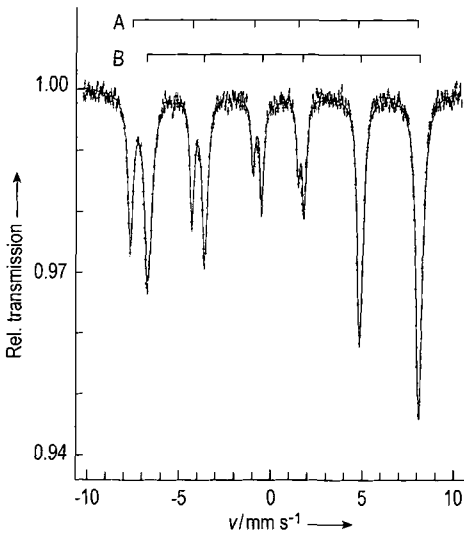
Figure 13. Room temperature Mössbauer spectrum of magnetite: subspectra of lattice sites A and B are indicated by the stick diagram

Figure 14. Mössbauer spectra of the most important corrosion products

The statistical distribution of the di- and trivalent ions of iron on equivalent lattice sites (B) explains many unusual properties of magnetite, in particular its extremely high electrical conductivity comparable to that of metals. Investigations have shown that the conductivity is caused by electron and not by ion transport. A fast electron-transfer process (electron hopping) between ferrous and ferric ions on the octahedral B-sites takes place [27]. The Mössbauer spectrum of magnetite at room temperature (cf. Figure 13) confirms this interpretation.

The spectrum consists of two more or less overlapping six-line spectra with internal magnetic fields of 49 and 46 T and a relative area ratio of approximately 1 : 2, which originate from the iron ions on sites A (Fe(III)) and B (average field of Fe(II) + Fe(III)). The Mössbauer parameters (δ, ΔE_Q and H_{int}) for the sextet stemming from the site B ions are intermediate between those of oxidic Fe(II) and Fe(III) compounds; i.e., concerning the iron ions on lattice site B, we are dealing with a mixed-valent system of Class III in the Robin–Day classification. At ca. 120 K, magnetite shows the so-called Verwey transition. The rate of the electronic hopping between the Fe(II) and Fe(III) ions on the lattice B sites slows down such that it becomes slower than the reciprocal of the Mössbauer time window. This causes a significant change in the magnetically split subspectra [28], [29].

19.7.2. Corrosion Products

Mössbauer spectroscopy of ^{57}Fe has proven to be an extremely powerful tool for the analysis of iron containing corrosion products on steel and technical alloys. Depending on the corrosion conditions a variety of iron oxides and oxyhydroxides may be formed. Mössbauer spectroscopy enables one to differentiate between the various phases on the basis of their different magnetic behavior. Figure 14 illustrates this with six different corrosion products. The Fe_3O_4 (magnetite), shows the superposition of two six-line spectra as described in Section 19.7.1.

The α-Fe_2O_3 (haematite), γ-Fe_2O_3 (maghemite), and α-FeOOH (goethite) differ in the size of the internal magnetic field yielding different spacings between corresponding resonance lines. A problem seems to arise with β- and γ-FeOOH, because they have nearly identical Mössbauer parameters at room temperature and cannot be distinguished by their Mössbauer spectra. However, lowering the temperature solves the problem: β-FeOOH appears to be magnetically ordered at 80 K, whereas γ-FeOOH still remains in the para-magnetic state.
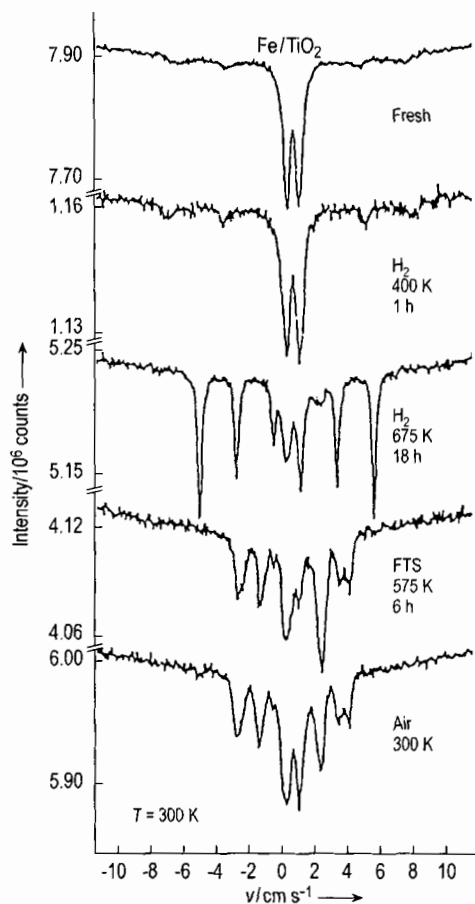
Figure 15. ^{57}Fe Mössbauer spectra at room temperature (from top to bottom) of the catalyst system Fe/TiO$_2$ after different treatments (cf. text). Reproduced from Applied Catalysis [29]

Such Mössbauer studies may be carried out in a nondestructive manner and are even possible with highly dispersed particles which are no longer amenable to X-ray diffraction measurements.

19.7.3. Catalysis

The successful use of Mössbauer spectroscopy for studies of catalysts is demonstrated in an important technical example (catalyst system). In Figure 15 a series of spectra of the Fe/TiO$_2$ catalyst system [30] is shown. The top spectrum represents the freshly prepared catalyst; it is obvious that iron is present in the 3+ oxidation state, possibly in the form of finely dispersed oxide or oxyhydroxide on TiO$_2$. Under a reducing atmosphere
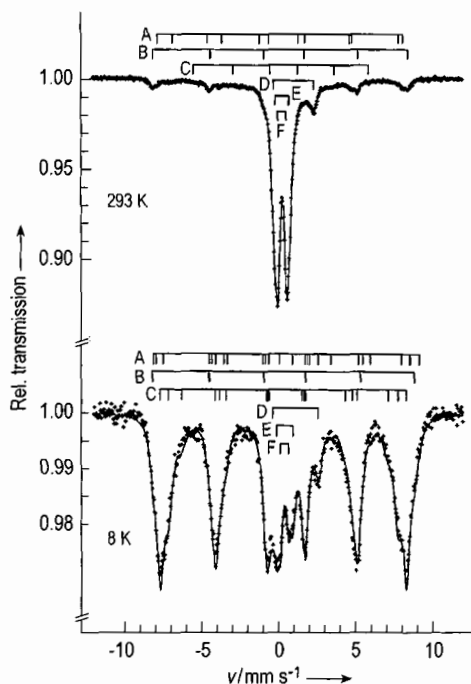
Figure 16. 293 K (top) and 8 K (bottom) Mössbauer spectra of the aerosol sample; the decomposition into the spectral components is indicated by the stick bars, where A stands for magnetite (Fe$_3$O$_4$), B for hematite (α-Fe$_2$O$_3$), C for goethite (α-FeOOH), D for Fe^{2+}-, E and F for Fe^{3+}-compounds

(H$_2$ at 675 K) this iron phase is essentially converted into metallic iron (six line pattern). Some residues of Fe^{2+} and Fe^{3+} contribute to the spectrum as doublets. Under the conditions of the Fischer–Tropsch synthesis (FTS) at 575 K in CO and H$_2$ the metallic iron is transformed into the Hägg carbide χ-Fe$_5$C$_2$. The unreduced iron is now present as Fe^{2+}, which is oxidized to ferric iron when the catalyst is exposed to air at ambient temperature. The carbide phase is left unchanged.

19.7.4. Atmospheric Aerosol Samples

Iron is one of the most abundant elements in solid and aqueous atmospheric samples and is usually introduced into the atmosphere as soil dust, fly ash from power plants, exhaust from combustion engines, and from industrial operations. The role of iron in important atmospheric chemical reactions is subject to numerous investigations. Mössbauer spectroscopy proves to be a useful analytical method to determine the iron containing species in the atmosphere. In Figure 16
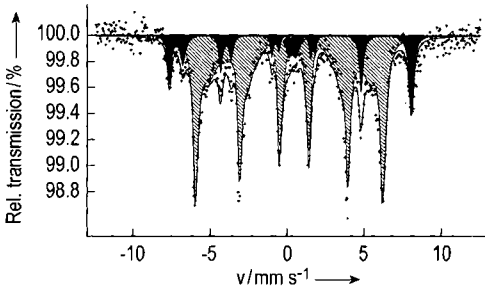
Figure 17. ^{57}Fe Mössbauer spectrum of a one US-$ bill. The different spectral components are marked as: black = magnetite (Fe_3O_4), dashed lines = goethite (α-FeOOH), white = spm Fe^{3+} component

the result of the Mössbauer analysis [31] of an aerosol sample taken from an air-conditioning device of a building is shown.

The typically very small crystals of the different components in the aerosol sample give rise to the so-called superparamagnetic (spm) behaviour (cf. BRONGER et al. [32] and references therein) resulting in the pronounced resonance doublet for Fe^{3+} compounds in the room temperature spectrum. The low temperature spectrum (8 K) revealed that this doublet component mainly originates in superparamagnetic microcrytals of goethite and hematite.

19.7.5. US-$ Bill

Figure 17 shows the Mössbauer spectrum of a one dollar bill. For the measurement the note was folded in such a manner that approximately 10 layers of the bill were analyzed. The result of the analysis: the banknote contains about 20 wt-% of magnetite and 80 wt-% of goethite as iron containing color pigments.

19.8. References

[1] R. Mössbauer, Z. Phys. **151** (1958) 124.
[2] R. Mössbauer, Naturwissenschaften **45** (1959) 538.
[3] R. Mössbauer, Z. Naturforsch. **14a** (1959) 211.
[4] H. Frauenfelder: The Mössbauer Effect, Benjamin, New York 1962.
[5] G. K. Wertheim: Mössbauer Effect: Principles and Applications, Academic Press, New York 1964.
[6] H. Wegener: Der Mössbauer-Effekt und seine Anwendung in Physik und Chemie, Bibliographisches Institut, Mannheim 1965.
[7] V. I. Goldanskii, R. Herber (eds.): Chemical Applications of Mössbauer Spectroscopy, Academic Press, New York 1968.
[8] N. N. Greenwood, T. C. Gibb: Mössbauer Spectroscopy, Chapman and Hall, London 1971.
[9] G. M. Bancroft: Mössbauer Spectroscopy. An Introduction for Inorganic Chemists and Geochemists, McGraw-Hill, New York 1973.
[10] U. Gonser (ed.): "Mössbauer Spectroscopy," in Topics in Applied Physics, vol. 5, Springer, Berlin, Heidelberg, New York 1975.
[11] T. C. Gibb: Principles of Mössbauer Spectroscopy, Chapman and Hall, London 1976.
[12] P. Gütlich, R. Link, A. X. Trautwein: "Mössbauer Spectroscopy and Transition Metal Chemistry," in Inorganic Chemistry Concepts, vol. 3, Springer, Berlin, Heidelberg, New York 1978.
[13] J. G. Stevens: Mössbauer Effect Reference and Data Journal, Mössbauer Data Center, University of North Carolina, Asheville, USA.
[14] W. Kuhn, Philos. Mag. **8** (1929) 625.
[15] D. Barb: Grundlagen und Anwendungen der Mössbauerspektroskopie, Akademie-Verlag, Berlin 1980.
[16] E. Kankeleit: Rev. Sci. Instrum. **35** (1964) 194.
[17] U. Shimony, Nucl. Instrum. Methods **37** (1965) 348.
[18] S. Nagy, B. Levay, A. Vertes, Acta. Chim. Acad. Sci. Hung. **85** (1975) 273.
[19] R. E. Watson, Phys. Rev. **118** (1960) 1036.
[20] G. K. Shenoy, F. E. Wagner: Mössbauer Isomer Shifts, North Holland Publishing Company, Amsterdam, New York, Oxford 1978.
[21] A. Abragam: The Principles of Nuclear Magnetism, Oxford University Press, Oxford 1961.
[22] W. Marshall, Phys. Rev. **110** (1958) 1280.
[23] W. Marshall, Phys. Rev. **105** (1957) 158.
[24] R. L. Collins, J. Chem. Phys. **42** (1965) 1072.
[25] T. Viegers: Thesis, University of Nijmegen, Netherlands 1976.
[26] A. Gedikli, H. Winkler, E. Gerdau, Z. Phys. **267** (1974) 61.
[27] E. J. W. Vervey, Nature (London) **144** (1939) 327.
[28] R. S. Hargrove, W. Kundig, Solid State Commun. **8** (1970) 330.
[29] M. Rubinstein, D. M. Forester, Solid State Commun. **9** (1971) 1675.
[30] A. M. van der Kaan, R. C. H. Nonnekens, F. Stoop, J. W. Niemantsverdriet, Appl. Catal. **27** (1986) 285.
[31] P. Hoffmann, A. N. Dedik, J. Ensling, S. Weinbruch, S. Weber, T. Sinner, P. Gütlich, H. M. Ortner, J. Aerosol. Sci. **27** (1996) 325.
[32] A. Bronger, J. Ensling, P. Gütlich, H. Spiering, Clays Clay Miner. **31** (1983) 269.

20. Mass Spectrometry

MICHAEL LINSCHEID, Department of Chemistry, Humboldt University. Berlin, Germany

Abbreviations

AMS   accelerator mass spectrometry
API   atmospheric pressure ionization
APCI   atmospheric pressure chemical ionization
CAD   collision activated dissociation
CIMS   chemical ionization mass spectrometry
CZE   capillary zone electrophoresis
DADI   direct analysis of daughter ions
EIMS   electron impact (ionization) mass spectrometry
ESI   electrospray ionization
FABMS   fast atom bombardment mass spectrometry
FDMS   field desorption mass spectrometry
FIMS   field ionization mass spectrometry
FTMS   Fourier transform mass spectrometry
GC/MS   gas chromatography/mass spectrometry
GDMS glow discharge mass spectrometry
HPLC   high performance liquid chromatography
ICPMS inductively coupled plasma mass spectrometry
ICR   ion cyclotron resonance
IDMS   isotope dilution mass spectrometry
IMMA ion microprobe mass analyzer
IRMS   isotope ration mass spectrometry
LC/MS liquid chromatography/mass spectrometry
LD   laser desorption
LAMMA   laser microprobe mass analyzer
LAMS laser ablation mass spectrometry
LSIMS liquid secondary ion mass spectrometry (see FAB)
MALDI   matrix assisted laser desorption
MID   multiple ion detection
MIKES   mass analyzed ion kinetic energy spectrome-
           try
MS/MS   mass spectrometry/mass spectrometry
MUPI multiphoton ionization
NCI   negative chemical ionization
PD   plasma desorption
PFK   perfluorokerosene
REMPI   resonance multi photon ionization
RIMS   resonance ionization mass spectrometry
SEM   secondary electron multiplier
SIMS   secondary ion mass spectrometry
SIR   selected ion recording
SNMS sputtered neutral mass spectrometry
TCDD tetrachlorodibenzodioxin
TIMS   thermal ionization mass spectrometry
TOFMS   time-of-flight mass spectrometry

20.1. Introduction

A mass spectrometer can be defined as any instrument capable of producing ions from neutral species and which provides a means of determining the mass of those ions, based on the mass-to-charge ratio (m/z, where z is the number of elemental charges) and/or the number of ions. Therefore, a mass spectrometer has an ion source, an analyzer, and a detector. All further details are largely dependent on the purpose of the mass spectrometric experiment.

Historically, mass spectroscopy was developed to separate atoms and to determine the masses of isotopes accurately [1]. The predecessor of magnetic sector field mass spectrometers was built in 1918 by DEMPSTER [2]. Later, some of the first instruments for the analysis of molecules [3] were used by analytical chemists in refineries to determine hydrocarbons (the first article in the first issue of the journal *Analytical Chemistry* dealt with mass spectrometry of a mixture of hydrocarbons [4]). Combinations of infrared and mass spectroscopy were reported shortly afterwards [5].

Today, mass spectrometers may be used to determine the isotopic distribution of an element, the elemental or molecular composition of a sample, or the structure of a compound or its molecular mass. They also make it possible to study the kinetics and thermodynamics of gasphase processes or the interactions at phase boundaries. Mass spectrometers can also serve to accurately determine physical laws and natural constants.

In recent years, mass spectrometry has become one of the key technologies in proteome and genome research because of the rapid development of new components — from the ion sources to the detector — and intergration into the World Wide Web has also helped to make this possible.

20.2. General Techniques and Definitions

20.2.1. Resolution

Generally, resolution is defined by $R = m/\Delta m$, where m is the mass under consideration. This can be separated from a mass $m + \Delta m$, assuming a peak overlap of 10%. The equation can be used either to calculate the resolution of two signals in a spec-
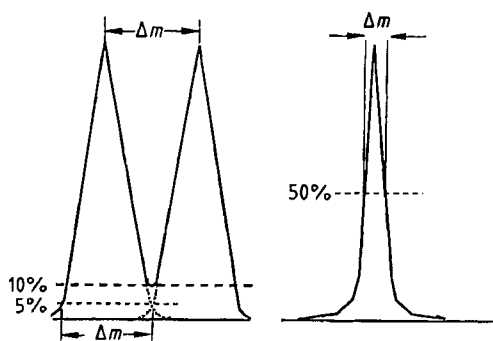
Figure 1. Resolution of two peaks, defined for the 10 % height of the valley, for one signal at the 5 % (sector field MS), or the 50 % (all others) height

trum or to calculate the resolution necessary to separate two species of different elemental composition. For example, consider C_7H_{16} and $C_6H_{12}O$. The mass difference at $m/z = 100$ is 0.0364 amu. Thus, a resolution of ca. 3000 is required to resolve the two signals. The same result is obtained when the peak width at 5 % overlap is used for the calculation.

A different definition uses the peak width at half height as Δm (50 % definition) for the calculation (see Fig. 1). This approach is used often for Fourier transform mass spectroscopy (FTMS) and time-of-flight (TOF) instruments. In modern mass spectrometers, the resolution is calculated directly by the computer from the peak width.

More information can be extracted if the elemental composition of fragments or molecular ions can be determined. For this purpose two different requirements have to be fulfilled. First, the signal must be homogeneous, i.e., even isobaric ionic species have to be separated (resolved) and, secondly, the accuracy of the mass measurement has to be sufficient to allow a calculation with small errors.

High Resolution. The separation of isobaric ionic species (i.e., high resolution, with $R > 10\,000$), is a relative requirement. It depends on various factors, such as the type of heteroatoms, but with $R = 10\,000$ most common species can be resolved. Exceptions can be ions containing, e.g., sulfur. Increasing the resolution requires reducing the width of the ion beam in the mass spectrometer. Thus, in theory, the height of the signal is inversely proportional to the resolution. In reality, this relation is not linear over the full

range of possible resolutions, and generally the loss of peak height becomes much greater when the resolution is increased above $10\,000$. Merely instruments like the ion trap (with very large molecules) or especially the FT mass spectrometers are capable of achieving resolution exceeding 10^5, even up to 10^6 under optimal circumstances without deterioration in detection power. The scan speed and the scan width may be affected. Thus, for applications in analytical problem solving, the resolution should be set as high as necessary yet as low as possible.

Accurate Mass Measurements. For accurate mass measurements a signal containing sufficient ions must be recorded on a precisely calibrated mass scale. In the past this was a cumbersome undertaking because each peak had to be determined separately by comparison to a known reference signal. The accelerating voltage U_0 was changed by the amount necessary to shift the signal of the mass m_x to the position of the reference m_0 on the monitor. This can be performed very accurately by switching between the two voltages U_0 and U_x (for the unknown). Since $U_0 m_0 = U_x m_x$, m_x can be calculated from $m_x = m_0 \cdot (U_0/U_x)$. This technique, known as peak matching, is still the most precise method, and today computerized versions are available.

Techniques were then developed that allow accurate mass measurement of many signals in a spectrum simultaneously. A reference compound such as perfluorokerosene (PFK) is introduced into the ion source together with the analyte in such a concentration that signals of both can be acquired simultaneously. The reference compound is chosen so that its signals are easilyresolved from those of the analyte. The computer then searches for the reference signals in the spectrum, interpolates between them, and calculates the correct position of the unknowns on the mass scale. Sufficient precision can be achieved to allow the calculation of possible elemental compositions for the signal, provided the number of ions is sufficient. It has been shown [6] that the number of ions N required to define the peak position is a function of resolution R, where σ is the standard deviation of the mass measurement in parts per million

$$N = \frac{1}{24}\left(\frac{10^6}{R\sigma}\right)^2$$

Thus, to obtain an accuracy of 2 ppm, the number of ions required for $R = 10\,000$ is $N = 105$, but for
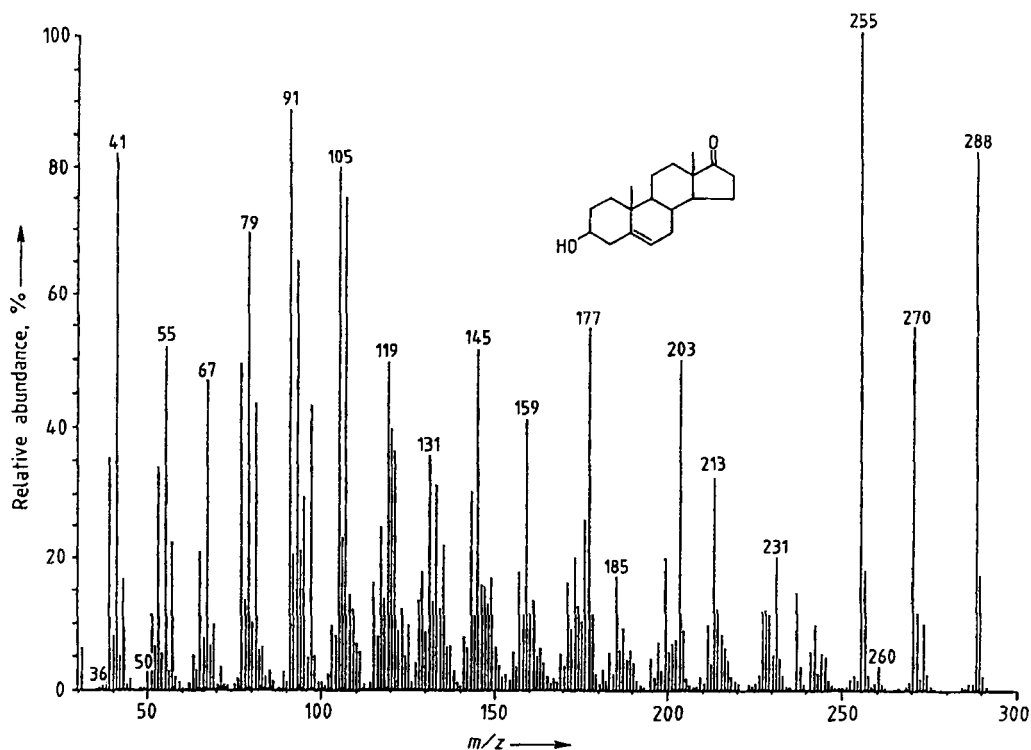
Figure 2. Low resolution ($R = 1000$) spectrum of dehydro-epi-androsteron, direct inlet using a VG 7070E, electron impact mass spectrometer (EIMS)

$R = 5000$, $N = 420$. As long as the reduction of signal intensity with increasing resolution is linear, higher resolution can be beneficial, because fewer ions are required for a given accuracy. However, if the intensity of the signal is also of relevance (e.g., for quantitative determinations) more ions must be detected. Thus, in this case the optimal resolution may be lower to allow better detection power.

20.2.2. Tools for Structure Elucidation

20.2.2.1. Full Spectra: Low Resolution

In the structure determination of natural products, electron impact ionization not only produces ions of the compound itself ("molecular ions") but also results in formation of fragments, often with typical patterns (Fig. 2).

The spectra can be used to identify a compound or at least the class of compound. The spectrum of, for example, a steroid can reflect its

chemical nature to such an extent that the probability of correct identification of the type of compound or even the compound itself may be acceptable. To achieve this it is sufficient to register the spectrum of fragments with integer precision. However, this means that isobaric molecules containing different heteroatoms are not separated. The success of this method for the elucidation of structures of, say, terpenes, steroids, complex lipids, and saccharides is documented in many publications and textbooks [7], [8]. The typical features of such spectra are described, and explanations are given for the characteristic fragmentation patterns. This is still the most widely used application of mass spectrometry, for which there are several reasons. First, most mass spectrometers are low resolution instruments such as quadrupoles. Second, the most common inlet systems or interfacing techniques, namely the direct probe and GC, are compatible with electron impact, giving highly reproducible spectra, largely independent of the type of mass spectrometer. Even the more modern interface for supercritical fluid chroma-

tography or the particle-beam LC/MS interface can be used. Finally, modern library searches or intelligent data handling systems make efficient use of such data and can retrieve much of the information.

20.2.2.2. Elemental Compositions of Ions

The determination of the elements contained in an ion is a major aid in structure analysis. This information could be derived in principle from the isotope pattern of the molecular ion, because the isotope distribution is typical for each element. The ratios of $^{12}C/^{13}C$ and sometimes $^{16}O/^{18}O$ are the most significant parameters as well as the distinctive patterns of sulfur, silicon, the halogens, and metals in the mass range below ca. 1000 amu. Generally, however, the accuracy of the intensity measurement is not sufficient to distinguish between the various possibilities because the differences may be very small. As a rough indication, the height of the $M+1$ signal can be used to estimated the number of carbon atoms, since ca. 1.1 % of carbon is ^{13}C. Only elements such as sulfur, silicon, bromine, and chlorine with typical isotope patterns can be identified easily. For the calculation of the isotope pattern from a known elemental composition, many programs are available and can be very helpful.

Another way of obtaining information about the elemental composition is to accurately determine the mass of an ion. It is then possible to propose possible elemental compositions. These proposals can be used to support suggestions from the low-resolution data and from other means or even to elucidate the structure of a compound. However, with increasing mass (above about $m/z = 500$) the total number of possible elemental compositions for the measurements even under the assumption of very small errors becomes so great that without further information interpretation becomes impossible.

In such cases, information about the unknown compound from other sources is required. Other spectroscopic techniques or chemical evidence may be used, because the knowledge about the absence of certain elements can reduce the number of possible combinations to a manageable level. Generally, ions above $m/z = 1000$ cannot be analyzed in this way, even when the errors are very small, because the number of possible elemental compositions is far too large. There are many successful examples in the literature, mostly involving the structure elucidation of natural compounds. Accurate mass measurement has also been used in GC/MS analyses with a resolution of 10000; in one example the structures of photoadducts of psoralens to nucleosides were determined. Since a portion of the molecules was known (thymine was the nucleoside) it was possible to ascertain the nature of the previously unknown product [9].

20.2.3. Fragmentation in Organic Mass Spectrometry

The formation of fragments is one of the most useful characteristics of organic mass spectrometry. This becomes particularly apparent when no fragments are available, as with soft-ionization techniques, which produce almost exclusively molecular ions with negligible fragmentation. With electron impact, still the most important ionization technique, the majority of molecules form a large number of fragments.

Organic mass spectrometry can be regarded as gas-phase chemistry. Knowledge about the chemical reactions leading to specific fragments is largely based on isotopic labeling studies, often employing metastable ions, linked scans, or MS/MS experiments. These reactions are described by the so-called quasi-equilibrium theory. This statistical theory was formulated originally in 1952 [10] to characterize qualitatively the fragmentation behavior of gas-phase ions by using physical parameters. In Figure 3 a brief summary is given in terms of the so-called Wahrhaftig diagram. A treatment based on thermal reactions gave a similar description, today often referred to as the Rice – Ramsberger – Kassel – Marcus (RRKM) theory [11]. The most important unimolecular reactions can be explained in terms of reaction kinetics and thermochemistry.

More simplified concepts for explaining the fragmentation patterns of complex molecules such as steroids and terpenes are based on charge- or radical-induced reactions ("charge localization"). By considering the stability of intermediate products it is possible to construct reasonable fragmentation pathways, even though the true energetic, kinetic, and thermochemical background is unclear. The greatest problem here is that many rearrangements take place, some of which are well known and discernable [e.g., the McLafferty rearrangement (Fig. 4), retro-Diels – Alder rearrangement, hydrogen shifts], but others are not obvious in the mass spectra (e.g., hydrogen scram-
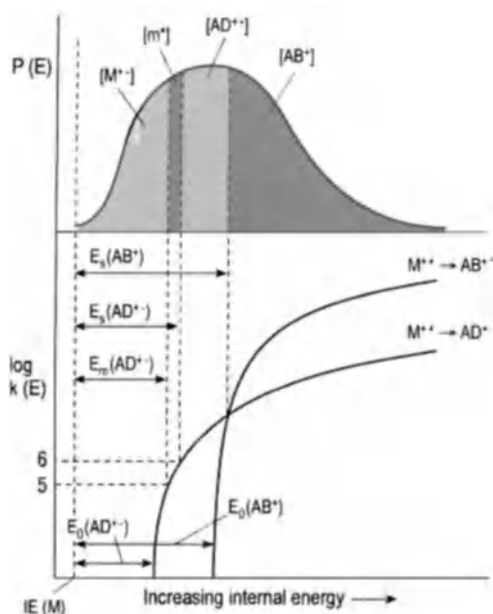
Figure 3. The Wahrhaftig diagram shows the relationship of probability function $P(E)$ (upper half) and two hypothetical functions of the rate constants $k(E)$ (lower half) for unimolecular descompositions of a molecule ABCD$^+$
IE(M) = Ionization energy; E_0(AD$^+\cdot$) = Critical reaction value for the rearrangement M$^+\cdot$ to AD$^+\cdot$; and E_0(AB$^+$) = Critical reaction value for the decomposition M$^+\cdot$ to AB$^+$
For energies between E_m(AD$^+\cdot$) and E_s (AD$^+\cdot$) corresponding to values between log $k = 5$ and 6, metastable ions are formed (lifetimes longer than 10^{-6} s; from then on the rearrangement dominates until the function for the decomposition crosses at E_s(AB$^+$); the formation of the fragment AB$^+$ is now favored. Adapted from [13] with permission

bling, skeleton rearrangements, silyl group shifts), since the mass-to-charge ratio remains the same. Some of them can be studied by using techniques such as isotope labeling.

The large number of rearrangement reactions, controlled by small differences in activation energy, make the prediction of mass spectra and the understanding of the bidirectional relation between spectra and structures impossible. For a thorough discussion of these problems and related topics see [12] and [13].

20.2.4. Quantitative Analysis

Generally stated, mass spectrometers are unstable detectors. The rather high noise level of the baseline and the signal is a problem. The noise can
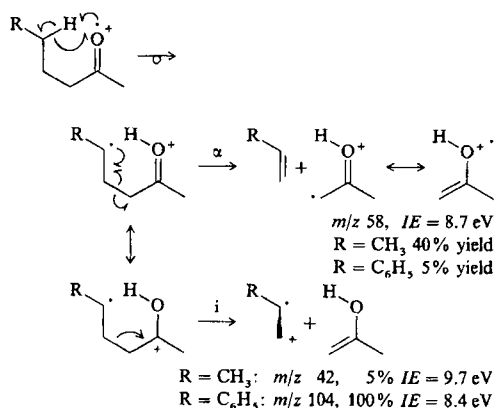
Figure 4. The McLafferty rearrangement with examples for R = methyl or phenyl and the corresponding ionization energies IE for charge retention (α cleavage) and charge migration (inductive fission) [13]

be of chemical origin (i.e., undefined ions throughout the spectrum) or may have its origin in electronic components such as the voltages of the detection system. In the long term, drift of the calibration, contamination of the source or the analyzer, and decreasing efficiency of the detector can slowly degrade the performance of the instrument, resulting in poor reproducibility over long periods. In addition, especially for small signals, the ion statistics may be poor. If N is the number of ions in a peak, than the percentage variation coefficient for the measurement of the area is

$$v = 100 \sqrt{\frac{1}{N}}$$

Even with 1000 ions in the signal, the area can only be measured with a variation coefficient of 1 %. Clearly the only way to improve the ion statistics for a given ion current is to increase the time spent at the top of the peak. This is possible if the associated chromatography allows slow scanning or the use of selected ion recording. For quantitative determinations, often the peak areas are used. If the separation of signals is not complete, peak heights rather than peak areas may give better results, because the overlap should be smaller.

Generally speaking, quantitative measurements require most meticulous attention to obtain high precision and good reproducibility. This requires carefully chosen acquisition conditions and, in most cases, use of internal standards.

Selected Ion Recording. Most mass spectrometers in practical work are scanning type instruments, which screen the mass range for ions by varying the voltage, frequency, or magnetic field. Therefore, a major fraction of the signal remains unrecorded, giving away the sensitivity necessary to obtain signals with a satisfactory number of ions. Techniques are available which circumvent this problem by observing only a few ions or even only one type of ion. This increases the sensitivity by two orders of magnitude but also results in a considerable loss in specificity, since a signal with a single m/z value makes distinction of unknown compounds impossible. Therefore, chromatographic information from the GC or LC is necessary. The technique, termed selected ion recording (SIR) or multiple (single) ion detection (MID), is in general applicable for known compounds with known chromatographic properties. In some cases, the observation of isotopic patterns can improve the certainty of identification and may help to rule out cross contamination. When the mass of one of the ions in the mass spectrum is accurately known, high resolution can be used. This brings back some specificity, although the power of detection of the mass spectrometer is reduced. Often, the improved signal-to-noise ratio due to the reduction of chemical noise more than makes up for the reduction in detection power. In cases where the compound shows strong fragmentation it may be possible to record the metastable transition signal for this reaction. The technique, termed "reaction monitoring," is specific for a given compound.

Internal and External Standards. Quantitative determinations require standards, either external or internal. External standardization using samples with known amounts of analyte ("standard addition") is possible, if a reproducibility and accuracy of ca. 10 % is acceptable. In pharmaceutical studies or routine environmental work such an accuracy is sufficient, since individual variations are much larger. When the required precision must be higher, internal standards must be used. As internal standard a compound with similar chromatographic and, if possible, similar mass spectrometric properties as the analyte can be employed. This may be a homologue, an analogue with, e.g., a different heteroatom, or a positional isomer. The best choice is an isotopically labeled compound containing stable isotopes.

Isotope Dilution. This technique is based on the fact that most elements have several stable isotopes; if only one stable isotope exists, as for phosphorus or fluorine, the technique cannot be used. Let the ratio of two isotopes of an element in the sample be R_s, and the ratio for the same isotopes in the internal standard R_I. From the ratios, the percentage peak heights can be calculated for the sample (s_1 and s_2) and the standard (i_1 and i_2). Then, a known amount I of the standard is added to the sample S and the two signals are measured (Fig. 5). The ratio of the two peaks is then

$$R = \frac{Ss_1 + Ii_1}{Ss_2 + Ii_2}$$

Rearranging for the unknown amount S of the sample gives

$$S = I\frac{i_1 - Ri_2}{Rs_2 - s_1}$$

Isotope dilution for the analysis of organic compounds is often accomplished by deuterium labeling at positions not prone to rapid exchange in the ion source. The synthesis of such compounds is normally straightforward (e.g., by exchange processes). Care must be taken, however, where the mass difference between the labeled and the unlabeled compound must be large because the isotope cluster is broad (several mass units). If many hydrogens are replaced with deuterium, the labeled compound may have a different chromatographic behavior and the retention times can differ. If deuterium labeling is difficult or not possible, ^{13}C, ^{15}N, or ^{18}O enrichment or combinations thereof can be used although the synthesis may be laborious.

20.3. Sample Inlets and Interfaces

20.3.1. Direct Probe

For many years the direct probe for introducing samples into the spectrometer has dominated all the others. The sample is transferred into a small cup with a volume of a few microliters, made of gold, silica, or aluminum. Some of them have a cap with a small hole to exploit the Knudsen effect for controlled evaporation. The cup is placed in the tip of the removable probe. The probe is introduced into the ion source housing with the sam-
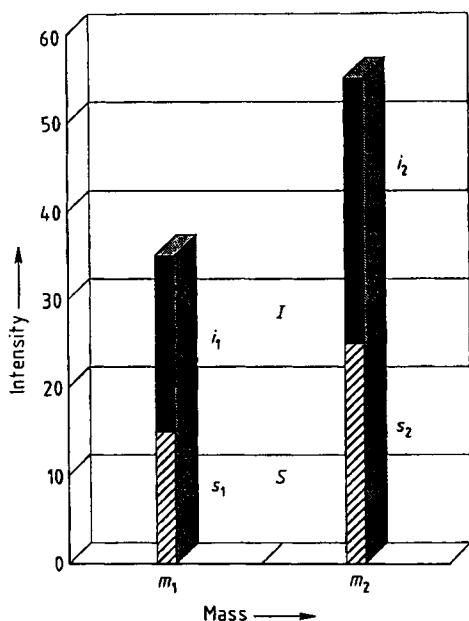
Figure 5. Isotope dilution

ple cup inside the ion volume without interfering with the electron beam. It can be heated and cooled directly under computer control, when necessary as a function of the total ion current or of a single ion current.

20.3.2. Batch Inlets

Batch inlets used to be important for the analysis of gases, volatile liquids, and mixtures such as hydrocarbons in oil samples. The inlet consists of a capillary, which connects a heated reservoir with a volume of several hundred milliliters, normally glass-lined to prevent cracking and pyrolysis, to the ion source. The capillary acts as a throttle, controlling the flow into the ion source; alternatively, a diaphragm can be used. The sample is introduced in the reservoir through a septum by means of a syringe or via a small sample inlet connected to the reservoir. This type of inlet has been largely replaced by GC/MS interfaces.

20.3.3. Pyrolysis

Pyrolysis is a powerful tool for the analysis of very high molecular mass samples, where there is no possibility for separating mixtures and deter-

mining the components or where the compounds are too complex to give mass spectra [14]. Examples are the attempts to characterize different fractions of humic acids, strains of bacteria, and unusual components in intact DNA of a variety of animals. In all such cases, mixtures of smaller molecules are prepared by carefully controlled pyrolysis (e.g., Curie point pyrolysis). The distribution of the resulting mixture of thermally stable compounds is characteristic of the starting material. Since the identity of the pyrolytic decomposition products is not known in most instances and the chemistry involved is complex, only the appearance of the particular distribution is significant, and techniques for cluster analysis based on pattern recognition must be used. Field desorption, field ionization, direct chemical ionization, and low-energy EI are mainly used with pyrolysis, because molecular ions are mostly formed and the pattern is not further complicated by fragmentation in the ion source. The desorption process in FDMS (see Section 20.4.8) can be considered a pyrolysis technique. Often, the pyrolysis products are injected into a GC, separated, and soft ionization techniques such as FIMS, low energy EI, or photoionization [15] are used to detect the components of the mixture.

20.3.4. GC/MS Interfaces

The most successful hyphenated method continues to be GC/MS, generally with EI ionization, but particularly the inductively coupled plasma MS (ICPMS) has become a successful detector as well. In the form of benchtop instruments, this combination is found in the chromatography laboratory rather than in mass spectrometry laboratories. Instruments such as mass selective detectors and ion traps are dedicated devices for gas chromatography and have developed into robust, dependable instruments with high detection power and selectivity. With modern computerized instruments and the extensive storage capabilities of data systems, series of spectra can be acquired rapidly, and elaborate data extraction procedures have been devised for the identification of important spectra with relevant data.

The reasons for this success are numerous: Gas chromatography is a well-developed technique with high separation power and excellent stability, and its specificity can easily be adjusted to analytical problems (see → Gas Chromatography). Further, the gas flow in capillary GC is a reason-

able match to the pumping requirements in the mass spectrometer. The GC/MS interface has now been developed to such an extent that compromises are no longer necessary, allowing both to be optimized separately. Today, the mass spectrometer can be regarded as a GC detector with high detection power and specificity. Compared with flame ionization or electron capture detectors, for example, it has many advantages, and the disadvantages such as higher expenses and lack of long-term stability have become less significant.

Separators. As long as packed columns with high gas-flow rates were used, it was necessary to separate the carrier gas from the analytes. This was accomplished by a variety of devices called separators; examples include the Biemann – Watson separator and the jet separator. The latter was subsequently used in LC/MS interfaces. Due to the replacement of packed GC columns by capillary columns, separators are no longer used; they are described in detail in [16].

Direct Coupling. In direct coupling, the GC capillary is introduced directly into the ion source. This technique has the advantage of simplicity in construction and it avoids losses of analyte. The disadvantage is that the chromatographic separation is influenced by the vacuum of the ion source. Furthermore, the flow of gas into the ion source is a function of the GC oven temperature. The ion source pressure and the sensitivity may vary with the temperature program. Recently, this problem has been addressed using flow and temperature programming simultaneously. It is therefore not straightforward to transfer a separation developed for GC alone to the GC/MS system, because the parameters may be vastly different. But the simplicity of construction makes this version very popular.

Open Coupling. In open coupling, the end of the capillary from the chromatograph is inserted into a small glass tee piece and the interface capillary, with a slightly wider inner diameter, is located opposite to it, separated by a small gap. Through the third entrance of the tee piece, an additional flow of inert gas (the make-up gas) protects the gap from air and adds to the flow into the spectrometer. If the dimensions are right, the complete effluate from the separation column, together with some make-up gas, enters the ion source and no loss of analyte occurs. Since the

end of the GC capillary remains at atmospheric pressure, the separation integrity is maintained and analytical procedures can be transferred.

Reaction Interfaces. For special purposes it can be advantageous to add a reaction zone between the GC and the MS to convert the analytes into species that give characteristic mass spectra or more precise quantitative information. Examples are the hydrogenation of double bonds in unsaturated hydrocarbons and the combustion of organic materials to CO_2 for the accurate determination of the $^{13}C/^{12}C$ ratio. Generally, the analyte is allowed to contact a reactive surface or catalyst, or a gaseous reagent is mixed with the gas stream from the GC. The analyzer can be any conventional sector or quadrupole spectrometer. For accurate isotope analysis, specially designed versions can be used. Since this method is important in analytical applications such as the determination of the origin of sugars in wine, dedicated instruments are commercially available. For a recent review with well chosen instructive examples, see reference [17].

20.3.5. LC/MS Interfaces

LC/MS interfaces can be divided into three groups: the first comprises sputtering and desorption processes, the second group has an aerosol generating step and separate ionization by EI or CI, and the third group — thermospray and electrospray — has an integrated ionization/aerosol generation step. The areas of applications for the three groups are: natural products for the first; smaller molecules, metabolites, and industrial compounds for the second; and biological oligomers and polymers for the third. Recent years have seen dramatic developments towards very stable and sensitive interfaces based on electrospray and atmospheric pressure ionization, whereas all others, except for the particle beam interface, have disappeared almost completely. The latter is still in use in industrial laboratories and for environmental studies where a detector for less polar compounds, is required.

20.3.5.1. Moving Belt

The moving belt system, invented by W. H. McFADDEN, was one of the first LC/MS interfaces. The stream of eluent from the LC is directed onto the surface of a belt made of pure

nickel or polyimide, which runs through a series of vacuum stages into the ion source of the mass spectrometer. To minimize memory effects (signals from not completely removed compounds) and to clean the belt, infrared heaters and washing steps are included. The analyte was transferred into the source by thermal desorption using direct or indirect heating, by SIMS, FAB, or laser desorption. However, the technical requirements and the rather high costs (additional vacuum systems, maintenance of the belt), and the high background and memory effects limited the number of users from the beginning. The interface has the advantage of a complete separation of chromatography and mass spectrometry, allowing a wide range of chromatographic techniques (HPLC, micro HPLC, SFC) to be used.

20.3.5.2. Continuous Flow FAB

The continuous-flow FAB interface (online FAB, frit FAB) was developed by R. CAPRIOLI et al. [18] and Y. ITO et al. [19]. The technique, based on the conventional FAB source (see Section 20.4.8.3), is straightforward and is compatible with many mass spectrometers. The exit of a capillary, which is coupled to a micro HPLC system, is connected to an FAB target and the effluent is sputtered from the target as usual. Since the gas load to the ion source is high due to the evaporating effluent, glycerol, and the FAB gas, the sensitivity is enhanced by adding high pumping capability to the source housing using an additional cryopump. Glycerol is used to enhance the desorption, as with FAB; it can be added directly to the eluent (1–5 %) or after separation, by means of concentric capillaries   [20]. Flow rates of 15 µL/min are acceptable, but the performance is better with lower rates. Recently, capillary zone electrophoresis (CZE) has been interfaced [21], [22], although serious problems arise due to the very small amounts of sample and narrow peaks from the electrophoretic separation. Mass spectrometers with simultaneous detectors compare very favorably [23] (see Section 20.8.4).

20.3.5.3. Direct Liquid Introduction

Direct introduction of liquids was probably the first LC/MS interface to be used in modern organic mass spectrometry [24], [25]. A flow of typically 10–20 µL/min is allowed to enter the ion source without separation of eluent and analyte. The construction is relatively simple, and

conventional electron impact ion sources can be used, even without modifying the pumping systems [26]. The main disadvantage is the inflexibility with respect to the ionizing method (for CI only the eluent is available as reactant gas), while the compounds, that can be analyzed are similar to those amenable to GC/MS.

20.3.5.4. Supercritical Fluid Interface

The SFC/MS interface is a direct-introduction system with careful temperature control up to the very tip of the capillary [27]. The end of the capillary contains a frit or a diaphragm that keeps the phase supercritical until it leaves the capillary and enters the ion source. Otherwise, the dissolved analyte would clog the end of the capillary. Since the gas load is rather high, chemical ionization is normally used with capillary SFC, although EI is possible. When packed columns are used, momentum separators can reduce the pressure in the ion source, making electron impact possible. Limitations of the method with respect to classes of compounds are mainly due to SFC itself, not to the interface or the MS.

20.3.5.5. Particle Beam Interface

The particle beam interface (Fig. 6) was created under the acronym MAGIC (monodisperse aerosol generator interface for chromatography) [28]. Now, the aerosol is produced by a variety of means (with auxiliary gas, thermospray, or ultrasonic nebulizers) at atmospheric pressure and a uniform distribution of the droplets results in particles of a narrow size distribution, which can be handled more efficiently by the separator. The droplets are dried to particles in a heated expansion chamber, and a momentum separator isolates the particles from the gas. In the source, the particles are destroyed by impact and the sample is released and ionized by using EI, CI, or even FAB. The appearance of the EI spectra is almost identical to conventional EI spectra obtained by direct probe or GC/MS. Therefore, library searches are possible, which is the major advantage of this interface.

The main problem of the technique is its unsatisfactory basic power of detection. Also, sensitivity can vary widely even for similar compounds, and the difficulties in obtaining a linear response over a wider range of concentrations make quantitative determinations difficult. In addition, components eluting at the same time may interfere
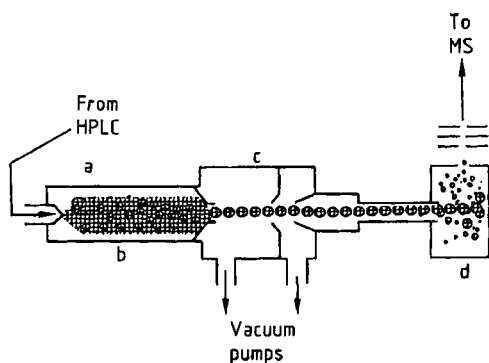
Figure 6. Particle beam interface
a) Nebulizer; b) Desolvation chamber; c) Momentum separator; d) Ion source

with one another, leading to unpredictable effects. For applications where the sensitivity is not of importance (e.g., monitoring of synthesis), this technique can be employed with success, because typical, interpretable mass spectra can be obtained.

20.3.5.6. Chemical Ionization at Atmospheric Pressure

Atmospheric-pressure chemical ionization (APCI) was described by HORNING et al. in 1960 [29], and they recognized the potential of this approach for LC/MS applications. An aerosol is generated by spraying an LC eluate within a heated sheath gas flow to give a spray stable at atmospheric pressure. In the spray, ions are generated via a chemical ionization plasma, created by a corona discharge from a needle. The analyte is ionized by the reactant ions in the plasma and subsequently transferred into the mass spectrometer. This is achieved by a sampler – skimmer system with an intermediate pumping stage to separate the high vacuum from the atmosphere. A curtain of drying gas is used in front of the sampler to reduce the background ions.

In general, basic compounds with high proton affinity can be detected with the highest limit of detection [30], and in some cases, a strong temperature dependence has been observed [31]. In principle, APCI is compatible with every separation technique which allows the effluent to be sprayed in an aerosol. Thus, not only HPLC but also SFC can be interfaced to this ion source.

20.3.5.7. Thermospray

Thermospray, developed in 1980 by M. VESTAL [32] rapidly found its way into routine applications, but now the atmospheric pressure interfaces have replaced it completely. The eluent of an LC is rapidly heated in a steel capillary either by direct resistive heating or by indirect heating using a cartridge system. A fine aerosol of charged droplets is generated in the ion source at reduced pressure, although the details of the formation of ions in the source is still subject to discussion [33] – [36]. The ions are transferred into the mass spectrometer by a sampling cone reaching into the center of the inner chamber of the ion source, but most of the solvent is rapidly pumped away. The source has to be very tight to stabilize the pressure inside and the vacuum outside. The method is shown schematically in Figure 7 [37].

It is possible to include a Townsend discharge or a filament and a repeller electrode to allow plasma chemical ionization to be performed. For some compounds, the sensitivity, selectivity, and fragmentation behavior can be modified. The advantage of the technique is the simplicity of transferring previously developed LC procedures since the normal flow rates and some of the common buffers (e.g., ammonium acetate, formate) can be used.

20.3.5.8. ESI Interface

The combination of the electrospray ion source with HPLC has without a doubt become "the" LC/MS interface in recent years. It is a particularly powerful combination, since this ionization technique covers a wide range of samples [38] that are commonly separated by HPLC [39] or electrophoresis [40] – [44]. The ESI source exhibits concentration-dependent behavior and thus gives optimal signals at most flow rates. The principle of the ionization process is discussed in Section 20.4.9. The most important feature of this interface is aspray needle which can be connected directly to the separation column, if the flow rates are compatible. Initially the major limitation was that only low flow rates (a few µL/min) could be used, but now flow rates of 1 mL/min or more are possible by using heated sprayers or ultrasonic devices. Splitting of the flow is possible as well, allowing two detectors to be used simultaneously. Since buffers can be used as long as they are volatile and not too concentrated, a sheath flow
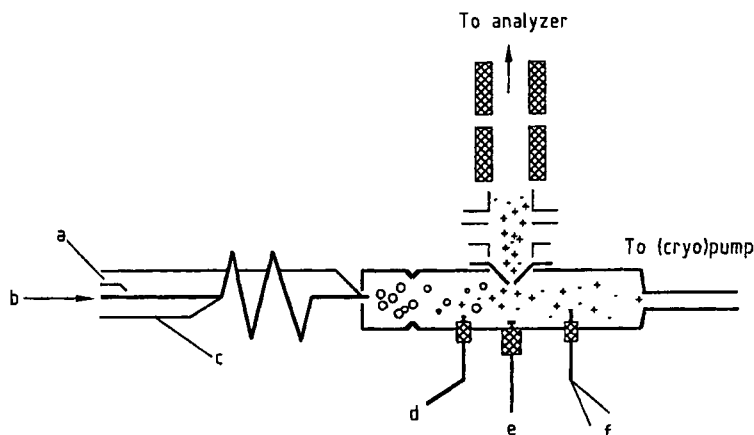
Figure 7. Thermospray interface/ion source
a) Connections for direct heating: b) HPLC flow: c) Thermocouple: d) Townsend discharge: e) Repeller: f) Thermocouple (aerosol)

of methanol can be added to enhance the ion yield and reduce background due to buffer salt clusters. Recently, interfacing with capillary zone electrophoresis (CZE) has been demonstrated [45] and this combination appears to be a powerful tool for the analysis of trace organic compounds [46], although the amount of analyte which can be separated in a CZE capillary and transferred to the mass spectrometer is very small and may even lie below the detection capabilities of the mass spectrometer [44], [47] – [50]. The near future will see devices used in many fields based on a chip with separation by CE or HPLC and an integrated sprayer, especially in biochemical and clinical analysis [51], [52]. In Figure 8 such a device is shown (a), and a photograph of the Taylor cone developed from the chip (b).

20.3.6. TLC/MS

Two-dimensional separations such as thin layer chromatography (TLC) or gel electrophoresis in combination with SIMS [53], FAB [54], and laser desorption [55] have recently given impressive results and are promising for the future [56]. In principle the technique is identical to FAB or laser desorption, with the crucial difference being that the target is the device used for the separation of the sample mixture. There are problems associated with this difference: the sample may be rather dilute, not in an appropriate matrix and, because the layers are rather thick, much of the sample is protected from the sputtering beam. It is

therefore necessary to optimize the TLC procedure, e.g., by using micro rough surfaces for the separation or blotting techniques to transfer the analyte to a different matrix. To allow the scanning of larger TLC plates, the incorporation of sample stages with x,y-adjustments would be necessary. The major advantage of the technique is that method development of TLC methods is rapid, efficient, and cost effective, and the plates can be investigated with several spectroscopic techniques sequentially after separation, making this setup very attractive for routine analysis (e.g., in environmental chemistry) [57], [58]. For biochemistry, matrix assisted laser desorption of peptides, proteins, and oligonucleotides separated in thin-layer electrophoresis and often transferred to a more suitable matrix by blotting techniques appears promising, e.g., for screening purposes and the identification of modifications and mutants [59] – [62], even though excision of the spots and off line analysis is often favored. This strategy allows direct enzymatic digest ion in the gel avoiding sample loss in the course of transfers.

20.4. Ion Generation

20.4.1. Electron Impact

Electron impact is still the most widely used technique for generating ions. It uses electrons, emitted from a rhenium or tungsten wire or ribbon and accelerated with typically 70 V, to ionize the
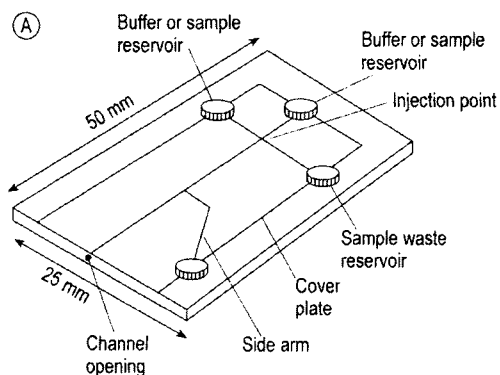
Figure 8. (A) Schematic diagram of microchip used to pump fluids electroosmotically and to generate the electrospray; (B) Photomicrograph of the Taylor cone and electrospray generated from a 60 % H_2O – 40 % methanol solution that was electroosmotically pumped using a microchip (reproduced from RAMSEY [51])

vapor of a compound. In this process, molecular ions are formed together with a pair of electrons after impact. The molecular ion may be stable or it may dissociate immediately into a neutral and an ionic species (dissociative ionization) or into two ions of opposite charge (ion-pair formation). The positive ions may have sufficient internal energy to undergo further fragmentation. Generally, the fragments are formed in the ion source within a few microseconds and are responsible for the normal mass spectrum. If ions are stable enough to dissociate later in one of the field-free regions between analyzers, they appear as "metastable ions" or may be detected by using MS/MS experiments. These metastable ions have been studied in great detail, since they facilitate the understanding of many gas-phase processes. The EI ion source is a small chamber with a filament (cathode), an electron trap to regulate the emission, and a pusher or repeller, which controls the space charge distribution in the chamber. Small permanent magnets perpendicular to the electron beam broaden the electron cloud, which enhances the sensitivity. The ions are extracted by means of a set of extraction plates and focused with several lenses and steering plates onto the entrance of the analyzer.

20.4.2. Chemical Ionization

Chemical ionization can be carried out either at atmospheric pressure or under reduced pressure in an ion source chamber; in both cases ion/molecule reactions are used to produce the ions of the analyte [63]. At atmospheric pressure, the plasma is generated with a discharge needle and the ions come mainly from the surrounding air, water, and from components of the analyte spray, which can contain organic solvents and buffers. Mainly protonation occurs. Under reduced pressure, a reactant gas must be added to the mixture. For this purpose, a gas such as methane, isobutane, ammonia, or water is introduced into a gas-tight ion source. This gas is then ionized by electron impact (or, sometimes, by a Townsend discharge) to give a reactive plasma. The nature of the plasma ions determines which type of spectrum is formed; therefore, the different gases can be used to control the specificity of the ionization process. The reactions most often observed are proton transfer from the reagent ion to the analyte, charge exchange, electrophilic addition of a plasma ion to the analyte, and abstraction of a negative leaving group from the analyte. When ammonia is used, only substances with high proton affinity are protonated, and hydrocarbons, for example, do not show up in the mass spectra. With CO_2 as reagent gas, the spectrum resembles a normal EI spectrum due to the high energy associated with the electron transfer in the CI plasma. The important feature of chemical ionization with respect to analytical applications is that the ratio of molecular ions to fragments is usually much higher than with EI, thus yielding information about the molecular mass.

20.4.3. Negative Chemical Ionization (Electron Capture)

Negative chemical ionization is similar to CI only in so far as the same gases are used in the same ion source, although the pressure is usually
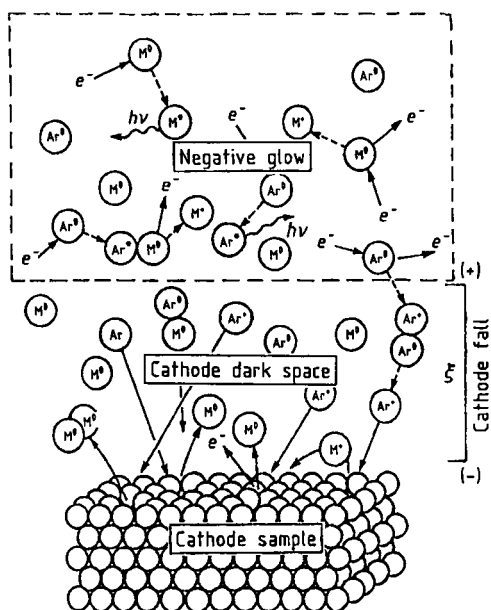
Figure 9. Principles of GDMS illustrating the main processes of sputtering, excitation, and ionization
M^0 = Sample atom; M*/Ar* = Excited states; M^+ = Sample atom
Only neutrals can cross into the negative glow, where ions are formed via electron impact or in collisions with excited argon molecules (Reproduced with permission from [68])

lower than in CI. The purpose of the gas is to moderate the energy of the electrons emitted from the filament to thermal energies. With such slow electrons, electron capture becomes possible, resulting in the formation of negative ions. If the analyte can stabilize the negative charge either in electronegative moieties, by charge delocalization, or by dissociation to give a stable fragment, the cross section of the ionizing process can be high, and the technique very sensitive. For the analysis of substances unable to accommodate negative charge, special derivatization techniques such as electrophoric labeling are used. An example is the derivatization of adducts of carcinogenic metabolites to nucleosides with perfluorobenzyl bromide yielding GC/MS detection limits in the attomole range [64].

20.4.4. Field Ionization (FI)

Field ionization, developed by H. D. BECKEY [65], was of importance only for a short time. In the ion source, a strong electric field gradient (10^7 V/cm) is formed by means of a potential ap-

plied to a sharp edge, a tip, or an emitter with whiskers; the potential difference between the acceleration voltage and field potential is ca. 11 kV. In the gradient, the vaporized molecules of the analytes travel to the regions of increasing field strength. When the probability of an electron leaving the molecule by tunneling through the internal energy barrier becomes sufficiently high, ions can be observed. The ionization process is very soft. Since the technique is technically difficult and needs skilled operators, only a few laboratories have used FI for practical work. Its strength is that only molecular ions are formed, allowing, for example, the distribution analysis of a series of homologous hydrocarbons in oil samples. A special version of the technique, called field ion kinetics, has been used for some time to study the very fast fragmentation processes of several classes of compounds [65].

20.4.5. Plasma Ionization

Plasma ionization can, in general, only be applied to the analysis of elements, since the energy of the plasma is too high for most molecules to survive. However, molecules formed from the plasma gas and the analytes in the plasma are one of the main obstacles of this ion source type. Interferences in the lower mass regions — others come from doubly charged ions and oxides — complicate the spectra and may even obscure some elements completely if only trace amounts are present. High-resolution mass spectrometers can be used to solve this problem. Today, there are three major types of plasma source in use: the glow discharge source, the inductively coupled plasma source, and sputtered neutral mass spectrometry [66].

Glow Discharge (GDMS). Glow discharge ion sources have been built for quadrupole and sector field instruments. The solid sample is part of the source chamber and operates as cathode. The source chamber serves as anode and is filled with noble gas (generally argon) at a reduced pressure (1 kPa) that allows a stable plasma to be formed. A high voltage is applied between cathode and anode resulting in a self-sustaining discharge. Two main regions exist in the plasma: the cathode fall next to the sample and the negative glow (Fig. 9).

In the cathode fall, argon ions are accelerated toward the sample surface and atoms from the sample are sputtered into the plasma. Upon elec-

Table 1. Some ICPMS interferences and the resolution required to resolve the signals

Analyte species	Exact mass	Interfering species	Exact mass	Resolution required for separation
^{55}Mn	54.938	^{40}Ar^{14}N^{1}H	55.090895	359
^{28}Si	27.976929	^{14}N$_2$	28.006146	960
^{31}P	30.9737634	^{14}N^{16}O^{1}H	31.005813	966
^{48}Ti	47.9479467	^{32}S^{16}O	47.986986	1228
^{44}Ca	43.95549	^{12}C^{16}O^{16}O	43.98982	1280
^{32}S	31.9720727	^{16}O$_2$	31.989828	1802
^{56}Fe	55.934938	^{40}Ar^{16}O	55.957299	2504
^{51}V	50.944	^{35}Cl^{16}O	50.96376	2580
^{62}Ni	61.9283	^{46}Ti^{16}O	61.94754	3220
^{75}As	74.921596	^{40}Ar^{35}Cl	74.9312358	7771
^{80}Se	79.91647	^{40}Ar$_2$	79.92477	9638
^{48}Ti	47.94795	^{48}Ca	47.95253	10466
^{58}Ni	57.9353	^{58}Fe	57.9333	29000
^{115}In	114.903871	^{115}Sn	114.903346	218900
^{87}Rb	86.90916	^{87}Sr	86.908892	325000

tron impact, the sputtered atoms are ionized and extracted into the mass spectrometer by application of a potential [67]. The technique is considered to suffer only minor matrix effects thus giving reliable quantitative results. Since the sample is one of the discharge electrodes, the thermal and electrical conductivity of the sample may influence the plasma and the sputtering process. For the analysis of nonconducting materials it may be necessary to mix the sample with a good conductor.

Sputtered Neutral MS. In sputtered neutral MS (SNMS) an electron cyclotron resonance in a static magnetic field is used. Under these conditions, an inductively coupled RF plasma is formed in high vacuum with argon as plasma gas; since no electrodes are involved the electrical properties of the sample do not influence the plasma, allowing the analysis of metals, particulate materials, and ceramics. The sample is connected to an adjustable high voltage, and argon ions from the plasma are accelerated onto the surface of the sample. The intensity of the sputtering process can be varied by means of the high voltage and thus the depth resolution obtainable can be optimized for the analytical problem. Since only atoms can penetrate into the potential — sputtered ions being reflected — the electron impact ionization step is virtually independent of the sputtering process and the technique is therefore free of matrix effects. In addition, ions formed at plasma potential are suppressed by an energy filter. Therefore, the technique gives reliable quantitative results when calibrated with standard reference materials. The best

resolution that has been achieved is 4 nm. Since the sputtering process is slow, the usable range for maximum depth is in the micrometer range. Consequently, the technique has some attraction for the analysis of semi- and superconductors [69].

Inductively coupled plasma mass spectrometry (ICPMS) is very sensitive and allows the analysis of solutions of metal ions at pg/mL levels [66]. In principle, the ion source consists of an aerosol generator, a desolvation unit, the inductively coupled plasma, and the ion transfer and focusing optics. Several types of aerosol generator, such as the Babington and Meinhard types, ultrasonic nebulizers, and thermospray are used [70] . For interfacing ICPMS to LC/MS, efficient drying of the aerosol is necessary to stabilize the plasma. The most effective method appears to be heating and subsequent cooling of the aerosol, sometimes supported by a counterflow of gas. The ions are generated in an argon plasma, which burns in the high-frequency field of the induction coil just in front of the transfer optics. Since this all takes place at atmospheric pressure, the sampling into the mass spectrometer by means of a skimmer sampler arrangement, similar to that of other API sources, requires very effective pumping to ensure sufficient sensitivity, and the outer skimmer must withstand the temperature of the plasma. ICPMS instruments are available with quadrupole and sector field analyzers [71] ; the latter is of interest, when higher resolution (typically ca. 7000–10000) is necessary to separate molecular interferences due to oxides and argides [66], [72]–[74] (see Table 1 and Fig. 10).
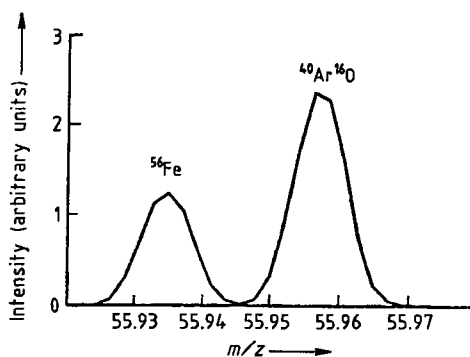
Figure 10. The spectrum shows the ^{56}Fe and ^{40}Ar^{16}O, resolved with a resolution of $R = 3000$, using a laboratory prototype of the ELEMENT, Finnigan MAT (high resolution ICPMS) [75]

20.4.6. Thermal Ionization

Thermal ionization is a technique for elemental analysis that uses thermal energy to ionize elements. The sample is deposited on a ribbon which can be heated to very high temperature, and a second ribbon a few millimeters from the first is used to ionize the atoms in the gas phase. This ribbon is made from rhenium or tungsten, sometimes treated with carbon or thorium, depending on the nature of the element under investigation. Heavy metals such as lead, uranium, thorium, and others have been determined as positive ions in very different matrices such as geological samples, microelectronic chips, and ice cores from the antarctic, in conjunction with isotope dilution (IDMS) techniques. Iodine, bromine, chlorine, selenium, and also some heavy metals have been determined by using negative thermal ionization and IDMS [76]. Thermal ionization with isotope dilution is considered one of the reference techniques allowing the most precise quantitative determination of elements.

20.4.7. Optical Methods

Photo-Ionization. Photo-ionization has been used mainly to study basic principles of the ionization process. The light source is usually a helium lamp emitting 21.22 eV photons; other noble gases can be used to produce other photon energies [77]. When the energy deposited in the molecules is sufficient to exceed the first ionization potential, ions of the analyte are formed. By analyzing the energy distribution of the photo-

electrons, the energy distribution of the molecular ions can be deduced. It has been shown that the distribution formed by electron impact may be similar.

Multiphoton Ionization. The specificity of multiphoton ionization (MUPI, sometimes termed resonance multiphoton ionization, REMPI) is based on the existence of well-defined electronic states at low temperature [78], [79]. If a molecule is cooled to a few kelvin, a first photon can be absorbed upon resonance and excites the electrons to a state below the ionization potential. This absorption process is specific for a compound or class of compounds; for example, isomeric methylphenols can be distinguished with such a technique. A second photon is necessary to excite the molecules above the ionization threshold. By varying the laser power, the fragmentation can be controlled (see the example given in [80]). The more energy is available the greater the fragmentation.

In the ion source, a supersonic beam of noble gas is generated and the sample is seeded into it (e.g., by laser desorption). Together with the beam, the sample is carried through a skimmer system that allows only the isentropic core to enter the vacuum chamber where the ions are generated by two tunable lasers. The technique is still in the development stage.

20.4.8. Desorption Methods

Desorption methods have been developed since around 1970. They have not only expanded the scope of mass spectrometry, but have also changed the strategies of sample introduction and brought attention to solid and liquid phase chemistry as well as to particle and gas phase chemistry. They have also stimulated the rapid development of new ion sources, mass analyzers, and detectors with higher mass range capabilities and improved sensitivity.

20.4.8.1. Secondary Ion Mass Spectrometer (SIMS)

Secondary ion mass spectrometry, pioneered by A. BENNINGHOVEN et al. [81], can be regarded as one of the first desorption techniques. A primary ion beam (typically ions as Cs$^+$ and O$^-$) is focused onto a target; the sample can be deposited on the surface of the target or the surface is the
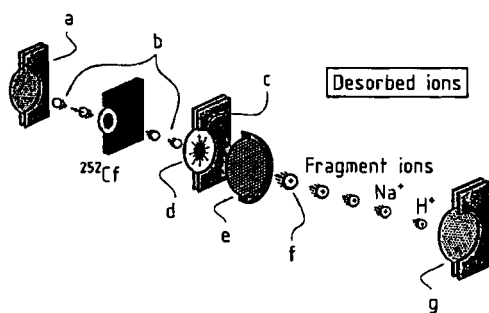
Figure 11. Plasma desorption MS
a) Fission fragment detector; b) Fission fragments; c) Excitation area (100 nm²); d) Sample foil; e) Acceleration grid; f) Molecular ion; g) Desorbed ions detector

sample itself. Upon impact, the sample is sputtered and ionized simultaneously, and the secondary ions are extracted into a time-of-flight instrument. Since the primary beam can be focused to a small spot, the technique can be used in principle to achieve spatial resolution as well (ion micro probe) [82]. SIMS is used in material science to analyze the surface of metals, but due to the combination of the sputtering and ionization processes, matrix effects can be strong and atomic mixing occurs. Although this approach is used primarily for the determination of elements, the investigation of organic surfaces is also possible [83], and organic materials deposited on a appropriate surface such as silver have been studied [84].

20.4.8.2. Field Desorption (FD)

Field desorption is technically closely related to field ionization, although the desorption/ionization process is different [85]. The sample is deposited directly on the emitter to produce the high field gradient and, therefore, materials with virtually no vapor pressure can be investigated, in contrast to field ionization. The emitter can be heated either by direct current or by using external sources (e.g., lasers). The sample is generally desorbed as so-called "preformed ions" (protonated or with alkali cations attached, both or either type are always present in samples). The ion currents are generally very weak and often erratic; therefore, integrating devices (photoplates, multichannel analyzers) have been used to overcome these problems. Since FDMS and FIMS are similar methods, they have the same technical problems. FDMS was the first desorption technique to allow

mass spectrometric analysis of peptides, oligonucleotides, and polysaccharides. Today the technique is used for the analysis of industrial polymers.

20.4.8.3. Fast Atom Bombardment (FAB)

Fast atom bombardment, developed by M. BARBER in 1980 [86], was the first technique which could be used with virtually every mass spectrometer, since the ion source can be retrofitted to a given mass spectrometer without problems. Furthermore, sample preparation is straightforward. The technique is based on SIMS, with the major difference being that the sample on the target is dissolved in a liquid matrix, usually glycerol. In the meantime, many other matrices have been examined for the analysis of different classes of compounds. The most widely used are glycerol, thioglycerol, "magic bullet" (a 5:1 mixture of dithiothreitol/dithioerythritol), 3-nitrobenzyl alcohol, 2-nitrophenyl octyl ether and triethanolamine. It can be beneficial to add acid (e.g., phosphoric acid, sulfonic acids) to glycerol to increase the abundance of analyte ions. The second difference to SIMS, the use of accelerated atoms instead of ions, is of importance only in so far as an atom gun can be added to the source housing of even sector field instruments. Because atoms can penetrate the high electric field but cannot be focused on a small area, the atom gun has often been replaced by a high-energy ion gun in a technique known as liquid SIMS (LSIMS). The advantage is that the primary ions can be focused and deliver a higher ion density in comparison to the saddle field atom guns commonly used with FAB MS.

20.4.8.4. ²⁵²Cf Plasma Desorption

Plasma desorption (PD), first described as a tool for protein analysis in 1976 [87], exploits the fact that ^{252}Cf "explodes" spontaneously into smaller nuclides, the lighter ones centered around ^{108}Tc, and the heavier ones around ^{144}Cs. The products from each unsymmetric fission contain most of the high energy as kinetic energy and travel in opposite directions. The californium source is located between a thin foil loaded with the sample and a target, which serves to start the clock of a detector in a time-of-flight mass spectrometer [88] (see Fig. 11). When the heavy atom hits the foil, a cloud of atoms and neutrals, including intact large molecules, are desorbed in a very short pulse. The ions observed are preformed

ions (protonated or with alkali metal ions), which are either desorbed directly or formed in the dense cloud of desorbed material.

The technique gives poor resolution, but good mass accuracy and high detection capability, although strong background problems must be solved. The time required to record a spectrum depends on several factors (sample preparation, amount of ^{252}Cf available) and can vary from a few minutes to several hours. The technique has already experienced the same fate as FDMS. When it was developed, it was highly praised for its unprecedented high molecular mass determination capabilities, its sensitivity, and simplicity. The results for peptides [89] and other high mass polymers were equally impressive. This type of work is now largely done by matrix-assisted laser desorption and electrospray.

20.4.8.5. Laser Desorption/Ionization

Lasers have been used in mass spectrometry for many years. Trace elements in biological samples [90] can be determined by using laser microprobes (LAMMA, laser microprobe mass analyzer) or a combination of laser ablation with ICPMS. For the analysis of bulk materials, techniques such as resonance ionization mass spectrometry (RIMS) and laser ablation MS (LAMS) are employed; for a review see [91].

The desorption of organic substances from surfaces without destruction is difficult and the duration of ion production is short. Therefore, the mass limitation of this approach without a matrix is around 1000 amu [92]. Recently, molecular ions of complete proteins with $M_r > 200\,000$ [93] have been generated, which is mainly due to the use of appropriate matrices for matrix-assisted laser desorption/ionization [94]. For sample preparation, the analytes are dissolved and crystallized in a matrix on a target. The matrix for the first experiment using a UV laser for protein analysis was nicotinic acid [95], but the search for other matrices is continuing, since the nature of the matrix seems to be crucial. A similar approach with a metal powder (10 nm diameter) mixed with a liquid matrix (glycerol) to form a colloid gave comparable results [96]. The crucial point is the absorption of the matrix, which protects the large molecules from the laser energy. Apparently, a variety of ions and radicals are formed initially from the matrix by photoionization and, in the dense expanding plume of desorbed material, protons become attached to the analyte molecules.

These ions are then extracted in the TOF analyzer. The first lasers used were UV lasers tuned to the absorption of nicotinic acid, but later it was shown that IR lasers give similar spectra [97]. Due to the pulsed ion generation, TOF instruments are the analyzers of choice, although recent attempts have used sector field instruments with simultaneous detectors. The signals obtained are rather broad, probably due to matrix – molecule adducts of the analytes, but the detection potential is very good. Since the accuracy of the mass measurement is ca. $1 \times 10^{-3} - 5 \times 10^{-3}$, yielding an error of ca. $100 - 500$ amu at $M_r = 100\,000$, the precision of the technique is superior to other methods such as gel chromatography or centrifugation. Today, benchtop instruments are available with easy sample handling and short analysis times, making the technique suitable for biochemical laboratories. The development of delayed extraction has improved the resolution considerably [98] – [100] and with post-source decay [101] – [103] the detection of fragments has also become possible.

20.4.9. Electrospray

Electrospray was described in 1968 by DOLE et al. [104] and the ions were detected by ion mobility. FENN and coworkers [105], [106], and, independently, ALEXANDROV et al. [107] re-adopted the idea as a means of creating ions for mass spectrometry and since then rapid development of this technique has occurred. The fundamental difference to thermospray is that the spray is formed by charging droplets to such an extent that they explode by coulomb repulsion into smaller and smaller droplets. The droplets are further decreased in size by collisions with gas and, finally, highly charged ions are liberated. The process in which ions are created from the droplets is still not clear and several partial models exist [108].

The electrospray source is shown schematically in Figure 12. A capillary is held under high voltage (3 – 5 kV) at atmospheric pressure to generate a spray of charged droplets. To obtain a dry aerosol, either a heated steel capillary [109] or a hot gas curtain [110] is used in conjunction with collisions in the first vacuum stage. These collisions are necessary to decluster the ionic species, but may also be used to make fragments in a process equivalent to collision-induced dissociations for MS/MS experiments.
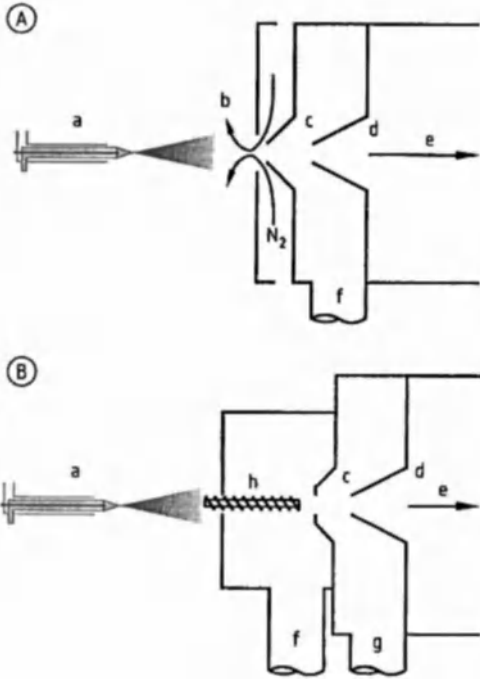
Figure 12. Electrospray ion source types
A) Desolvation region with gas curtain; B) Desolvation region with heated capillary
a) Sprayer with three concentric capillaries: the inner flow is the eluent from the separation (CZE, HPLC), the next flow is sheath liquid such as methanol or isopropanol to enhance desolvation, and the outer flow is sheath gas, normally nitrogen; b) Gas curtain; c) Nozzle; d) Skimmer; e) To ion optics and analyzer; f) To roughing pump; g) To turbomolecular pump; h) Heated capillary

For quadrupole mass spectrometers, the ions from the first stage are extracted and pass directly into the mass spectrometer through a skimmer. For sector field mass spectrometers, a second vacuum stage is required to minimize high-energy collisions in the acceleration region, which could destroy the ions. The main advantage of electrospray is that even molecules with very high molecular mass can be detected at low m/z values, because the ions are highly charged (charges of up to 100 have been observed). Thus, the use of inexpensive quadrupole instruments is possible, but the more powerful analyzers such as FTMS or the new hybrid spectrometers have demonstrated superior performance (vide infra). The range of commercially available ion sources with new design features such as the orthogonal sprayer, developed by Hewlett Packard, (Fig. 13) or similar devices allow connection with HPLC even with high salt
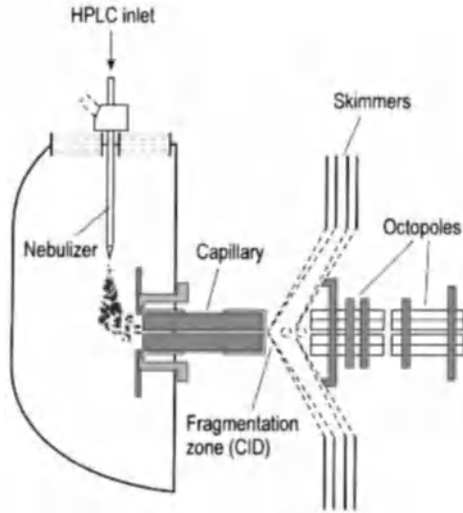
Figure 13. Schematic of the orthogonal sprayer assembly ESI Ion Source and transfer optics (Hewlett Packard)

loads. The new low flow sprayer has increased the sensitivity for microHPLC and capillary electrophoresis and the nanospray sources for biomedical work [111] have become pivotal in proteomic research.

20.5. Analyzers

20.5.1. Electromagnetic Sector Fields

The first mass spectrometer [112] was a single focusing instrument with a magnet deflecting the ions through 180°. Since then, magnetic fields alone and in combination with electrostatic fields have been used most successfully to separate ions. When an ion of mass m has the charge ze where z is the number of charges and e is the elemental charge, its kinetic energy after acceleration through a potential V is

$$\frac{1}{2}mv^2 = zeV$$

where v is the velocity after acceleration. In a magnetic field of strength B, the force is $Bzev$. For the stable radius r

$$Bzev = \frac{mv^2}{r}$$
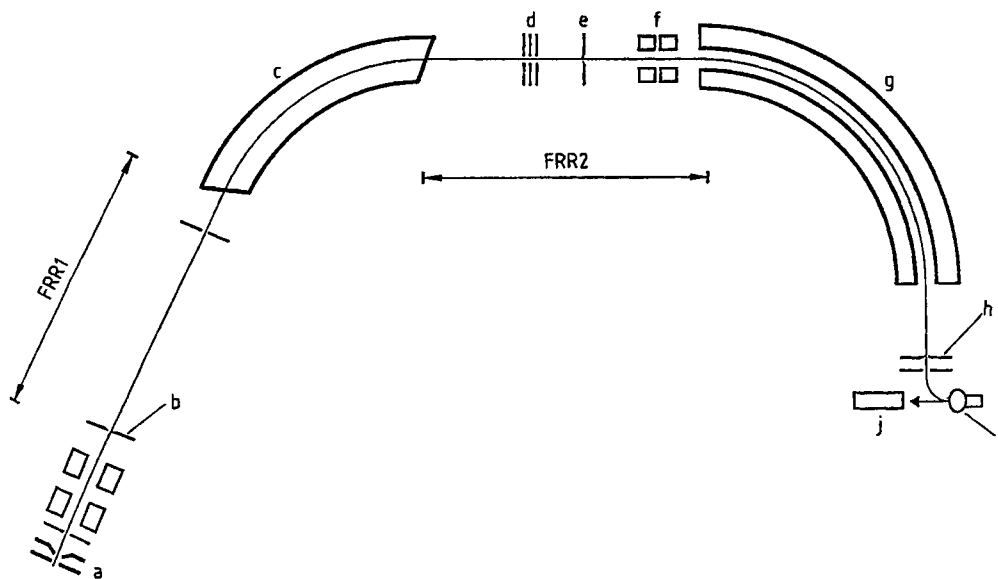
Figure 14. Double focusing MS (reversed Nier–Johnson geometry)
a) Ion source; b) Entrance slit; c) Magnetic sector; d) Focusing elements; e) Energy slit; f) Beam Shaping static quadrupoles;
g) Electrostatic analyzer; h) Exit slit; i) Conversion dynode; j) Multiplier; FFR1, FFR2, Field-free regions

Combining the equations yields

$$\frac{m}{z} = \frac{B^2 r^2 e}{2V}$$

Thus, for any given m/z value a certain r is valid, if B and V are constant. Scanning the magnetic field or the voltage, on the other hand, allows different ions to travel the same path through the magnet and reach the detector consecutively. The resolution attainable with a magnet is limited to ca. 5000 by the initial energy spread of ions due to the Boltzmann distribution and as a result of inhomogeneous and fringe fields. It can be increased by adding an electrostatic field so that the ions are focussed in both direction and energy ("double focusing"). An example is given in Figure 14.

Various designs exist, with the electrostatic field preceding or following the magnet [Nier–Johnson (BE), reversed geometry (EB), Herzog–Mattauch], and all of them differ in performance. The advantage in all cases is a far superior resolution (10 000 – 100 000) compared to magnets alone. But another degree of complexity is added to the mass spectrometer and there are some other handicaps as well. In comparison, the instruments are expensive, and more difficult to operate and maintain. In addition the long-term stability of

calibration is rather poor and some of the interfacing techniques are quite cumbersome to carry out. Nevertheless, sector field instruments are, generally speaking, still the most versatile mass spectrometers with strengths in detection power, mass accuracy, resolution, and flexibility [113].

20.5.2. Quadrupoles

Quadrupoles are the most successful concept from a broad range of so-called dynamic mass spectrometers [114]. They differ from static analyzers in that they use a high-frequency voltage to disperse the ions. In a quadrupole four poles (rods) form a hyperbolic field through which the ions travel with low kinetic energies. The superposition of high frequency and a variable d.c. voltage enables the quadrupole to separate the ions according to their mass. The regions of stability and instability are shown in Figure 15. The coefficients a and q are derived from the Matthieu equations. They are functions of the d.c. voltage U and the a.c. voltage $V_0 \cos \omega t$. For a fixed frequency, m is proportional to both U and V_0 and when the ratio $a/q = 2U/V_0$ is held constant, the masses follow stable paths and a spectrum may be recorded.
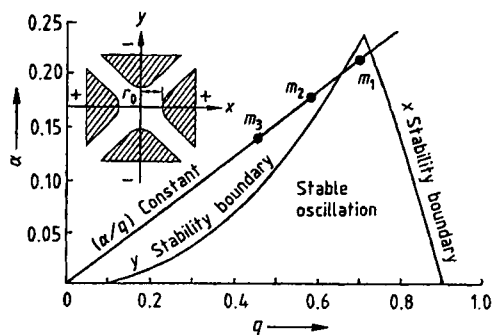
Figure 15. Quadrupole stability diagram indicating the conditions for stable and unstable oscillations

Although mathematical treatment of the separation process is complicated, the operation of such a spectrometer is simple. They have quickly become established in GC/MS systems, since they can be very compact, easy to handle, and less expensive than other instruments. In addition, computer control is simple. This advantage became even more significant with the development of triple quadrupole instruments (see Section 20.7.1).

20.5.3. Time-of-Flight (TOF) Mass Spectrometer

A few years ago, time-of-flight instruments were used almost exclusively for surface analysis. Then, the development of plasma desorption and laser desorption/ionization changed the situation dramatically. It is now conceivable that in the future TOF analyzers may become routine analyzers. The mass analysis is based on the fact that after uniform acceleration in the ion source, small ions arrive earlier at the detector than heavy ions. This is because all particles acquire the same energy resulting in different velocities. If the acceleration voltage is V, the energy is eV. For an analyzer of length L, the time to travel to the detector is

$$t = \sqrt{\left(\frac{m}{z}\right)\left(\frac{1}{2\,eV}\right)}\,L$$

and, solved for m/z

$$\frac{m}{z} = \frac{2eVt^2}{L^2}$$

The detectors, normally multichannel arrays, are capable of resolving the resulting small differences in time. Almost every ion leaving the source can be recorded, giving TOF instruments excellent sensitivity. To enhance resolution, a reflector device can be added to the linear flight tube. This reflectron focuses the ion beam on the detector, compensating for the initial energy spread of the ions, because the faster ions penetrate more deeply into the reflector than the slower ions and thus have a somewhat longer path to the detector. Resolutions in the range of 5000 – 10 000 have been obtained and ions with $m/z > 100\,000$ have been recorded. Alternatively, the reflector can be used to determine the sequence of biopolymers by means of laser desorption MS. Many metastable ions dissociate after acceleration, and fragments can be generated by collisions (psd, post-source decay) . These normally arrive simultaneously with the precursor ions, since no field is present to induce separation. If the voltage of the reflectron is then changed stepwise, such ions can be separated prior to arrival at the detector [115], since their masses are different. Recently orthogonal injection of ions into the TOF instrument was used to interface electrospray and electron impact sources, also giving excellent performance (for a review see [116]). The acquisition speed of the TOF instruments renders them excellent detectors for rapid separations, such as microCE. An impressive example has been published recently [117] and in Figure 16 spectra of such a combination are shown.

20.5.4. Fourier Transform Mass Spectrometry

Fourier Transform MS was developed on the bases of ion cyclotron (ICR) mass spectrometry [118]. The principle is shown in Figure 17. The cubic cell comprises two trapping plates (T), two excitation plates (E), and two detector plates (D) and is located in the center of a high-field magnet, generally a superconducting magnet with up to 7 Tesla. Trapping voltages applied to two opposite plates keep the ion in the center. After application of a radiofrequency (RF) voltage at two plates perpendicular to the others, the ions are excited and travel on circular trajectories with the same radius, but with frequencies that depend on their mass. The third pair of plates detects the RF image current induced by the rotating ions. The different frequencies and amplitudes in the image current

Figure 16. TOFMS of cytochrome *c*, concentration 0.1 μM in the solvent; Data aquisition: 5000 Hz; Spectrum A 3200, B 400, and C 100 summed scans

Figure 17. Schematic of a Fourier transform ion cyclotron resonance mass spectrometer with external ion source, transfer optics, and cubic FTMS cell
Ions can be generated using EI (even in combination with GC), FAB (LC/MS), or laser desorption/ionization; the transfer optics allow the required vacuum to be maintained.

[120]. For small molecules, resolutions in excess of 10^8 have been established. Further, MS/MS is rather easy to perform and operation is straightforward [121]. FTMS is the only competitor for sector field instruments with respect to resolution and accuracy.

However, its usefulness in analytical mass spectrometry still seems uncertain, partly because the very high cost prevents widespread use. Furthermore, optimal vacuum conditions ($< 10^{-6}$ Pa) are necessary in the ICR cell, which is difficult to maintain during GC/MS analyses. Attempts to solve the vacuum problem have been made with dual cell systems, where the ions are transported from an ionizer cell to an analyzer cell [122]. Two quadrupoles for mass selection and an RF-only quadrupole for focused injection [123] have been used, as has a time-of-flight arrangement [124]. They all serve to isolate the "dirty" ion source or the sample matrix from the analyzer to maintain optimal pressure conditions. With the last mentioned arrangement, the GC inlet replaced by a FAB target, and an atom gun, a spectrum of an enkephalin (Tyr-GlyGly-Phe-Leu) was obtained [125] with normal FAB features, but with almost no matrix background. Newly designed instruments seem to have solved the technical problems to a great extent and both, MALDI and electrospray spectra have been published with exorbitantly high resolutions (in excess of 500 000) and with impressive sensitivity.

signal can be extracted by Fourier transformation and converted into mass spectra. Theoretically, FT ICR MS has almost unmatched high-mass and high-resolution capabilities, the latter at the expense of very long acquisition times. CsI clusters up to $m/z = 31\,830$ have been observed [119], with rather low resolution, using an external ion source and quadrupole transfer ion optics; with a different instrument, a resolution of 60 000 at $m/z = 5922$ has been demonstrated using laser desorption

Figure 18. Schematic representation of the ion trap components, the scan function, the electron beam, and the resulting ion signals on the time axis (Adapted from [113] and [127])

Figure 19. Cut view of the LCQ$_{Deca}$ (Thermoquest)

20.5.5. Ion Traps

Another ion storage device in rapid development is the ion trap. Although it has long been used by physicists [126] it was only much later that it was converted into a powerful GC detector [127]. In this case, only high frequencies and no magnetic field are necessary to create a cage for the ions in a closed quadrupole field. The ions are ejected several times per second from the trap, hitting the detector sequentially. The principles of operation are depicted in Figure 18. The first commercial version had a limited mass range of 600 amu and low resolution, but quite convincing detection capabilities.

The mass range has since been dramatically increased (up to 40 000 amu), although the resolution and particularly the mass accuracy is still fairly poor. In recent versions, an additional scan ramp is used to evaluate the ion population of the trap for optimization of the residence time. However, high-resolution experiments using electrospray have now been described [128] and the detection capabilities have been steadily improved by optimizing the ion transfer into the trap. A modern version that is used as a detector for HPLC is shown in Figure 19. A weak point is still quan-

titative reproducibility. The storage capability is limited because when many ions are stored, coulomb repulsion destroys the performance of the trap. Therefore, software has been developed to monitor the number of ions in the trap and adjust the storage time accordingly to avoid unwanted ion – molecule reactions. Ion traps are the analyzer of choice in all cases where structure elucidation is the major application and several commercial versions from different verndors are on the market today.

20.5.6. Isotope Mass Spectrometer

The reliable determination of isotope ratios of elements normally needs specialized instrumentation. The ion sources depend on the type of element or molecule; for many elements the thermal ionization source has been employed, but for others and for small molecules, EI can be used as well. Although conventional analyzers can be used in such instruments, for precise determination, simultaneous detection of more than one ion beam is necessary and the instruments must be optimized for this purpose. Isotope ratio mass spectrometry (IRMS) is a specialized technique designed for precise determination of isotope ratios such as $^{12}C/^{13}C$, $^{1}H/^{2}H$ in organic compounds. The analysis requires the separation of compounds by GC, combustion of the compound to small molecules such as CO_2 and H_2O, ionization and accurate measurement of the signals due to the isotopically pure fragments. The analyzers allow simultaneous detection of the signals at low levels with high precision by using several Faraday cups [129] (Fig. 20). With such instruments, an abundance sensitivity greater than 10^{-8} was obtained for $^{230}Th/^{232}Th$ ratios [129].
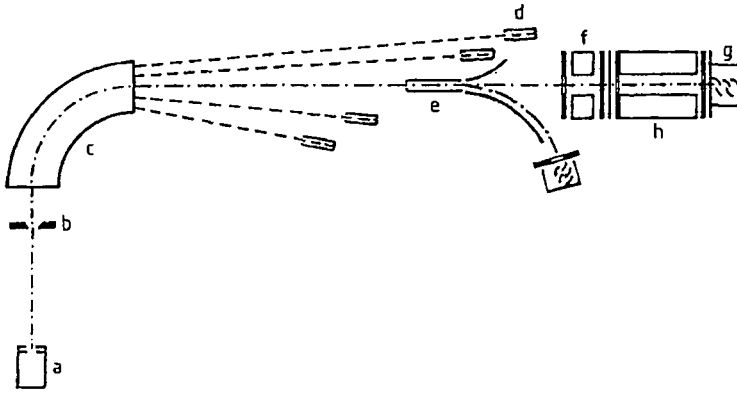
Figure 20. Isotope mass spectrometer with variable faraday cups and a quadrupole mass filter for MS/MS experiments where a deceleration lens is necessary between the analyzers
a) Ion source; b) Aperture slit; c) Magnet; d) Variable Faraday cups; e) Center slit; f) Deceleration lenses; g) Counter-SEM; h) Quadrupole mass filter
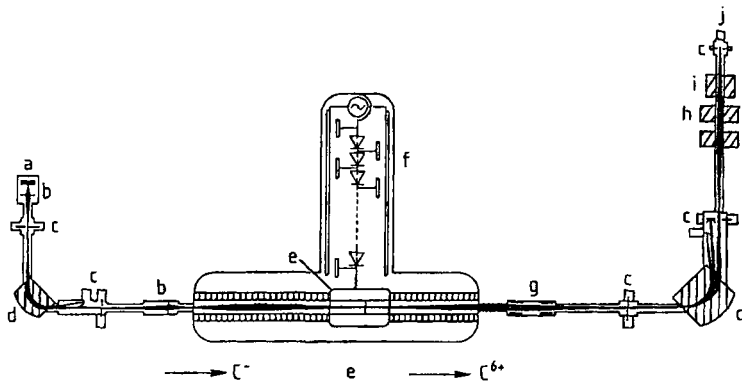
Figure 21. Accelerator mass spectrometry
a) Ion source; b) Lens; c) Slits; d) Beam bending magnet; e) Stripper foil charged to 2.5×10^6 V; f) High voltage source; g) Electrostatic lens; h) Magnetic lens; i) Velocity filter; j) Ion detector
(Reproduced with permission from [131])

To obtain the highest precision, further precautions must be taken with sample handling, instrumental parameters, and electronics to minimize the noise levels. This makes the procedure highly complicated, and only a few laboratories in the world are able to perform such precise measurements. An example is the correction of the Avogadro number, based on isotope mass spectrometry with silicon isotopes [130].

20.5.7. Accelerator Mass Spectrometry (AMS)

Accelerator mass spectrometry (Fig. 21) is designed for the most precise atom counting of cos-

mogenic radionuclides such as ^{10}Be, ^{14}C, ^{26}Al, ^{32}Si, and ^{36}Cl. It can measure these elements with unparalleled detection power in, for example, polar ice cores, deep sea sediments, or wood. The nuclide ^{14}C is used in the geosciences for dating of organic matter. The ions from the source (different sources may be used) are selected with a 90° magnet, accelerated toward the stripper foil, and converted into highly charged positive ions. These ions are focused again into a magnetic sector and detected within a gas-filled chamber by means of secondary ions produced there. The differentiation of, for example, ^{14}C^{6+} and N^{7+} is possible, since both species penetrate into the chamber with a different depth.

The detector is designed to identify species of nearly the same mass and velocity but different charges with very high selectivity.

20.5.8. Other Analyzers

The number of additional analyzer systems similar to those discussed above, but with only minor modifications, and others making use of different principles, is large, but most of them have no practical application. An exception may be the Wien filter, an analyzer with crossed electric and magnetic fields [132]. It has high-mass capabilities and high sensitivity but very poor resolution, which has inhibited its widespread application as an analyzer, although it is used to control ion energies in some instruments. The use of Wien filters in hybrid MS/MS instrumentation has been proposed [133].

Aside from quadrupoles and ion traps, other dynamic analyzers [114] such as the monopoles have been merely of theoretical interest.

20.6. Metastable Ions and Linked Scans

The determination of metastable ions has found much interest, because of their structural and theoretical significance. Several techniques have been developed and some of them are discussed here. Some of the techniques can be made more powerful by means of forced fragmentation in a field-free region (collision-induced dissociation). This process is essential for MS/MS experiments (see Chapter 20.7). Metastable ions are gaining new relevance in time-of-flight instruments with reflectrons (see Section 20.5.3).

20.6.1. Detection of Metastable Ions

In a magnetic sector instrument the signals of metastable ions appear as broad, mostly gaussian peaks in normal mass spectra. The location m^* may be calculated from the precursor ion mass m_p and the product ion mass m_d using the relation:

$$m^* = \frac{m_d^2}{m_p}$$

These signals are normally lost in double focusing instruments, although in Nier–Johnson type mass spectrometers, reactions in the field-free region between electrostatic analyzer and magnet deliver signals at the same positions as well. In one experiment the acceleration voltage is scanned. To detect a metastable transition, the spectrometer is focused for the product ion m_d with reduced acceleration voltage V_d. The acceleration voltage is then increased to V_p without changing the electrostatic analyzer and the magnet until the signal of the unknown precursor mass m_p appears. The ratio V_d/V_p is equal to the ratio of the masses m_d/m_p and m_p can be calculated. The resolution obtained with this so-called Barber–Elliott technique [134] or accelerating voltage scan is low and the instrument is continuously detuned, but for processes with rather small losses the detection capabilities can be great and kinetic energy release information is retained. This means that part of the excess energy of the ion is released as kinetic energy, leading to a range of translational energies yielding broad signals.

20.6.2. Mass Selected Ion Kinetic Spectrometry

Double-focusing instruments can be regarded as a combination of two mass spectrometers. For instruments with a reversed Nier–Johnson geometry, the magnet can serve as the first mass spectrometer (MS-1) that isolates one ion species from the mixture. This ion may undergo collisions in the field-free region following the magnet to form fragments, which can be separated by means of the electrostatic sector (MS-2). This technique, called mass selected ion kinetic spectrometry (MIKES) or direct analysis of daughter ions (DA-DI), has only the very limited resolution of the energy analyzer for the fragments and requires reversed geometry. Therefore, the number of applications in analytical problems is small; its importance lies in determining the energy released during fragmentation.

20.6.3. Linked Scans

A simple and inexpensive way of obtaining MS/MS information is the use of linked scans in sector field instruments. The advantage over the MIKES technique is that the resolution is generally better and that all double-focusing instruments can be used. The two sectors and the acceleration voltage are linked together by functions that allow
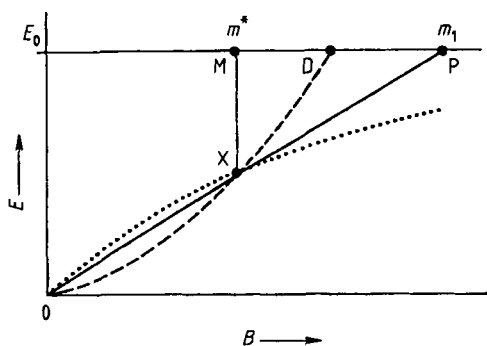
Figure 22. The B/E plane
—— for B/E = const.; ——— for B^2/E = const.; .... for neutral loss scan $B(1-E)^{1/2}/E$ = const.

only specific fragment types to pass both analyzers. These may be the product ions of a precursor, all precursor ions of a particular product, or all ions with a preselected mass difference of precursor and product ions (neutral loss scan). In Figure 22. the relation between the most common linked scans is shown.

Product (Daughter Ion) Scans. When the ratio of magnetic field B and acceleration voltage E is held constant ($B/E = 1/v$, with v as velocity) during a mass scan downward from a precursor ion, only those fragments of this preselected ion which are formed in the first field-free region are transmitted through the electrostatic sector. This is because the velocity is unique for each precursor and thus for all the fragments formed after acceleration. The resolution of the precursor ion is low, but the resolution of the fragment ions can be quite high and the kinetic energy release information is lost. If a collision cell is available, high-energy collisions may be employed to induce fragmentation.

This type of scan is often used to analyze the fragmentation pattern of a substance. The spectra obtained may serve as a "finger print" of an ion and may allow a fragmentation scheme to be established. For a known reaction of a known compound the instrument can be set to detect only this one signal on this particular B/E = const. surface, enhancing the specificity of detection (see "reaction monitoring," Chap. 20.2).

Precursor (Parent) Ion Scans. When a product ion appears with high frequency and the precursor is to be identified, a different scan is necessary. This precursor or parent ion scan requires a scan law in which $B^2/E = m$ with constant m. The ve-

locity spread of the ions is not filtered, thus having poor resolution for the precursors, but good detection capabilities. This approach has been employed successfully for the screening analysis of drugs and their metabolites (metabolic profiling).

Neutral Loss Scans. Among several other possibilities for linking the fields and voltages in a double-focusing mass spectrometer, the neutral loss scan is the only one which deserves a brief discussion here. The scan function is

$$\frac{B^2(1-E)}{E^2} = \text{const.}$$

This type of scan has a distinct analytical potential because the loss of a common neutral fragment can be taken as an indication of a particular class of compounds. Therefore, it has been applied in the search for specific compounds derived from a certain precursor compound (drug/metabolite monitoring).

20.7. MS/MS Instrumentation

The combination of two or more analyzers, commonly known as MS/MS or tandem mass spectrometry, is a highly specific means of separating mixtures, studying fragmentation processes, and analyzing gas-phase reactions. The field is reviewed thoroughly in [135]. With the first analyzer, one ion is isolated from all the others. In the next, reactions of that ion are studied further. Fragments may be formed in unimolecular reactions from metastable ions, but generally fragmentation is induced by a variety of means. Collisions with gases in a collision chamber (CAD: collision-activated decomposition), laser photodissociation [136], or surface collisions [137] can be employed to induce fission. Since MS/MS can be combined with chromatographic separation techniques, it has found a broad range of applications in analytical chemistry. Indeed, LC/MS interfaces such as thermospray or the API techniques depend largely on collision-induced dissociations and MS/MS. Since they produce only molecular ions, little structural information is provided. With CAD and MS/MS these details are obtainable. A recent review gives a fairly detailed discussion of the thermochemical aspects of the various fragmentation induction methods used in mass spectrometry [138].
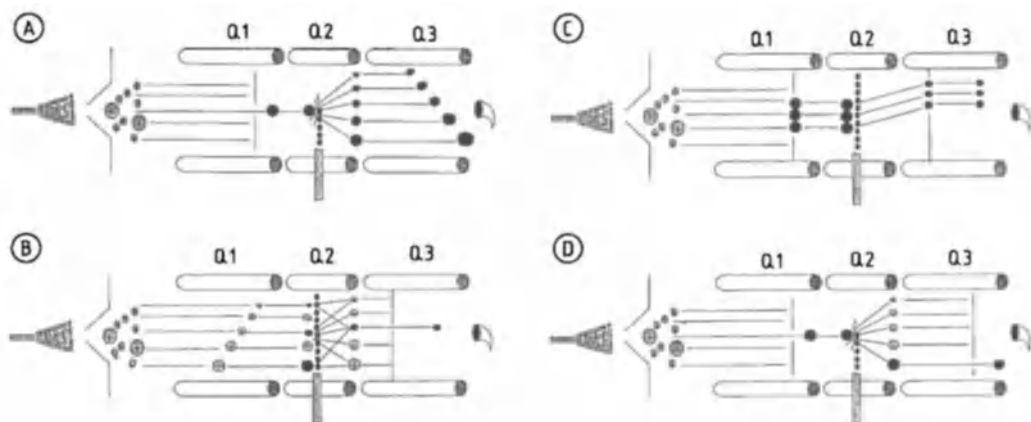
Figure 23. MS/MS experiments with the triple quadrupole MS
A) Daughter scan; B) Parent scan; C) Neutral loss scan; D) Reaction monitoring
Q1, Q3: mass analyzers; Q2: collision cell

20.7.1. Triple Quadrupoles

Triple quadrupoles for analytical MS/MS experiments have been built since the pioneering work of YOST and ENKE in 1978 [139] and they have found applications in all areas of mass spectrometry. The first and the third quadrupole are MS-1 and MS-2, whereas the second in the middle is an RF-only quadrupole acting as collision cell. In the collision cell, the transmitted ions undergo low energy collisions with a gas. For the product ion scan (daughter scan), only one precursor (parent) is selected and the second analyzer scans for the product ions. For the precursor scan, the first quadrupole is scanned and only the selected product ion is recorded. In the neutral loss scan, both analyzers are scanned with the selected mass difference, and for reaction monitoring only one precursor and one product ion species is permitted to travel through MS-1 and MS-2, respectively. In the meantime, instruments with different collision regions have been built. It was surprising that fragments are formed under the low energy regime in the collision cell. The fragment ion reactions are similar to those from high-energy collisions; exceptions include side-chain reactions of amino acids in peptides. The collision efficiency is quite high, particularly when the collision cell is an RF-only quadrupole [140] or octapole with nearly 100 % transmission [141]. The resolution of both quadrupoles as MS-1 and MS-2 can be set independently allowing, for example, unit resolution for precursor and product ions. Many different

experiments with scan functions similar to the linked scans have been made and the set up and computer control is quite straightforward. Figure 23 shows schematically the widely used MS/MS experiments. The triple quadrupole is the most common MS/MS instrument today.

20.7.2. Multiple-Sector Instruments

In cases where high-resolution MS/MS experiments are required, multiple-sector field instruments are the only choice. They have the advantages of potentially high resolution in both stages, a wide mass range, and high sensitivity. However, there is a considerable trade off due to complexity, space requirements, and high cost. This is true particularly for the four-sector field mass spectrometers. As an example, the schematic of a four-sector field instrument is given in Figure 24.

Such mass spectrometers are highly flexible since the possible combinations of sectors and the number of field-free regions allow many different experiments to be performed [142]. The limitation is that the fragmentation process becomes less probable with increasing mass and the upper limit seems to be around 2500 amu. To increase the dissociation efficiency not only collision cells, but also photo-dissociation and surface-collision-induced dissociation have been investigated as well as simultaneous recording with a zoom lens integrated into the second MS [143].
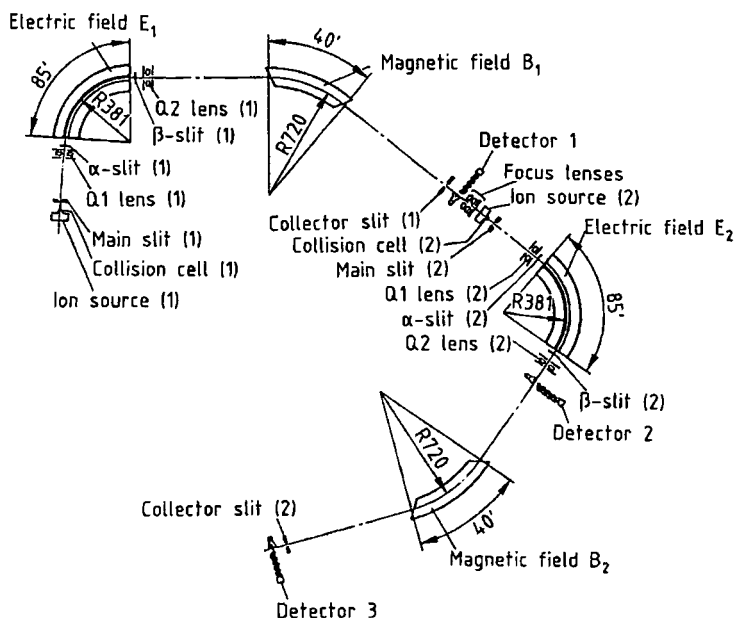
Figure 24. The schematic diagram and the ion optical components of the four-sector field mass spectrometer JE-OL HX110/HX110 (Reprinted from [142])

20.7.3. Hybrids (Sector – Quadrupoles)

The apparently logical combination of the high resolution capabilities of sector instruments with the simplicity of quadrupoles in hybrid instruments is complicated in practice. Since the quadrupole only accepts ions of low kinetic energy, it must be floated at the potential of the sector instrument or, if high-energy collisions are to be possible, a zoom lens allowing the deceleration of the ions must be incorporated between the mass spectrometers. The latter has been realized in only one commercial version, whereas the simpler first combination has been made by several manufacturers. Mostly the sequence is sector – quadrupole, but it is possible to arrange the spectrometers the other way around. Combinations of double focusing spectrometers with time-of-flight instruments [156] or with ion traps have been built and both are hybrids with great analytical potential. However, in general the sector instruments made these combinations too complicated and too expensive.

20.7.4. Ion-Storage Devices (FTMS, Ion Traps)

Two types of ion-storage device (ion traps) are in use today as mass spectrometers, as discussed already, with built in MS/MS capabilities: the magnetic trap as ion cyclotron resonance spectrometer, which is, in its modern version, the fourier transform mass spectrometer [144] – [147] and the electric field version, called ion trap [148] – [152]. Both types of instrument allow all but the one selected ion to be ejected, which then undergoes ion – molecule reactions, collisions, or photodissociation to yield fragments. This process may be repeated, allowing not only MS/MS but also MS/MS/MS and even MS^n, which allows the detailed study of fragmentation reactions. However, currently only daughter-ion scans are possible.

20.7.5. Quadrupole Time-of-Flight Tandem Mass Spectrometers

The development of combination sector field/TOF led almost instantaneously to the combination of quadrupoles with TOF spectrometers, since at the same time the TOF spectrometers were con-
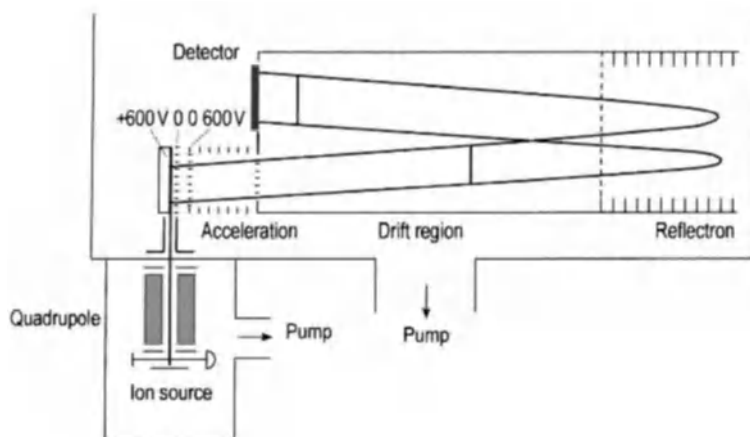
Figure 25. Q -TOF tandem mass spectrometer, schematic.

verted into high-resolution high-precision instruments (Fig. 25). A new aspect of this is that the ion injection into the TOF spectrometer is orthogonal, yielding an MS/MS combination with extremely high sensitivity for the fragments, without compromising scanning speed, resolution, mass range or the mass accuracy [153].

Thus this spectrometer has become one of the most successful new developments in the history of mass spectrometry (produced by two vendors) and today the instrument is one of the key tools in protein, peptide, and DNA research all over the world [154].

20.7.6. Others

Several TOF/TOF instruments have recently been described with impressive sensitivities, although the resolution in the second spectrometer is low due to the energy spread (similar to the situation in MIKE) [155]. The same is true of BE/TOF combinations, which seem to be highly suitable for electrospray and MALDI techniques [156], but it is too early to predict the future importance of these instruments.

20.8. Detectors and Signals

The detectors in mass spectrometry can be divided into two general types. One group is designed to detect one ion beam at a single location only (e.g., Faraday cage, secondary ion multiplier), and the second group can detect several ion beams simultaneously (e.g., photoplates and multiplier arrays). It is clear that the second group can be adapted only to specific analyzers, but the advantage in terms of detection power is so great that such instruments have been developed. The most common types of detector are shown in Figure 26.

20.8.1. Faraday Cage

One of the first detectors in mass spectrometry was the faraday cage. In principle it is simply a cup, which converts the current induced by impacting ions into a voltage pulse, the height of which is a measure of the number of ions reaching the detector per unit time. Since at that stage no amplification is achieved, its detection power is limited, whereas its linearity over the full dynamic range is still unsurpassed. The latter characteristic makes it even today a valuable tool, when high precision must be maintained (e.g., isotope mass spectrometry). In early instruments it was possible to switch between a faraday cup and a secondary electron multiplier. Conversion dynodes normally included into modern spectrometers, may be used as faraday cups as well, if they are positioned in line with the ion beam [157].

20.8.2. Daly Detector

The Daly detector [158] is still in use; the first version was able to distinguish between metastable and normal ions. The ions are converted into photons by means of a scintillator and the photons
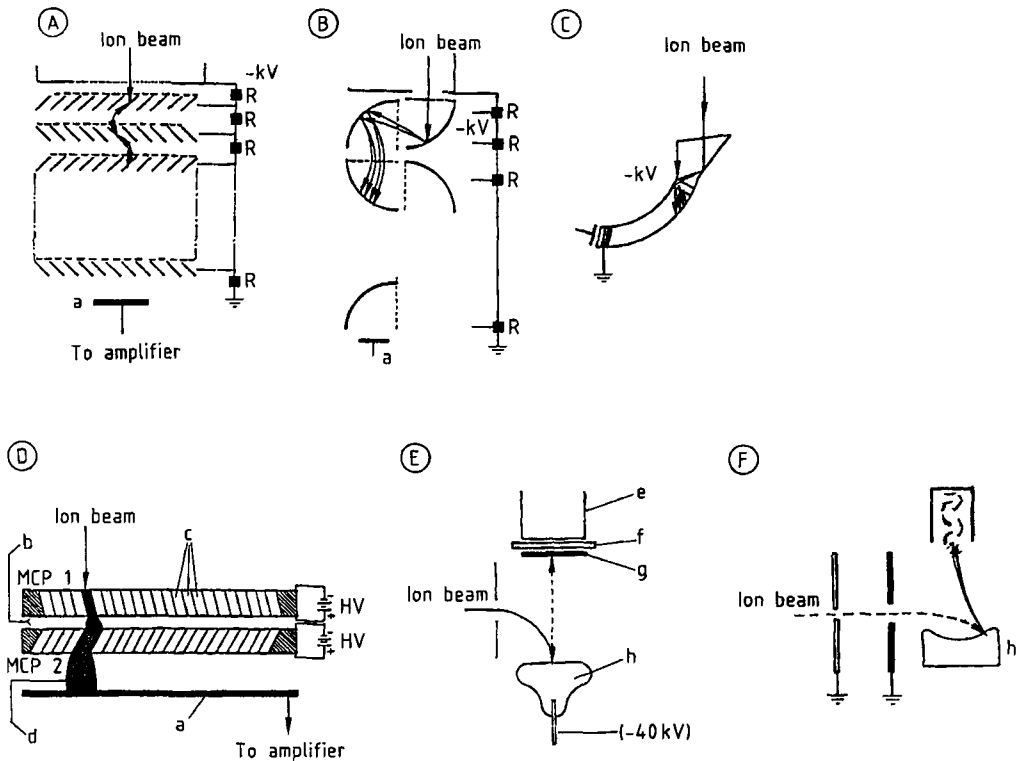
Figure 26. Types of detectors
A) Venetian blind multiplier; B) Box and grid multiplier; C) Channeltron multiplier; D) Microchannelplates; E) Daly detector;
F) Postacceleration/conversion dynode
a) Anode; b) Chevron; c) Channels; d) Electrons; e) Photomultiplier; f) Window; g) Scintillator (+ 10 kV); h) Dynode

are measured by a photomultiplier which is located outside the vacuum (see Fig. 26 E). It is easy to integrate a post-acceleration/conversion dynode opposite to the scintillator plate to enhance the sensitivity. The originalversion was arranged in beam and had an additional plate in front of the scintillator, serving as conversion dynode. With a potential at the ion energy, the normal ion would only gently touch the scintillator, giving no signal. Metastable ions with less energy will be reflected to the intermediate plate, giving secondary ions which are in turn accelerated toward the scintillator; thus, it is possible to distinguish metastable ions from "normal" ions.

20.8.3. Secondary Electron Multiplier (SEM)

The most common detector today is the secondary electron multiplier, which amplifies the first impulse from the impacting ion by orders of magnitude. Several different types have been developed, with specific properties. Multipliers with open dynodes, the surfaces of which are made from beryllium oxide, have high amplification, can be baked out and, depending on the number of stages, can have short blind times allowing ion counting. This is the period of time during which the electron cascade runs through the dynodes and when no new impact is seen. Similar are the venetian blind multipliers, which, however, cannot be used for ion counting. Each dynode consists of slats arranged like a venetian blind with a light grid in front to ensure a homogeneous field in the dynode. Due to the circular, wide open active area, the positioning is not critical, although the ion beam should not be too narrow (see Fig. 26 A). Multipliers based on semiconducting surfaces (channeltrons) are widely used in high-current systems and are incorporated in quadrupole and ion-trap instruments. Instead of discrete dynodes with
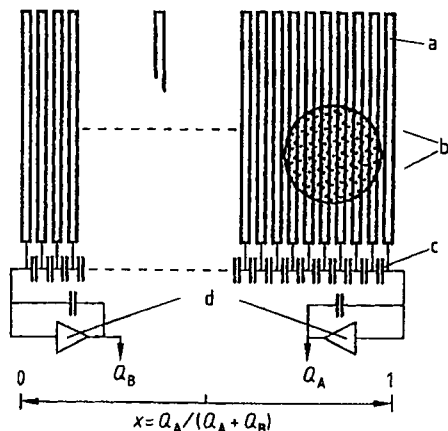
Figure 27. The Patric detector
a) Conductive strips; b) Electron cloud; c) Condenser chain;
d) Charge sensitive amplifiers (position and time information
for every ion)
(Reproduced with permission)

resistor chains, the channeltron uses a semiconducting material for the same effect and many different types and shapes are in use, giving similar performance as the dynode types (see Fig. 26 C).

When heavy molecules had to be detected, the need for post-acceleration to increase the probability for secondary electrons arose. Since then, conversion dynodes have been included in the detection systems to accelerate slow heavy ions (typical potentials between 15 and 30 kV), and thus enhance the sensitivity for both positive and negative ions.

20.8.4. Microchannel Plates

Microchannel plates are based on the same principle as the channeltron (Fig. 26 C) and can be regarded as an array of small channeltrons of rather short length (a few mm instead of cm) arranged side by side in one plate similar to the photoplates of Mattauch – Herzog instruments. They are used in TOF instruments and for the simultaneous detection of ions in sector field mass spectrometers. The major problem to be solved in the latter case was the fast readout of the signals, and various techniques have been developed for this purpose. A first, very elaborate version of such a device was described as early as 1974 [159]; more modern ones suitable for tandem mass

spectrometers use fiber optics to transmit the signals to photodiode arrays [160]. Detectors with zoom capabilities [143], static or scanning modes, and switching between analog and counting modes have been designed and successfully used [161], [162]; the gain in sensitivity can be two orders of magnitude depending on the system employed, although the resolution and sometimes the dynamic range is limited. The Patric detector (see Fig. 27) is a simultaneous detector with scanning option. The detector is capable of locating the impinging ion on the surface as long as the events are not exactly simultaneous; therefore, the dynamic range of the device is somewhat limited. The resolution obtainable is ca. 5000; since the mass spectrometer can switch between analog recording with a conventional SEM and the Patric detector even in a scan, the enhancement in sensitivity (ca. 100) can be fully utilized.

20.8.5. Signal Types

The shape of the signals depends largely on the type of instrument. Signals in sector field instruments are ideally triangular when the magnetic field is scanned and resolution as well as intensity are under optimal conditions. A rectangular beam profile (from the entrance slit) is moved across the exit slit, the width of which should be adjusted to match the width of the ion beam. This means in practice that the height of the signal is adjusted by the entrance slit. Then, the width of the final triangle is optimized by means of the exit slit without reducing the intensity further.

In quadrupole instruments the signals are in principle rectangular, since no slit can be used to define the beam shape precisely. In Fourier transform mass spectrometers, the peak profile should be gaussian when the signals are detected by resonance phenomena. The signals in time-of-flight instruments depend on the shape of the primary pulse (ions, neutrals, or laser photons), but even more on the desorption characteristics of the sample. Since the time for desorption and ionization of the sample is usually much longer than the primary pulse, the resulting signals have rather wide, poorly defined peak shapes with poor resolution. It is possible to shape the secondary pulse by means of ion optics to enhance the performance of the TOF instruments with respect to resolution.

20.9. Computer and Data Systems

The first applications of computers in mass spectrometry were merely the recording of signals from the multipliers, conversion to spectra, and some data handling. Today, powerful computers are integrated in the spectrometer. They control the scan, define the scan function, set the required voltages, tune the source, and supervise the experiments. The data are stored and interpreted with advanced tools, although at that stage the fully automated procedures are at an end.

20.9.1. Instrument Control, Automation

Modern mass spectrometers are completely dependent on computers and data systems. All the instrumental parameters of the inlet systems such as direct probe, GC, or LC (temperatures, gradients, times) are supervised by the computer. The ion source is tuned by automated procedures, the instrumental parameters (voltages, frequencies, resolution, etc.) are set by software procedures, and even the design of experiments (EI, CI, positive/negative ions, linked scans, MS/MS, etc.) can be started automatically, if specified conditions are met (e.g., signal height, retention times). Under such circumstances, the spectrometers can run a large number of samples unattended, since the analytical methods can be specified first, stored and activated upon request of the software. One advantage of this development is that a single man–machine interface is sufficient to work with such an MS. Further, most of the information necessary to describe an experiment from an instrumental point of view is stored automatically and is therefore available for reports and documentation.

20.9.2. Signal Processing

Most mass spectrometers produce analog signals at the detector. Thus it is necessary to convert the analog current or voltage signal into digital signals by means of analog-to-digital converters. This step has been a problem in the past, because the dynamic range and the speed of such devices was not always sufficient to maintain the integrity of resolution and dynamic range delivered by SEMs. In cases where ion counting is possible, different problems must be solved: the maximum counting speed is the limiting factor for the dy-

namic range. Since the design of the multiplier defines the pulse width and blind period, the dynamic range is a function of the performance of the multiplier and the pulse counting device. When pulses overlap, errors may occur, although correction is possible if the frequency is not to high.

Analog-to-Digital Conversion and Mass Calibration. As mentioned above, the analog signal is digitized by fast and accurate A/D converters (typically above 200 kHz) with sufficient dynamic range (≥ 16 bit). The limiting factor for the dynamic range is often the noise of the baseline. The digitized information about the ion current is fed into the computer, where the signals are analyzed and converted from "raw data" into peaks or noise, the intensities are calculated and placed on a mass scale. The conversion can be performed on the basis of time, magnetic field, or frequency, depending on the type of spectrometer. The mass scale must be calibrated to achieve the required precision. The most reliable way to calibrate the mass scale for all mass spectrometers is by means of reference compounds, whose mass spectrum in terms of accurate masses and approximate intensities should be known to the software. In addition, they should have signals evenly spaced over a large mass range.

Some reference compounds for mass calibration are listed below:

EI, CI, positive and negative
Perfluorokerosene (PFK, up to 1000 amu)
Perfluorotributylamine   (FC-43,   heptacosa, M_r 671)
Tris(perfluoroheptyl)-s-triazine   (up   to 1200 amu)
Ultramark 3200F (Fomblin oil, perfluoro polyethers, up to 3200 amu)
Ultramark 1621 (perfluorinated phosphazine, up to 1600 amu)

LC/MS (Thermospray)
Polyethylene glycols (up to 2000 amu)
Polypropylene glycols (up to 2500 amu)
Acetic acid cluster (up to 1500 amu)

FDMS
Polystyrenes (up to 10 000 amu)

FABMS
Glycerol (useful up to 1000 amu)
CsI (wide mass range, rather wide spacings)
CsI/RbI and NaI/CsI/RbI (wide mass range)

Ultramark 3200F (up to 3200 amu)
Phosphoric Acid (5%) in glycerol (up to 1500 amu, negative ions)

Electrospray
Polypropylene glycols (up to 2500 amu)

After acquisition of the spectrum, the computer tries to identify some starting pattern and then searches for all the other peaks. When the number of signals identified is sufficient, a calibration table is created which allows raw data to be converted into mass/intensity pairs. If high-resolution data have been acquired and accurate masses are to be calculated, additional internal standards are necessary for precise calculation of the signal location relative to known signals of the internal standard.

Peak Detection. One of the problems to be solved is the detection of true signals. It is necessary to determine the base line and to discriminate between noise signals of low intensity or insufficient width and real peaks. For this purpose, sophisticated software has been developed. Several parameters must be defined under a given set of instrumental conditions to ensure that no signal is overlooked and that as few noise signals as possible are recognized as peaks, obscuring the information contained in the mass spectrum. To do this the peak width, the minimum height and minimal number of counts, and, for unresolved signals, the depth of the valley between two adjacent peaks are calculated for a given scan speed, scan function, and number of digital points ("counts"). Only signals meeting all such requirements are accepted as peaks and if a problem is encountered during assignment, the peak can be flagged for classification.

Dynamic Range. The dynamic range of a multiplier in a mass spectrometer covers several orders of magnitude (up to 1×10^6) and is defined as the ratio between the smallest and the largest signal which can be produced. The digitalization system in the computer interface has to comply with this adequately, meaning the dynamic range of the A/D converter and its speed must be sufficient. With modern converters and sample and hold devices coupled with autoranging systems, this problem is solved. The simultaneous recording devices have rather low dynamic ranges due to saturation effects and readout problems. Since they are specifically designed to acquire signals of very small ion currents, this limitation is acceptable.

20.9.3. Data Handling

When the mass spectral data have been stored in the computer, the interpretation of the data is the next step. In most cases the data are spectra with peak intensities on a *m/z* scale. However, chromatograms from selected ion recordings and data from multichannel analyzers must also be handled. Sophisticated software has been written to develop increasingly automated procedures for quantitative analyses, mathematical treatment of data, display of spectra and structures, and to transport the data into laboratory data systems for integration into reports. While this aspect of data handling can be considered rather satisfactory today, the extraction of information contained in the spectra is under development. The deduction of possible structures of unknown compounds from mass spectra is demanding and often fails, and the calculation of spectra based on the structure is in its infancy. It is still uncertain whether this is at all possible, because the competitive reactions of the gas-phase ions are controlled by minute differences in energy between intermediate structures, and the kinetic behavior is difficult to model for complex molecules. Initial attempts have been published for both the modeling of reactions in the mass spectrometer [163] and the prediction of mass spectra from structures [164]. In both cases the results are far from application for analytical purposes.

20.9.4. Library Searches

So far, the most successful approach for extracting structural information from spectra is the library search. Libraries have been collected from different sources with large numbers of spectra; the most widely used are the Wiley data base with ca. 160 000 and the NIST data base with ca. 60 000 spectra; others, specializing in topics such as toxicology, agricultural chemicals, or foodstuffs, are also available [165]. The quality of the data contained in some of the large databases is still a substantial problem and must be improved, since spectra often strongly reduced in the number of peaks, incomplete, or incorrect have been taken from old collections and included in the new compilations. It is not trivial to identify and to replace them with better spectra, because wrong spectra
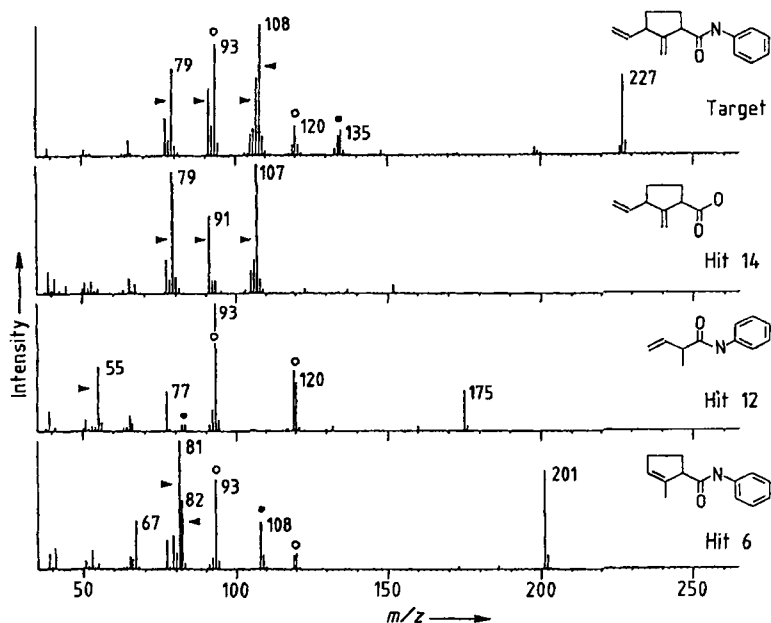
Figure 28. Spectra of similar compounds found by a similarity search for structures, not for spectra
▶ = Hydrocarbon fragment; ○ = Anilide moiety; ● = Loss of aniline radical or aniline molecule
The fragmentation pattern (bottom) explains why the analog reaction in Hit 6 yields m/z 82 instead of m/z 108; m/z 108 in spectrum Hit 6 must have a different origin. (Adapted from [171], with permission).

may be never found (similar spectra may no longer show up)and the definition of better spectra is also difficult. Quality is a relative attribute of a spectrum and sometimes two completely different spectra must be maintained, if they were recorded under diverse conditions. Library searches, with a variety of algorithms, have been developed, and the best known are SISCOM [166], [167], PBM [168], [169], and the INCOS type search [170]. The SISCOM approach also allows the identification of unknowns and the deduction of structural features, thus giving information about similar compounds. This is illustrated by the example given in Figure 28.

Recently the chemical structures of the compounds have been integrated as a means to enhance the power of the information retrieval. It has become possible to identify not only similar spectra, but also the spectra of similar structures,

revealing much of the complex relation between spectra and structures [171].

20.9.5. Integration into Laboratory Networks

Since in many laboratories, particularly in industrial environments, mass spectrometers are merely one tool among others such as NMR and IR spectrometers, it is becoming increasingly necessary to integrate the spectrometers into laboratory networks [172]. Advantages of this approach include:

1) The customer as a producer of a sample can use his terminal in the network to inspect all the spectroscopic data of a compound submitted for analysis or structure elucidation.

2) "Intelligent" interpretation, aided by computer software, can be initiated prior to examination.

3) It is possible to follow the fate of sample not only on the spectroscopic environment, but in other analytical, biological, or toxicological examinations as well. In the following step, all the data can be combined and the information may be integrated into a single report.

4) The thorough documentation and the assurance of quality becomes possible. The integrity of the data from the sample to the final report becomes increasingly important and is only possible with adequate use of the data systems.

20.9.6. Integration in the World Wide Web

As with many others examples, mass spectrometric data have been published through the WWW from many publishers and it is fairly easy today to access new bibliographic data via the internet. The homepages of most publishers provide either the bibliographic data or even the abstracts free and often the costs for the online publications are more or less the same as for the printed versions. Even more relevant may be the fact that in protein and nucleic acid research the most important databases are accessible over the net. MS datasystems have built-in software to search the databases directly or even via an automated strategy for known sequences in peptides or oligonucleotides, thus allowing the rapid identification of known proteins based on MALDI data of peptide mixtures and electrospray MS/MS data ("sequence tags"). It can be stated that the stunning speed in proteomic research is largely due to the immediate access to such databases and the retrieval software.

20.10. Applications

Typical areas for the application of mass spectrometry are:

1) Analysis of trace elements and compounds in environmental samples

2) Problem solving in biomedical research (protein identification/DNA), particularly the structure elucidation of natural compounds and determination of molecular mass

3) Control of synthetic procedures by rapid analysis of intermediates and products

4) Analysis of bulk materials and surfaces

One of the major driving forces for the rapid development of mass spectrometry during the last years is trace analysis in environmental chemistry and biochemistry. The search for sensitive, reliable methods that allow the determination of compounds at very low levels together with the option for mass spectra of highly polar and labile compounds motivated the development of new interfaces, ion sources, analyzers, and detectors. The second impetus is the search for techniques for the determination of very high molecular masses with better precision than is attainable by other known methods.

20.10.1. Environmental Chemistry

Mass spectrometry is a powerful tool in environmental chemistry because of its detection power, specificity, and structure analysis capabilities. Techniques have been developed for the determination of catabolites and metabolites of chemicals in practically all relevant matrices since no other comparable tool is available, even though technical problems can occur. Generally, sample preparation uses at least one type of chromatography, connected with MS either off-line or on-line. Of the on-line combinations ("hyphenated techniques"), GC/MS is the most successful, eventhough LC/MS is rapidly catching up in this area . The development of benchtop GC/MS and LC/MS instruments has made it possible to use mass spectrometry in routine laboratories and even in the field [173].

Air. The investigation of reactions in the atmosphere is one of the major fields of modern environmental chemistry. The processes responsible for the formation of smog in cities [174] as well as the interactions of man-made or natural hydrocarbons with ozone [175] are examples where mass spectrometric techniques have been applied. Usually, the chemicals under investigation must be trapped on a stationary phase such as charcoal or tenax. The samples are then transferred by thermodesorption to a GC/MS system [176]. For the investigation of the organic contents of single particles from aerosols, laser desorption has been employed successfully [177]. The potential of the new ion sources such as the atmospheric pressure ion (API) source for monitoring reactions directly or for detecting air pollutants at very low

levels has yet to be fully explored; however, initial experiments are encouraging [178]. The process of gas to particle conversion has been investigated using an adapted API source [179]. Volatile organic compounds in air and in water have been successfully analyzed by using membranes in various configurations to give enrichment factors in the range 40–780 [180]; with nitrogen as matrix, several halogenated or aromatic compounds have been detected in the ppb range. Recently, the specificity and sensitivity of REMPI mass spectrometry has been used to detect organic compounds in gases directly without trapping [181]–[183]; one additional benefit of this technique is the superior time resolution allowing rapid changes in concentration, to be followed.

Water. For water the situation is similar in that usually in the first step pollutants must be extracted from the matrix. Derivatization is often necessary to make MS analysis possible. Extraction and preconcentration is the critical step in the procedure, because the nature of the chemicals to be monitored can vary widely. LC/MS techniques that avoid derivatization have been used successfully. Pesticides such as triazines, organophosphorus compounds, and phenylureas have been determined in water samples by thermospray LC/MS [184]. Even though the well-developed GC/MS procedures are mostly used, they fail to detect some classes of compounds, such as plasticizers in drinking water, where they were first found by LC/MS [185]. A broad range of pollutants have been identified in wastewaters by particle beam LC/MS [186]. For the direct analysis of organic compounds in water an instrument has been designed with a membrane inlet, which allows certain chemicals to pass directly into the detection system, in this case, an ion trap. The detection limits for some of the compounds are low enough (e.g., benzene 1 ppb, toluene 10 ppb, ether 10 ppb; acetone and tetrahydrofuran 100 ppb) for routine monitoring of water [187].

For the determination of elements in water, ICPMS has demonstrated superior performance due to its sensitivity and multielement capabilities (see Section 20.11.6).

Soil. The analysis of herbicides, insecticides, and fungicides in soil is commonly based on extraction with liquids or supercritical fluids. The samples are then concentrated, derivatized if necessary, and transferred into the detection system. Often, GC/MS methods are the first step, although

MS can be replaced with a less complicated detector in the final routine version of the procedure. For developing new procedures and in cases where the identification of a compound is uncertain due to contamination, GC/MS may be the only choice. One of the most important applications is the analysis of tetrachlorodioxins (TCDDs), especially the toxic isomer 2,3,7,8-tetrachlorodibenzo-p-dioxin, and related compounds. These are formed as by-products of pesticide production, and occur in fly ashes from municipal incinerators and other sources [188]. All standard procedures for the detection of such compounds have as the key step GC/MS, normally with high resolution (10 000) [189], [190] to increase the specificity. Mass spectrometers have been designed solely for this purpose by several manufacturers and, since the detection limits are very low and the reliability of the procedures must be high, the use of reference compounds with ^{13}C labels is mandatory [191]. Another method of enhancing the specificity is the use of API sources [192]. For the investigation of metabolism or catabolism of herbicides (e.g., by photooxidation), LC/MS has become increasingly important [193]. The fate of such compounds (e.g., phenoxyalkanoic acids and their derivatives [194]) can be monitored in soil samples by analysis of their stability and by the emergence of catabolic and metabolic products. A further example is the analysis of phenols and their biological conjugates (e.g., sulfates, glucuronides) with LC/MS [195], which was previously very difficult and time consuming. The fate of elements in soil is difficult to investigate, since the interactions with fulvic and humic acids are complex and not well understood.

Waste Materials and Dump Sites. The major difficulty in MS analysis of wastes is often the highly complex matrix, which must be excluded from the analysis. One possibility is the use of multiple chromatographic steps prior to MS analysis. Alternatively the compounds of interest may have very different physical properties from the matrix. This is the case with polycyclic aromatic compounds, which have been determined directly by using thermal desorption with GC/MS [196] detection. Recently, small portable quadrupole mass spectrometers have been built as a means to identify contaminated sites such as abandoned industrial plant locations and old waste dumps.

N-terminal ions

a_n: H-(NH-CHR-CO)$_{n-1}$-N$\overset{+}{\text{H}}$=CHR$_n$

or

$$H-(NH-CHR-CO)_{n-1}-NH-\overset{\overset{\displaystyle CR_n^aR_n^b}{\|}}{CH}$$

with H^+ bracket over H-(NH-CHR-CO)$_{n-1}$

a_n+1: $$H-(NH-CHR-CO)_{n-1}-NH-\overset{\overset{\displaystyle R_n}{|}}{CH}\cdot$$

with H^+ bracket

b_n: H-(NH-CHR-CO)$_{n-1}$-NH-CHR$_n$-C≡O$^+$

c_n: $$H-(NH-CHR-CO)_n-NH_2$$

with H^+ bracket

d_n: $$H-(NH-CHR-CO)_{n-1}-NH-\overset{\overset{\displaystyle CR_n^b}{\|}}{CH}$$

with H^+ bracket

C-terminal ions

v_n: $$HN=CH-CO-(NH-CHR-CO)_{n-1}-OH$$

with H^+ bracket

w_n: $$\overset{\overset{\displaystyle CR_n^b}{\|}}{CH}-CO-(NH-CHR-CO)_{n-1}-OH$$

with H^+ bracket

x_n: $^+$O≡C-NH-CHR$_n$-CO-(NH-CHR-CO)$_{n-1}$-OH

or

$$O=C=N-CHR_n-CO-(NH-CHR-CO)_{n-1}-OH$$

with H^+ bracket

y_n: $$H-(NH-CHR-CO)_n-OH$$

with H^+ bracket

y_n-2: $$HN=CR_n-CO-(NH-CHR-CO)_{n-1}$$

with H^+ bracket

z_n: $$\overset{\overset{\displaystyle CR_n^aR_n^b}{\|}}{CH}-CO-(NH-CHR-CO)_{n-1}-OH$$

with H^+ bracket

z_n+1: $\cdot$CHR$_n$-CO-(NH-CHR-CO)$_{n-1}$-OH

with H^+ bracket

Figure 29. Most common peptide fragments from positive ions
R = Side chain of an amino acid; R^a and R^b = Beta substituents of the nth amino acid (From [207], with permission)

20.10.2. Analysis of Biomedical Samples

In the last few years the biomedical field has experienced a revolution of mass spectrometric possibilities [197].

MS has long been important for the detection of trace amounts of, for example, steroids [198], [199], drugs [200], and prostaglandins [201] in biological fluids such as urine, plasma, or saliva. Another area with many applications is the structure analysis of a wide variety of compounds, sometimes after derivatization or chemical modification. Micro-scale methylation [202], acetylation, and silylation have been used for the MS analysis of large molecules. For oligosaccharides, often not only the sequence and the type of monomer must be analyzed, but also the position of the glycosidic linkage and derivatization/cleavage schemes have been developed for this purpose [203], [204]; for details of the Lindberg method [205], see [197]. The sequence of peptides can be determined by cleavage to short oligomers and GC/MS, often in combination with FABMS for the determination of large fragments made enzymatically [206]. The reactions of the protonated ions produced with FAB (or liquid SIMS, LSIMS) have been studied extensively. Chemically labile bonds have the tendency to break more often, leading to analytically significant ions. Some of the most important are shown in Figure 29. The same type of ions are observed with electrospray CID (MS/MS) or MALDI PSD spectra which have revolutionized the analysis of proteins in the last few years, and proteomics without mass spectrometry would be unthinkable. The general strategy applied is shown in Figure 30.

Since such ionizing techniques are compatible with many MS/MS experiments, sequence information can be deduced from the spectra in great detail. When high-energy collisions are used, even leucine and isoleucine can be distinguished due to the different fragmentation of the side chain.

The analysis of complex lipids such as glycosphingolipids [208] and lipopolysaccharides [209] became even more complete with the advent of desorption techniques. But the localization of the position of double bonds in unsaturated fatty acids from complex lipids is still difficult. Several techniques were conceived using pyrolidides [210] or dimethylsulfides [211] and EIMS, CIMS [212], or MS/MS techniques [213]. As another example, electrophoretic labeling (perfluoroacylation) allows the determination of modifications in nucleobases at the attomole level [64] or the characterization of prostaglandins in biological fluids by means of negative chemical ionization (NCI) [214].

Field desorption mass spectrometry (FDMS) indicated for the first time [215] that the analysis
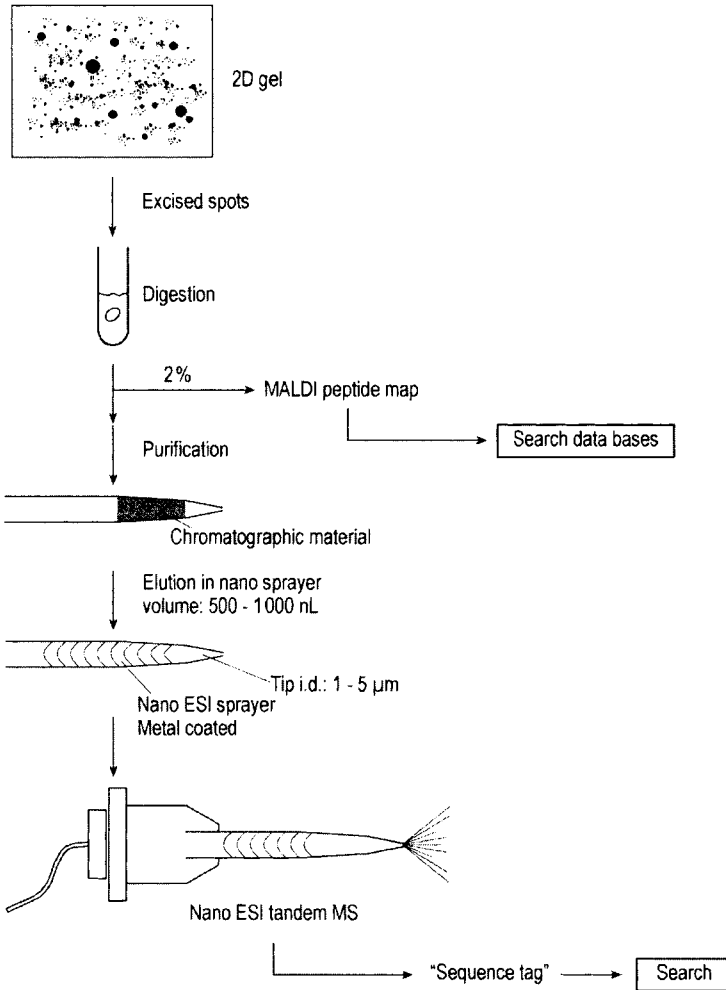
Figure 30. Strategy for the identification of proteins (after Protana, Odense)
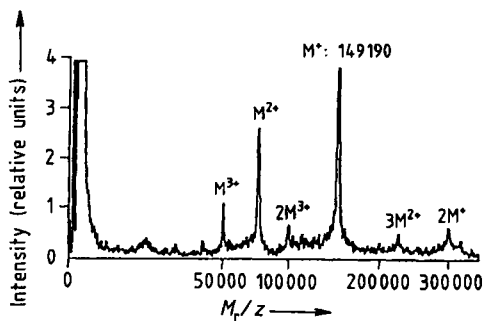
Figure 31. Matrix-assisted laser desorption (MALDI) spectrum of a carefully purified monoclonal antibody (IgG of mouse against a specific human lymphokine) (Reproduced with permission)

of large, labile, and highly polar compounds such as nucleotides [216] and oligosaccharides [217] might become possible. The next step was the development of fast atom bombardment, although it seems in retrospect that plasma desorption was the first method to allow molecular ions of very large molecules to be observed. Thus the whole range of biological polymers became accessible to mass spectrometry. This expectation was indeed largely fulfilled with laser desorption and electrospray, which appeared shortly afterwards. As an instructive early example, the spectrum of an intact protein is shown in Figure 31.

With these methods, large proteins, enzymes, or polynucleotides yield molecular ions, allowing the precise determination of molecular mass. In

combination with chemical modification of amino acid side chains, the secondary and even the tertiary structure of proteins could be studied [218]. The synthesis of peptides, for example, became much more efficient, because the time-limiting step was formerly the analysis of the products, which can now be performed in minutes.

In addition, electrospray is also well suited to the analysis of small, highly polar substances (e.g., drugs and metabolites [219]). Recently, even noncovalent, highly specific interactions [220] have been studied. Apparently, the conformation of some biological polymers can be retained in the gas phase, allowing receptor – ligand complexes to survive the ionizing step. Examples include the interaction of the synthetic macrolide FK506 with its cytoplasmic receptor [221] and the detection of selfcomplementary DNA duplexes [222].

Since some of the techniques can be interfaced to liquid chromatography (LC/MS) or capillary zone electrophoresis (CZE/MS), the analysis of complex mixtures is also possible. Thus, these methods are suitable for the efficient design of pharmaceuticals and agrochemicals where careful analysis of the metabolic processes is necessary, but was formerly difficult and time consuming.

The detection capability of this type of instrumentation allows the reactions of xenobiotic compounds and drugs with macromolecules such as proteins or nucleotides [223] to be studied and very low concentrations of platinum-drug-modified biopolymers [224] in living organisms to be investigated. The metabolic or catabolic fate of drugs can be studied (e.g., the abuse of steroid hormones by athletes).

20.10.3. Determination of High Molecular Masses

One of the most important applications of mass spectrometry has always been the determination of molecular mass. Since this can be difficult with EIMS, the search for soft ionization techniques is almost as old as organic mass spectrometry. The first step was chemical ionization and particularly direct chemical ionization (DCI), which allowed the samples to be desorbed directly from a thin wire into the CI plasma. It has been successfully used for natural compounds such as glycerides. Later, field ionization (FI) and especially field desorption (FD) opened up new possibilities for the analysis of labile natural compounds such as saccharides, peptides, and nucleotides. When fast

atom bombardment (FAB), plasma desorption (PD), electrospray ionization (ESI), and laser desorption (LD) mass spectrometry came into use, molecular ions of labile and very large molecules in excess of 200 000 amu could be determined with unparalleled accuracy. The electrospray spectrum of a protein allows its molecular mass to be calculated with high precision from a series of differently charged molecular ions (Fig. 32).

In this context, the discussion about resolution became a new relevance [225]. The use of nominal integral masses (e.g., 12 for C, 1 for H, 16 for O) is meaningless here as is true for the monoisotopic masses (e.g., 1.0078 for ^{1}H, 14.0030 for ^{14}N, 15.9949 for ^{16}O). Even with resolutions exceeding 10 000 it is impossible to resolve all the possible isobaric species in the molecular ion pattern of molecules such as peptides or oligonucleotides with hundreds of carbons and hydrogens, and tens of oxygen, nitrogen, and sulfur atoms. This is even more difficult with multiply charged ions as occurs regularly with electrospray. In such a case, the averaged chemical mass of the elements can be used for the calculation such as 12.011 for C, 1.008 for H, 14.007 for N, 15.999 for O, 32.060 for S, and 30.974 for P. The resolution power of the instrument should always be set to give maximum sensitivity and, if possible, only sufficient separation power to resolve the major peaks due to carbon isotopes for singly charged molecules (see Fig. 33). For multiply charged species, even this is impossible but, nevertheless, information about the chemical molecular mass can be deduced from such spectra (all the isotopes are essentially merged into one peak). It is only necessary to distinguish between different biological modifications (e.g., phosphorylation, glycosylation).

If instruments can, however, achieve much higher resolution (100 000 or more) without sacrificing sensitivity, the gain in information can be striking: since from m/z, the distance between isotope signals is $1/z$, the charge state of a multiply charged ion signal can be read directly simply by counting the peaks within one mass unit. For example, 10 peaks at m/z 800 – 801 means that this ion carries 10 charges. The distance between the isotope signals is, therefore, $^1/^{10}$ of the mass. This can be very important for MS/MS work, since the charge state of fragments from multiply charged ions is generally unknown and the information about the "true" molecular mass is obscured. When the charge is known, sequencing of peptides on the basis of MS/MS analysis of electrospray ions becomes feasible. This is the reason why in
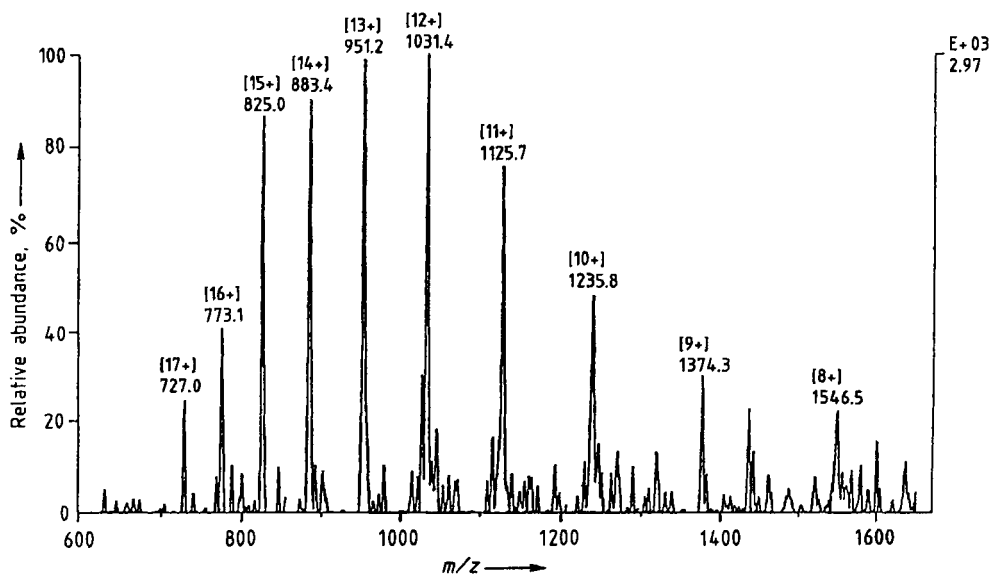
Figure 32. Electrospray spectrum of ca. 180 pmol cytochrome c using an MAT 90 sector field MS; homemade ESI source

the last few years FTMS with MALDI and electrospray have begun to be surprisingly employed in biomedical laboratories. The results reported on large peptides [145]–[147], [227]–[229] as well as on large oligonucleotides [147], [230]–[233] are so promising that even the high costs of such instruments cannot stop the development and not only research institutes but also industry have bought the FTMS instruments. As an illustration, data from the second isotope signal of a peptide show a resolution of 420 000, separating the isotope signals in this peak (Fig. 34).

In Figure 35, the product ions from a doubly charged tryptic digestion product illustrate the use. The product ion spectrum of the doubly protonated oligo-peptide Phe-Ser-Trp-Gly-Ala-Glu-Gly-Glu-Arg ($m/z = 520$) contains several significant singly charged sequence ions; this type of data is used for "mass tagging" in protein identification.

20.10.4. Species Analysis

The storage, transport, and action of metals in the environment is largely dependent on the chemical form such as oxidation state and association with organic ligands. Heavy metals in the environment are stored in complexes with humic acids, can be converted by microbes into different complexes, and may be transported into living animals

and human beings. This is true for many elements such as lead, mercury [234], arsenic [235], [236], astatine, tin, and platinum [237]. Several mass spectrometric techniques have been employed in the study of the fate of metals in the human body and other organic environments by the determination of the species formed in the biological or environmental matrices.

For example, tin and lead alkylates have been determined in soil, water, or muscle tissue with GC/MS after exhaustive alkylation or with thermospray [239], API [240], and ICP–LC/MS [244] methods. The ESI spectra of some tin compounds is given in Figure 36. The APIMS techniques have proven to be very successful, even with metals bound to proteins and enzymes [241] and, interfaced to microseparation techniques, even elements such as iodine [242] or phosphorus [243] can be determined quantitatively.

20.10.5. Determination of Elements

Mass spectrometry as a multielement technique has the advantage that many metal ions can be detected and quantitatively determined at once. In bulk materials such as steel [245] or refractory metals [246], elements can be determined by means of low-resolution glow-discharge
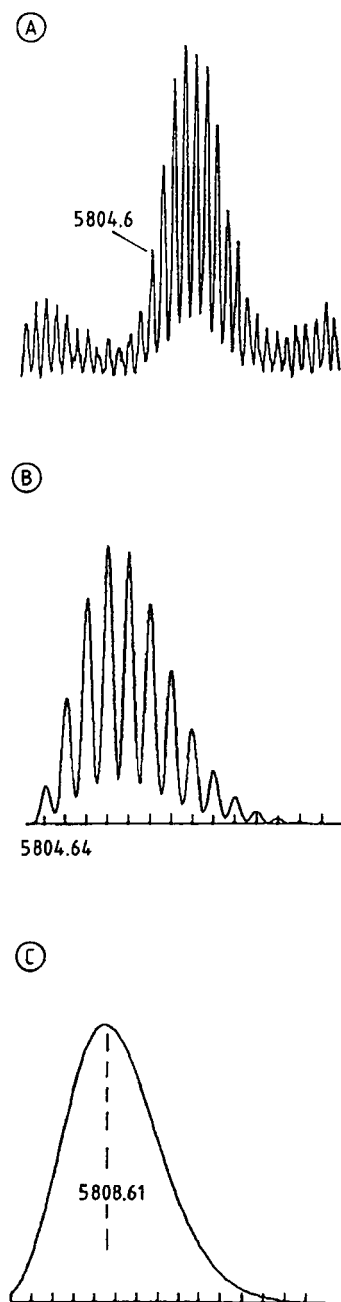
Figure 33. Effect of resolution on peak shapes
A) The protonated molecular cluster of human insulin
($C_{257}H_{384}N_{65}O_{77}S_6$) recorded with FABMS; B) Calculated at
a resolution of 6000 (10 % valley definition) with the
monoisotopic mass at $m/z = 5804.6455$; C) The unresolved
cluster with an average mass at $m/z = 5808.61$
(Reproduced from [226], with permission)

mass spectrometry. High-resolution GDMS has
been used to study semiconducting materials
[247]. GDMS is regarded as being virtually free
of matrix effects [248]. The state of the art in glow
discharge mass spectrometry has been reviewed
recently and from the data presented there it is
evident that the technique is a mature tool for
the material sciences [249]. If part of the material
can be dissolved prior to mass spectral analysis,
ICPMS or, for high precision isotope determina-
tion, thermal ionization mass spectrometry
(TIMS) can be applied. Detection limits in ICPMS
are listed in Table 2.

ICPMS and TIMS are also suitable for the
analysis of solutions such as biological fluids or
water samples. In combination with laser ablation,
bulk materials or small biological compartments
can be analyzed with ICPMS (see Section 20.4.5).
The desorbed neutrals, ions, and clusters are en-
trained in the gas, which carries the aerosol di-
rectly into the plasma. One important area of ap-
plication is geochronology where today ICPMS
and AMS are in use; for a recent comprehensive
description of this field see [250].

20.10.6. Surface Analysis and Depth Profiling

The characterization of material surfaces and
of the underlying layers is one of the most impor-
tant problems in modern material science, since
the properties of a material are not only dependent
on the bulk concentration of elements, but also on
the distribution on the surface, in the upper layers,
and at grain boundaries. Therefore, techniques
have been developed which can analyze the sur-
face and the next layers with very high resolution.
The first two monolayers, ca. 0.5 nm in depth, can
be investigated by secondary ion mass spectrome-
try (SIMS) [81]. Positive and negative ions from
the elements in the surface can be detected with a
sensitivity in the ppm range. It has found many
applications in the analysis of metals, alloys, semi-
conductors, ceramics, and glasses. The problem
with SIMS is that matrix effects can cause diffi-
culties in quantitation. It is possible to employ
SIMS for the analysis of organic compounds on
a metal surface. This technique, called static
SIMS, makes use of the sputtered organic ions
from material deposited on the surface (typically
Ag), in monolayers. Since the ion current is weak
and lasts for only a few seconds, time-of-flight
instruments are generally used.
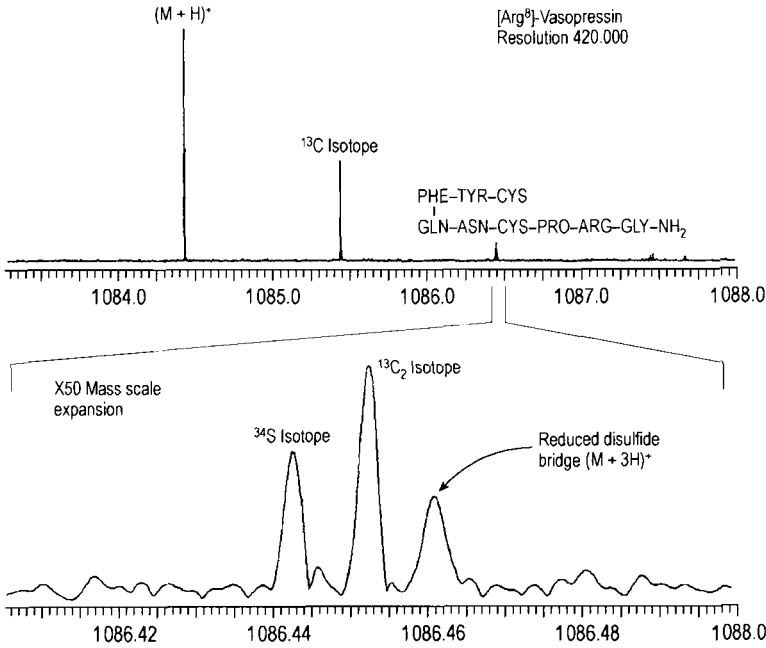
Figure 34. Top: 1 pmol of [arg8]-Vasopressin MALDI MS at a resolution of 420 000. Bottom: isobaric 34S (m/z 1086.4415) and 13C2 (m/z 1086.4524) and a third peak due to the reduced disulfide bridge are clearly distinguished
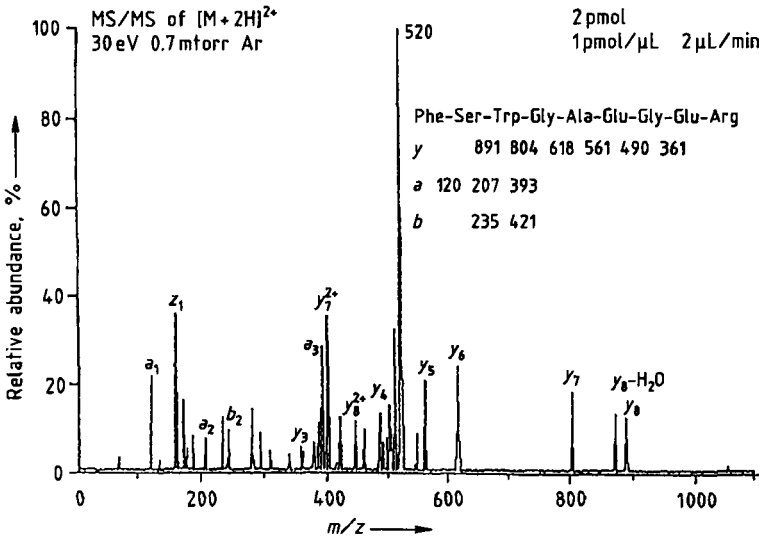
Figure 35. Doubly charged ions from tryptic digests of proteins
The fragments are labeled according to the scheme given in Figure 29 (Reproduced with permission from [238])

For the depth profiling of a metal, of particulate matter, or a ceramic with a resolution of a few nanometers, sputtered neutral mass spectrometry (SNMS) [251] is the best choice since the sputter-ing process is completely decoupled from the ionization, and quantitation is possible by calibrating the instrument with standard reference materials. The technique has found many applications for the
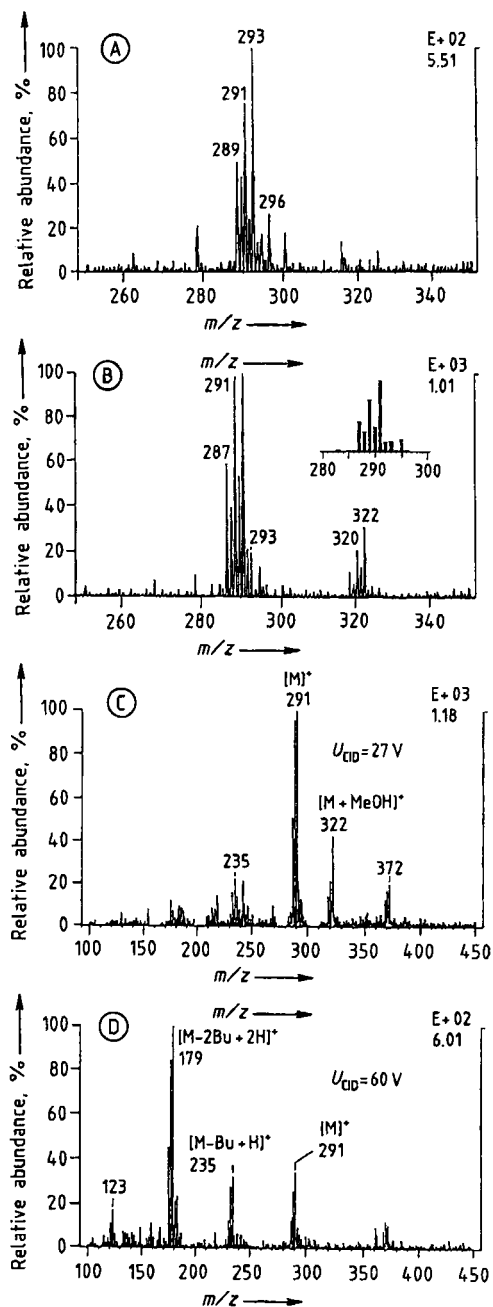
Figure 36. ESI spectra of some tin compounds
A) Dibutyl tin acetate (DBT) $(C_4H_9)_2Sn^{2+}(OCOCH_3)^-$;
B) Tributyl tin (TBT) $(C_4H_9)_3Sn^+$ cations with theoretical
distribution given in the inset; C) TBT with small CAD
voltage; D) TBT with large CAD voltage showing fragments
due to loss of the side chains
Instrument: MAT 90, electrospray MS, flow injection
Upper spectrum: CAD voltage 27 V; Lower spectrum: CAD
voltage 60 V

Table 2. Detection limits in ICPMS [65]

Element	Detection limit, pg/mL	Element	Detection limit, pg/mL
Li	100	Sn	60
B	400	Sb	60
F	1.1×10^5	Te	80
Mg	130	I	10
Al	160	Cs	20
P	2×10^4	Ba	150
S	1×10^5	La	10
Cl	1×10^3	Ce	10
Ti	320	Pr	10
V	80	Nd	45
Cr	10	Sm	30
Mn	30	Eu	15
Co	10	Gd	50
Ni	40	Tb	10
Cu	20	Dy	35
Zn	10	Ho	10
Ge	20	Er	35
As	40	Tm	10
Se	800	Yb	30
Br	1×10^3	Lu	10
Y	50	Hf	10
Zr	30	Ta	50
Mo	40	W	50
Ru	60	Au	60
Ag	30	Hg	20
Cd	30	Pb	10
In	60	Th	20
		U	10

characterization of semiconductors and implanted atoms. Figure 37 gives an example compared to a result using GDMS.

The determination of the spatial distributions of elements is becoming increasingly important since the grain boundaries in ceramics or the metal distribution in microelectronic devices have a major influence on the stability of the material and the performance of the device. With a highly focused primary ion beam it is possible to scan the surface in two dimensions and the distribution of the elements on the surface. Combinations of sputtering processes and microprobing allow this distribution to be investigated in depth. Instruments used for this type of analysis are called ion microprobe mass analyzers (IMMA). The spot of the primary beam has a diameter of less than 10 μm, thus defining the resolution obtainable.
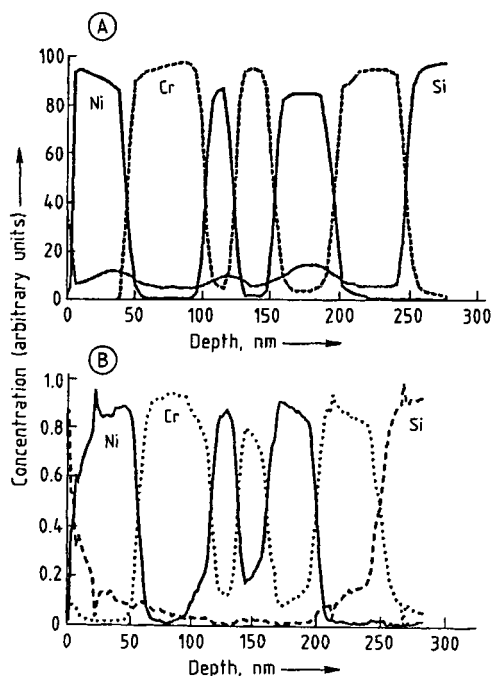
Figure 37. Depth profiles of a multilayer sample (Ni, Cr 50nm, Ni, Cr 25nm, Ni, Cr 50nm)
A) SNMS; B) GDMS
(Reproduced with permission from [252])

20.11. References

[1] F. W. Aston: *Mass Spectra and Isotopes*. Edward Arnold & Co, London 1933.

[2] A. J. Dempster. *Phys. Rev.* **11** (1918) 316.

[3] H. W. Washburn, H. F. Wiley, S. M. Rock, *Ind. Eng. Chem. Anal. Ed.* **15** (1943) 541–547.

[4] R. A. Brown, R. C. Taylor, F. W. Melpolder, W. S. Young, *Anal. Chem.* **20** (1948) 5–9.

[5] M. J. O'Neal, Jr., *Anal. Chem.* **22** (1950) 991–995.

[6] A. J. Campbell, J. S. Halliday in R. M. Elliott (ed.): *13th Annual Symposium on Mass Spectrometry*, vol. 2, Pergamon Press, Oxford 1962.

[7] H. Budzikiewicz, C. Djerassi, D. H. Williams: *Structure Elucidation of Natural Products by Mass Spectrometry*, Holden Day, San Francisco 1964.

[8] G. R. Waller (ed.): *Biochemical Applications of Mass Spectrometry*, Wiley Interscience, New York 1972.

[9] D. Kanne, K. S. Straub, J. E. Hearst, H. Rapoport, *J. Am. Chem. Soc.* **103** (1982) 6754–6764.

[10] H. M. Rosenstock, M. B. Wallenstein, A. L. Wahrhaftig, H. Eyring, *Proc. Natl. Acad. Sci. USA* **38** (1952) 667.

[11] R. A. Marcus, O. K. Rice, *J. Phys. Colloid. Chem.* **55** (1951) 894.

[12] K. Levsen: *Fundamental Aspects of Organic Mass Spectrometry*, VCH Verlagsgesellschaft, Weinheim 1978.

[13] F. W. McLafferty, F. Turecek: *Interpretation of Mass Spectra*, University Science Books, Mill Valley, California 1993.

[14] J. J. Boon, *Int. J. Mass Spectrom. Ion Processes* **118/119** (1992) 755–787.

[15] W. Genuit, J. J. Boon, O. Faix, *Anal. Chem.* **59** (1987) 508–513.

[16] W. H. McFadden: *Techniques of Combined Gas Chromatography/Mass Spectrometry*, Wiley & Sons, New York 1973.

[17] W. A. Brand, *J. Mass Spectrom.* **31** (1996) 225.

[18] R. M. Caprioli, T. Fan, *Biochem. Biophys. Res. Commun.* **141** (1986) 1058.

[19] Y. Ito, T. Takeuchi, D. Ishii, M. Goto, *J. Chromatogr.* **346** (1985) 161.

[20] M. A. Moseley et al., *Anal. Chem.* **61** (1989) 1577.

[21] R. M. Caprioli et al., *J. Chromatogr.* **480** (1989) 247.

[22] M. A. Moseley, L. J. Deterding, K. B. Tomer, J. W. Jorgenson, *Rapid Commun. Mass Spectrom.* **3** (1989) 87.

[23] N. J. Reinhoud et al., *J. Chromatogr.* **516** (1990) 147.

[24] M. A. Baldwin, F. W. McLafferty, *Org. Mass Spectrom.* **7** (1973) 1111.

[25] V. L. Tal'roze, G. V. Karpov, I. G. Gordetskii, V. E. Skurat, *Russ. J. Phys. Chem. (Engl. Transl.)* **42** (1968) 1658.

[26] J. D. Henion, *J. Chromatogr. Sci.* **19** (1981) 57.

[27] R. D. Smith, W. D. Felix, J. C. Fjedsted, M. L. Lee, *Anal. Chem.* **54** (1982) 1885.

[28] R. C. Willoughby, R. F. Browner, *Anal. Chem.* **56** (1984) 2625.

[29] E. C. Horning, M. G. Horning, D. I. Carroll, R. N. Stillwell, I. Dzidic, *Life Sci.* **13** (1973) 1331.

[30] J. Sunner, G. Nicol, P. Kebarle, *Anal. Chem.* **60** (1988) 1300.

[31] J. Sunner, M. G. Ikonomou, P. Kebarle, *Anal. Chem.* **60** (1988) 1308.

[32] C. R. Blakley, J. C. Carmody, M. L. Vestal, *Clin. Chem. (Winston Salem N.C.)* **26** (1980) 1467.

[33] J. V. Iribarne, B. A. Thomson, *J. Chem. Phys.* **64** (1976) 2287.

[34] B. A. Thomson, J. V. Iribarne, *J. Chem. Phys.* **71** (1979) 4451.

[35] G. Schmelzeisen-Redeker, L. Bütfering, F. W. Röllgen, *Int. J. Mass Spectrom. Ion Processes* **90** (1989) 139.

[36] M. L. Vestal, *Int. J. Mass Spectrom. Ion Phys.* **46** (1983) 193.

[37] J. Hau, W. Nigge, M. Linscheid, *Org. Mass Spectrom.* **28** (1993) 223–229.

[38] M. Linscheid, D. Westmoreland, *Pure Appl. Chem.*, (1994) in press.

[39] M. A. Niessen, J. van der Greef: *Liquid Chromatography-Mass Spectroscopy*, Marcel Dekker, New York 1992.

[40] R. D. Smith et al., *Anal. Chem.* **65** (1993) 574A–584A.

[41] W. M. Niessen, *J. Chromatogr. A* **856** (1999) 179.

[42] W. M. A. Niessen, A. P. Tinke, *J. Chromatogr. A* **703** (1995) 37.

[43] M. R. Emmett, F. M. White, C. L. Hendrickson, S.-H. Shi, A. G. Marshall, *J. Am. Soc. Mass Spectrom.* **9** (1997) 333.

[44] I. M. Lazar, A. L. Rockwood, E. D. Lee, J. C. H. Sin, M. L. Lee, *Anal. Chem.* **71** (1999) 2578.

[45] R. D. Smith, J. A. Olivares. N. T. Ngyen, H. R. Udseth, *Anal. Chem.* **60** (1988) 436–441.
[46] C. G. Edmonds et al., *Biochem. Soc. Trans.* **19** (1991) 943.
[47] S. A. Hofstadler, F. D. Swanek, D. C. Gale, A. G. Ewing, R. D. Smith, *Anal. Chem.* **67** (1995) 1477.
[48] J. C. Severs, S. A. Hofstadler, Z. Zhao, R. T. Senh, R. D. Smith, *Electrophoresis* **17** (1996) 1808.
[49] I. M. Johannson, E. C. Huang, J. D. Henion, J. Zweigenbaum, *J. Chromatogr.* **554** (1991) 311.
[50] A. L. Burlingame, J. A. McCloskey, *Biol. Mass Spectrom., Proc. Int. Symp. Mass Spectrom. Health Life Sci.*, Elsevier, Amsterdam 1990.
[51] R. S. Ramsey, J. M. Ramsey, *Anal. Chem.* **69** (1997) 1174.
[52] R. D. Oleschuk, D. J. Harrison, *Trends Anal. Chem.* **19** (2000) 379.
[53] K. L. Busch, *J. Planar Chromatogr.-Mod. TLC* **2** (1989) 355.
[54] A. Hayashi, T. Matsubara, Y. Nishizawa, T. Hattori, *Iyo Masu Kenkyukai Koenshu* **11** (1986) 147.
[55] L. Li, D. M. Lubman, *Anal. Chem.* **61** (1989) 1911.
[56] K. L. Bush, *TrAc Trends Anal. Chem. (Pers. Ed.)* **11** (1992) 314.
[57] I. D. Wilson, *J. Chromatogr. A* **856** (1999) 429.
[58] J. T. Mehl, D. M. Hercules, *Anal. Chem.* **72** (2000) 68.
[59] Y. Liu, J. Bai, X. Liang, D. M. Lubman, *Anal. Chem.* **67** (1995) 3482.
[60] M. Schreiner, K. Strupat, F. Lottspeich, C. Eckerskorn, *Electrophoresis* **17** (1996) 954.
[61] C. Eckerskorn et al., *Anal. Chem.* **69** (1997) 2888.
[62] D. Schleuder, F. Hillenkamp, K. Strupat, *Anal. Chem.* **71** (1999) 3238.
[63] A. G. Harrison: *Chemical Ionization Mass Spectrometry*, CRC Press, Boca Raton 1982.
[64] G. B. Mohamed et al., *J. Chromatogr.* **314** (1984) 211–217.
[65] H. D. Beckey: *Principles of Field Ionization and Field Desorption Mass Spectrometry*, Pergamon Press, Frankfurt 1977.
[66] G. M. Hieftje, L. A. Norman, *Int. J. Mass Spectrom. Ion Processes* **118/119** (1992) 519–573.
[67] W. W. Harrison, *J. Anal. At. Spectrom.* **7** (1992) 75–79.
[68] W. W. Harrison et al., "Glow Discharge Techniques in Analytical Chemistry," *Anal. Chem.* **62** (1990) 943A–949A.
[69] H. Oechsner, *Pure Appl. Chem.* **64** (1992) 615.
[70] A. Montaser, H. Tan, I. Ishii, S. H. Nam, M. Cai, *Anal. Chem.* **63** (1991) 2660.
[71] N. Jakubowski, L. Moens, F. Vanhaecke, *Spectrochim. Acta B: At. Spectrosc.* **53** (1998) 1739.
[72] M. Morita, H. Ito, M. Linscheid, K. Otsuka, *Anal. Chem.* **66** (1994) 1588.
[73] K. E. Milgram et al., *Anal. Chem.* **69** (1997) 3714.
[74] J. R. Eyler, R. S. Houk, C. H. Watson, K. E. Milgram: "Development and Application of Inductively Coupled Plasma (ICP)-FTICR Mass Spectrometry," *46th ASMS Conference*, Orlando, Fl 1998.
[75] N. Jakubowski, D. Stuewer, unpublished results.
[76] K. G. Heumann, *Int. J. Mass Spectrom. Ion Processes* **118/119** (1992) 575–592.
[77] W. Genuit, Chen He-Neng, A. J. H. Boerboom, J. Los, *Int. J. Mass Spectrom. Ion Phys.* **51** (1983) 207–213.
[78] U. Boesl, J. Grotemeyer, K. Walter, E. W. Schlag, *Anal. Instrum. (N. Y.)* **16** (1987) 151.
[79] J. Grotemeyer, U. Boesl, K. Walter, E. W. Schlag, *Org. Mass Spectrom.* **21** (1986) 645.
[80] J. Grotemeyer, E. W. Schlag, *Acc. Chem. Res.* **22** (1989) 399.
[81] A. Benninghoven, *Z. Phys.* **230** (1970) 403.
[82] L. Schwieters et al.: "Secondary Ion Mass Spectrometry," in A. Benninghoven, K. T. F. Janssen, J. Tümpner, H. W. Werner (eds.): *Proc. 8th Int. Conf. (SIMS VIII)* **1992**, 497.
[83] P. M. Thompson, *Anal. Chem.* **63** (1991) 2447–2456.
[84] B. Hagenhoff et al., *J. Phys. D, Appl. Phys.* **25** (1992) 818.
[85] F. W. Roellgen in A. Benninghoven (ed.): *Ion Formation from Organic Solids*, Springer-Verlag, Berlin 1983, p. 2.
[86] M. Barber, R. S. Bordoli, R. D. Sedgwick, A. N. Tyler, *J. Chem. Soc. Chem. Commun.* 1981, 325–327.
[87] R. D. McFarlane, D. F. Torgerson, *Science (Washington, D.C. 1883)* **191** (1976) 920.
[88] R. D. McFarlane: "Methods in Enzymology," vol. 193 in J. A. McCloskey (ed.): *Mass Spectrometry*, Academic Press, San Diego 1990, p. 263–280.
[89] G. Jonsson et al., *Rapid Commun. Mass Spectrom.* **3** (1989) 190.
[90] A. H. Verbueken, F. J. Bryunseels, R. E. van Grieken, *Biomed. Mass Spectrom.* **12** (1985) 438.
[91] J. S. Becker, H. J. Dietze, *Fresenius J. Anal. Chem.* **344** (1992) 69–86.
[92] F. Hillenkamp: "Ion Formation from Organic Solids II," in A. Benninghoven (ed.): *Springer Series in Chem. Phys.* **1983**, no. 25, p. 190.
[93] M. Karas, U. Bahr, A. Ingendoh, F. Hillenkamp, *Angew. Chem.* **101** (1989) 805–806.
[94] M. Karas, D. Bachmann, F. Hillenkamp, *Anal. Chem.* **57** (1985) 2935.
[95] M. Karas, F. Hillenkamp, *Anal. Chem.* **60** (1988) 2299–2301.
[96] K. Tanaka et al., *Rapid Commun. Mass Spectrom.* **8** (1988) 151.
[97] F. Kirpekar, S. Berkenkamp, F. Hillenkamp, *Anal. Chem.* **71** (1999) 2334.
[98] R. Kaufmann, T. Wingerath, D. Kirsch, W. Stahl, H. Sies, *Anal. Biochem.* **238** (1996) 117.
[99] P. Juhasz et al., *Anal. Chem.* **68** (1996) 941.
[100] U. Bahr, J. Stahl-Zeng, E. Gleitsmann, M. Karas, *J. Mass Spectrom.* **32** (1997) 1111.
[101] R. Kaufmann, *J. Biotechnol.* **41** (1995) 155.
[102] B. Sprengler, *J. Mass Spectrom.* **32** (1997) 1019.
[103] P. Chaurand, F. Luetzenkirchen, B. Spengler, *J. Am. Soc. Mass Spectrom.* **10** (1999) 91.
[104] M. Dole et al., *J. Chem. Phys.* **49** (1968) 2240.
[105] M. Yamashita, J. B. Fenn, *J. Phys. Chem.* **88** (1984) 4451–4459.
[106] C. M. Whitehouse, R. N. Dreyer, M. Yamashita, J. B. Fenn, *Anal. Chem.* **57** (1985) 675–679.
[107] M. L. Alexandrov, L. N. Gall', N. V. Krasnov, V. I. Nikolaev, V. A. Shkurov, *Zh. Anal. Khim.* **40** (1985) 1227–1236.
[108] L. Tang, P. Kebarle, *Anal. Chem.* **65** (1993) 3654–3668.
[109] S. K. Chowdhury, V. Katta, B. T. Chait, *Rapid Commun. Mass Spectrom.* **4** (1990) 81–87.
[110] A. P. Bruins, T. R. Covey, J. D. Henion, *Anal. Chem.* **59** (1987) 2642–2646.
[111] M. Wilm, M. Mann, *Anal. Chem.* **68** (1996) 1.
[112] A. J. Dempster, *Phys. Rev.* **11** (1918) 316.
[113] C. Brunnée, *Int. J. Mass Spectrom. Ion Processes* **76** (1987) 121–237.

[114] E. W. Blauth: *Dynamische Massenspektrometer*, Vieweg & Sohn, Braunschweig 1965.

[115] R. Kaufmann, B. Spengler, F. Lützenkirchen, *Rapid Commun. Mass Spectrom.* **7** (1993) 902.

[116] I. V. Chernushevich, W. Ens, K. G. Standing, *Anal. Chem.* **71** (1999) 452A.

[117] I. M. Lazar, R. S. Ramsey, S. Sundberg, J. M. Ramsey, *Anal. Chem.* **71** (1999) 3627.

[118] M. B. Comisarow, A. G. Marshall, *Chem. Phys. Lett.* **26** (1974) 489;M. B. Comisarow, A. G. Marshall, *J. Chem. Phys.* **64** (1976) 110.

[119] C. B. Lebrilla, D. T. S. Wang, R. L. Hunter, R. T. McIver, Jr., *Anal. Chem.* **62** (1990) 878–880.

[120] C. F. Ijames, C. F. Wilkins, *J. Am. Chem. Soc.* **110** (1988) 2687–2688.

[121] A. G. Marshall, *Acc. Chem. Res.* **18** (1985) 316–322.

[122] C. Giancaspro, F. R. Verdun, *Anal. Chem.* **58** (1986) 2097.

[123] R. T. McIver, R. L. Hunter, W. D. Bowers, *Int. J. Mass Spectrom. Ion Processes* **64** (1985) 67.

[124] P. Kofel, M. Allemann, H. Kellerhals, K. P. Wanczek, *Int. J. Mass Spectrom. Ion Processes* **72** (1986) 53.

[125] P. Grossmann, P. Caravatti, M. Allemann, H. P. Kellerhals, *Proc. 36th ASMS Conf. Mass Spectrom. All. Top.* 1988, 617.

[126] W. Paul, H. Steinwedel, *Z. Naturforsch. A*: Phys. Phys. Chem. Kosmophys. 8 (1953) 448.

[127] G. C. Stafford et al., *Int. J. Mass Spectrom. Ion Processes* **60** (1984) 85.

[128] B. D. Nourse, R. G. Cooks, *Anal. Chim. Acta* **228** (1990) 1.

[129] H.-J. Laue, K. Habfast, *Proc. 37th ASMS Conf. Mass Spectrom. All. Top.* 1989, p. 1033.

[130] P. De Bievre, *Fresenius J. Anal. Chem.* **337** (1990) 767.

[131] R. E. M. Hodges in J. F. J. Todd (ed.): *Proc. 10th Int. Mass. Spectrom. Conf. Swansea 1985*, Wiley, London 1986, p. 185.

[132] W. Aberth, *Biomed. Mass Spectrom.* **7** (1980) 367.

[133] H. Matsuda, H. Wollnik, *Int. J. Mass Spectrom. Ion Processes* **86** (1988) 53.

[134] M. Barber, R. M. Elliott, paper presented at ASTM E-14 Conference on Mass Spectrometry, Montreal 1964.

[135] K. L. Busch, G. L. Glish, S. A. McLuckey: *Mass Spectrometry/Mass Spectrometry: Techniques and Applications of Tandem Mass Spectrometry*, VCH Verlagsgesellschaft, Weinheim 1988.

[136] M. A. Seeterlin, P. R. Vlasak, D. J. Beussmann, R. D. McLane, C. G. Enke, *J. Am. Soc. Mass Spectrom.* **4** (1993) 751–754.

[137] S. R. Horning, J. M. Wood, J. R. Gord, B. S. Freiser, R. G. Cooks, *Int. J. Mass Spectrom. Ion Processes* **101** (1990) 219.

[138] S. A. McLuckey, D. E. Goeringer, *J. Mass Spectrom.* **32** (1997) 461.

[139] R. A. Yost, C. G. Enke, *J. Am. Chem. Soc.* **100** (1978) 2274.

[140] R. E. March, R. J. Hughes, *Quadrupole Storage Mass Spectrometry*, Wiley Interscience, New York 1989.

[141] C. Hagg, I. Szabo, *Int. J. Mass Spectrom. Ion Processes* **73** (1986) 295.

[142] K. Sato et al., *Anal. Chem.* **59** (1987) 1652.

[143] J. A. Hill et al., *Int. J. Mass Spectrom. Ion Processes* **92** (1989) 211.

[144] I. J. Amster, *J. Mass Spectrom.* **31** (1996) 1325.

[145] V. C. M. Dale et al., *Biochem. Mass Spectrom. Biochem. Soc. Trans.* **24** (1996) 943.

[146] M. W. Senko, J. P. Speir, F. W. McLafferty, *Anal. Chem.* **66** (1994) 2801.

[147] E. R. Williams, *Anal. Chem.* **10** (1998) 179A.

[148] A. Ingendoh et al., *J. Mass Spectrom. Soc. Jpn.* **45** (1997) 247.

[149] R. E. March, *J. Mass Spectrom.* **32** (1997) 351.

[150] R. J. Strife, M. Bier, J. Zhou, *Proc. 43rd ASMS Conference on Mass Spectrometry and Allied Topics, (Atlanta),* (1995) 1113.

[151] J. C. Schwartz, M. Bier, D. M. Taylor, J. Zhou, *Proc. 43rd ASMS Conference on Mass Spectrometry and Allied Topics, (Atlanta),* (1995) 1114.

[152] E. R. Badman, J. M. Wells, H. A. Bui, R. G. Cooks, *Anal. Chem.* **70** (1998) 3545.

[153] H. R. Morris et al., *Rapid Commun. Mass Spectrom.* **10** (1996) 889.

[154] C. Borchers, C. E. Parker, L. J. Deterding, K. B. Tomer, *J. Chromatogr. A* **854** (1999) 119.

[155] T. J. Cornish, R. J. Cotter, *Anal. Chem.* **65** (1993) 1043–1047.

[156] F. H. Strobel, L. M. Preston, K. S. Washburn, D. H. Russell, *Anal. Chem.* **64** (1992) 754–762.

[157] M. Linscheid, *Fresenius Z. Anal. Chem.* **337** (1990) 648–661.

[158] N. R. Daly, *Rev. Sci. Instrum.* **34** (1963) 1116.

[159] C. E. Griffin, H. G. Boettger, D. D. Norris, *Int. J. Mass Spectrom. Ion Phys.* **15** (1974) 437.

[160] J. S. Cottrell, S. Evans, *Anal. Chem.* **59** (1987) 1990–1995.

[161] C. E. D. Ouwerkerk, A. J. H. Boerboom, T. Matsuo, T. Sakurai, *Int. J. Mass Spectrom. Ion Phys.* **70** (1986) 79–96.

[162] R. Pesch, G. Jung, K. Rostand, K.-H. Tietje, *Proc. 37th ASMS Conf. Mass Spectrom. All. Top.* 1989, 1079.

[163] S. Bauerschmidt, W. Hanebeck, K.-P. Schulz, J. Gasteiger, *Anal. Chim. Acta* **265** (1992) 169–182.

[164] J. Gasteiger, W. Hanebeck, K.-P. Schulz, *I. Chem. Inf. Comp. Sci.* **32** (1992) 264–271.

[165] K. Pfleger, H. H. Maurer, A. Weber: *Mass Spectral and GC Data of Drugs, Poisons, Pesticides, Pollutants and their Metabolites,* 2nd ed., VCH Verlagsgesellschaft, Weinheim 1992.

[166] H. Damen, D. Henneberg, B. Weimann, *Anal. Chim. Acta* **103** (1978) 289.

[167] D. Henneberg, *Anal. Chim. Acta* **150** (1983) 37.

[168] G. M. Pesyna, R. Venkataraghavan, H. G. Dayringer, F. W. McLafferty, *Anal. Chem.* **48** (1976) 1362.

[169] B. L. Atwater, D. B. Stauffer, F. W. McLafferty, D. W. Peterson, *Anal. Chem.* **57** (1985) 899.

[170] S. Sokolov, J. Karnovsky, P. Gustafson, *Finnigan Appl. Rep.* no. 2, 1978.

[171] D. Henneberg, B. Weimann, U. Zalfen, *Org. Mass Spectrom.* **28** (1993) 198–206.

[172] M. J. Hayward, P. V. Robandt, J. T. Meek, M. L. Thomson, *J. Am. Soc. Mass Spectrom.* **4** (1993) 742–750.

[173] T. Kotiaho, *J. Mass Spectrom.* **31** (1996) 1.

[174] P. Ciccioli, A. Cecinato, E. Brancaleoni, M. Frattoni, A. Liberti, *HRC CC J. High Resolut. Chromatogr.* **15** (1992) 75–84.

[175] T. Hoffmann, P. Jacob, M. Linscheid, D. Klockow, *Int. J. Environ. Anal. Chem.* **52** (1993) 29–37.

[176] A. Robbat, C. J. Liu, T. Y. Liu, *J. Chromatogr.* **625** (1992) 277–288.

[177] P. J. McKeown, M. V. Johnston, D. M. Murphy, *Anal. Chem.* **63** (1991) 2073–2075.

[178] K. G. Asano, G. L. Glish, S. A. McLuckey, *Proc. 36th ASMS Conf. Mass Spectrom. All. Top.* 1988, 637.

[179] T. Hoffmann, R. Bandur, U. Marggraf, M. Linscheid, *J. Geophys. Res. D: Atmospheres* **103** (1998) 25.

[180] M. A. LaPack, J. C. Tou, C. G. Enke, *Anal. Chem.* **62** (1990) 1265–1271.

[181] R. Zimmermann, H. J. Heger, A. Kettrup, U. Boesl, *Rapid Commun. Mass Spectrom.* **11** (1997) 1095.

[182] R. Zimmermann et al., *Rapid Commun. Mass Spectrom.* **13** (1999) 307.

[183] U. Boesl, *J. Mass Spectrom.* **35** (2000) 289.

[184] G. Durand, N. de Bertrand, D. Barcelo, *J. Chromatogr.* **554** (1991) 233–250.

[185] H. Fr. Schröder, *J. Chromatogr.* **554** (1991) 251–266.

[186] M. A. Brown, I. S. Kim, F. I. Sasinos, R. D. Stephens: "Liquid Chromatography/Mass Spectrometry, Applications in Agricultural, Pharmaceutical and Environmental Chemistry," in M. A. Brown (ed.): *ACS Symp. Ser.* **420** (1990) 199–214.

[187] A. K. Lister, R. G. Cooks, *Proc. 36th ASMS Conf. Mass Spectrom. All. Top.* 1988, 646.

[188] C. Rappe, S. Marklund, R.-A. Berqvist, M. Hanson in G. Choudhary, L. H. Keith, C. Rappe (eds.): *Chlorinated Dioxins and Dibenzofurans in the Total Environment*, Butterworth, Stoneham 1983, chap. 7.

[189] R. M. Smith et al., *Environ. Sci. Res.* **26** (1983) 73–94.

[190] L. Q. Huang, A. Paiva, H. Tond, S. J. Monson, M. L. Gross, *J. Am. Soc. Mass Spectrom.* **3** (1992) 248–259.

[191] E. R. Barnhart et al., *Anal. Chem.* **59** (1987) 2248–2252.

[192] W. A. Korfmacher, G. F. Moler, R. R. Delongchamp, R. K. Mitchum, R. L. Harless, *Chemosphere* **131** (1984) 669–685.

[193] M. Linscheid, *Int. J. Environm. Anal. Chem.* **49** (1992) 1–14.

[194] T. L. Jones, L. D. Betowski, J. Yinon: "Liquid Chromatography/Mass Spectrometry, Applications in Agricultural, Pharmaceutical and Environmental Chemistry," in M. A. Brown (ed.): *ACS Symp. Ser.* **420** (1990) 62.

[195] W. M. Draper, F. R. Brown, R. Bethem, M. J. Miille: "Liquid Chromatography/Mass Spectrometry, Applications in Agricultural, Pharmaceutical and Environmental Chemistry," in M. A. Brown (ed.): *ACS Symp. Ser.* **420** (1990) 253.

[196] A. Robbat, T. Y. Liu, B. M. Abraham, *Anal. Chem.* **64** (1992) 1477–1483.

[197] J. A. McCloskey (ed.): "Mass Spectrometry," *Methods in Enzymology*, **vol. 193**, Academic Press, San Diego 1990.

[198] H. M. Leith, P. L. Truran, S. J. Gaskell, *Biomed. Environm. Mass Spectrom.* **13** (1986) 257–261.

[199] C. H. L. Shackleton in H. Jaeger (ed.): *Glass Capillary Chromatography in Clinical Medicine and Pharmacology*, Marcel Dekker, New York 1985.

[200] A. Frigerio, C. Marchioro, A. M. Pastorino in D. D. Breimer, P. Speiser (eds.): *Top. Pharm. Sci. Proc. Int. Congr. Pharm. Sci. F.I.P.* **45th** (1985) 297–310.

[201] H. Schweer, C. O. Meese, O. Fuerst, P. K. Gonne, H. J. W. Seyberth, *Anal. Biochem.* **164** (1987) 156–163.

[202] H. R. Morris, D. H. Williams, R. P. Ambler, *Biochem. J.* **125** (1971) 189.

[203] V. N. Reinhold in S. J. Gaskell (ed.): *Mass Spectrometry in Biomedical Research*, J. Wiley & Sons, Chichester 1986, p. 181.

[204] D. Rolf, G. R. Gray, *Carbohydr. Res.* **137** (1985) 183–196.

[205] H. Björndal, C. G. Hellerqvist, B. Lindberg, S. Svensson, *Angew. Chem.* **82** (1979) 643.

[206] K. Biemann, *Pract. Spectrosc.* **8** (1990) 3–24.

[207] K. Biemann in J. A. McCloskey (ed.): "Mass Spectrometry," *Methods in Enzymology*, **vol. 193**, Academic Press, San Diego 1990, p. 886–887.

[208] S. A. Carr, V. N. Reinhold, *Biomed. Mass Spectrom.* **11** (1984) 633–642.

[209] A. Dell, C. E. Ballou, *Carbohydr. Res.* **120** (1983) 95–111.

[210] B. A. Andersson, W. W. Christie, R. T. Holman, *Lipids* **10** (1975) 215–219.

[211] P. Sperling, M. Linscheid, S. Stöcker, H.-P. Mühlbach, E. Heinz, *J. Biol. Chem.* **268** (1993) 26935–26940.

[212] M. Suzuki, T. Ariga, M. Sekine, E. Araki, T. Miyatake, *Anal. Chem.* **53** (1981) 985–988.

[213] N. J. Jensen, K. B. Tomer, M. L. Gross, *J. Am. Chem. Soc.* **107** (1985) 1863.

[214] H. Schweer, H. W. Seyberth, C. O. Meese, O. Fürst, *Biomed. Environm. Mass Spectrom.* **15** (1988) 143–151.

[215] H.-R. Schulten in H. R. Morris (ed.): *Soft Ionization of Biological Substrates*, Heyden & Sons, London 1981.

[216] M. Linscheid, G. Feistner, H. Budzikiewicz, *Isr. J. Chem.* **17** (1978) 163.

[217] M. Linscheid, J. D'Angona, A. L. Burlingame, A. Dell, C. Ballou, *Proc. Natl. Acad. Sci. USA* **78** (1981) 1471–1475.

[218] D. Suckau, M. Mak, M. Przybylski, *Proc. Natl. Acad. Sci. USA* **89** (1992) 5630–5634.

[219] L. Weidolf, T. R. Covey, *Rapid Commun. Mass Spectrom.* **6** (1992) 192–196.

[220] R. D. Smith, K. J. Light-Wahl, *Biol. Mass Spectrom.* **22** (1993) 493–501.

[221] B. Ganem, Y.-T. Li, J. D. Henion, *J. Am. Chem. Soc.* **113** (1991) 6294–6296.

[222] B. Ganem, Y.-T. Li, J. D. Henion, *Tetrahedron Lett.* **34** (1993) 1445–1448.

[223] M. G. Ikonomou, A. Naghipur, J. W. Lown, P. Kebarle, *Biomed. Environm. Mass Spectrom.* **19** (1990) 434–446.

[224] C. E. Costello, K. M. Comess, A. S. Plaziak, D. P. Bancroft, S. J. Lippard, *Int. J. Mass Spectrom. Ion Processes* **122** (1992) 255–279.

[225] J. Yergey, D. Heller, G. Hansen, R. J. Cotter, C. Fenselau, *Anal. Chem.* **55** (1983) 353–356.

[226] K. Biemann in J. A. McCloskey (ed.): "Mass Spectrometry," *Methods in Enzymology*, vol. 193, Academic Press, San Diego 1990, p. 295–305.

[227] J. Yao, M. Dey, S. J. Pastor, C. L. Wilkins, *Anal. Chem.* **67** (1995) 3638.

[228] Y. Li, R. T. McIver, R. L. Hunter, *Anal. Chem.* **66** (1994) 2077.

[229] C. C. Stacey et al., *Rapid Commun. Mass Spectrom.* **8** (1994) 513.

[230] Y. Li, K. Tang, D. P. Little, H. Köster, R. L. Hunter, R. T. M. Jr, *Anal. Chem.* **68** (1996) 2090.

[231] E. A. Stemmler, M. V. Buchanan, G. B. Hurst, R. L. Hettich, *Anal. Chem.* **67** (1995) 2924.

[232] E. A. Stemmler, M. V. Buchanan, G. B. Hurst, R. L. Hettich, *Anal. Chem.* **66** (1994) 1274.

[233] H. Yoshida. R. L. Buchanan. *Radiat. Res.* **139** (1994) 271.

[234] R.-D. Wilken, R. Falter. *Appl. Organomet. Chem.* **12** (1998) 551.

[235] L. Ebdon, A. Fischer, N. B. Roberts, M. Yaqoob, *Appl. Organomet. Chem.* **13** (1999) 183.

[236] Y. Inoue et al., *Appl. Organomet. Chem.* **13** (1999) 81.

[237] C. Siethoff, B. Pongraz, M. Linscheid, *Proc. 44th ASMS Conference on Mass Spectrometry and Allied Topics* (Portland, USA), (1996) 987.

[238] I. Jardine in J. Villafranca (ed.): *Current Research in Protein Chemistry,* Academic Press, San Diego 1991.

[239] N. Blaszkiewicz, G. Baumhoer, B. Neidhart, R. Ohlendorf, M. Linscheid, *J. Chrom.* **439** (1988) 109–119.

[240] K. W. M. Siu, G. J. Gardner, S. S. Berman, *Rapid Commun. Mass Spectrom.* **2** (1988) 201–204.

[241] Y. Morita, K. Takatera, T. Watanabe, *Seisan Kenkyu* **43** (1991) 367.

[242] B. Michalke, P. Schramel, *Electrophoresis* **20** (1999) 2547.

[243] C. Siethoff, I. Feldmann, N. Jakubowski, M. Linscheid, *J. Mass Spectrom.* **34** (1999) 421.

[244] I. Tolosa, J. M. Bayona, J. Albaigés, L. F. Alencastro, J. Taradellas, *Fresenius J. Anal. Chem.* **339** (1991) 646–653.

[245] N. Jakubowski, D. Stuewer, W. Vieth, *Anal. Chem.* **59** (1987) 1825.

[246] N. Jakubowski, D. Stuewer, W. Vieth, *Mikrochim. Acta* **1** (1987) 302.

[247] A. P. Mykytiuk, P. Semeniuk, S. Berman, *Spectrochim. Acta Rev.* **13** (1990) 1.

[248] D. Stüwer in G. Holland, A. Eaton (eds.): *Application of Plasma Source Mass Spectrometry,* Royal Society of Chemistry, Cambridge 1990, p. 1 ff.

[249] F. L. King, J. Teng, R. E. Steiner, *J. Mass Spectrom.* **30** (1995) 1061.

[250] J. R. d. Laeter, *Mass Spectrom. Rev.* **17** (1998) 97.

[251] H. Oechsner, W. Gerhard, *Phys. Lett.* **40A** (1972) 211.

[252] N. Jakubowski, D. Stuewer in D. Littlejohn, D. T. Burns (eds.): *Rev. Anal. Chem., Euroanalysis VIII,* Royal Society of Chemistry, (1994) 121–144.

21. Atomic Spectroscopy

José A. C. Broekaert, Universität Leipzig, Institut für Analytische Chemie, Leipzig, Germany

E. Hywel Evans, University of Plymouth, Plymouth, United Kingdom

Symbols :

A	absorbance, $A = \log I_0/I$ with I_0 the intensity of the incident beam and I the intensity of the sorting beam
A_{qp}	Einstein transition probability for spontaneous emission of radiation corresponding with a transition from the higher level q to the lower level p (s^{-1})
B_{pq}	Einstein transition probability for absorption corresponding with a transition from the lower level p to the higher level q (s^{-1})
B_{qp}	Einstein transition probability for stimulated emission corresponding with a transition from the higher level q to the lower level p (s^{-1})
c	velocity of light in vacuum (3×10^{10} m/s)
c_L	detection limit in concentration units or in terms of absolute amounts of analyte
c_U	background equivalent concentration, calibration constant in optical atomic spectrometry for trace analysis
C	capacitance (Farad)
d_0	Sauter diameter; droplet diameter for an aerosol at which the volume-to-surface ratio of the particles equals the volume-to-surface ratio for the whole aerosol
D_{hkl}	lattice constant of crystal
f_k	focal length of collimator (m)
g_q	statistical weight of an atomic (molecular) energy level q
G	optical conductance
h	isotopic abundance in %
j	total internal quantum number
k	Boltzmann's constant (1.38×10^{-23} J/K)
k_{12}	velocity constant for collisional excitation (s^{-1})
k_{21}	velocity constant for collisional de-excitation (s^{-1})
l	orbital quantum number
L	inductance (Henry)
L	Lorentz unit ($e/4\pi m \cdot c^2$)
m_1	magnetic quantum number
M	magnetic quantum number
n	principle quantum number
$P(v)$	profile function of a line
q	sputtering rate (μg/s)
QE	quantum efficiency of a photocathode
R	Rydberg constant (109.677 cm^{-1})
R_0	theoretical resolving power of optical spectrometer
RSF	relative sensitivity factor in elemental mass spectrometry
s	spin quantum number
s_a	exit slit width (μm)
s_e	entrance slit width (μm)
s_{eff}	effective slit width of an optical spectrometer in μm; $s_{eff}^2 = s_e^2 + s_L^2 + s_0^2 + s_z^2$ with s_L and s_z the contributions of the physical width of the line measured and the optical aberrations, respectively (all in μm)
s_0	diffraction slit width (μm)
$s_r(c_x)$	relative standard deviation of the estimate
S	blackening of a photographic emulsion
S	sputtering yield in sputtered ions/incident ion
$S_{nj}(T)$	Saha function for neutrals n of an element j
S/N	signal-to-noise ratio
t_r	response time of a decay process for an excited level (s)
t_p	pulse duration (s)
U	potential (V)
v'	wavenumber (cm^{-1})
Δv	frequency broadening contribution (cm^{-1})
w	penetration rate (μm/s)
Z	charge of nucleus or ion (in elementary charge units)
Z_{ij}	partition function of an element j in the ith ionization state
α_D	divergency of laser beam with $\alpha_D = \lambda/D$ with λ the wavelength and D the beam diameter
γ	contrast of photographic emulsion
$\Delta\lambda_L$	physical halfwidth of a spectral line (nm or pm)
$\theta(V)$	amplification of photomultiplier
$\lambda/\Delta\lambda$	practical resolving power
ϱ_v	radiant density of a spectral feature with frequency v
$\sigma(v)$	collision cross section for species with velocity v
φ	radiant flux $\varphi = \tau B G$ with τ the transmittance and B the radiant density of the source in W m^{-2} sterad^{-1}

21.1. Introduction

Atomic spectroscopy is the oldest principle of instrumental elemental analysis, going back to the work of BUNSEN and KIRCHHOFF in the mid-nineteenth century [1]. The optical radiation emitted from flames is characteristic of the elements present in the flame itself or introduced by various means. This discovery depended on the availability of dispersing media, such as prisms, which spectrally resolved the radiation to produce the

line spectra of the elements. It was also found that radiation with the same wavelength as the emitted lines is absorbed by a cold vapor of the same element. This observation is the basis for atomic emission and absorption spectrometry.

Flames proved to be suitable sources for the analysis of liquids, but the need for direct chemical analysis of solids soon arose, including the need for rapid analysis of more than one element in a solid sample. The development of electrical discharges fostered the growth of atomic spectrometry; arc and spark sources proved useful for analyte ablation and excitation. Low-pressure discharges proved to be powerful radiation sources (a similar development has been observed more recently, with the availability of laser sources).

The limitations stemming from the restricted temperatures of flames led to the development of high-temperature plasma sources for atomic emission spectrometry. The development of high-frequency inductively coupled plasmas and microwave plasmas has led to the widespread use of these methods for routine analytical work.

Parallel with these developments, there have been advances in spectrometer design, from the spectroscope, through photographic spectrographs, sequential spectrometers, and simultaneous multichannel spectrometers with solid state detectors.

The use of flames as atomizers for atomic absorption spectrometry (AAS) has also made its impact on analytical methodology, largely as the result of work by WALSH [2]. L'VOV and MASSMANN [3], [4] introduced the graphite furnace atomizer, now widely available in commercially available AAS with electrothermal atomization, especially for ultratrace analysis.

The sources used for optical atomic spectrometry are also powerful sources for elemental mass spectrometry, from the classical spark source mass spectrometer to present-day plasma mass spectrometric methods such as glow discharge and inductively coupled plasma mass spectrometry.

21.2. Basic Principles

21.2.1. Atomic Structure

The basic processes in atomic spectrometry involve the outer electrons of atoms. Accordingly, the possibilities and limitations of atomic spectroscopy depend on the theory of atomic structure.

On the other hand, the availability of optical spectra was decisive in the development of the theory of atomic structure and even for the discovery of a number of elements.

In 1885, BALMER reported that for a series of atomic emission lines of hydrogen, a simple relationship existed between the wavelengths:

$$\lambda = k \frac{n^2}{n^2 - 4} \tag{1}$$

where $n = 2, 3, 4, \ldots$ for the lines H_α, H_β, H_γ, $\ldots$

Equation (1) can be written in wavenumbers:

$$v' = \frac{1}{\lambda} = R \left(\frac{1}{2^2} - \frac{1}{n^2} \right) \tag{2}$$

where v' is the wavenumber (cm^{-1}) and R the Rydberg constant ($109.677\ cm^{-1}$). The wavenumbers of all series in the hydrogen spectrum are given by:

$$v' = R \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right) \tag{3}$$

where n_2 is a series of numbers $> n_1$, with $n_1 = 1, 2, 3, 4, \ldots$ for the Lyman, Balmer, Paschen, and Brackett series, respectively.

Rydberg extended the Balmer formula to:

$$v' = RZ^2 \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right) \tag{4}$$

where Z is the effective charge of the atomic cleus, allowing calculation of wavelengths for elements other than hydrogen. The wavenumber of each line in an atomic spectrum can thus be calculated from a formula involving two positive numbers, called spectral terms.

The significance of the spectral terms emerges from Bohr's theory of atomic structure, in which the atom has a number of discrete energy levels related to the orbits of the electrons around the nucleus. As long as an electron is in a defined orbit no energy is emitted, but when a change in orbit occurs, another energy level is reached and the excess energy is emitted in the form of electromagnetic radiation. The wavelength is given by Planck's law:

$$E = hv = hc/\lambda \tag{5}$$

Here $h = 6.626 \times 10^{-34}\ J \cdot s$, v is the frequency (s^{-1}), $c = 3 \times 10^{10}\ cm/s$ is the velocity of light, and λ is the wavelength (cm). Accordingly:

$$\overset{.}{\nu} = \frac{1}{\lambda} = \frac{E}{hc} = \frac{E_1}{hc} - \frac{E_2}{hc} = T_1 - T_2 \qquad (6)$$

where T_1 and T_2 are the Bohr energy levels, the complexity of emission spectra reflects the complex structure of atomic energy levels.

For an atom with nuclear charge Z and one valence electron, the energy of this electron is given by:

$$E = -\frac{2\pi Z^2 e^4 \mu}{n^2 h^2} \qquad (7)$$

where $\mu = mM/(m+M)$, with m the mass of the electron and M the mass of the nucleus; n is the principal quantum number ($n = 1, 2, 3, \ldots$) on which the order of the energy levels depends. For the orbital angular momentum L:

$$\|L\| = \frac{h}{2\pi}\sqrt{l(l+1)} \qquad (8)$$

l is the orbital quantum number and has values of: $0, 1, 2, \ldots, (n-1)$; $l = 0$ for a circular orbit, $l = 1, 2, \ldots$ for elliptical orbits.

The possible orientations of the elliptical orbits with respect to an external electric or magnetic field:

$$L_z = \frac{h}{2\pi}m_l \qquad (9)$$

where L_z is the component of the orbital angular momentum along the field axis and $m_l = \pm 1, \pm (l-1), \ldots, 0$ is the magnetic quantum number; for each value of l it has $(2l+1)$ values.

When a spectral line source is subject to a magnetic field, the spectral lines display hyperfine structure (Zeeman effect). In order to explain hyperfine structure it is postulated that the electron rotates on its axis with spin angular momentum S:

$$\|S\| = \frac{h}{2\pi}\sqrt{S(S+1)} \qquad (10)$$

The spin quantum number m_s determines the angles between the axis of rotation and the external field as:

$$s_z = \frac{h}{2\pi}m_s \qquad (11)$$

where $m_2 = \pm 1/2$.

The orbital and spin angular momenta determine the total angular momentum of the electron J:

$$\vec{J} = \vec{L} + \vec{S} \text{ with } \|J\| = \frac{h}{2\pi}\sqrt{j(j+1)} \qquad (12)$$

where $j = l \pm s$ is the total internal quantum number.

Atomic spectral terms differ in their electron energies and can be characterized by the quantum numbers through the term symbols:

$$n^m l_j \qquad (13)$$

Here $l = 0, 1, 2, \ldots$ and the corresponding terms are given the symbols s (sharp), p (principal), d (diffuse), f (fundamental), etc., relating originally to the nature of different types of spectral lines; n is the principal quantum number, m is the multiplicity ($m = 2s + 1$), and j is the total internal quantum number. The energy levels of each element can be given in a term scheme, in which is also indicated which transitions between energy levels are allowed and which forbidden. This is reflected by the so-called selection rules; only those transitions are allowed for which Δn has integral values and at the same time $\Delta l = \pm 1$, $\Delta j = 0$ or ± 1, and $\Delta s = 0$. The terms for an atom with one outer (valence) electron can easily be found, e.g., for Na ($1s^2 2s^2 2p^6 3s^1$), in the ground state: $3^2 S_{1/2}$ [$l = 0$ (s), $m = 2(1/2) + 1 = 2$ ($s = 1/2$), and $j = 1/2, (j = |l \pm s|)$]. When the $3s$ electron is promoted to the $3p$ level, the term symbol for the excited state is: $3^2 P_{1/2,3/2}$ ($l = 1$ (p), $m = 2(1/2) + 1 = 2$, and $j = 1/2, 3/2$).

The term schemes of the elements are well documented in the work of GROTRIAN [5]. The term scheme for Na is shown in Figure 1.

For atoms with more than one valence electron, Russell–Saunders (L–S) coupling applies. The orbital momenta of all electrons have to be coupled to the total orbital momentum, like the spin momentum. The fact that each electron has a unique set of quantum numbers is known as the Pauli exclusion principle. The total quantum number L is obtained as $L = l$, $S = s$, and $J = L - S, \ldots, L + S$. The term symbol accordingly becomes:

$$^M L_J \qquad (14)$$

For the case Mg ($1s^2 2s^2 2p^6 3s^2$), the ground state is $3^1 S_0$ ($L = 0$ as $l_1 = 0$ and $l_2 = 0$, $S = 0$ as $s_1 = 1/2$ and $s_2 = -1/2$, and $J = 0$ as both L and $S = 0$). The first excited state ($1s^2 2s^2 2p^6 3s3p$) is characterized by the terms: $3^1 P_1$ ($L = 1$ as $l_1 = 0$ and $l_2 = 1$, $S = 0$ as $s_1 = 1/2$ and $s_2 = -1/2$, and $J = |L \pm S| = 1$), but also
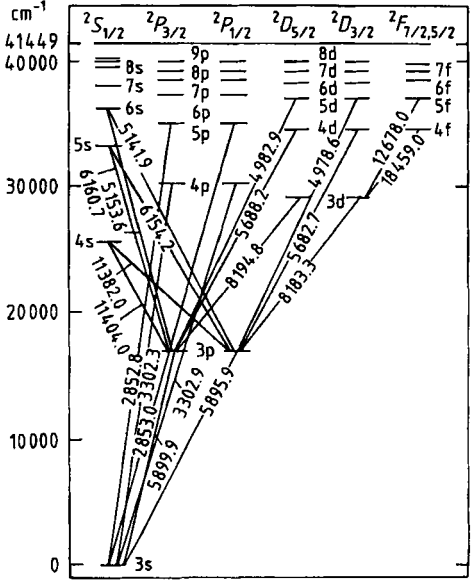
Figure 1. Atomic energy level diagram for the sodium atom [5]

3^3P_2, 3^3P_1, and 3^3P_0 (as for the parallel spins $s_1 = 1/2$ and $s_2 = 1/2$, $S = 1$, and further $J = 0, 1, 2$).

As the number of electrons increases, the coupling becomes more complex, increasing the number of spectral terms and thus the number of lines in the spectrum.

21.2.2. Plasmas [6]

Partially ionized gases are usually referred to as plasmas. They contain molecules, radicals, atoms, ions, and free electrons, and result from the interaction of energy with matter in the gaseous state. Just as for atoms, the radicals, molecules, and ions in a plasma can be in their ground or excited states and radiation can be emitted or absorbed when transitions occur from one state to another. The wavelengths of spectral lines can be obtained from Planck's law, but their intensities depend on the number densities of the species and states involved.

Energy transfer in a plasma results from non-radiative processes (involving collisions) as well as from radiative processes (emission, absorption, and fluorescence). The efficiency of collision processes is described by the cross section $\sigma(v)$. It reflects the change in momentum of a particle with mass m and velocity v when it collides with a

particle with mass M. Apart from the cross section, the velocity distribution for a given species affects energy transfer in a plasma.

For a Maxwell distribution, the velocity distribution is given by:

$$\frac{dn}{n} = \frac{2}{\sqrt{\pi}} \sqrt{u'} \exp(-u') \, du' \tag{15}$$

For the so-called Druyvenstein distribution which pertains at low pressure:

$$\frac{dn}{n} = 1.039 \sqrt{u'} \exp(-0.548u'^2) \, du' \tag{16}$$

where $u' = E/kT$, E is the mean energy of the particles, and T the absolute temperature.

For a plasma contained in a closed system and in thermal equilibrium, the population of the excited levels for all particles is given by Boltzmann's law:

$$\frac{n_q}{n_0} = \frac{g_q}{g_0} \exp(-E_q/kT) \tag{17}$$

where n_q is the number density of particles in the excited state, n_0 is the number density of particles in the ground state, g_q and g_0 are the statistical weights of the corresponding levels, E_q is the excitation energy of the state q, k is Boltzmann's constant (1.38×10^{-23} J/K), and T is the absolute temperature. Equation (17) can be transformed when n_q is expressed as a function of the total number density of the species. Then:

$$\frac{n_q}{n} = \frac{g_q \exp(-E_q/kT)}{\sum_m g_m \exp(-E_m/kT)} \tag{18}$$

as $n = \sum_m n_m$. The sum $Z_m = \sum_m g_m \exp(-E_m/kT)$ is the partition function. It is a function of temperature and its coefficients for a large number of neutral and ionized species are listed in the literature, e.g., [7]. When E_q is expressed in electron volts, Equation (18) can be written:

$$\log n_{uq} = \log n_u + \log n_q - (5040)/TV_q - \log Z \tag{19}$$

21.2.3. Emission and Absorption of Radiation [8]

In a steady-state plasma, the number of particles leaving an energy level per unit time equals the number returning to this level. The most im-

portant energy exchange processes in a plasma are:

1) Collisions where atoms are excited to a higher level by collision with energetic neutrals (collisions of the first kind)
2) Collisions where excited neutrals lose energy through collisions without emission of radiation (collisions of the second kind)
3) Excitation by collision with electrons
4) De-excitation where energy is transferred to electrons
5) Excitation of atoms by the absorption of radiation
6) De-excitation of atoms by spontaneous or stimulated emission

When n is the number density of one type of particle and N is that of a second species, present in excess ($n << N$), the following equilibria are to be considered:

$$\alpha N n_0 = \beta N n_q \tag{20}$$

$$\alpha_e n_e n_0 = \beta_e n_e n_q \tag{21}$$

$$B' \varrho_v n_0 = (A + B \varrho_v) n_q \tag{22}$$

where n_e is the electron number density; A, B, and B' are the Einstein transition probabilities for spontaneous emission, stimulated emission, and absorption, respectively; and α_e, α, β_e, and β are functions of the cross sections for the respective processes, as well as of the velocity distribution of the particles involved. The radiation density at frequency v is ϱ_v. When the system is in thermodynamic equilibrium, the neutrals and electrons have the same Maxwell velocity distribution and at temperature T:

$$n_q/n_0 = \alpha/\beta = \alpha_e/\beta_e = B'/(A/\varrho_v + B)$$
$$= \frac{g_q}{g_0} \exp\left(-E_q/kT\right) \tag{23}$$

Thus, each process is in equilibrium with its inverse and the Boltzmann distribution of each state is maintained by collisions of the first and second kind, including collisions with electrons.

In a real radiation source, emission and absorption of radiation have to be considered as well, but contribute little to the energy balance. The system is in so-called local thermal equilibrium (LTE) and:

$$\alpha N n_0 + \alpha_e n_e n_0 + B' \varrho_v n_0$$

$$= \beta N n_q + \beta_e n_e n_q + (A + B \varrho_v) n_q \tag{24}$$

from which n_q/n_0 can be calculated. The population of excited states is determined by the excitation processes in the radiation source, as reflected by the coefficients in Equation (24). For a d.c. arc, $\alpha N >> \alpha_e n_e + B' \varrho_v$ and $\beta N >> \beta_e n_e + (A + B \varrho_v)$. This leads to:

$$\frac{n_q}{n_0} = \frac{\alpha}{\beta} = \frac{g_q}{g_0} \exp\left(-E_q/kT\right) \tag{25}$$

As the radiation density is low, the d.c. arc plasma can be considered to be in thermal equilibrium. The simplification which leads to Equation (24) does not apply for low-pressure discharges, where collisions with electrons are very important as are radiative processes; moreover, the velocity distributions are described by the Druyvenstein equation. These sources are not in thermal equilibrium.

Excited states are prone to decay because of their high energy. This can take place by collisions with molecules, atoms, electrons, or ions, or by emission of electromagnetic radiation. In the latter case, the wavelength is given by Planck's law. For levels q and p, the number of spontaneous transitions per unit time is given by:

$$-dN_q/dt = A_{qp} N_q \tag{26}$$

where A_{qp} is the Einstein coefficient for spontaneous emission (s^{-1}). When several transitions can start from level q, Equation (26) becomes:

$$-dN_q/dt = N_q \sum_p A_{qp} = N_q v_q \tag{27}$$

where v_q is the lifetime of excited state q. If decay can take place by an allowed radiative transition, the lifetime is of the order of 10^{-8} s. When no radiative transitions are allowed levels are metastable (e.g., Ar 11.5 and 11.7 eV), which can decay only by collisions. In low-pressure discharges such levels may have very long lifetimes (up to 10^{-1} s).

For absorption of electromagnetic radiation with frequency v_{qp} and radiation density ϱ_v, the number density of N_q increases as:

$$dN_q/dt = B_{qp} N_q v_\varrho \tag{28}$$

For stimulated radiation, atoms in excited state q decay only when they interact with radiation of wavelength λ_{qp} and:

$$- \mathrm{d}N_q/\mathrm{d}t = B_{qp}N_q\varrho_v \tag{29}$$

For thermal equilibrium:

$$g_q B_{qp} = g_p B_{pq} \tag{30}$$

where g_p and g_q are the degeneracies of the respective levels.

The intensity I_{qp} of an emitted spectral line (a) is proportional to the number density of atoms in state q:

$$I_{qp} = A_{qp}n_{aq}h\nu_{qp} \tag{31}$$

or after substitution of n_{aq} (n_q for the atomic species according to Eq. 18):

$$I_{qp} = A_{qp}h\nu_{qp}n_a(g_q/Z_a)\exp\left(-E_q/kT\right) \tag{32}$$

where T is the excitation temperature which can be determined from the intensity ratio for two lines (a and b) of the same ionization state of an element as:

$$T = [5040(V_a - V_b)]/\{\log\left[(gA)_a/(gA)_b\right] -$$
$$log(\lambda_a/\lambda_b) - \log(I_a/I_b)\} \tag{33}$$

Often the line pair Zn 307.206/Zn 307.59 nm is used. This pair is very suitable because ionization of zinc is low as a result of its relatively high ionization energy, the wavelengths are close to each other, which minimizes errors introduced by changes in the spectral response of the detector, and the ratio of the gA values is well known.

The excitation temperatures can also be determined from the slope of the plot $\ln[I_{qp}/(g_q A_{qp} \nu_{qp})]$ or $\ln(I_{qp} \lambda/gA_{qp})$ against E_q, which is $-1/kT$. The λ/gA values for a large number of elements and lines are available [8]. Spectroscopic measurement of temperatures from line intensities may be hindered by deviations from ideal thermodynamic behavior in real radiation sources, and by inaccuracies of transition probability estimates. Determination of excitation temperatures in spatially inhomogeneous plasmas is treated extensively by BOUMANS [9].

According to classical dispersion theory, the relation between absorption and the number density of the absorbing atoms is given by:

$$\int K_v\,\mathrm{d}v = \frac{\pi e^2}{mc}N_v f \tag{34}$$

where K_v is the absorption coefficient at frequency v, m is the mass and e the charge of the electron, c is the velocity of light, N_v is the density of atoms with frequency between v and $v+\mathrm{d}v$ (practically the same as N), and f is the oscillator strength. This relation applies strictly to monochromatic radiation, so use of a primary source which emits very narrow lines is advantageous. The relation between absorption A and the concentration of the absorbing atoms in an atomizer is given by the Lambert–Beer law. If I_0 is the intensity of the incident radiation, l the absorption path length, and I the intensity of the transmitted radiation, the change in intensity $\mathrm{d}I$ resulting from absorption within the absorption path length $\mathrm{d}l$ is:

$$-\mathrm{d}I = k\,I_0 c\,\mathrm{d}l \tag{35}$$

or

$$\int_{I_0}^{I} -\,\mathrm{d}t/t = k\,c\int_0^l \mathrm{d}l \tag{36}$$

$A = \log(I_0/I) = \log(1/T)$, in which A is the absorbance and T the transmittance. Accordingly, Equation (36) becomes:

$$A = k\,c\,l \tag{37}$$

The absorbances are additive. The Lambert–Beer law, however, is valid only within a restricted concentration range; in atomic absorption spectrometry deviations from linearity are common.

Line Broadening. Atomic spectral lines have a physical width resulting from several broadening mechanisms [10]. The *natural width* of a spectral line results from the finite lifetime of an excited state, τ. The corresponding half-width in terms of frequency is:

$$\Delta v_N = 1/(2\pi\tau) \tag{38}$$

This results for most spectral lines in a halfwidth of the order of 10^{-2} pm.

The *Doppler spectral width* results from the velocity component of the emitting atoms in the observation direction. The corresponding half-width is:

$$\Delta v_D = \left[2\sqrt{(\ln 2)}/c\right]v_0\sqrt{[2(RT)/M]} \tag{39}$$

where c is the velocity of light, v_0 is the frequency of the line maximum, R is the gas constant, and M
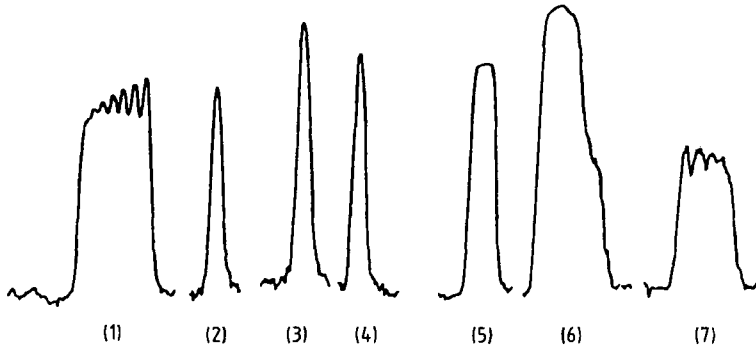
Figure 2. Profiles of some rare earth atomic emission lines in inductively coupled plasma atomic emission spectrometry (photographic measurements obtained with a high-resolution grating spectrograph; theoretical resolving power 460 000 [12]

the atomic mass. Doppler broadening thus depends strongly on temperature. Accordingly, it is often referred to as temperature broadening and reflects the kinetic energy of the radiating atoms, ions, or molecules. For the Ca 422.6 nm line in the case of a hollow cathode discharge at pressures of the order of 0.1 kPa, the Doppler broadening at 300 K is 0.8 pm whereas at 2000 K it is 2 pm [11].

Lorentz or pressure broadening results from the interaction between the emitting atoms and atoms of other elements. The half-width is:

$$\Delta v_{\mathrm{L}} = (2/\pi)\sigma_{\mathrm{L}}^2 N \sqrt{\left[2\pi R T\left(\frac{1}{M_1} + \frac{1}{M_2}\right)\right]} \qquad (40)$$

where M_1 and M_2 are the atomic masses, N is the concentration of the foreign atoms, and σ_{L} is their cross section. Pressure broadening is low in low-pressure discharges. For the Ca 422.6 nm line, this type of broadening at 300 K and 0.9 kPa is only 0.02 pm. At atmospheric pressure, however, this type of line broadening is predominant.

Isotopic and *hyperfine structure, resonance broadening* (resulting from interaction between radiating and nonradiating atoms of the same species), and *Stark broadening* (resulting from interaction with electric fields), contribute to the widths of spectral lines.

Natural and Lorentz broadening have a Lorentzian distribution:

$$I(v) = I_0 / \left\{1 + [2(v - v_0)/\Delta v]^2\right\} \qquad (41)$$

Doppler broadening has a Gaussian distribution:

$$I(v) = I_0 \exp - \left\{[2(v - v_0)/2\,\Delta v_{\mathrm{D}}]\sqrt{(\ln 2)}\right\} \qquad (42)$$

The combination of both types of profile (normally both pressure and temperature broadening are important) results in a so-called Voigt profile:

$$V(\alpha, \omega) = \alpha/\pi \int_{-\infty}^{+\infty} \left\{[\exp(-y^2\,dy)]/[\alpha^2 + (\omega - y)^2]\right\} \qquad (43)$$

where $\omega = 2(v - v_0)/\Delta v_{\mathrm{D}}\sqrt{(\ln 2)}$ and $\alpha = (\Delta v_{\mathrm{L}}/\Delta v_{\mathrm{D}})\sqrt{(\ln 2)}$, when the contribution of the natural width is neglected. From the widths of spectral lines, in most cases between 1 and 20 pm, and knowledge of the distribution functions, the contributions of the different broadening processes can be calculated by deconvolution methods. This allows calculation of the so-called gas temperature from the Doppler broadening component, giving an estimate of the kinetic energy of neutrals and ions.

The profiles of spectral lines from plasma sources at atmospheric pressure are illustrated by the high-resolution spectra of a number of rare-earth elements in an inductively coupled plasma (ICP) (Fig. 2) [12].

Self-Absorption. The radiation emitted by the source is absorbed by ground state atoms of the same species. This phenomenon is known as self-absorption [13]. As the chance that an absorbed photon is re-emitted is less than unity, the observed radiation is reduced. When I_0 is the intensity emitted at the line maximum and $P_{\mathrm{E}}(v)$ is the profile function, the intensity distribution for an emission line is $I_0 P_{\mathrm{E}}(v)$ and the intensity observed after the radiation passes through a layer with number density of absorbing atoms of n_{A} is:

$$I(v) = I_0 P_E(v) \exp \left[- \varrho P_A(v)/P_A(v_0) \right] \tag{44}$$

where v_0 is the frequency at the line center, $P_A(v)$ is the absorption profile function, $P_A(v_0)$ is its value at the line center, and ϱ is an absorption parameter:

$$\varrho \sim B P_A(v) n_A \tag{45}$$

ϱ increases with the transition probability for absorption and thus is larger for resonance lines which involve transitions to the ground state. It is also larger in sources with high analyte number density n_A. As the absorption maximum is in the center of the line, self-absorption always leads to flatter line profiles. When a minimum occurs in the absorption profile, the line undergoes so-called self-reversal and $\varrho > 1$. Self-reversal only occurs when there is a strong temperature gradient in the radiation source and when the analyte number densities in both the hot and the cooler zones of the plasma are high.

21.2.4. Ionization

If sufficient energy is transferred to a plasma, atoms may be ionized. This depends on the temperature of the plasma, but also on the ionization energy of the elements. As ions have discrete energy levels between which transitions are possible, ionic spectra are also important when considering the emission of radiation by a plasma. The ionization of atoms a of the element j into ions i is an equilibrium:
$$n_{aj} \leftrightarrows n_{ij} + n_e \tag{46}$$
and the equilibrium constant $S_{nj}(T)$, known as the Saha constant, is:

$$S_{nj}(T) = (n_{ij} n_e)/n_{aj} \tag{47}$$

The degree of ionization α_j for element j is:

$$\alpha_j = n_{ij}/n_j = n_{ij}/(n_{aj} + n_{ij}) \tag{48}$$

where n_{aj} and n_{ij} are the concentrations of atoms and ions and can be expressed as a function of the total number of atoms n_j by:

$$n_{aj} = (1 - \alpha_j) n_j$$

and

$$n_{ij} = \alpha_j n_j \tag{49}$$

In the notation used by BOUMANS [9], the intensity of an atom line can be written:

$$I_{qp} = A_{qp} h v_{qp} (g_q/Z_{aj})(1 - \alpha_j) n_j \exp \left(- E_q/kT \right) \tag{50}$$

and the intensity of an ion line is:

$$I_{qp}^+ = A_{qp}^+ h v_{qp}^+ \left(g_q^+/Z_{ij} \right) \alpha_j n_j \exp \left(- E_q^+/kT \right) \tag{51}$$

These expressions for the intensities contain three factors which depend on temperature; i.e., the degree of ionization, the Boltzmann factor, and the partition functions. The degree of ionization can be written as a function of the electron number density and the Saha function:

$$[\alpha_j/(1 - \alpha_j)] n_e = S_{nj}(T) \tag{52}$$

However, the latter is also given by the Saha equation. When the Saha function is expressed in partial pressures, the Saha equation is:

$$S_{pj}(T) = [p_{ij} p_e]/[p_{aj}] = \frac{\left[(2\pi m)^{3/2} (kT)^{5/2} \right]}{h^3 (2Z_{ij}/Z_{aj}) [\exp(-E_{ij}/kT)]} \tag{53}$$

The factor 2 is the statistical weight of the free electron (two spin orientations), $k = 1.38 \times 10^{-23}$ J/K, the mass of the electron $m = 9.11 \times 10^{-28}$ g, $h = 6.62 \times 10^{-34}$ J $\cdot$ s, and 1 eV $= 1.6 \times 10^{-19}$ J. This leads to the following expression for the Saha equation:

$$\log S_{pj} = \frac{5}{2} \log \frac{T - 5040}{T} V_{ij} + \log \frac{Z_{ij}}{Z_{aj}} - 6.18 \tag{54}$$

where V_{ij} is the ionization energy (in eV).

The Saha equation is valid only for a plasma in local thermal equilibrium; then the temperature in the equation is the ionization temperature. When this condition is not fulfilled, the equilibrium between the different states of ionization is given by the so-called Corona equation [14].

Accordingly, the degree of ionization in a plasma can be determined from the intensity relation of atom and ion lines of the same element:

$$\log \left[(\alpha_j)/(1 - \alpha_j) \right] = \log \left(I_{qp}^+/I_{qp} \right) -$$
$$\log \left[\left(g_q^+ A_{qp}^+ v_{qp}^+ \right)/(g_q A_{qp} v_{qp}) \right] +$$
$$(5040/T)(V^+ - V_q) + \log (Z_{ij}/Z_{aj}) \tag{55}$$

This method can only be applied for a plasma in local thermal equilibrium at known temperature. The partition functions Z_{aj} and Z_{ij} for the atom and ion species are functions of the temperature, and their coefficients have been calculated for many elements [7]. The accuracy of the gA values and of the temperatures affects the accuracy of the degree of ionization. The line pairs Mg II 279.6 nm/Mg I 278.0 nm, and Mg II 279.6 nm/Mg I 285.2 nm are often used for determinations of the degree of ionization of an element in a plasma.

Once α_j is known, the electron pressure can be determined. From

$$\log\left[\alpha_j/(1 - \alpha_j)\right] = \log\left[S_{Pj}(T)/p_e\right] \tag{56}$$

one can calculate:

$$\log p_e = - \log\left[\alpha_j/(1 - \alpha_j)\right] + \log S_{Pj}(T) \tag{57}$$

Taking into account Equations (54) and (55), this results in:

$$\log p_e = - \log\left(I_{qp}^+/I_{qp}\right) + \log\left(g_q^+ A_{qp}^+ v_{qp}^+/g_q A_{qp} v_{qp}\right)$$
$$- (5040/T)\left(V_{ij} + V_q^+ - V_q\right) + 5/2\,(\log T) - 6.18 \tag{58}$$

This reflects the fact that the intensity ratio of the atom and ion lines of an element changes considerably with the electron pressure in the plasma, particularly for elements with low ionization energy, such as Na. This is analytically very important as it is the cause of so-called ionization interferences in classical d.c. arc emission spectrometry, atomic absorption, and plasma optical emission, as well as mass spectrometry.

When the plasma is not in LTE, the electron number densities cannot be determined from the Saha equation, but have to be derived directly from the Stark broadening of the H$_\beta$ line or of a suitable argon line. This contribution to broadening can be written [15]:

$$\delta\lambda = 2\left[1 + 1.75\alpha(1 - 0.75\varrho)\right]\omega \tag{59}$$

where ϱ is the ratio of the distance between the ions and the Debye path length, ω is the broadening due to the interaction of the electrons ($\sim n_e$), and α is the contribution of the interaction with the quasi-static ions ($\sim n_e^{1/4}$). The value of $\delta\lambda$ can be calculated as a function of n_e. The electron number density can be determined directly from the widths of the Ar I 549.59 or Ar II 565.07 nm lines, which are mainly due to Stark broadening. The temperature can also be determined by combining measurements of the intensities of atom and ion lines of the same element:

$$\log n_i/n_a = - \log n_e + 3/2\,(\log T) - (5040/T)V_{ij} + \log\left(Z_{ij}/Z_{aj}\right) + 15.684 \tag{60}$$

which can be combined with Equation (58).

From Equations (50) and (51), which give the intensity of a line, and from the Saha equation (54), it can be seen that, for each spectral line emitted by a plasma source, there is a temperature where its emission intensity is maximum. This is the so-called standard temperature. To a first approximation [16], it can be written:

$$T = \left(0.95V_{ij} \times 10^3\right)/\left[1 - 0.33\left(V_{ij}/10\right) + 0.37\log\left(V_{ij}/10\right) - 0.14\log P_e^*\right] \tag{61}$$

where V_{ij} is the ionization energy and V_e the excitation energy; P_e^* is the electron pressure (in atm; 1 atm $\approx$ 101 kPa) and is a function of T and n_e. In the case under discussion, the standard temperature for a line of an element which is present as an impurity in a plasma (e.g., in a noble gas), dilution in the plasma must also be considered. The standard temperatures indicate which types of line will be optimally excited in a plasma of given temperature, electron pressure, and gas composition, and are thus important indications for line selection. Atom lines often have standard temperatures below 4000 K, especially when analyte dilution in the plasma is high, whereas ion lines often have values of 10 000 K. Atom and ion lines are called "soft" and "hard" lines, respectively.

Because of the interaction of free and bound electrons in a plasma continuous radiation is also emitted over a wide wavelength range. The intensity distribution for this background continuum can be expressed:

$$I_\nu \, d\nu = K \, n_e n_r r^2 / \left(T_e^{1/2}\right) \exp\left(-h\nu/kT_e\right) \cdot d\nu (\text{free} - \text{free})$$

$$+ \, K'\left(1/j^3\right) n_e n_Z Z^4 / \left(T_e^{3/2}\right) \cdot \exp\left(U_j - h\nu\right)/(k\,T_e)(\text{free} - \text{bound}) \qquad (62)$$

where n_r is the concentration of atoms with charge r, Z is the nuclear charge, and U_j is the ionization energy of the level with quantum number j. This background continuum consists of so-called bremsstrahlung, caused by the interaction of free electrons with each other, and of recombination radiation, which is due to the interaction of free and bound electrons. The latter is especially important in the UV region. It is continuous radiation with spectral lines superimposed. As T_e is the electron temperature, absolute measurements of the background continuum emission in a plasma, e.g., for hydrogen, allow the determination of the electron temperature in a plasma, whether in local thermal equilibrium or not.

21.2.5. Dissociation

The dissociation of molecular plasma gases or analyte molecules within the radiation source is an equilibrium reaction. Accordingly, highly stable radicals or molecules are always present in a radiation source. They emit molecular bands which are superimposed on the atomic and ionic line spectra in the emission spectrum. Radicals and molecules may also give rise to cluster ions which may be detected in mass spectra. Common species in plasma gases are: CN, NH, NO, OH, and N_2 (or N_2^+). From the analytes, highly stable oxides may persist (e.g., AlO^+, TiO^+, YO^+). A thorough treatment of molecular spectra is available [17], [18].

Vibrational and Rotational Hyperfine Structures. Molecules or radicals have various electronic energy levels ($^1\Sigma, \, ^2\Sigma, \, ^2\Pi, \ldots$), which have vibrational fine structure ($v = 0, 1, 2, 3, \ldots$), which, in turn, have rotational hyperfine structure ($J = 0, 1, 2, 3, \ldots$). The total energy of a state may be written:

$$E_1 = E_{el} + E_{vib} + E_{rot} \qquad (63)$$

E_{el} is of the order of $1-10$ eV, the energy difference between two vibrational levels of the same electronic state is of the order of 0.25 eV, and the separation of rotational levels is of the order of 0.005 eV. When the rotational levels considered belong to the same electronic level, the emitted radiation is in the infrared region. When they be-

long to different electronic levels, they occur in the UV or visible region. Transitions are characterized by the three quantum numbers of the states involved: n', v', j' and n'', v'', j''. All lines which originate from transitions between rotational levels belonging to different vibrational levels of two electronic states form the band: $n', v' \longrightarrow n'', v''$. For these band spectra the selection rule is $\Delta j = j' - j'' = \pm 1, \, 0$. Transitions for which $j'' = j' + 1$ give rise to the P-branch, $j'' = j' - 1$ to the R-branch, and $j' = j''$ to the Q-branch of the band. The line corresponding to $j' = j'' = 0$ is the zero line of the band. When $v' = v'' = 0$ it is also the zero line of the system. The difference between the wavenumber of a rotation line and the wavenumber of the zero line in the case of the P- and the R-branch is a function of the rotational quantum number j and the rotational constant B_v for which:

$$E_j/(hc) = B_v j(j + 1) \qquad (64)$$

The functional relation is quadratic known as the Fortrat parabola.

As in the case of atomic spectral lines, the intensity of a rotational line can be written:

$$I_{nm} = (N_m A_{nm} h\nu_{nm})/2\pi \qquad (65)$$

where N_m is the population of the excited level and ν_{nm} the frequency of the emitted radiation. The transition probability for dipole radiation is:

$$A_{nm} = \left(64\pi^4 \nu_{nm}^3\right)/(3k)[1/(g_m)] \sum \|Rn_i m_k\|^2 \qquad (66)$$

where i and k are degenerate levels of the upper (m) and the lower state (n), $Rn_i m_k$ is a matrix element of the electrical dipole moment and g_m is the statistical weight of the upper state; N_m is given by the Boltzmann equation:

$$N_m = N(g_m)/Z(T) \exp\left(-E_r/kT\right) \qquad (67)$$

where E_r is the rotational energy of the excited electronic and vibrational level, given by:

$$E_r = hc B_v J'\left(J' + 1\right) \qquad (68)$$

where $B_{v'}$ is the rotational constant and J' is the rotational quantum number of the upper state m.

For a $^2_g - {}^2_u$ transition, between a so-called "gerade" (g) and "ungerade" (u) level, the term $\Sigma |Rn_i m_k|^2 = J' + J'' + 1$, where J' and J'' are the rotational quantum numbers of the upper and lower state. Accordingly:

$$I_{nm} = (16\pi^3 c N v_{nm}^4)/3Z(T)(J' + J'' + 1)\exp\left(-hc B_r J'(J' + 1)/kT\right) \tag{69}$$

or

$$\ln\left[I_{nm}/(J' + J'' + 1)\right] = \ln\left[16\pi^3 c N v_{nm}^4\right]/[3Z(T)] - \left[hc B_r J'(J' + 1)\right]/kT \tag{70}$$

By plotting $\ln\left[I_{nm}/(J' + J'' + 1)\right]$ against $J'(J' + 1)$ for a series of rotational lines a so-called rotational temperature can be determined, reflecting the kinetic energy of neutrals and ions in the plasma.

Molecular Bands. Spectral lines of molecular bands emitted by molecules and radicals in a plasma often interfere with atomic emission lines. In atomic absorption spectrometry, absorption by molecular bands arising from undissociated molecules in the atomizer may also interfere. Therefore, it is important to study the dissociation of molecular species in high-temperature radiation sources or atomizers. In plasma sources, band-emitting species may stem from the working gas. In this respect, N_2, N_2^+, CN, OH, and NH band emission are important. Undissociated sample and analyte species are also present in the plasma. Highly stable molecules such as Al_2O_3, La_2O_3, BaO, AlF_3, CaF_2, MgO, may be present in atomic spectrometric sources. It is important to know their dissociation as a function of plasma temperature and composition. This dependence can be described by a dissociation equation, similar to the Saha equation:

$$K_n = \left[(2\pi/h^2)(m_X m_Y/m_{XY})(kT)^{3/2}\right] \cdot (Z_X Z_Y/Z_{XY})[\exp\left(-E_d/kT\right)] \tag{71}$$

where:

$$K_n = (n_X n_Y)/n_{XY} \tag{72}$$

Thus, for a metal oxide the degree of dissociation can be calculated when the plasma temperature, the partial pressure of oxygen in the plasma, and the dissociation constant are known.

21.2.6. Sources and Atom Cells

For the purposes of the following discussion it is useful to distinguish between the terms "source" and "atom cell". Sources are the generators of the radiation (narrow line or continuum) which is eventually to be measured by the detector. Atom cells are the devices for producing the population of free atoms (or ions) from a sample. In atomic absorption and fluorescence spectrometry these two devices are separate. In atomic emission spectrometry the source is also the atom cell. In atomic mass spectrometry the terms source and atom cell are often used interchangeably to denote the "ion source".

In atomic spectrometry, the sample is introduced, by means of a sampling device, into a high-temperature source or atom cell (plasma, flame, etc.). Here, the sample is vaporized, e.g., by thermal evaporation or sputtering. It is important to supply as much energy as possible, so that the volatilization processes, which involve a physical or chemical equilibrium, result in complete atomization, irrespective of the state of aggregation, solid state structure, or chemical composition of the sample. This is very important, both to ensure maximum sensitivity and to minimize matrix interference in the analysis. The effectiveness of the volatilization processes involved, the plasma temperature, and the number densities of the various plasma components will all influence sample atomization.

Rotational temperatures are relevant for all processes in which molecules, radicals, and their dissociation products are involved. They can be obtained from the intensity distribution for the rotational lines in rotation–vibration spectra. Molecules such as OH or CN have often been used as "thermometers" (Eqs. 63–70) [19]–[21].

Table 1. Temperatures of sources and atom cells in atomic spectrometry

Source	Temperature, K				State
	Rotational, T_{rot}	Excitation, T_{ex}	Electron, T_e	Ion T_i	
Arc (d.c.)	5000	5 000	5 500	5 000	LTE
Spark		20 000	20 000	20 000	LTE
Inductively coupled plasma	4800	5 000	6 000	6 000	$\approx$ LTE
Microwave plasma	2000	4 000	6 000	6 000	non-LTE
Low-pressure discharge	600	20 000	30 000	30 000	non-LTE

The *gas temperature* is determined by the kinetic energy of neutral atoms and ions. It can be estimated from the Doppler broadening (Eq. 39). However, the contributions of Doppler and temperature broadening have to be separated, which involves the use of complicated deconvolution procedures [22].

The *electron temperature* is a measure of the kinetic energy of the electrons. It is relevant in the study of excitation and ionization by collisions with electrons, which is an important process for analyte signal generation. The electron temperature can be determined from the intensity of the recombination continuum or of the bremsstrahlung (Eq. 62).

The *excitation temperature* describing the population of the excited levels of atoms and ions is important in studies of the dependence of analyte line intensities on various plasma parameters in analytical emission spectrometry. It can be determined from the intensity ratio of two atomic emission lines of the same element and state of ionization (Eq. 33) or from plots of the appropriate function for many atomic emission lines against their excitation energies.

The *ionization temperature* is relevant for all phenomena involving equilibria between analyte atoms, ions, and free electrons in plasmas. In the case of thermal equilibrium, it occurs in the Saha equation (Eqs. 53, 54) and can be determined from the intensity ratio of an ion and an atom line of the same element. In all other cases, ionization temperatures can be determined from the n_e value obtained from Stark broadening (Eqs. 59, 60). Temperatures for the most important sources in atomic spectrometry are listed in Table 1.

In a plasma which is at least in local thermal equilibrium, all these temperatures are equal. This implies that the velocity distribution of all types of particle in the plasma (molecules, atoms, ions, and electrons) at any energy level can be described by the Maxwell equation (15). For all species, the population of the different levels is then given by the Boltzmann equation (Eq. 17). Further, the ionization of atoms, molecules, and radicals can be described by the Saha equation (Eqs. 53, 54) and the related equations of chemical equilibrium. Finally, the radiation density in the source obeys Planck's law, and the exchange of both kinetic energy between particles and electromagnetic radiation are in equilibrium with each other. Real plasma sources used in atomic spectrometry are, at best, approximations to local thermal equilibrium. Although processes between the particles may be in thermal equilibrium, those involving electromagnetic radiation are not, as the plasma cannot be considered as a closed system. Moreover, real plasma sources are extremely inhomogeneous with respect to temperature and number density distributions. Accordingly, the equilibria occur only within small volume elements of the sources, in which gradients can be neglected. Many plasmas, however, have cylindrical symmetry and can be observed laterally, integrating the information from many volume elements along the observation direction. From this integral value, the values at a defined radial distance from the plasma center can be calculated. This necessitates side-on observation along several axes, which are equidistant with respect to each other, and a so-called Abel inversion. If $I'(x)$ is the first derivative of function $I(x)$, describing the variation of a measured value as a function of the lateral position x during side-on observation (Fig. 3), the radial values at distance r are:

$$I(r) = -(1/\pi) \int_r^x [I'(x)\,dx]/\sqrt{(r^2 - x^2)} \qquad (73)$$
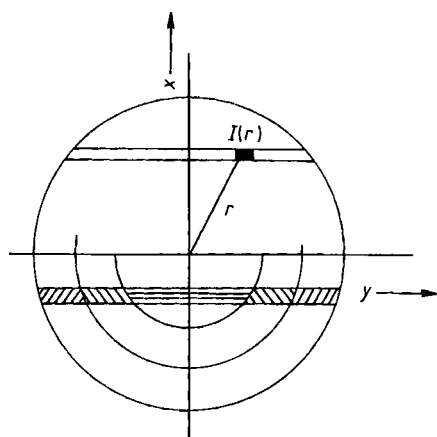
Figure 3. Abel inversion procedure for the determination of radial distributions in plasma sources
$I(r)$ = Intensity at radial distance r in volume element $dx \times dy$

This integral, known as the Abel integral, can only be solved numerically, writing each laterally measured intensity as a sum of the different volume elements with coefficients denoting their contribution. For certain numbers of lateral observation positions (3, 10,...) these Abel coefficients are listed in the literature [23]. By inversion of the corresponding matrix the radial values are obtained. This allows determination of radial distributions of emissivities, absorbances, temperatures, or number density distributions in plasmas. Repeating this procedure at different heights in a plasma is equivalent to plasma tomography. Similar results, which do not require the assumption of cylindrical symmetry in the source can be obtained by imaging spectrometry with two-dimensional detectors [24].

In real plasmas, departures from thermal equilibrium often occur. In the extreme case, as encountered in plasmas under reduced pressure, emission or absorption of radiation becomes so important that there is no longer a clear relation between the mean kinetic energies of the species and the excitation temperatures. The latter then lose the physical meaning of temperature. The absence of local thermal equilibrium in these non-LTE plasmas relates to the existence of high field gradients or a.c. fields; only the electrons, not the heavy atoms and ions, can follow the field changes and fully take up the dissipated energies. Accord-

ingly, the mean kinetic energy of the electrons and thus the electron temperature is much higher than the gas temperature.

Sources and atom cells for atomic spectrometry include flames, arcs, sparks, low-pressure discharges, lasers as well as d.c., high-frequency, and microwave plasma discharges at reduced and atmospheric pressure (Fig. 4). They can be characterized as in Table 2. *Flames* are in thermal equilibrium. Their temperatures, however, are at most 2800 K. As this is far below the standard temperature of most elemental lines, flames have only limited importance as sources for atomic emission spectrometry, but are excellent atom cells for absorption, fluorescence, and laser-enhanced ionization work. *Arcs and sparks* are well-known sources for atomic emission spectrometry. The high temperature obtained in spark sources leads to the excitation of ion lines, for which the standard temperatures are often beyond 10 000 K, whereas in arc sources atom lines predominate. In *plasma sources* under reduced pressure the gas kinetic temperatures are low and their atomization capacity is limited. When these sources are used as atomic emission sources or as primary sources in atomic absorption, line widths are very narrow. Moreover, especially with gases having high ionization energy such as helium, high-energy lines such as those of the halogens can be excited. Low-pressure discharges are valuable ion sources for mass spectrometry as analyte ionization takes place. Because of the low pressure, coupling with a mass spectrometer is possible. So-called plasmajet and plasma sources at atmospheric pressure are especially useful for the emission spectrometric analysis of solutions. Their gas kinetic temperatures are high enough to achieve complete dissociation of thermally stable oxides and both atom and ion lines occur in the spectra. As reflected by the fairly high ionization temperature, they are powerful ion sources for mass spectrometry; plasma mass spectrometry is now one of the most sensitive methods in atomic spectrometry. Lasers are very suitable for ablation of solids. Owing to their high analyte number densities, plasmas are subject to high self-absorption so it is better to use them solely for volatilization, and to introduce the ablated material in a second source for signal generation.

Table 2. Applications of sources and atom cells for atomic spectrometry

Source	Spectroscopy	Sample	Concentration range [a]
Flames	AES [b]	liquid	t (alkali)
	AAS [c]	liquid (solid)	t
	AFS [d]	liquid	ut
	LEI [e]	liquid	ut
Arc (d.c.)	AES [b]	solid, liquid	t
	MS [f]	solid	ut
Spark	AES [b]	solid	t
	MS [f]	solid	ut
Glow discharge	AES [b]	solid, liquid, gas	t
	AAS [c]	solid, liquid, gas	t
	AFS [d]	solid, liquid, gas	ut
	MS [f]	solid, gas	ut
	LEI [e]	solid, gas	ut
Laser	AES [b]	solid	m
	AAS [c]	solid	t
	AFS [d]	solid	ut
	MS [f]	solid	ut
	LEI [e]	solid	ut
Inductively coupled plasma	AES [b]	liquid (solid, gas)	t
	AFS [d]	liquid (solid, gas)	t
	MS [f]	liquid (solid, gas)	ut
Microwave plasma	AES [b]	gas (liquid)	t
	MS [f]	gas (liquid)	ut
Furnace	AES [b]	liquid, solid	t
	AAS [c]	liquid, solid	ut

[a] m = Minor; t = Trace; ut = Ultra-trace.
[b] AES = Atomic emission spectroscopy.
[c] AAS = Atomic absorption spectroscopy.
[d] AFS = Atomic fluorescence spectroscopy.
[e] LEI = Laser-enhanced ionization.
[f] MS = Mass spectrometry.

21.2.7. Analytical Atomic Spectrometry

Analytical atomic spectrometry nowadays includes the use of flames and plasma discharges for optical and for mass spectrometry. These are used directly as emission sources, as atomizers for atomic absorption or atomic fluorescence, or for ion production. In a source for atomic spectrometry atomic vaporization and signal generation take place together. The first processes require high energy for complete atomization, whereas the signal generation processes often require discrete, selective excitation. Therefore, the use of so-called tandem sources is preferred, where analyte vapor generation and signal generation take place separately [26].

For atomic emission spectrometry, selectivity is achieved by isolation of the spectral line at the exit slit of the spectrometer. This puts high demands on the optical quality of the spectral apparatus. In atomic absorption and fluorescence, selectivity is partly controlled by the radiation source delivering the primary radiation, which in most cases is a line source (hollow cathode lamp, laser, etc.). Therefore, the spectral bandpass of the monochromator is not as critical as it is in atomic emission work. This is especially true for laser-based methods, where in some cases of atomic fluorescence a filter is sufficient, and in laser-induced ionization spectrometry where no spectral isolation is required at all. In the case of glow discharges and inductively coupled high-frequency plasmas, ion generation takes place in the plasmas. In the first case, one can perform direct mass spectrometry on solids, and in the second case on liquids or solids subsequent to sample dissolution. In the various methods of atomic spectrometry, samples have to be delivered in the proper form to the plasma source. Therefore, in the treatment of the respective methods attention will be given to techniques for sample introduction.
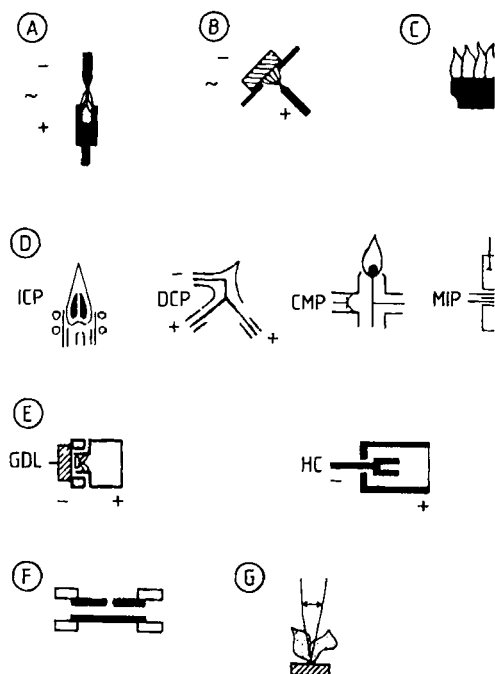
Figure 4. Sources for atomic spectrometry [25]
A) Arc; B) Spark; C) Flame; D) Plasma sources; E) Low-pressure discharges; F) Graphite furnace; G) Laser plume
ICP = Inductively coupled plasma; DCP = Direct current plasmajet; CMP = Capacitively-coupled microwave plasma; MIP = Microwave-induced plasma; GDL = Glow discharge lamp; HC = Hollow cathode

21.3. Spectrometric Instrumentation

Atomic spectrometric methods of analysis essentially make use of equipment for spectral dispersion to achieve their selectivity. In optical atomic spectrometry, this involves the use of dispersive as well as nondispersive spectrometers, whereas in the case of atomic spectrometry with plasma ion sources mass spectrometric equipment is used. In both cases, suitable data acquisition and processing systems are built into the instruments.

In the design of instruments for atomic spectrometry the central aim is to achieve the full potential of each method. This involves the power of detection and its relation to precision, freedom from spectral interference, and the price –performance ratio.

21.3.1. Figures of Merit of an Analytical Method

In practice, atomic spectrometric methods are relative methods and need to be calibrated. The calibration function describes the relationship between the analytical signals (absorbances, absolute or relative radiation intensities, or ion currents). In its simplest form, the calibration function can be written (→Chemometrics):

$$Y = ac + b \tag{74}$$

The inverse function is often used:

$$c = a'Y + b' \tag{75}$$

This is known as the analytical evaluation function. Calibration curves are often nonlinear, so a polynomial function is needed:

$$Y = a_0 + a_1c + a_2c^2 + \dots \tag{76}$$

where $a_0, a_1, a_2, \dots$ are determined by multivariate regression. Normally, a second-degree polynomial is sufficient to describe the calibration function over a large concentration range. Alternatively, segmented calibration curves can be used if the calibration function is nonlinear.

For a series of analytical signal measurements for a well-defined analyte concentration, a statistical uncertainty exists which stems from fluctuations in the analytical system. The precision achievable is an important figure of merit of an analytical procedure, for the statistical evaluation of analytical data see, →Chemometrics, and classical textbooks [27] – [29]. The precision of a measurement procedure is expressed in terms of the standard deviation:

$$\sigma = \sqrt{\left[\sum (Y_m - Y_i)^2 / (n - 1) \right]} \tag{77}$$

where $Y_m = \sum (Y_i/n)$ is the mean value, Y_i an individual measurement, and n the number of measurements. The precision of a certain measurement system depends on the noise in the system. Different types of noise can be distinguished [30].
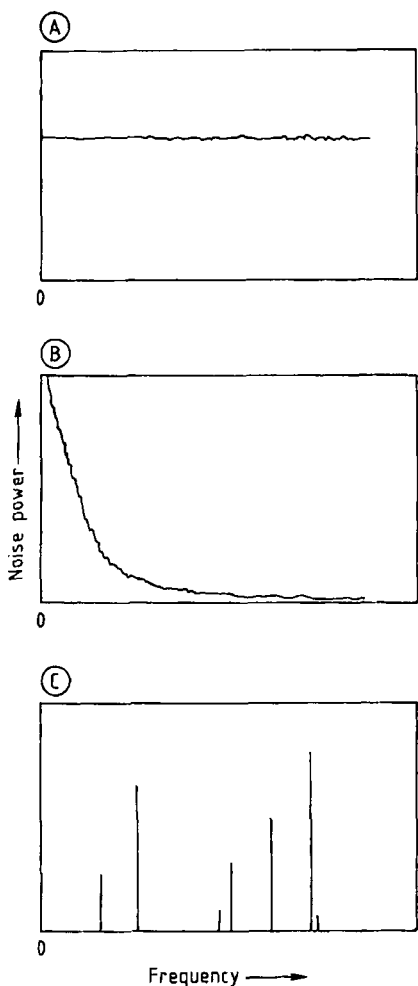
Figure 5. Noise power spectra commonly found in chemical instrumentation
A) White noise; B) Flicker ($1/f$) noise; C) Interference noise [31]

Fundamental or Random Noise. This is statistically distributed and its amplitude as a function of frequency can be written as a sum of sinusoidal functions. This type of noise is related to the discrete nature of matter and the quantization of radiation, and cannot be completely eliminated.

Nonfundamental, Flicker, or Excess Noise. Here, the sign or the magnitude can correlate with well-defined phenomena.

Interference Noise. This is observed at well-defined frequencies and mostly stems from components in the system.

Both nonfundamental and interference noise can often be eliminated by appropriate filtering.

In a noise spectrum (Fig. 5) the noise amplitude is plotted as a function of frequency [31]. White noise (Fig. 5 A) occurs over all frequencies and is almost always fundamental in origin, whereas for $1/f$ noise (Fig. 5 B) the amplitude decreases with the frequency and it is nonfundamental in origin. Discrete noise bands with well-defined causes may also occur (Fig. 5 C). These may stem from the source, caused by gas flow dynamics or contributions from the vacuum line, etc. Noise spectra are a powerful diagnostic tool to trace the sources of noise, and to study instrumental limitations to the power of detection. In atomic spectrometry, it is important to determine if noise from the detector is predominant; if so, it can be described by a Poisson distribution:

$$\sigma^2 = n \tag{78}$$

where n is the number of events per unit time and σ the standard deviation. Alternatively, the background noise of the source may be much more important, or Flicker noise or frequency-dependent noise are predominant. In the latter case, overtones often occur.

For the precision of an analytical method, not only the reproducibility of single measurements, but also calibration errors have to be considered. This is a complex problem as, depending on the nature of the noise, precision may vary considerably with concentration.

When a linear regression is performed, the standard deviation of the regression can be defined:

$$s(Y) = \sqrt{\left[\sum (Y_i - Y)^2 / (n - 2)\right]} \tag{79}$$

where Y_i is the signal obtained for a standard sample with concentration c from the regression equation. The latter is calculated by a least-squares procedure from the pairs (c, Y), where Y are the measured signals, and n is the total number of measurements. The standard deviation for the concentration of the analytical sample c_x can be calculated through propagation of errors:

$$s_r(c_x) = \ln_{10} s(c_x) = \ln_{10} a' s(Y) \tag{80}$$

where a' is the slope of the calibration curve. The magnitude of the concentrations of the analytical sample with respect to those of the standard samples has to be considered, and can be included in Eq. (80):

$$s_r(c_x) = \ln_{10} a' s(Y) \cdot \sqrt{\left\{(1/n) + (1/m) + (c_x - c_m)^2 / \left[\sum (c - c_m)^2\right]\right\}} \tag{81}$$

where m is the number of replicates for the analytical sample and c_m is the mean of all the standard concentrations measured.

Equations (74) and (75) are valid only for a limited concentration range, known as the linear dynamic range. This range is limited at the upper end by physical phenomena, such as detector saturation, and at the lower end by the limit of determination. This limit is typical of a given analytical procedure and is the lowest concentration at which a determination can be performed with a certain precision [32] (→Chemometrics; →Trace Analysis, →Trace Analysis). For 99.86% certainty, and provided the fluctuations of the limiting noise source can be described by a normal distribution, the lowest detectable net signal Y_L is three times the relevant standard deviation:

$$Y_L = 3\sigma^* \tag{82}$$

The net signal is determined from the difference between a background signal and a gross signal including analyte and background contributions, so a factor of $\sqrt{2}$ has to be introduced.

In many cases the limiting standard deviation is often the standard deviation obtained for a series of blank measurements [33]. From the calibration curve the detection limit is:

$$c_L = a' \left(3\sqrt{2}\right)\sigma \tag{83}$$

The detection limit is thus closely related to the signal-to-background and the signal-to-noise ratios. It is the concentration for which the signal-to-background ratio is $3\sqrt{2}$ times the relative standard deviation of the background, or at which the signal-to-noise ratio is $3\sqrt{2}$. The signal-to-noise ratio itself is related to the types of noise occurring in the analytical system. From a knowledge of the limiting noise sources, well-established signal acquisition measures can improve the signal-to-noise ratio and hence the power of detection.

The analytical signals measured often include contributions from constituents other than the analyte (e.g., matrix constituents). This is known as spectral interference and can be corrected by subtracting contributions calculated from the magnitude of interference and the concentration of the interfering species. A special type of interference influences the background signal on which the analyte signals are superimposed; a number of correction methods exist. Freedom from interference is an important figure of merit for an analytical method.

21.3.2. Optical Spectrometers

In optical atomic spectrometry, the radiation emitted by the radiation source, or the radiation which comes from the primary source and has passed through a separate atomizer, must be fed into a spectrometer; this radiation transfer should be as complete as possible. The amount of radiation passing through an optical system is expressed by its optical conductance G_0:

$$G_0 = \int_A \int_B (\cos\alpha_1 \cos\alpha_2 \, dA \, dB)/a_{12}^2 \tag{84}$$

$$\sim (A\,B)/a_{12}^2 \tag{85}$$

where dA and dB are surface elements of the entrance and exit apertures. a_{12} is the distance between them, and α_1, α_2 are the angles between the normals of the aperture planes and the radiation. If n is the refractive index of the medium, the optical conductance is:

$$G = G_0 n^2 \tag{86}$$

The radiant flux through an optical system is:

$$\varphi = \tau B\,G \tag{87}$$

where τ is the transmittance, determined by reflection or absorption losses at the different optical elements, and B is the radiant density of the source ($W\,m^{-2}\,sterad^{-1}$). For optimal optical illumination of a spectrometer, the dispersive element should be completely illuminated to obtain full resolution. However, no radiation should bypass the dispersive element, as stray radiation may produce anomalous signals. Further, the optical conductance at every point in the optical system should be maximum.

21.3.2.1. Optical Systems

The type of illumination system needed to fulfill these conditions depends on the source and detector dimensions, the homogeneity of the source, the need to fill the detector homogeneously with radiation, the distance between the source and the entrance aperture of the spectrometer, and the focal length of the spectrometer.

In conventional systems, lenses and imaging mirrors are used. Glass lenses can be used only at wavelengths above 330 nm. For quartz, radiation with wavelengths down to 165 nm is transmitted, but evacuation or purging the illumination system and spectrometer with nitrogen or argon is required below 190 nm, as a result of the absorption of short-wavelength radiation by oxygen. At wavelengths below 160 nm, the system must be evacuated and MgF_2 or LiF optics must be used. With lenses, three main illumination systems are used.

Imaging on the Entrance Collimator. In this system a lens is placed immediately in front of the entrance slit to image the relevant part of the radiation source on the entrance collimator (Fig. 6 A). This has the advantage that the entrance

slit is homogeneously illuminated; however, stray radiation may occur inside the spectrometer. The distance between the source and the entrance slit a is given by the magnification required:

$$x/W = a/f_k \tag{88}$$

where x is the width of the source, W is the width of the entrance collimator, and f_k its focal length. The f-number of the lens is:

$$1/f = 1/a + 1/f_k \tag{89}$$

The focal length of a lens depends on the wavelength:

$$f(\lambda_1)/F(\lambda_2) = n_1/n_2 \tag{90}$$

where n_1, n_2 are the refractive indices of the lens material at wavelengths λ_1, λ_2. For quartz, a factor of 0.833 has to be applied when passing from the Na 583 nm D-line to 200 nm.

Illumination with Intermediate Image. In the second illumination system, a field lens is used to produce an intermediate image on a diaphragm. A suitable zone can be selected to fully illuminate the collimator mirror with the aid of a lens placed immediately in front of the exit slit. A third lens is used to illuminate the entrance slit homogeneously (Fig. 6 B). The magnification is divided over all three lenses, so that chromatic aberrations are minimized, but the set-up is inflexible. The distances between the three lenses must be selected to achieve the required magnifications:

$$x/I = a_1/a_2 \tag{91}$$

$$D/s_h = a_2/a_3 \tag{92}$$

and

$$I/W = a_3/f_k \tag{93}$$

where I is the diameter of the intermediate image, D the diameter of the field lens, s_h the entrance slit height, and a_1, a_2, a_3 are the distances between the lenses. Further

$$a_1 + a_2 + a_3 = A \tag{94}$$

is mostly fixed by constructional factors, and:

$$1/f_1 = 1/a_1 + 1/a_2 \tag{95}$$

$$1/f_2 = 1/a_2 + 1/a_3 \tag{96}$$
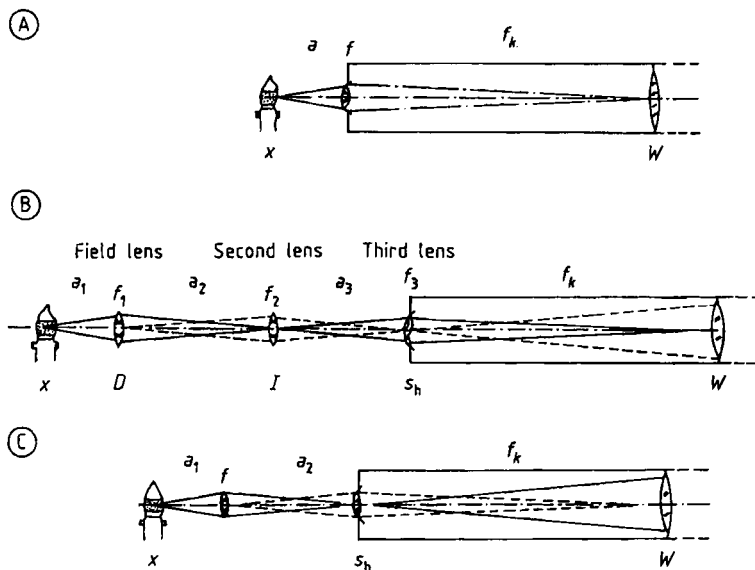
Figure 6. Illumination of the optical spectrometer with lenses
A) Imaging on the entrance collimator; B) Illumination with intermediate image; C) Imaging on the entrance slit
For explanation of symbols see text

$$1/f_3 = 1/a_3 + 1/f_k \qquad (97)$$

give the *f*-numbers of the respective lenses. Only one parameter (e.g., the width of the intermediate image) can be freely selected because x, W, s, A, and f_k are fixed.

Image on the Entrance Slit. The third illumination system produces an image on the entrance slit with the aid of one lens; the structure of the source is imaged on the entrance collimator and the detector. This allows spatially resolved line intensity measurements when a detector with two-dimensional resolution, such as a photographic emulsion or an array detector (see Section 21.3.2.2), is employed. This type of imaging is often used for diagnostic studies.

Fiber optics are very useful for radiation transmission. Small lenses are necessary to accommodate the entrance angle of the fiber, which depends on the refractive index of the material (usually quartz), and is typically 30 – 40°. A typical illumination of a spectrometer with an optical fiber (Fig. 7) uses a lens (diameter d) for imaging the source on the fiber. Bundles of fiber beams with diameter $D = 600$ µm are often used. The magnification x/D as well as the entrance angle $\alpha = \mathrm{arc}$

$\tan(d/2)/a_2$ and the lens formula determine the *f*-number of the lens and d. At the fiber exit, a lens allows the radiation to enter the spectrometer, without causing stray radiation.

With quartz fibers, radiation is efficiently fed into spectrometer, but the transmittance decreases seriously below 220 nm. This may lead to detector noise limitations for analytical lines at lower wavelengths.

Spectrometers are used to produce the spectrum or to isolate narrow spectral ranges, without further deconvolution. In dispersive apparatus, the spectrum is produced with a prism or diffraction grating. In nondispersive apparatus, spectral regions are isolated from the radiation beam by reflection, interference, or absorption in interferometers or filter monochromators. The latter are of use only in flame emission spectrometry.

A *dispersive spectrometer* contains an entrance collimator, a dispersive element, and an exit collimator [34], [35].

The entrance collimator produces a quasi-parallel beam from the radiation coming through the entrance aperture, of width s_e and height h_e. The entrance collimator has focal length f_k and width W. The diffraction slit width s_0 and height h_0 are the half-widths:
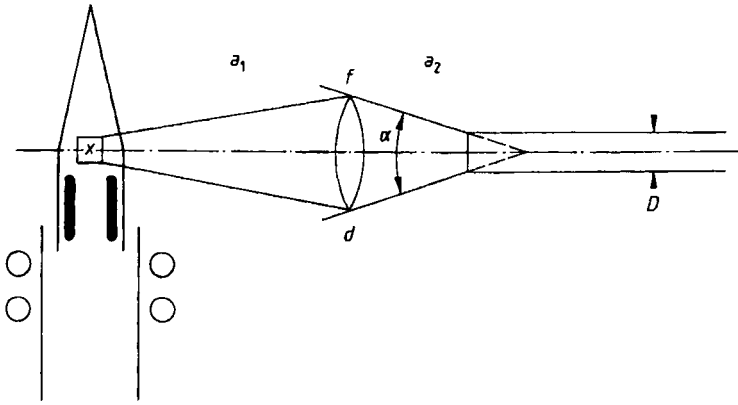
Figure 7. Use of optical fibers for illumination of an optical spectrometer
α = Entrance angle of the fiber; d = Lens diameter; D = Fiber diameter; x = Diameter of the zone in the radiation source to be selected; α_1 = Distance between lens and radiation source; α_2 = Distance between fiber entrance and lens

$$s_0 = \lambda f / W$$

and

$$h_0 = \lambda f / h \tag{98}$$

The entrance aperture dimensions should not be smaller than s_0 and h_0, so that diffraction does not limit resolution. The value of f/W is a measure of the amount of radiant energy entering the spectrometer. The exit collimator images the monochromatized radiation leaving the dispersive element on the exit slit, giving a series of monochromatic images of the entrance slit. In a monochromator, a single exit slit allows isolation of one line after another by turning the dispersive element. In a polychromator the dispersive element is fixed and there are many exit slits placed where the images of the lines of interest occur. They are often on a curved surface with a radius of curvature R (Rowland circle). Simultaneous measurement of several lines and thus simultaneous multielement determinations are possible. With a spectrographic camera, lines are focused in a plane or on a slightly curved surface, where a two-dimensional detector can be placed. With such a detector (photographic plate, diode array detector, etc.), the signal intensities of part of the spectrum over a certain wavelength range or intensities of a spectral feature at several locations in the source can be simultaneously recorded. The energy per unit area at the detector is given by the irradiance:

$$E = \varphi \cos\alpha / A \tag{99}$$

Diffraction Gratings. Diffraction gratings are used almost exclusively as the dispersive element in modern apparatus; prisms are still used as predispersors and crossdispersors . Both plane and concave gratings are used; the latter have imaging qualities. Because of the profile of the grooves, holographically ruled gratings have a rather uniform radiant output over a large spectral area compared with mechanically ruled gratings which always have a so-called blaze angle, and hence a blaze wavelength where radiant energy is maximum. Modern spectrometers tend to use reflection gratings. As a result of interference, a parallel beam of white radiation at incident angle φ_1 with the grating normal is dispersed through angle φ_2; radiation with wavelength λ is sorted (Fig. 8) according to Bragg's equation:

$$\sin\varphi_1 + \sin\varphi_2 = m\lambda / a \tag{100}$$

where m is the order of interference and $a = 1/n_G$ is the grating constant, where n_G is the number of grooves per unit length. When B is the grating width, the total number of grooves $N = B\,n_G$ determines the theoretical resolving power R_0:

$$R_0 = B\,n_G m \tag{101}$$

The angular dispersion can be obtained by differentiation of Equation (101) with respect to λ:

$$d\varphi_2 / d\lambda = m/(a \cos\varphi_2) \tag{102}$$

The reciprocal linear dispersion is:
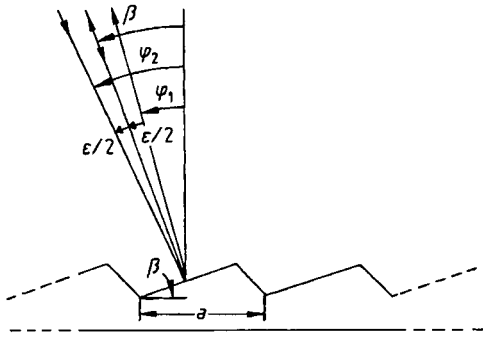
Figure 8. Spectral dispersion at a diffraction grating
α = Grating constant; β = Blaze angle; φ_1 = Angle of incident radiation; φ_2 = Angle of diffracted radiation with wavelength λ; $\varepsilon = \varphi_1 - \varphi_2$

$$\mathrm{d}x/\mathrm{d}\lambda = (\mathrm{d}x/\mathrm{d}\varphi_2)(\mathrm{d}\varphi_2/\mathrm{d}\lambda)$$

$$= (f_\mathrm{k}/\cos\theta')(m/a\cos\varphi_2) \tag{103}$$

where θ' is the slope of the surface of the radiation detector.

The form and depth of the grooves determine the blaze angle β defined earlier. The blaze wavelength for order m is:

$$\sin\beta = \lambda_\mathrm{B} m/[2a\cos(\varepsilon/2)] \tag{104}$$

where $(\varphi_1 - \varphi_2) = \varepsilon$.

In stigmatic spectrometers the height and width of the slit are imaged in the same plane. Several mountings (Fig. 9) are used with a plane gratings.

In the Czerny–Turner mounting (Fig. 9 B), two spherical mirrors with slightly different focal lengths are positioned at slightly different angles to correct for spherical aberration. This mounting is used for high-luminosity monochromators with short focal length and highly dispersive gratings (more than 1800 grooves per millimeter). In the Ebert mounting (Fig. 9 A), there is only one mirror serving for both entrance and exit collimator; accordingly, aberrations occur and require the use of curved slits, to achieve maximum resolution. The Fastie–Ebert mounting is used for large spectrographs. The entrance collimator lies under the exit collimator plane. The spectrum is focused in a plane and the f/W value can be as low as 1/30, which may require the use of a cylindrical lens to increase the detector irradiance. Photographic plates or films are used as detectors.

Normally, gratings are used at low orders; different orders can be separated by special photomultipliers. With a solar-blind photomultiplier
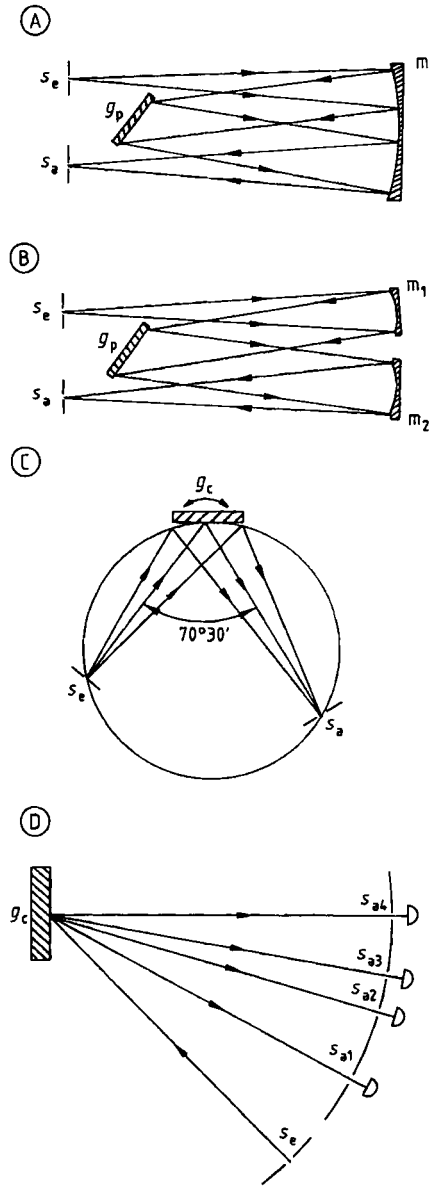
Figure 9. Optical mountings for optical spectrometers with a plane grating
A) Ebert; B) Czerny–Turner; C) Seya–Namioka; D) Paschen–Runge
s_e = Entrance slit; s_a = Exit slit; g_P = Plane grating; g_c = Concave grating; m = Mirror

only radiation with wavelength below 330 nm can be detected. This allows separation of the first order radiation at 400 nm from the second order at 200 nm. This can be used in polychromators to double the effective resolution. So-called echelle

gratings have low groove density (up to 1/100 mm) and therefore enable the use of orders up to 70 [36]. Here, the orders overlap and must be separated by a second dispersive element (e.g., a prism) either with its axis parallel to that of the echelle grating or in the so-called crossed-dispersion mode. In the latter case, the spectrum occurs as a number of dots of height equal to that of the entrance slit. The radiant flux is:

$$\varphi = (B_\lambda \tau s h W H \cos\beta)/f^2 \tag{105}$$

where B_λ is the spectral radiance, τ is the transmittance of the optics, s is the slit width and h is height, W is the width, H the height of the grating, and β is the angle between the collimator plane and the radiation beam. Echelle spectrometers often use an Ebert mounting for high resolution at low focal length. They are thermally very stable and are used as both monochromators and polychromators.

For concave gratings, the radius of the grating determines the Rowland circle. In the direction of dispersion spectral lines are focused on the Rowland circle as monochromatic images of the entrance slit. In the Paschen – Runge mounting (Fig. 9 D), grating, entrance slit, and all the exit slits are fixed on the Rowland circle. This mounting is most often used in large simultaneous polychromators with photoelectric detection. In the Eagle mounting, the grating can be simultaneously turned around its axis and moved along the radiation direction. This mounting can be easily housed in a narrow tank; it is mechanically rather simple and is much used in vacuum monochromators. In the Seya –Namioka mounting (Fig. 9 C), the directions of incident and dispersed radiation beams are kept constant at an angle of $70°30'$. Wavelength selection at the exit slit is performed by rotating the grating. At large angles serious defocusing occurs and the aberrations are larger than in the Eagle mounting. The Seya – Namioka system is often used in vacuum monochromators.

In most sequential spectrometers, switching from one line to another is achieved by turning the grating. From Equation (100) it can be seen that, as the spectral lines have widths of $1 - 3$ pm, angle selection must be very accurate. For a grating with $a = 1/2400$ mm and a line width of 1 pm, the grating must be positioned with an accuracy of $0.0001°$.

It is possible to step through the line profile or to achieve random access by using angle encoders.

Computer-controlled stepper motors are used to turn the grating. Their angular resolution is often above 1000 steps per turn. For a so-called sinebar drive, the number of steps performed by the motor is directly proportional to the wavelength displacement. A pen recorder is running off on the spindle driven by the motor. A further approach is to use a Paschen – Runge spectrometer with equidistant exit slits and a detector moving in the focal plane. Fine adjustment of the lines is by computer-controlled displacement of the entrance slit.

21.3.2.2. Detectors

Photographic emulsions or photoelectric devices can be used as detectors for electromagnetic radiation between 150 and 800 nm. Among the latter, photomultipliers are the most important, but solid state devices are now becoming very common.

Photographic emulsions allow recording of the whole spectrum simultaneously and, accordingly, they are useful for multielement survey analysis. However, their processing is slow, precision is low, and they are not capable of on-line data processing. Quantitative treatment of photographically recorded spectra is of limited importance for analytical use and current interest in photographic detection is limited to research in which documentation of the whole spectrum is required. Therefore, it is treated only briefly.

When a photographic emulsion is exposed to radiation blackening S occurs, which is a function of the radiant energy accumulated during the exposure time. It is given by:

$$S = \log(1/\tau) = -\log\tau = \log\varphi_0/\varphi \tag{106}$$

where τ is the transmission, φ_0 is the flux through a nonirradiated part of the emulsion, and φ is the flux for an exposed part of the emulsion, when, in a densitometer, a beam of white light passes through the emulsion. The emulsion characteristic for a selected illumination time gives the relation between the logarithm of the intensity ($Y = \log I$) and the blackening, and is S-shaped (Fig. 10). Originally, only the linear part of the characteristic was used for quantitative work. However,techniques have been developed,which linearize the characteristic at low exposures. The P-transformation has been described by KAISER [37]:

$$P = \gamma(Y - Y_0) = S - \varkappa D \tag{107}$$
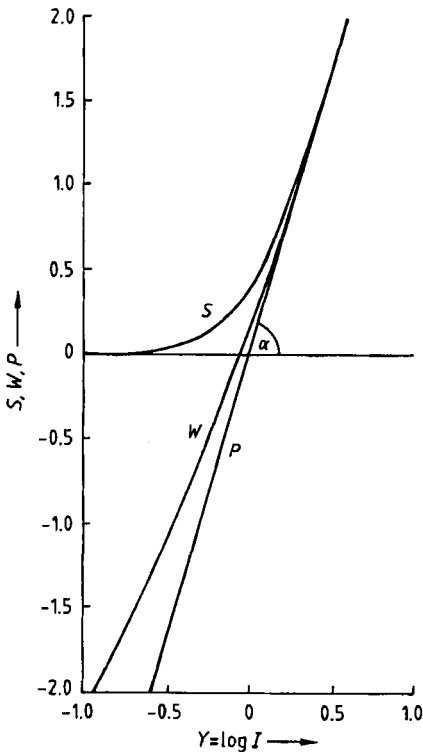
Figure 10. Emulsion characteristic (S) and transformation functions (P, W)

where $D = \log[1/(1-\tau)]$.

The Seidel function is:

$$W = \log(10^S - 1) \tag{108}$$

and thus:

$$P = (1 - \varkappa)S + \varkappa W \tag{109}$$

where Y_0 is the inertia (speed) of the emulsion and constants κ and γ describe the properties of the emulsion, but also depend on the densitometer. In most cases, Y_0 is unknown so ΔP or ΔY values are used. $\gamma = \tan\alpha$ is the contrast of the emulsion. The emulsion can be calibrated with the aid of a step filter placed in front of the spectrometer [38]. When the ratio of the intensities passing through two sectors of the filter is:

$$\Delta Y_m = \log\left(I_{m(1)}/I_{m(2)}\right) \tag{110}$$

the transformation equation becomes:

$$\Delta S = \gamma \Delta Y_m + \varkappa \Delta D \tag{111}$$

where ΔD is the independent variable and ΔS the function. A linear regression for a number of line pairs gives κ as slope and $\gamma \Delta Y_m$ as intercept on the x-axis. Values of $Y = \log I$ for a given blackening can be calculated:

$$Y = (\varkappa/\gamma) - \log(1 - \tau) - (1/\gamma)\log\tau + Y_0 \tag{112}$$

and

$$\Delta Y = -(\varkappa/\gamma)\Delta D + (1/\gamma)\Delta S \tag{113}$$

Errors in the intensity measurement stem from the graininess of the emulsion. According to SIEDENTOPF [39], the standard deviation at low exposure is:

$$\sigma(S) = \sqrt{[0.5\,(a/F)\,S]} \tag{114}$$

where a is the surface area of a grain and F the area of the densitomer slit (e.g., $10\,\mu\text{m}\times 1$ mm). From $\sigma(S)$ it is possible to calculate the relative standard deviation of the intensities $\sigma_r(I)$:

$$\sigma_r(I) = \ln 10\sigma(S)(1/\gamma\{1 + [\varkappa/(10^S - 1)]\}) \tag{115}$$

Typical values for γ are $1-2.5$ and for κ, $0.5-2$. Nonsensitized emulsions can be used in the wavelength range $220-450$ nm. At longer wavelengths green- or red-sensitized emulsions can be used. They have a large γ and accordingly a small dynamic range. At UV or vacuum UV wavelengths the gelatin of the emulsion absorbs significantly, and low-gelatin emulsions (which are rather insensitive) are used; alternatively, radiation can be transformed to longer wavelengths with the aid of a scintillator, such as sodium salicylate.

Photomultipliers [40] are most commonly used for precise measurements of radiant intensities. The photomultiplier has a photocathode and a series of dynodes. The radiation releases photoelectrons from the photocathode as a result of the Compton effect. These photoelectrons impact after acceleration on the dynodes so that a large number of secondary electrons is produced, which after successive action on a number of dynodes lead to a measurable photocurrent I_a at the anode. For a photon flux N_φ through the exit slit of the spectrometer and impacting on the photocathode, the flux of photoelectrons produced N_e is:

$$N_e = N_\varphi \, QE(\lambda) \tag{116}$$

where QE is the quantum efficiency (up to 20%).
The cathode current I_c is:

$$I_c = N_e \, e \tag{117}$$

where $e = 1.9 \times 10^{-19}$ A·s and:

$$I_a = I_c \theta(V) \tag{118}$$

where $\theta(V)$ is the amplification factor (up to 10^6).
The dark current I_d is the value of I_a measured
when no radiation enters the photomultiplier (ca.
1 nA). The quantities QE, $\theta(V)$, and I_d are char-
acteristics for a certain type of photomultiplier.
The photocurrents are fed into a measurement
system. First, I_a is fed to a preamplifier containing
a high-ohmic resistor. The voltage produced is fed
to an integrator. By selection of the resistors and
capacitors, a high dynamic measurement range
can be realized. When I_a is accumulated in a ca-
pacitor with capacitance C, during an integration
time t the potential obtained is:

$$U = I_a t / C \tag{119}$$

This voltage is digitized. The dynamic range is
limited at the lower end by the dark current of the
photomultiplier and at the high end by saturation
at the last dynode.

For a photomultiplier, the relative standard de-
viation of the measured intensities $\sigma_r(I)$ is deter-
mined by the noise of the photoelectrons and the
dark current electrons. The quantum efficiency
and the amplification at the dynodes are also im-
portant:

$$\sigma_r(I) \approx 1.3 \left(N_{e,d} \right)^{-1/2} \tag{120}$$

where $N_{e,d}$ is the total number of photoelectrons
and dark current electrons in measurement time t.

Photomultipliers for the spectral range
160–500 nm usually have an S 20 photocathode;
$QE(\lambda) \approx 5–25\%$, the amplification factors are
3–5 at dynode voltages 50–100 V, and there
are 9–15 dynodes mounted in head-on or side-on
geometry. For selected types, the dark current may
be below 100 photoelectrons per second
($I_a < 10^{-10}$ A). Red-sensitive photomultipliers
have a so-called bialkali photocathode. Their dark
current is higher, but can be decreased by cooling.
For wavelengths below 160 nm, photomultipliers
with MgF_2 or LiF windows are used. Solar-blind

photomultipliers are only sensitive for short-wave-
length radiation (e.g., below 330 nm).

Solid-state detectors are increasingly important
for optical atomic spectrometry. They are multi-
channel detectors and include vidicon, silicon-in-
tensified target (SIT) vidicon, photodiode array
detectors, and image dissector tubes [41]. Charge-
coupled devices (CCD) and charge-injection de-
vices (CID) [42] have been introduced. Photodiode
arrays (PDA) may consist of matrices of up to 512
and even 1024 individual diodes (e.g., Reticon)
with individual widths of 10 µm and heights of
up to 20 mm. The charge induced by photoelectric
effects gives rise to photocurrents which are
sequentially fed into a preamplifier and a multi-
channel analyzer. They are rapid sequential de-
vices, with the advantage that memory effects
due to the incidence of high radiant densities on
individual diodes (lag) as well as cross-talk be-
tween individual diodes (blooming) are low. Their
sensitivity in the UV and vacuum UV is low com-
pared with photomultipliers. Nevertheless, they
are of interest for atomic emission spectrometry,
especially as they can be coupled with microchan-
nel plates, giving additional amplification. Signal-
to-noise ratios can be considerably improved by
cooling the array with the aid of a Peltier element
or liquid nitrogen. Photodiode arrays have been
used successfully in a segmented echelle spec-
trometer. Here, spectral segments around the ana-
lytical lines are sorted out by primary masks sub-
sequent to dispersion with an echelle grating, and
are detected after a second dispersion on a diode
array. More than ten analytical lines and their
spectral background contributions can be meas-
ured simultaneously. This is shown for a glow
discharge lamp in Figure 11 [43], but restrictions
arise due to detector noise limitations at low wave-
lengths.

Image dissector tubes (Fig. 12) make use of an
entrance aperture (d) behind the photocathode (a),
by which photoelectrons from different locations
on the photocathode can be scanned and measured
after amplification in the dynode train, as in a
conventional photomultiplier. Though used in
combination with an echelle spectrometer with
crossed dispersion for flexible rapid sequential
analyses [45], the system has not been successful.
This may be due to its limited cathode dimensions,
but also to stability problems.

Charge transfer devices (CTDs) are solid state,
silicon-based detectors which store charge as a
result of photons impinging on the individual
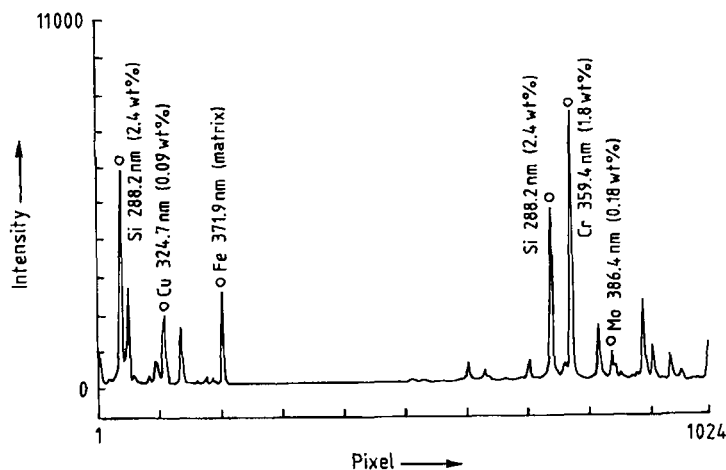
Figure 11. Segmented spectrum of glow discharge [43]
Grimm-type glow discharge lamp with floating restrictor 8 mm diameter, 50 mA, 3.5 torr, 2 kV [44], plasma-array spectrometer (LECO), steel sample 217 A
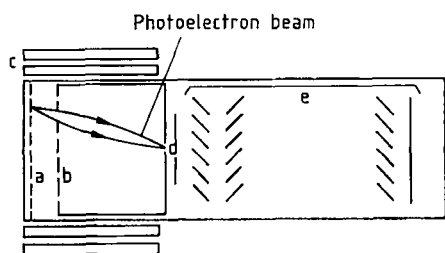
Figure 12. Image dissector tube
a) Photocathode; b) Accelerating electrode (mesh); c) Deflection coils; d) Aperture; e) Electron multiplier

CTD elements, called pixels, and can be used to monitor large portions of the spectrum in multiple orders by taking an "electronic photograph". There are two main types, namely charge injection device (CID) and charge-coupled device detectors (CCD). They differ in that the CID can be read randomly and non-destructively, whereas the CCD can only be read sequentially and destructively. A typical CCD detector [46] is the segmented array CCD used in the PerkinElmer Optima ICP-AES instrument. This consists of 224 linear photodetector arrays on a silicon chip with a surface area of 13 mm by 18 mm (Fig. 13). The array segments detect three to four analytical lines of high analytical sensitivity, large dynamic range and which are free from spectral interferences. Each subarray is comprised of pixels. The pixels are positioned on the detector at $x - y$ locations that correspond to the locations of the desired emission lines, usually

generated by an echelle spectrometer. The emission lines are detected by means of their location on the chip and more than one line may be measured simultaneously. The detector can then be electronically wiped clean and the next sample analyzed. A number of newer CCD detectors are capable of a pixel resolution of 0.009 nm and can record the continuous first order spectrum from 175 to 800 nm, giving simultaneous access to 10 000 emission lines. Compared with many PMTs, a CCD offers improvement in quantum efficiency and lower dark current.

21.3.2.3. Nondispersive Spectrometers

Filter monochromators are of use only for flame photometry. They make use of interference filters, which often have a spectral bandpass of a few nanometers or less. Multiplex spectrometers include Hadamard transform spectrometers and Fourier transform spectrometers, and are especially useful where very stable sources are needed. Hadamard transform instruments make use of a coding of the spectrum produced by recombining the information with the aid of a slit mask which scans the spectrum [48].

Fourier transform spectrometry [49] often uses a Michelson interferometer (Fig. 14) to produce the interferogram. With the aid of a beam splitter, the radiation is split into two parts which are each directed to a mirror. Shifting the mirror in one of

Table 3. Evaluation of detectors for atomic emission spectrometry

Detector	Dimensions, mm	Spectral region, nm	Sensitivity vs. photomultiplier
Vidicon [41]	12.5×12.5[a]	with scintillator down to 200	poorer, especially at < 350 nm; dynamics $< 10^2$; lag/blooming
Diode array [47]	up to 25 (up to 1024 diodes)	especially for > 350	poorer in UV; dynamics 10^3
Diode array MCP[b] [41]	12.5	200 – 800	similar; linear; dynamic range $> 10^3$
Dissector tube [45]	up to 60 mm[a]	200 – 800	similar; linear; dynamic range $> 10^3$
CCD [42]	up to 13×18	200 – 800	similar; linear; dynamic range $> 10^5$

[a] Coupling with crossed-dispersion echelle spectrometer possible.
[b] MCP = Microchannel plate.
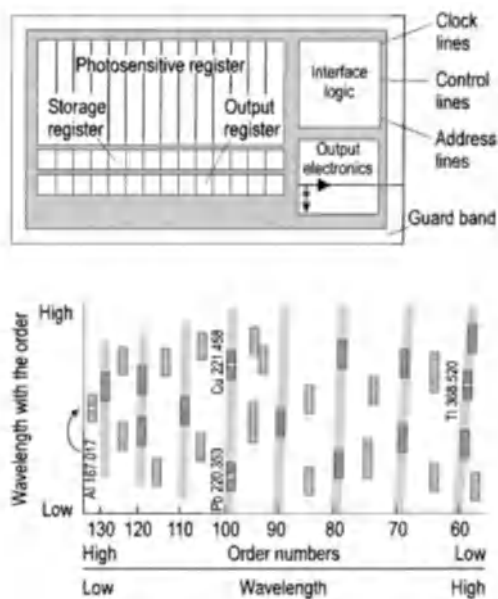
Figure 13. Schematic of a CCD detector (top) and a segmented two-dimensional detector array with respect to wavelength and order (bottom)

Figure 14. Fourier transform emission spectrometry
a) Source; b) Predispersion; c) Interferometer; d) Detector; e) FT in computer

the side arms, gives the interference for each wavelength:

$$I(x) = B(\sigma)[1 + \cos(2\pi\sigma x)] \tag{121}$$

where x is the optical path difference, σ is the wavenumber of the radiation, I is the intensity measured with the detector, and B is the radiation density of the source. A polychromatic radiation source gives an interferogram where the intensity of each point is the sum of all values resulting from Equation (121). The central part contains the low-resolution information and the ends con-

tain high-resolution information. The resolution depends on the recording time, the spectral bandwidth, and the number of scans. By applying a Fourier transform to the signal for each point of the interferogram:

$$I(x) = \int_{-\infty}^{+\infty} B(\sigma)\,d\sigma + \int_{-\infty}^{+\infty} B(\sigma)\cos(2\pi\sigma x)\,d\sigma \tag{122}$$

$$= C + \int_{-\infty}^{+\infty} B(\sigma)\cos(2\pi\sigma x)\,d\sigma \tag{123}$$

where C is a constant which must be subtracted before the transformation. The end result is:

$$B(\sigma) = \int_{-\infty}^{+\infty} I(x)\cos(2\pi\sigma x)\,dx \tag{124}$$
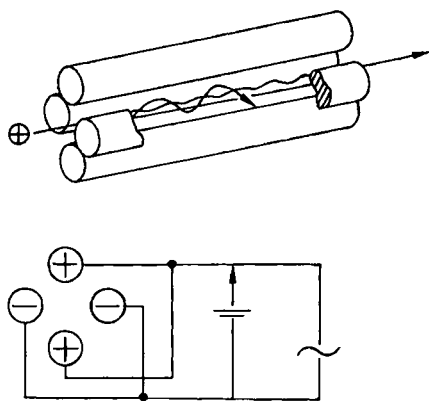
Figure 15. Quadrupole mass spectrometry

By digitizing the interferogram, rapid Fourier transform is possible with a powerful computer. For complex spectra, this is possible by small, but highly accurate stepwise displacements of the side arm. Repetitive scanning intensifies the image, and a reference laser is used to make the mirror positioning reproducible. The technique is well known in infrared spectroscopy and can now be used in the visible and UV regions. Fourier transform atomic emission spectrometry is suitable for sources with a low radiance with a high stability detector to avoid detector noise. Wavelength calibration must be very accurate to achieve maximum resolution. Reasonable signal-to-noise ratios are only achieved with low-noise sources.

21.3.3. Mass Spectrometers

In the sources used in optical atomic spectrometry, considerable ionization takes place, so they are useful ion sources for mass spectrometry [50]. Although, an overall treatment of instrumentation for mass spectrometry is given elsewhere (→ Mass Spectrometry), the most widely used types of mass spectrometer are briefly outlined here. In particular, the new types of elemental mass spectrometry have to be considered, i.e., glow discharges, and inductively coupled and microwave plasmas. In contrast to classical high voltage spark mass spectrometry [51] or thermionic mass spectrometry [52], which are beyond the scope of this article, the plasma sources are operated at considerably higher pressure between some 10 Pa and atmospheric pressure than the pressure in the mass spectrometer itself (10^{-3} Pa). Con-

sequently, there has to be an interface with the appropriate apertures and high capacity vacuum pumps to bridge the pressure difference between source and spectrometer.

In commercial instrumentation both low-cost quadrupole and time-of-flight (TOF) mass spectrometers, and expensive double-focusing sector field mass spectrometers are used [53].

21.3.3.1. Types of Mass Spectrometer

In all mass spectrometers a vacuum of 10^{-3} Pa or better is maintained to avoid further ion formation from residual gas species or collisions of analyte ions with these species. Nowadays, turbomolecular pumps are preferred over diffusion pumps as their maintenance is easier and oil backflow does not occur.

Quadrupole Mass Spectrometers. In a quadrupole instrument (Fig. 15), the spectrometer consists of four equidistant, parallel rods (diameter 10 – 12 mm) between which a d.c. voltage and a high-frequency field (up to 1 MHz) are maintained. The d.c. voltage at the quadrupole should be slightly below the energy of the entering ions (usually below 30 eV). The ions enter through the so-called skimmer and have to pass the ion lenses and then a mask which prevents UV from entering the spectrometer (beam-stop). By changing the voltages at the ion lenses, at the beam-stop, at the quadrupole rods, and on the spectrometer walls, both the transmission of the spectrometer and the resolution of ions of a given mass and energy, can be optimized. As these parameters are all interdependent, optimization is complex. By changing the quadrupole field, the transmission for a certain ion changes. Accordingly, ions of a certain mass can be filtered out by manual setting of the field, or the spectrum can be scanned. The scanning velocity is usually limited to 30 000 mass numbers per second, because of the high-frequency and d.c. components in the spectrometer. Accordingly, the mass interval of 0 – 300 can be scanned in about 30 ms. Mass resolution is mainly determined by the quality of the field, and in quadrupoles for plasma mass spectrometry it is optimally about 1 atomic mass unit. Line profiles are approximately triangular, but sidewings occur as well; together with the isotopic abundances they determine the magnitude of spectral interference. Quadrupole mass spectrometers are rapid sequential measurement systems. In the case of isotopic dilution techniques for calibration,

both precision and figures of merit can be limited by noise in the source. The price of quadrupole instruments is reasonable and their transmission is high. This is important because it allows their use in plasma mass spectrometry, where the source itself can be kept at earth potential. This is an advantage in inductively coupled plasmas.

Time-of-Flight (TOF) Mass Spectrometers. In quadrupole and sector field mass analysers the ion signal is a continuous beam. In TOF mass spectrometry the ion beam is pulsed so that the ions are either formed or introduced into the analyser in "packets". These ion packets are introduced into the field free region of a flight tube 30–100 cm long. The principle behind TOF analysis is that, if all ions are accelerated to the same kinetic energy, each ion will acquire a characteristic velocity dependent on its m/z ratio [54]. The ion beam reaches its drift energy (2700 eV) in less then 2 cm. The ions are then accelerated down the TOF tube with whatever velocity they have acquired. Because all ions have essentially the same energy at this point, their velocities are inversely proportional to the square roots of their masses. As a result, ions of different mass travel down the flight tube at different speeds thereby separating spatially along the flight tube with lighter, faster, ions reaching the detector before the heavier ions. Hence, the m/z ratio of an ion, and its transit time (T, in microseconds) through a flight distance (L, in cm) under an acceleration voltage (V) are given by

$$t = L \sqrt{\left[\left(\frac{m}{z}\right)\left(\frac{1}{2V}\right)\right]} \quad or \quad \frac{m}{z} = \frac{2Vt^2}{L^2} \qquad (124a)$$

TOF mass analyzers are calibrated using two ions of known mass, so exact values of L and V need not be known. The mass calibration is based on the conditions of the analyzer during the entire period of measurement.

While simple in theory, the TOF analyzer caused numerous problems when coupled with a plasma source such as an ICP. The spread in ion kinetic energies caused by the ion sampling process resulted in the ions entering the field free region of the flight tube at different angles [55]–[57]. Another difficulty with ICP-TOF-MS is that ions have to be introduced in "packets". This can be achieved, for example, by using an orthogonal interface with a pulsed repeller plate. Background noise can be reduced by using a com-

bination of quadrupole ion optics and an energy discriminator before the detector.

Sector Field Mass Spectrometers. In a magnetic sector analyzer, ions are subjected to a magnetic field which causes the ions to be deflected along curved paths. The ions are introduced into the analyzer via a series of electrostatic slits which accelerate and focus the ions into a dense ion beam. The velocity of the ions is controlled by the potential (V) applied to the slits. As the ions enter the magnetic sector analyzer they are subjected to a magnetic field parallel to the slits but perpendicular to the ion beam. This causes the ions to deviate from their initial path and curve in a circular fashion. A stable, controllable magnetic field (H) separates the components of the beam according to momentum. This causes the ion beam to separate spatially and each ion has a unique radius of curvature, or trajectory (R), according to its m/z. Only ions of a single m/z value will posses the correct trajectory that focuses the ion on the exit slit to the detector. By changing the magnetic field strength, ions with differing m/z values are brought to focus at the detector slit.

The ion velocity (v) in the magnetic field is given by:

$$\frac{1}{2}mv^2 = zV \quad or \quad v = \sqrt{\frac{2zV}{m}} \qquad (124b)$$

As the ions enter the magnetic field, they are subjected to a magnetic force at right angles to both the magnetic lines of force and their line of flight. This leads to a centrifugal force leading to curvature of the ion beam.

$$\frac{mv^2}{R} = Hzv \qquad (124c)$$

The radius of curvature of the flight path is proportional to its momentum and inversely proportional to the magnetic field strength:

$$R = \frac{mv}{zH} \qquad (124d)$$

By eliminating the velocity term (v) between Equations (124b) and (124d) we get

$$R = \frac{1}{H}\sqrt{2V\left(\frac{m}{z}\right)} \qquad (124e)$$

Hence, ions accelerated through an electrostatic field, uniform in nature, and then subjected
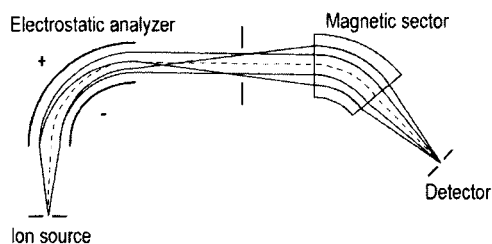
Figure 16. Sector field mass spectrometer

to a uniform magnetic field have different radii of curvature. This leads to ions of a specific m/z value being focused on to the detector slit while all other ions hit the side of the analyzer. Thus, the magnetic field classifies and separates ions into a spectrum of beams with each part of the spectrum having a different m/z ratio, where:

$$\frac{m}{z} = \frac{H^2 R^2}{2V} \qquad (124f)$$

To obtain a complete mass spectrum from a magnetic sector analyzer, either the accelerating voltage (V) or the magnetic field strength (H) is varied. Each m/z ion from light to heavy is sequentially focused on the detector, and hence a mass spectrum is obtained.

Single focusing magnetic sectors, as described above, have the disadvantage that ion energies vary depending on their point of formation in the ion source. The difference in ion energy is accentuated by the accelerating voltage, which leads to peak broadening and low resolution in the single focusing mass analyser.

Double focusing magnetic/electrostatic sector instruments use magnetic and electrical fields to disperse ions according to their momentum and translational energy (Fig. 16). An ion entering the electrostatic field travels in a circular path of radius (R) such that the electrostatic force acting on it balances the centrifugal force. The equation of motion or transmission is:

$$\frac{mv^2}{R} = Ez \qquad (124g)$$

where (E) is the electrostatic field strength. Hence the radius of curvature of the ion path in the electrostatic sector is dependent on its energy and not its mass [58]. A narrow slit placed in the image plane of the electrostatic sector can be used to transmit a narrow band of ion energies. If this sort of analyzer was placed in front of a magnetic

sector analyzer an increase in resolution would result but a decrease in detection would be inevitable due to the decrease in ions exiting the electrostatic sector in comparison with ions exiting the ion source. This loss in sensitivity can be compensated for by the choice of a suitable combination of electrostatic and magnetic sectors, such that the velocity dispersion is equal and opposite in the two analyzers. The narrow ion energy range transmitted the full length of a sector instrument would suggest a loss in sensitivity, compared with quadrupole mass analyzers when both are operated at the same resolution, this has not been observed in practice. This has been mainly attributed to the ion focusing requirements for the different analyzers. Double focusing mass analyzers require a high energy (3 – 8 kV) ion beam for effective ion transmission and resolution. Whereas, if such a high energy beam was focused down a quadrupole, the ions would have obtained a velocity profile too great to be affected by the hyperbolic energy fields of the quadrupole analyzer, which generally requires ion energies of less than 10 eV for effective mass separation. The higher ion energy beam is less affected by space charge interference, which causes scattering of the ion beam, which has lead to increased sensitivity of double focusing instruments.

Other Mass Spectrometers. Spectrometers such as ion traps and ion cyclotron resonance have also been used in atomic mass spectrometry, but have not yet been incorporated into a commercial instrument. The former instrument has the advantage of ion storage and ion – molecule reaction capability, while the latter instrument is capable of extremely high resolution.

21.3.3.2. Ion Detection

For ion detection, several approaches are possible. Classical spark source mass spectrometry has developed the use of photographic plates. Very hard emulsions with low gelatin content are required. Emulsion calibration is similar to that described in optical emission spectrography, but usually varying exposure time are applied instead of using optical step filters. Provided an automated microdensitometer is used, mass spectrography is still a useful tool for survey analysis of solids down to the sub-μg/g level.

Faraday cup detectors consist of a collector electrode which is surrounded by a cage. The elec-
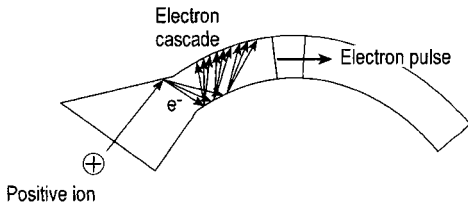
Figure 17. Continuous dynode electron multiplier

trode is positioned at an angle with respect to the ion beam, so that ions exiting the analyzer strike the electrode but secondary emissions are reflected away from the detector entrance. The function of the cage is to prevent detected ions and secondary electrons escaping from the detector. The electrode is connected to ground via a resistor. The ion current striking the electrode is neutralized by electron flow from ground through the resistor. This causes a potential drop which is amplified to create a signal. Currents as low as 10^{-15} A have been successfully measured in this way.

Electron Multipliers. For currents of less then 10^{-15} A an electron multiplier is necessary for detection. When the ion beam exits the analyzer it strikes a conversion plate which converts ions into electrons. The ions are drawn towards the plate by a strong voltage applied to the conversion plate. On striking the conversion plate the ions stimulate the ejection of electrons which are accelerated by the voltage applied to the plate. The electrons are multiplied using either continuous or discrete dynodes. A continuous dynode channel electron multiplier shown in Figure 17. This consists of a curved glass tube of approximately 1 mm in internal diameter with an inner resistive coating and a flared end. The multiplier can be operated in one of two modes. In pulse counting mode — the most sensitive mode of operation — a high voltage of between -2600 V and -3500 V is applied to the multiplier, which attracts ions into the funnel opening. When a positive ion strikes the inner coating the collision results in the ejection of one or more secondary electrons from the surface, which are accelerated down the tube by the potential gradient, and collide with the wall resulting in further electron ejection. Hence, an exponential cascade of electrons rapidly builds up along the length of the tube, eventually reaching saturation towards the end of the tube, resulting in a large electron pulse and a consequent gain of $10^7 - 10^8$ over the original ion collision. The electron pulses

are read at the base of the multiplier and are approximately $50 - 100$ mV and 10 ns in duration. Continuous dynode multipliers consist of a leaded glass tube which contains a series of metal oxides. The tube is curved and electrons are drawn down the tube by the potential gradient established by the resistivity of the glass. The tube is curved to stop electron feedback. A $10^5 - 10^7$ increase in signal is expected; however, the major constricting factor is the background noise of the system. Alternatively, the multiplier can be operated in analogue mode with a gain of only $10^3 - 10^4$ so that the multiplier does not become saturated and the pulses vary greatly in size. In this mode the applied voltage is between -500 V and -1500 V and the electron pulses are read at the collector electrode where they are amplified and averaged over a short time interval to allow rapid data acquisition. The greatest sensitivity is achieved with the detector in pulse counting mode, but the detector will become saturated at counting rates above 10^6 Hz, which are encountered when the analyte is at a high concentration in the sample. If the detector is switched into analogue mode it is less sensitive, but can be used for analyte concentrations which are much higher, typically up to three orders of magnitude higher than for pulse counting. Such dual mode operation results in an extremely large linear dynamic range of up to nine orders of magnitude. A discrete dynode detector consists of an array of discrete dynode multipliers, usually containing $15 - 18$ dynodes, coated with a metal oxide that has high secondary electron emission properties. The dynodes are placed in one of two configurations, either venetian blind or box and grid fashion. Secondary electrons emitted by the metal oxide are forced to follow a circular path, by a magnetic field, so they strike successive dynodes thereby multiplying the signal.

21.3.3.3. Ion Extraction

An important aspect of plasma mass spectrometry is the sampling interface. Ions are physically extracted from the plasma into a mass spectrometer which is required to be at extremely low pressure, so the sampling interface must be in direct contact with the plasma (usually an ICP). The problem of extracting ions from an extremely hot plasma at atmospheric pressure into a mass spectrometer at $\sim 10^{-4}$ Pa (10^{-9} atm) is overcome by making use of a series of differentially pumped vacuum chambers held at consecutively lower pressures. A schematic diagram of the ICP–MS
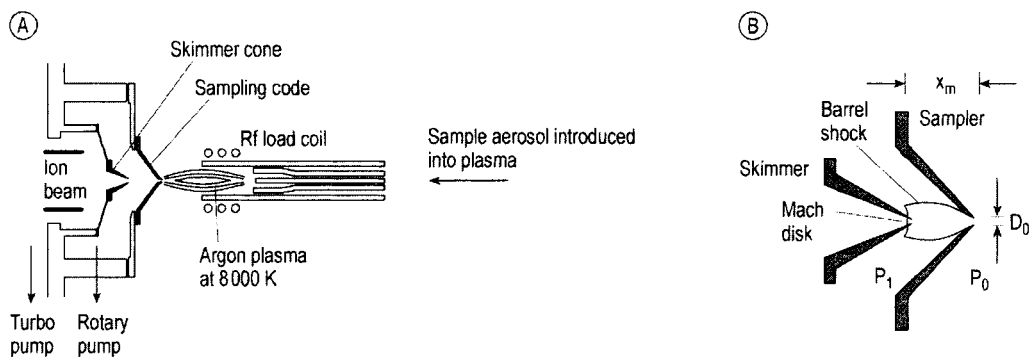
Figure 18. Ion extraction interface for ICP–MS

sampling interface is shown in Figure 18. The ICP is aligned so that the central channel is axial with the tip of a water-cooled, sampling cone, made of nickel or copper, which has an orifice of approximately 1 mm in diameter. The pressure behind the sampling cone is reduced, by means of a vacuum pump, to ~ 200 Pa (2×10^{-3} bar) so the plasma gases, together with the analyte ions, expand through the sampling orifice to form a shock-wave structure. This expansion is isentropic (i.e., no change in the total entropy) and adiabatic (i.e., there is no transfer of energy as heat), resulting in a supersonic expansion accompanied by a fall in temperature. This supersonic expansion takes the form of a cone with a shock-wave stucture at its base called a Mach disk. The region within the expansion cone is called the "zone of silence" which is representative of the ion species to be found in the ICP, i.e., the ionization conditions have been "frozen". The skimmer cone is another metal cone, the tip of which has an orifice of approximately 0.7 mm in diameter, that protrudes into the "zone of silence", and is axially in line with the sampling orifice. The ions from the zone of silence pass through the orifice in the skimmer cone, into a second intermediate vacuum chamber held at $< 10^{-2}$ Pa (10^{-7} bar), as an ion beam. The ion beam can then be focused by means of a series of ion lenses, which deflect the ions along a narrow path and focus them onto the entrance to the mass analyzer.

21.3.4. Data Acquisition and Processing

Modern optical and mass spectrometers are controlled by computers, responsible for wavelength or ion mass presetting, parameter selection, as well as control of the radiation or ion source and the safety systems. The software for data acquisition and processing may comprise routines for the calculation of calibration functions with the aid of linear or higher order regression procedures, for calculation of sample concentrations by standard addition techniques or by calibration with synthetic samples, for calculation of statistical errors, and drift correction. Computers also offer graphical display of parts of the spectrum, or superimposition of spectra of several samples for recognition of spectral interference or background correction.

In most cases, a linear calibration function can be used (see Section 21.3.1). If c_i is the concentration of a standard sample and x_i is the radiation intensity, absorption, or ion current for the element to be determined, the linear regression has the form:

$$x_i = a c_i \qquad (125)$$

where a is the sensitivity, i.e., the slope of the calibration curve. To overcome source fluctuations and for maximum precision, it may be useful to plot signal ratios as the analytical signal:

$$x_i/R = a c_i \qquad (126)$$

where R is the signal obtained for a reference element, which may be the matrix element, a major component, or a foreign element added in known concentration to all samples to be analyzed. This approach is known as internal standardization. If the calibration function is nonlinear, it may be described by a second- or third-degree polynomial:

$$x_i/R = a_0 + a_1 c_i + a_2 c_i^2 \qquad (127)$$

or

$$x_i/R = a_0 + a_1c_i + a_2c_i^2 + a_3c_i^3 \qquad (128)$$

The coefficients a are calculated by regression procedures. Alternatively, different linear calibration functions may be used for different concentration ranges.

Interference or matrix effects occur when the analytical signals depend not only on the analyte element concentrations, but also on the concentrations of other sample components. Additive interference may be avoided by estimation of the magnitude of the interfering signal from a scan of the spectral background in the vicinity of the analytical line. Mathematical interference correction expresses the net analytical signal as:

$$x_i/R = (x_i/R)_{measured} - [d(x_i/R)/dc_{interferent}] \times c_{interferent} \qquad (129)$$

where $c_{interferent}$ must be determined separately and $d(x_i/R)/dc_{interferent}$ is the interference coefficient for a certain sample constituent, determined from a separate series of measurements. One can also express the calibration function as:

$$x_i/R = a_1c_1 + a_2c_2 \qquad (130)$$

where a_1 and a_2 are determined by regression, which is straightforward for a linear relationship.

Certain matrix constituents may produce signal enhancement or reduction. These are multiplicative in nature and can be due to influences on sampling efficiency, analyte transport into the source, or generation of the species delivering the analytical signals. They can be written as:

$$Y = a(c_{interferent})c \qquad (131)$$

where the sensitivity a is a function of the concentration of the interferent $c_{interferent}$. This relation can be linear or of higher order. Calibration by standard addition allows correction of errors arising from signal reduction or enhancement insofar as the spectral background is fully compensated.

When calibrating by standard addition, the concentration of the unknown sample can be determined graphically. This is done by extrapolating to zero the curve of the signals from samples to which known amounts of analyte have been added plotted against the added amounts. Standard addition can be very useful in atomic absorption and

plasma mass spectrometry, which are zero-background methods. However, for atomic emission, where the lines are superimposed on a spectral background from the atomic emission source itself, and which is highly dependent on the sample matrix, it is more difficult. Here, calibration by standard addition can be used only when a highly accurate background correction is applied. It is important that the curve remains within the linear dynamic range of the method. This can be problematic atomic absorption spectrometry.

The precision obtainable in standard addition methods is optimum when the added amounts of analyte are of the same order of magnitude as the analyte concentrations present in the sample. An estimate of the analytical error can be made from the propagation of the errors of the measured intensities [64]. If X is a function of n measured values, each resulting from N replicate measurements, the standard deviation of the estimate $s(X)$ can be calculated from their estimates x_k ($k = 1, 2, 3, \ldots, n$) and standard deviations s_k:

$$s^2(X) = \sum_{k=1}^{n} [(\delta X/\delta x_k)s_k]^2 \qquad (132)$$

For one standard addition [65], the intensities I_x and $I_{x+\Delta}$ for the unknown sample c_x and the standard addition sample $c_{x+\Delta}$ are:

$$c_x = aI_x \qquad (133)$$

$$c_x + \Delta c = aI_{x+\Delta} \qquad (134)$$

and

$$c_x = \Delta cI_x/(I_{x+\Delta} - I_x) \qquad (135)$$

The standard deviation is obtained by applying Equation (132) to c_x. If I_x and $I_{x+\Delta}$ scatter independently and the added concentration Δc is free of error [66]:

$$s^2(c_x) = \{c_x[I_{x+\Delta}/I_x][s(I_x)]/[I_{x+\Delta} - I_x]\}^2 + \{c_x[s(I_{x+\Delta})]/[I_{x+\Delta} - I_x]\}^2 \qquad (136)$$

or

$$s_r^2(c_x) = [I_{x+\Delta}^2/(I_{x+\Delta} - I_x)][s_r^2(I_x) + s_r^2(I_{x+\Delta})] \qquad (137)$$
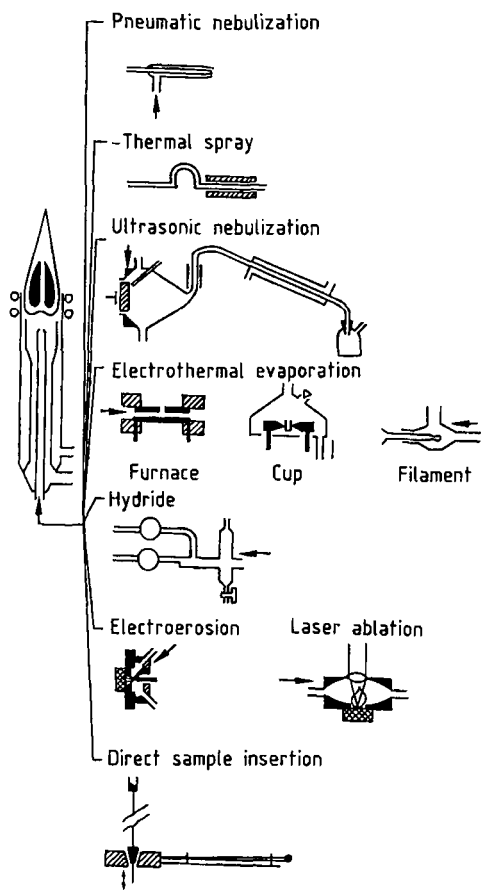
Figure 19. Sample introduction in plasma spectrometry [72]

21.4. Sample Introduction Devices

In atomic spectrometry, the sample must be converted to aerosol form and atomized before excitation. Accordingly, sample introduction is a very important step in all atomic spectrometric methods. The approach varies for different types of sample (size, state of aggregation stability, etc.) [67] – [71].

If sample volatilization and signal generation take place in a single source, both thermal evaporation and cathodic sputtering may occur. The latter is especially important in low-pressure discharges, as the ions can reach high energies when passing through the high-energy zones of the discharge and energy losses are small, because of the low number of collisions. In plasma sources at atmospheric pressure or flames, aerosol generation is often performed in a separate device. Figure 19

shows a number of sampling techniques for the analysis of solutions and solids. The techniques used for plasma spectrometry [73] and for sampling by cathodic sputtering, as performed in low-pressure discharges, are treated in detail in what follows.

21.4.1. Pneumatic Nebulization

For analysis of solutions, pneumatic nebulization is well known from early work on flame emission spectrometry. Pneumatic nebulizers must produce a fine aerosol with a gas flow, low enough to transport the aerosol droplets through the source with low velocity; this is essential for complete atomization. Nebulization of liquids is based on the viscosity drag forces of a gas flow passing over a liquid surface, producing small, isolated droplets. This may occur when the liquid is passed through a capillary tube, with the exit gas flows concentric around the tube, or perpendicular to the liquid stream. It may also occur at a frit, which is continuously wetted and through which the gas passes. Nebulizers are mounted in a nebulization chamber, which separates off the larger particles. Impact surfaces are often provided, fractionate the aerosol droplets further, and increase the aerosol production efficiency. The liquid flow is typically a few milliliters per minute and the maximum efficiency (a few percent) can be achieved at liquid flows of 2 mL/min; even below this flow rate, droplet diameters of the order of 10 μm and injection velocities of the order of 10 m/s are obtained. A number of nebulizer types are used. For flame emission, atomic absorption, and for plasma spectrometry, they include concentric nebulizers, cross-flow nebulizers, Babington nebulizers, and fritted-disk nebulizers (Fig. 20) [68].

Concentric nebulizers date back more than a century, to Gouy [75], and in principle are self-aspirating. They can be made of metal, plastic, or glass. In atomic absorption spectrometry, Pt – Ir capillaries are often used to allow the aspiration of highly acid solutions. In plasma atomic spectrometry, concentric glass nebulizers (Meinhard) [74] are well known. They use a rather lower gas flow (1 – 2 L/min) than atomic absorption types (up to 5 L/min). In both cases, aspiration rates are 1 – 2 mL/min and aerosol efficiencies 2 – 5 %. Direct injection nebulizers are a useful alternative for the introduction of small-volume samples [76].

Cross-flow nebulizers are less sensitive to salt deposits at the nozzle. They are used in plasma
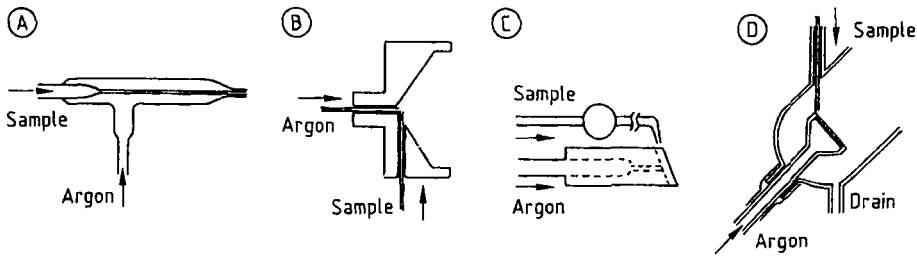
Figure 20. Pneumatic nebulizers for plasma spectrometry [68]
A) Concentric glass nebulizer [74]; B) Cross-flow nebulizer; C) Babington nebulizer; D) Fritted-disk nebulizer

spectrometry and flame work, and require a peristaltic pump for sample feeding. Cross-flow nebulizer capillaries can also be made of metal, glass, or plastic and have similar characteristics to those of concentric nebulizers. Materials such as poly(*p*-phenylene sulfide) (Ryton) enable aspiration of solutions containing HF.

In a *Babington nebulizer*, the sample solution is pumped through a wide tube and over a small gas orifice, so slurries can be analyzed without the risk of clogging.

Fritted-disk nebulizers have high nebulization efficiencies and low sample consumption (down to 0.1–0.5 mL/min). They are especially useful for plasma spectrometry of organic solutions, or for coupling with liquid chromatography [77]. However, in comparison with other types of nebulizer they suffer from memory effects; rinsing times are much higher than the 10–20 s required for other pneumatic nebulizers.

With pneumatic nebulizers, samples can be taken up continuously. After a few seconds of aspiration, a continuous signal is obtained, and can be integrated over a certain time. Alternatively, small sample aliquots can be injected directly into the nebulizer [78]–[81] or into a continuous flow of liquid. The latter is known as *flow injection analysis* (*FIA*) [82]. The signals in both cases are transient; they have a sharp rise and some tailing, resulting from displacement of the aerosol cloud at the signal onset and dilution of the aerosol at its end. Signals last for a few seconds and with sample volumes 10–20 µL are capable of giving sufficient signal intensities. These techniques enable analysis of small samples (down to some 10 µL), reduce salt deposits, and maximize the absolute power of detection. FIA are also amenable to automatation and on-line sample preparation steps, such as preconcentration and matrix removal by coupling to column chromatography [83]–[85].

Optimization of pneumatic nebulizers is concentrated on selection of the working conditions for optimum droplet size and efficiency. The so-called Sauter diameter d_0, i.e., the diameter for which the volume-to-surface ratio equals that of the complete aerosol, is given by the Nukuyama–Tanasawa equation [86]:

$$d_0 = (C/v_G)(\sigma/\varrho)^{0.5} + C'\left\{\eta/\left[(\sigma\varrho)^{0.5}\right]\right\}^{C''} [1000\,(Q_L/Q_G)]^{C'''} \tag{138}$$

where v_G is the gas velocity, Q_G the gas flow, Q_L the liquid flow, η the viscosity, ϱ the density, σ the surface tension of the liquid, and C, C', and C'' and C''' are constants.

When the nebulizer gas flow increases, d_0 becomes smaller, the sample introduction efficiency increases, and so do the signals. However, as more gas is blown through the source, it is cooled, and the residence time of the droplets decreases, so that atomization, excitation, and ionization also decrease. These facts counteract increase of the signals as a result of the improved sampling; maximum signal intensity and power of detection are achieved at a compromise gas flow.

The physical properties of the sample liquid (σ, η, ϱ) also influence d_0, the efficiency, and the analytical signals. Differences in the physical properties of the sample liquids thus lead to so-called nebulization effects. Viscosity has a large influence through the second term in Equation (138). However, in free sample aspiration, it also influences d_0 through Q_L, which is given by the Poiseuille law:

$$Q_L = \left[(\pi R_L^4)/(8\eta l)\right] \Delta P \tag{139}$$

where R_L is the diameter and l the length of the capillary. ΔP is the pressure difference:
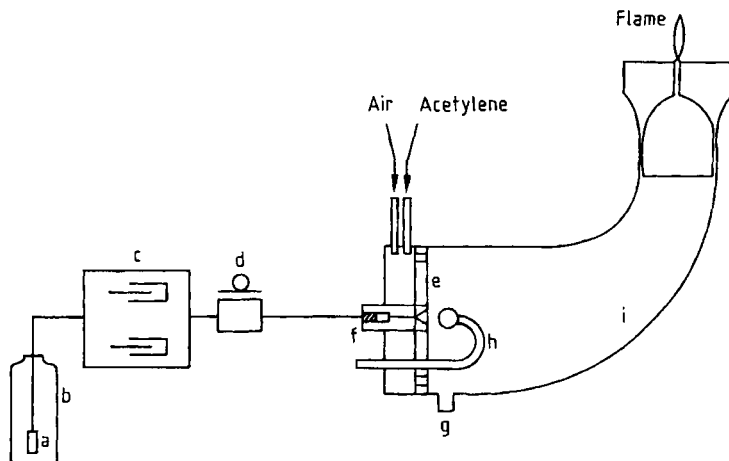
Figure 21. Hydraulic high-pressure nebulization system [94]
a) Solvent filter (20 μm); b) Solution reservoir (water, organic solvents;c) High-pressure pump; d) Sample feed valve/sample loop; e) Plate with concentric holes; f) High-pressure nebulization nozzle (10 – 30 μm) with integrated protection filter; g) Drain; h) Impact bead; i) Gas mixing chamber

$$\Delta P = \Delta P_U - g h_\varrho - f(\sigma) \qquad (140)$$

where ΔP_U is the difference between the pressure inside and outside the nebulizer, h is the height and $f(\sigma)$ a correction factor. In order to minimize nebulization effects, forced feeding with a peristaltic pump can be applied, so that Q_L is no longer a function of η. The nebulizer can also be operated at high gas flow, so that Equation (138) reduces to:

$$d_0 = C/v_G(\sigma/\varrho)^{0.5} \qquad (141)$$

when $Q_G/Q_L > 5000$. However, the gas flow may cool the source, so that incomplete atomization, or signal reduction/enhancement may lead to even stronger interference.

Aerosol droplet size distributions can be determined by stray-light measurements [87], [88]. Data obtained from sampling with cascade impactors may suffer from evaporation of the droplets on their way through the device, especially as sampling may require the application of underpressure [89]. Data for different nebulizers are available [90]. Particle size distributions depend on the nebulizer and its working conditions, on the liquid nebulized, and on the nebulization chamber. These relations have been studied in depth for organic solutions [91], [92], for which, at low gas flow, pneumatic nebulizers often

produce an aerosol with a lower mean particle size compared with aqueous solutions, and thus a higher efficiency and power of detection.

Apart from pneumatic nebulization by gas flow over a liquid surface, other methods have been developed. It is possible to use the impact of a high pressure jet of liquid on a surface (*jet impact nebulization*) [93], or of the expansion of a liquid jet under high pressure at a nozzle (*high-pressure nebulization*), as described by BERNDT [94] (Fig. 21). Much higher nebulization efficiencies can be obtained, even for solutions with high salt content or oils. This is due to the small droplet size, ca. 1 – 10 μm [95]. The system is ideally suited for speciation work as it is the appropriate interface for HPLC and atomic spectrometry [96]. A further advantage lies in the fact that the aerosol properties are independent of the physical properties of the liquid. In plasma spectrometry, however, it is necessary to desolvate the aerosol, as high water loadings cool the plasma too much and decrease the analytical signals; they may also lead to increased matrix effects. Desolvation can be done by heating the aerosol followed by solvent condensation, as applied in atomic absorption by SYTI [97]. New techniques involve the use of Peltier elements for cooling in aerosol desolvation. High-pressure nebulization is very promising for on-line preenrichement or separation by high-performance liquid chromatography.
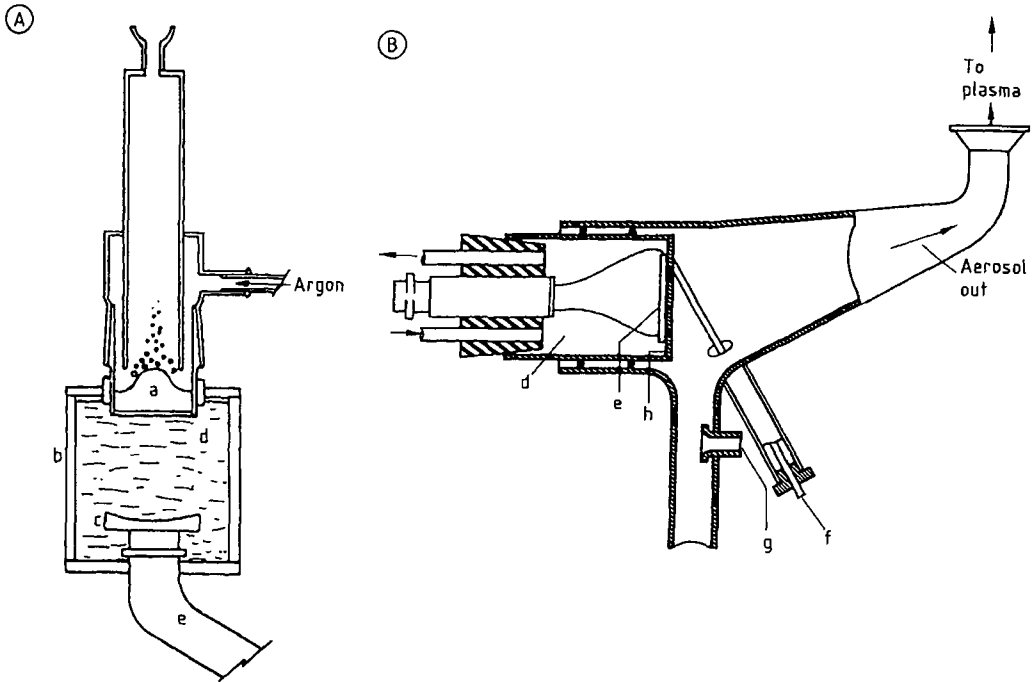
Figure 22. Ultrasonic nebulization
A) Discontinuous system [231]; B) System with liquid flowing over the transducer [99] a) Sample solution; b) Focusing chamber; c) Focusing lens; d) Cooling water; e) Ultrasonic transducer; f) Sample introduction; g) Gas inlet; h) Transducer plate

21.4.2. Ultrasonic Nebulization

By the interaction of sufficiently energetic acoustic waves of suitable frequency with the surface of a liquid, vortices are formed which produce an aerosol. The diameter of the aerosol droplets depends on the frequency and the physical properties of the liquid. For water, aerosol droplets formed at a frequency of 1 MHz have a diameter of ca. 4 μm. The energy can be focused with the aid of a liquid lens on the surface of the sample solution, or the sample liquid can be pumped continuously over the transducer, which must be cooled efficiently (Fig. 22).

Ultrasonic nebulization has two advantages over pneumatic nebulization. The aerosol particles have a lower diameter and a narrower particle size distribution compared with pneumatic nebulization (< 5 compared with 10–25 μm). Therefore, aerosol production efficiency may be up to 30 %, and analyte introduction efficiency is high. No gas flow is required for aerosol production, the trans-

port gas flow can be selected completely freely. However, when applying ultrasonic nebulization to plasma spectrometry, it is necessary to desolvate the aerosol, to prevent too intensive cooling of the plasma. After this measure, ultrasonic nebulization leads to an increase in the power of detection.

Memory effects in ultrasonic nebulization are generally higher than in the case of pneumatic nebulization. The nebulization of solutions with high salt concentrations may lead to salt deposition; special attention has to be paid to rinsing, and precision is generally lower than in pneumatic nebulization. Its state of the art is summarized in [98].

21.4.3. Hydride Generation

For elements with volatile hydrides, sampling efficiency can be increased by volatilizing the hydrides. This applies to elements As, Se, Sb, Te, Bi,

Sn, and some others. By in-situ generation of the hydrides of these elements from sample solutions, sampling efficiency can be increased from a few percent in the case of pneumatic nebulization to virtually 100%.

Hydride generation can be efficiently performed by reduction with nascent hydrogen, produced by the reaction of zinc or $NaBH_4$ with dilute acids. In the latter case, the sample can be brought into contact with a solid pellet of $NaBH_4$, which may prove useful as a microtechnique [100]. Alternatively, a stream of $NaBH_4$ solution stabilized with NaOH can be added to an acidified sample. This can be done in a reaction vessel, as in atomic absorption work [101], or in a flow cell, as in plasma atomic emission spectrometry [102], [103]. The hydrides produced are separated from the liquid and subsequently transported by a carrier gas into the source. The hydrides are accompanied by excess hydrogen. In weak sources, such as microwave discharges, this may disturb discharge stability. To avoid this the hydrides can be separated off, e.g., by freezing them out and sweeping them into the source by subsequent heating [104]. Permselective membranes may also be useful [105].

Hydride generation increases the power of detection of atomic spectrometric methods for the determination of certain elements, and allows their matrix-free determination. However, the technique is prone to a number of systematic errors. First, the hydride-forming elements must be present as inorganic compounds in a well-defined valence state. This may require sample decomposition prior to analysis. In water analysis, treatment with H_2SO_4/H_2O_2 may be effective [106]. Traces of heavy metals such as Cu^{2+} may have a catalytic influence on the formation and dissociation of the hydrides, as investigated by WELZ et al. [107] in atomic absorption with quartz cuvettes. These interferents can be masked by complexation with tartaric acid or coprecipitated with $La(OH)_3$. Calibration by standard addition is advisable.

Mercury compounds can also be reduced to metallic mercury with nascent hydrogen or Sn(II). The released Hg can be transferred to an absorption cell or plasma source. It can also be trapped on a gold substrate, from which it can be swept into the source after release by heating. This approach allows effective preconcentration and, in the case of optical and mass spectrometry, leads to very sensitive determinations of Hg.

21.4.4. Electrothermal Evaporation

Thermal evaporation of analyte elements from the sample has long been used in atomic spectrometry. In 1940, PREUSS [108] evaporated volatile elements from a geological sample in a tube furnace and transported the released vapors into an arc source. Thermal evaporation has also been used in double arc systems, where selective volatilization gives many advantages for direct solid analysis. Electrothermal evaporation became especially important with the work of L'VOV [3] and MASSMANN [4], who introduced electrothermally heated systems for the determination of trace elements in dry solution residues by atomic absorption spectrometry of the vapor cloud.

21.4.4.1. The Volatilization Process

Electrothermal evaporation can be performed with dry solution residues as well as with solids. In both cases, the analyte evaporates and the vapor is held for a long time inside the atomizer, from which it diffuses. The high analyte concentration in the atomizer results from a formation and a decay function. The formation function is related to the production of the vapor cloud. After matrix decomposition, elements are present in the furnace as salts (nitrates, sulfates, etc.). They dissociate into oxides as a result of the temperature increase. In a device made of carbon (graphite), the oxides are reduced in the furnace:

$$MO_{(s/l)} + C_{(s)} \rightleftharpoons M_{(g)} + CO_{(g)}$$
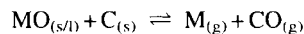
However, a number of metals tend to form carbides, which can be very stable (thermodynamic control) or refractory (kinetic control). In this case, no analyte is released into the vapor phase. Decay of the vapor cloud is influenced by several processes:

1) Diffusion of the sample liquid into the graphite, which can often be prevented by the use of tubes, pyrolytically coated with carbon
2) Diffusion of the sample vapor
3) Expansion of the hot gases during the temperature increase (often at a rate of more than 1000 K/s)
4) Recombination processes, minimized by transferring the sample into the electrothermal device on a carrier platform with low heat capacity [109]
5) Action of purging gases
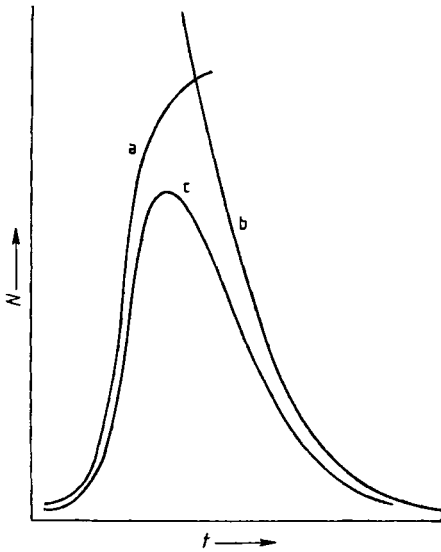
Figure 23. Signal form in graphite furnace atomic absorption spectrometry
a) Supply function; b) Removal function; c) Overall response function

Therefore, in electrothermal devices, transient signals are obtained. They have a sharp increase and a more or less exponential decay, lasting 1 – 2 s. Their form has been studied for volatilization processes [110]. The signal form (Fig. 23) is also influenced by adsorption and resorption of analyte inside the electrothermal device.

21.4.4.2. Types of Electrothermal Device
(Fig.24)

Graphite Furnaces. In most cases, the furnace is made of graphite, which has good thermal and corrosion resistance [3], [111]. As a result of its porosity, it can take up the sample without formation of appreciable salt deposits at the surface. Atomizers made of refractory metals such as tungsten have been used [112]. [113].

Graphite tubes with internal diameters of some 4 mm, wall thickness 1 mm, and a length up to 30 – 40 mm are mostly used. Filaments enabling the analysis of very small sample volumes and mini-cups of graphite held between two graphite blocks have also been described [111]. When graphite tube furnaces are used for optical spectrometry, the optical path coincides with the axis of the tubes, whereas in filaments and cups, the radiation passes above the atomizer. Furnace currents up to 100 A and potentials of a few volts are
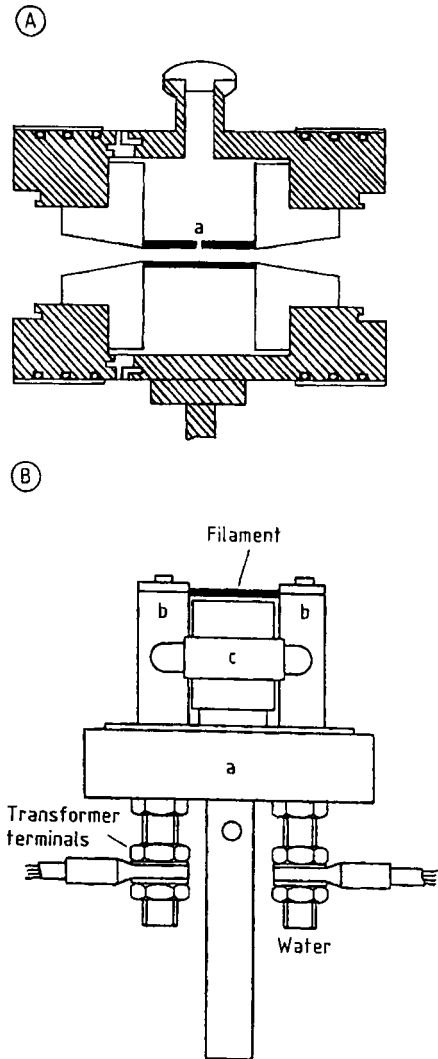
Figure 24. Graphite atomizers used in atomic spectrometry
A) Original graphite tube furnace according to Massmann [4]; B) Carbon-rod atomizer system according to West [109]a) Sampling hole; b) Heating blocks; c) Graphite rod or cup

used, giving temperatures of 3000 K or more. The tube is shielded by an argon flow, assuring stable working conditions for up to several hundreds of firings. With filaments, a lower power is used and the system can be placed in a quartz enclosure to keep out the air. Cups and tubes are nowadays often made of pyrographite, which prevents the analyte solution entering the graphite. Accordingly, chromatographic effects leading to selective volatilization are avoided, and the formation of

refractory carbides, which hampers the volatilization of elements such as Ti or V, decreases.

Refractory Metal Furnaces. Electrothermal furnaces made of refractory metals (tungsten especially) have been described by SYCHRA et al. [112] for use in atomic absorption work. Their heat capacity is generally smaller that of graphite tubes, which results in steeper heating and cooling. This may be extreme in the case of tungsten probes and cups [113], which are mechanically more stable than graphite probes. The signals are extremely short and the analyte is released in very short times, which leads to high signal-to-background ratios and extremely low detection limits, as in wire loop atomization in atomic absorption spectrometry [114] or wire loops used for microsample volatilization in plasma spectrometry [115]. Therefore, metal devices used for electrothermal evaporation find it hard to cope with high analyte loading. Small amounts of sample are easily lost during heating, as the sample is completely at the surface and thermal effects cause stress in the salt crystals, formed during the drying phase. In graphite, the sample partly diffuses into the graphite, suppressing this effect. Graphite is further advantageous, as for many elements, reduction occurs, which leads to free element formation as required for atomic absorption spectrometry.

21.4.4.3. Temperature Programming

When electrothermal evaporation is used for sample introduction, the development of a suitable temperature program is of prime importance. In liquid samples, a small sample aliquot ($10-50\ \mu L$) is transferred to the electrothermal device with a syringe or with the aid of an automated sampler and several steps are performed.

Drying. The solvent is evaporated and the vapor allowed to escape, e.g., through the sampling hole in the case of a graphite furnace. This step may last from tens of seconds to several minutes, and may take place at a temperature near the boiling point, e.g., 105 °C for aqueous solutions. This procedure often comprises several steps, advisable in the case of serum samples to avoid splashing.

Matrix Destruction. During this step, the matrix is decomposed and removed by volatilization. Chemical reactions are often used to enable volatilization of matrix constituents or their compounds. The temperature must be selected so that the matrix, but not the analyte, is removed. This is often achieved by applying several temperatures or even by gradually increasing the temperature at one or more ramping rates. Temperatures during this step are between 100 and 1000 °C, depending on the matrix to be removed and the analyte elements. Thermochemical reagents are often used to assist matrix destruction chemically and to achieve complete matrix removal at a lower temperature. Quaternary ammonium salts are often employed for the destruction of biological matrices at relatively low temperatures. Matrix decomposition is very important to avoid the presence of analyte in different chemical compounds, which can lead to transient peaks with several maxima.

Evaporation. The temperature selected for analyte evaporation depends strongly on the analyte elements and may range from 1000 K for volatile elements (e.g., Cd, Zn) to 3000 K for refractory elements (Fe). This step normally lasts no longer than 10 s, to maximize the number of firings which can be performed with a single tube.

Heating. The temperature is increased to its maximum (e.g., 3300 K over ca. 5 s with graphite tubes) to remove any sample residue from the evaporation device and to minimize memory and cross-over effects.

Direct solid sampling in electrothermal evaporation can be performed by dispensing an aliquot of a slurry prepared from the sample into the furnace. The analytical procedure is completely analogous to that with solutions. Powders can also be sampled with special dispensers, e.g., as described by GROBENSKI et al. [116]. They allow sampling of a few milligrams of powder reproducibly. Analyses with such small amounts make high demands on sample homogeneity. In direct powder sampling, the temperature program often starts with matrix decomposition and proceeds as for dry solution residues. In modern atomic spectrometers, sampling temperature programming, selection of gas flow, and visualization of the temperature program and signals are mostly controlled by computer.

The development of the temperature program is most important in the methodology of a spectrometric method using electrothermal evaporation. It should be documented completely in an analytical prescription for the determination of a given series of elements in well-defined samples.
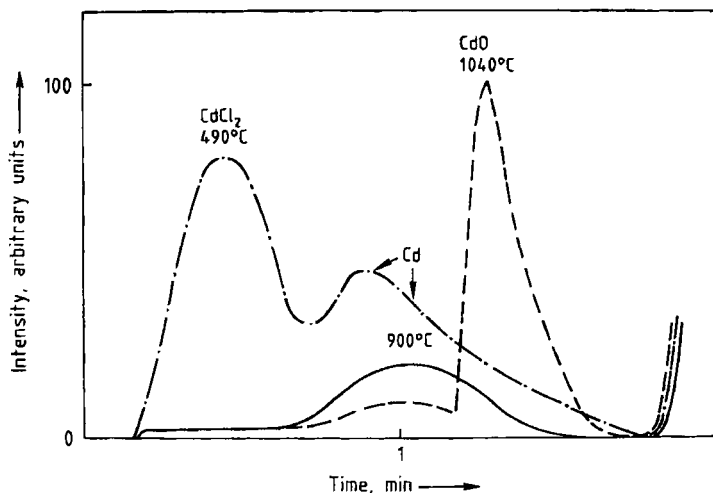
Figure 25. Thermochemical behavior of Cd and its compounds in graphite furnace evaporation [66]
ICP3kW argon – nitrogen ;samples 10 µg/mL Cd(NO$_3$)$_2$ in aqueous solution (solid line), added to 0.1 g/mL of SRM 1571 (dashed line) and SRM 1577 (NaCl) (dash-dotted line); Cd II 226.5 nm line; sampling volume 50 µL; temperature program: drying 95 °C, gradual temperature increase from 350–2400 °C (600 °C/min)

21.4.4.4. Analytical Performance

Electrothermal atomization, because of its high analyte vapor generation efficiency (in theory 100%), can yield extremely high absolute and relative detection power with any type of atomic spectrometry. In two-stage procedures, where the analyte vapor has to be transported into the signal generation source, diffusional losses of analyte vapor may occur. This has been described in detail for Cd [117], but is a general problem.

Owing to the transient nature of the analytical signals, precision is generally lower than with nebulization of liquids. Relative standard deviations are 3 – 5% for manual injection of microaliquots, and 1 – 3% for automated sample dispensation, whereas in pneumatic nebulization they are below 1%.

Interference in electrothermal evaporation may stem from differences in the physical properties of the sample liquids from one sample to another. This may influence the wetting of the graphite or the metal of the electrothermal device. When the latter is temperature dependent, it leads to differences in volatilization. Differences in the anions present may severely influence evaporation (chemical matrix effect). The boiling points of the compounds formed dictate the volatilization. Accordingly, the occurrence of double peaks is easily understood. In the case of Cd, this is documented for a sample rich in chloride (CdCl$_2$ evaporates at

480 °C) and for a sample rich in nitrate (Cd volatilizes first at 900 °C and then at 1200 °C as CdO; Fig. 25) [66].

21.4.5. Direct Solid Sampling

These techniques are especially useful in two-stage procedures for solid analysis. A separate source is used for the generation of a sufficiently concentrated sample vapor cloud, the composition of which truly reflects the sample. Several approaches may be used according to sample properties.

21.4.5.1. Thermal Methods

For powders, electrothermal evaporation may be used. It is often hampered by sample inhomogeneity, as it is generally a micro method, as well as by anion effects, causing chemical interference [118]. Transport losses may also occur. These vanish when the sample is inserted directly in the signal generation source and evaporated thermally. This approach is known from work with graphite or metal probes in atomic absorption, where tungsten wire cups and loops are used. In spectrometry with inductively coupled plasmas, the technique is used both in atomic emission [119] – [121] and in mass spectrometry [122]. Its absolute power of detection is extremely high and

the technique can be used both for the analysis of dry solution residues as well as for the volatilization of micro-amounts of solids.

In-situ thermal destruction of the matrix or selective volatilization can be applied. The latter has proved useful in geological samples [123]. Selective volatilization of volatile elements from refractory matrices is useful; spectral interference from matrices with complex spectra can be avoided. The approach can also be used for the volatilization of refractory oxide- or carbide-forming elements, which form volatile halogenides (e.g., Ti). Here, substances such as AgCl or polytetrafluoroethylene (PTFE) powder can be used as thermochemical reagents [121]. In the case of Ti in the presence of C and PTFE, the following reactions have to be considered:

$$TiO_2 + 2\,C \rightleftharpoons Ti + 2\,CO$$
$$Ti + C \rightleftharpoons TiC$$
$$Ti + 4\,F \rightleftharpoons TiF_4$$
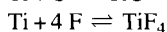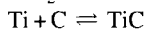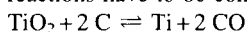
At high temperature, the latter reaction is favored thermodynamically as well as kinetically (TiF_4 is volatile).

21.4.5.2. Slurry Atomization

Powders often form stable slurries in suitable dispersing fluids. The latter should be able to wet the powder and to form stable suspensions, both of which depend on their physical properties (viscosity, density, surface tension). Surface tension, which is generally a function of pH, is known to influence the surface charge of suspended particles and hence the stability of the slurry (for ZrO_2 powders see [124]). The stability of suspensions depends greatly on the particle size of the powder and its tendency to form agglomerates. The latter depends on the preparation method applied, as known from synthetic ceramic powders [125], [126]. The particle size should be in the low micrometer range at most. The particle size distribution can be determined by laser diffractometry or by automated electron probe microanalysis [127], and is particularly critical when the aerosol is produced by pneumatic or ultrasonic nebulization. Measurements with powders of different particle size have shown that particles larger than 12 – 15 μm cannot be held in aerosol form with gas flows ≤ 2 L/min, as in the nebulizers used in plasma spectrometry [198]. In the other techniques of sample introduction, particle size is not so critical. This applies when slurries are analyzed by graphite furnace atomic absorption spectrometry. Volatile elements can be evaporated selectively from some types of sample, as recently shown for ceramic powders [128]. In any case, it is often necessary to apply ultrasonic stirring to destroy agglomerates and to disperse powders optimally, prior to slurry analyses. The addition of surface active substances such as glycols [129] has been proposed, but may introduce contamination when trace determinations are required.

In pneumatic nebulization, it is advisable to stir the sample continuously during slurry aspiration. A Babington nebulizer should be used, as other types are prone to clogging. The aerosol loaded with the solid particles, with or without a solvent layer, is transported to the source. Complete drying and evaporation on its way through the source is necessary for calibration by standard addition with aqueous solutions of the analyte. From knowledge of the temperature distribution in the source, the trajectory of the particles through the source, and the thermochemical properties of the sample material, it is possible to calculate the degree of volatilization for particles of a given size and composition at a certain point in the source, or the maximum size for particles to be completely volatilized [126], [130], [131].

Slurry analysis can also be applied to compact samples which are difficult to dissolve after pulverization. Attention must be paid to possible contamination resulting from abrasion in the mill. For mills of very hard materials such as tungsten carbide abrasion in the case of ceramics may amount to serveral percent of the sample mass. Slurry sampling has been reviewed comprehensively in [132].

21.4.5.3. Arc and Spark Ablation

Arc ablation has long been used for producing an aerosol at the surface of electrically conducting samples, in combination with various sources. In the version described by JOHNES et al. [133], the metal sample acts as the cathode of a d.c. arc discharge. With an open circuit voltage of 600 V and currents of 2 – 8 A, a broad pulse spectrum (mean frequency up to 1 MHz) is observed and rapid movement of the discharge across the cathode produces uniform sampling over a well-defined area. A gas stream can then transport the aerosol particles to the signal generation source. Remote sampling at up to 20 m makes the technique attractive for the analysis of large samples. On the other hand, it is also useful for precision analyses, both by atomic absorption and plasma spectrometric methods. The analysis of electri-

cally insulating powders is possible. A mixture of the powder to be analysed and copper powder in a ratio of about 1:5 can be made and formed into a pellet at pressures up to 80 t/cm^2 (ca. 80 MPa) [134]. The mass ratio of analyte to copper may differ considerably from one sample to another.

Spark Ablation. When sparks are used, the ablation of electrically conducting solids is less dependent on matrix composition. This applies both to high-voltage sparks (up to 10 kV) and to medium-voltage sparks, as used for emission spectrometric analysis of metals. Voltages of 500–1000 V with repetition rates up to 1 kHz are used. Increase in voltage and spark repetition rate leads to a considerable increase of the ablation rate. For aluminum samples, trapping the ablated material in concentrated HNO$_3$ increases the ablation rate from 1–5 μg/min for a 25-Hz spark. However, the particle diameter also increases significantly hampering volatilization of the material in a high-temperature source such as ICP, and leading to flicker noise [135]. It is better to use a medium-voltage spark at a high sparking frequency [136]. After aerosol sampling on Nucleopore filters, X-ray fluorescence spectrometry shows that the composition of the aerosol produced in aluminum samples agrees with the composition of compacted samples [137]. The particle size in the case of 400-Hz spark is 1–2 μm and the particles are mostly spherical, which argues for their formation by condensation outside the spark channel. In the case of supereutectic Al–Si alloys ($c_{Si} > 11$ wt%) some elements (especially Si) are enriched in the smaller particles.

Spark ablation is very abrasive when performed under liquid. BARNES and MALMSTADT [138] used this effect to reduce errors stemming from sample inhomogeneity in classical spark emission spectrometry. The approach is also useful for the dissolution of refractory alloys, which are highly resistant to acids [139]. When sparking under liquid, ablation rates up to ca. 3 mg/min can easily be achieved. In high-alloy Ni–Cr steels, electron probe micrographs show that selective volatilization can become problematic. Very stable colloids can be formed when a ligand such as ethylenediaminetetraacetic acid is present in the liquid.

21.4.5.4. Laser Ablation   (→ Laser Analytical Spectroscopy)

The interaction of the radiation from high-power laser sources with solids can lead to evaporation. As this is independent of the electrical conductivity of the sample, laser ablation is increasingly important for solid analysis [140]–[142].

Sources. Laser sources make use of population inversion. When radiation enters a medium, both absorption and stimulated emission of radiation may occur and the change in flux at the exit is:

$$dF = \sigma F(N_2 - N_1)dz \qquad (142)$$

where dz is the length of the volume element in the z direction, σ is the cross section for stimulated emission or absorption and F is the flux. The sign of $(N_2 - N_1)$ is normally negative, as the population of the excited state N_2 is:

$$N_2 = N_1 \exp\left[-(E_2 - E_1)/kT\right] \qquad (143)$$

and is smaller than the population of the ground state N_1. In the case of population inversion $N_2 > N_1$, the medium acts as an amplifier. When it is positioned between two mirrors, one of which is semitransparent, energy can leave the system. If R_1 and R_2 are the reflectances of the mirrors, the minimum theoretical population inversion is:

$$(N_2 - N_1)_{th} = 1/(2\sigma l) \ln\left[1/(R_1 R_2)\right] \qquad (144)$$

as the losses by reflection in a double pathway $2l$ are compensated. Population inversion requires the existence of a three- or four-level energy diagram. Laser radiation is monochromatic and coherent. The beam has a radiance–divergency relation α_D ($\approx \lambda/D$, where D is the beam diameter). The medium is in a resonator of which the length d determines the resonance frequency $v = n\,c/(2\,d)$, where n is the mode.

Solid state lasers are especially of interest for laser ablation. The laser medium is a crystal or a glass, doped with a transition metal. The medium is pumped optically by flash-lamps (discharges of 100–1000 J over a few ms) or continuously with a tungsten–halogen lamp. The resonator may be the space between two flat mirrors, or an ellipsoid, with the laser rod at one focus and the flash-lamp at the other. The ruby laser emits in the visible region (694.3 nm). It is thermally very robust, but

requires a high pumping energy and its energy conversion efficiency is low. The neodymium – yttrium aluminum garnet (Nd – YAG) laser is very widely used. Its wavelength is 1.06 μm and it has a much higher energy conversion. Gas lasers with CO_2 or Ar can be pumped electrically have a high power output, whereas dye and semiconductor lasers are mostly used for selective excitation.

Lasers can be operated in the Q-switch mode, so as to deliver their energy spikes very reproducibly. Optoacoustic or electroacoustic switches are often used. The first make use of a pressure-dependent change of refractive index of gases such as SF_6, which may be produced periodically at a suitable frequency (Boissel switch), whereas the latter are based on periodic changes of the transmittance of crystal in an a.c. electric field. As the interaction of laser radiation with solids highly depends on wavelength, frequency doubling, resulting from nonlinear effects, e.g., in $LiNbO_3$, is often used.

Interaction with Solids [142], [143]. When a laser beam with small divergence impinges on a solid surface, part of the energy ($10-90\%$) is absorbed and material evaporates locally. The energy required varies between ca. 10^4 W/cm^2 (for biological samples) and 10^9 W/cm^2 (for glasses). A crater is formed, of minimum diameter determined by diffraction of the laser radiation:

$$d \approx 1.2 f \alpha_D v$$

where f is the focal length of the lens used for focusing the laser radiation on the sample (5 –50 mm, minimum dictated by the risk of material deposits) and α_D is the divergency (2 – 4 mrad). Typical crater diameters are of the order of 10 μm. They also depend on the energy and the Q-switch used. The depth of the crater relates to the laser characteristics (wavelength, radiant density, etc.) as well as to the properties of the sample (heat conductance, evaporation energy, reflectance, etc.). The ablated material is ejected from the surface at high velocity (up to 10^4 cm/s); it condenses and is volatilized again by absorption of radiation. In this way, a laser vapor cloud is formed of which the temperature and the optical density are high, and the expansion velocity, composition, and temperature depend on the laser parameters and the gas atmosphere and pressure. Favorable working conditions in laser ablation work involve reduced pressures (1 –10 kPa) [144] – [146]. Optical and mass spectrometry can be performed directly on the laser vapor cloud, or the ablated material can be transferred to another source.

Analytical Performance. Laser ablation is a microsampling technique and thus enables local distribution analyses. Laterally resolved measurements can be made with a resolution of ca. 10 μm [144]. Sampling depth can be varied from 1 to ca. 10 μm by adjusting the laser. The amounts of material sampled can be of the order of 0.1 – 10 μg. As a result of the high signal generation efficiencies possible with mass spectrometry or laser spectroscopy, its absolute power of detection is very high. The laser sources now available allow high precision (relative standard deviations ca. 1 %), limited only by variations in reflectivity along the sample surface, and sample homogeneity.

21.4.6. Cathodic Sputtering

In low-pressure discharges, the atoms and ions undergo few collisions in the gas phase. Therefore, an electric field can induce in them the high energies required to remove material from solid lattices by impact and momentum transfer. Positive ions in particular can acquire the necessary energies; the phenomenon takes place when the sample is used as cathode and is known as cathodic sputtering. Its nature can be understood from the properties of low-pressure discharges. Models developed in physical studies reveal the analytical features of cathodic sputtering [147].

Low-Pressure Discharges. When a voltage is applied across two electrodes in a gas under reduced pressure, the ionization in the vicinity of the electrodes produces electrons, which may gain energy in the field and cause secondary ionization of the gas by collision. Secondary electron emission occurs when they impact on the electrodes. Field emission may take place near the electrodes (at field strengths above 10^7 V/cm) and, when the cathode is hot, glow emission can take place as well. In such a discharge, energy exchange takes place as a result of elastic and inelastic collisions, of which the latter may lead to ionization and recombination. With increasing voltage across the electrodes, a current builds up.

The characteristic (Fig. 26) includes the region of the corona discharge and the normal region, where the discharge starts to spread over the elec-
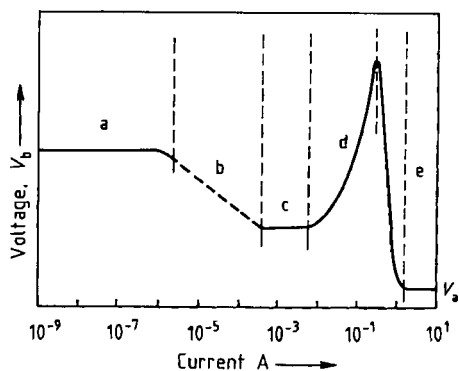
Figure 26. Current – voltage characteristic of a self-sustaining d.c. discharge
a) Townsend discharge; b) Corona range; c) Glow discharge with normal cathode fall; d) Glow discharge with abnormal cathode fall; e) Arc discharge
V_a = Arc voltage [148]; V_b = Breakdown voltage

Figure 27. Geometry (A) and potential distribution (B) of a d.c. electrical discharge under reduced pressure
a) Aston dark space; b) Hittorf dark space; c) Negative glow; d) Faraday dark space; e) Positive column; f) Anode region
V_A = Anode fall; V_C = Cathode fall

trodes. Here, the current can increase at practically constant voltage, whereas once the discharge covers the entire electrodes (restricted discharge), the current can only increase when the voltage is increased drastically (abnormal part of the characteristic). Here, the burning voltage, the positive space charge, and the field gradient in front of the cathode are very high.

Once the cathode has been heated to a sufficiently high temperature, thermal emission may start and the characteristic enters the arc discharge region, where burning voltage and space charge decrease rapidly. Such a discharge displays a cathode dark space where the energies are too high for efficient collisions, the cathode layer where intensive emission takes place, a further dark space, and the negative glow where the negative space charge is high (Fig. 27). The plasma is not in local thermal equilibrium, as the number of collisions is too low to thermalize the plasma. Electron temperatures are high (5000 K for the electrons involved in recombination and > 10 000 K for the high-energy electrons responsible for excitation by electron impact), but the gas temperatures are low (< 1000 K). In radio frequency (r.f.) discharges, which are increasingly important for atomic spectrochemical analysis, a barrier-layer and a bias voltage are built up in the vicinity of a metal electrode as a result of the different mobilities of electrons and ions in the r.f. field. Accordingly, the energy situation becomes very similar to that of d.c. discharges and similar sputtering phenomena can occur.

Ablation. The models developed for cathodic sputtering start from ideal solids, i.e., single, perfect crystals; whereas real samples in atomic spectrochemical analyses are polycrystalline or even chemically heterogeneous. The available models are valid only for monoenergetic beams of neutrals impacting on the sample, whereas in fact both ions and atoms of widely different energies impact at different angles.

The ablation is characterized by a sputtering rate q (µg/s) and a penetration rate w (µm/s). The latter is the thickness of the layer removed per unit time and relates to the sputtering rate as:

$$w = (10^{-2}q)/(\varrho s) \qquad (145)$$

where s is the target area (cm^2) and ϱ its density (g/cm³). The sputtering yield S indicates the number of sputtered atoms per incident particle:

$$S = (10^{-6}q N_A \, e)/(M \, i^+) \qquad (146)$$

where N_A is the Avogadro constant, e the charge on the electron (C), M the atomic mass, and i^+ the ion current (A).

Classical sputtering experiments with a monoenergetic ion beam in high vacuum show that the sputtering yield first increases with the mass of the incident ions and the pressure, but then decreases. For polycrystals, it is maximum at incident angle 30°. For single crystals, it is maximum in the direction perpendicular to a densely packed plane.

The results of these experiments can only be explained by the impulse theory. According to this theory a particle can be removed from a lattice site when the displacement energy E_d (the sum of the covalent or electrostatic binding energies) is delivered by momentum transfer from the incident particles. The maximum fraction of the energy transferrable from an incident particle is:

$$E_{max}/E = 4mM/\left[(m + M)^2\right] \qquad (147)$$

The ablation rate is thus proportional to the number of particles which deliver an energy equal to the displacement energy. It should, however, be taken into account that some incident particles are reflected (f_r) or adsorbed at the surface. Particles of low mass can penetrate into the lattice and be captured (f_p). Other particles enter the lattice and cause a number of collisions until their energy is below the displacement energy. The overall sputtering yield, accounting for all processes mentioned, is:

$$S = \left[(\alpha E)/E_d\right]^{1/2}(f_p + Af_r)\varphi$$
$$\alpha = 2mM/(m + M)^2$$

and

$$\varphi = f(m,M) \qquad (148)$$

Accordingly, cathodic sputtering increases with the energy of the incident particles and is inversely proportional to the displacement energy E_d. It is maximum when $m = M$. This explains why sputtering by analyte particles which diffuse back to the target is very efficient.

The dependence of sputtering yield on the orientation of the target with respect to the beam can be easily explained. In a single crystal, there is a focusing of momentum along an atom row in a direction of dense packing. If D_{hkl} is the lattice constant and d the smallest distance between atoms (or ions) during a collision, the angle θ_0, at which particles are displaced from their lattice sites relates to the angle θ_1, between the direction of the atom row and the line between the displaced atom and its nearest neighbor in the next row, according to the equation:

$$\theta_1 = \theta_0(D_{hkl}/d - 1) \qquad (149)$$

This focusing of momentum, referred to as Silsbee focusing, takes place when:

$$f = \theta_{n+1}/\theta_n$$

or

$$D_{hkl}/d < 2$$

Analytical Performance. The model described explains the features of cathodic sputtering as a technique for sample volatilization and is very helpful for optimizing the performance of cathodic sputtering sources [149]. When using sputtering for sample volatilization, it should be noted that some features can be realized only when working at sputtering equilibrium. On initiating a discharge, the burning voltage is normally high in order to break through the layer of oxides and of gases adsorbed at the electrode surface. When these species are sputtered off and the breakdown products are pumped away, the burning spot (crater) can start to penetrate with a constant velocity in the sample, so that the composition of the ablated material may become constant. The time required to reach sputtering equilibrium (burn-in time) depends on the nature and pretreatment of the sample as well as on the filler gas used and its pressure. All measures which increase ablation rate will shorten burn-in times.

The topography of the burning crater depends on the solid-state structure of the sample. It reflects the graininess, chemical homogeneity, and the degree of crystallinity. Inclusions and defects may locally disturb the sputtering. These effects can be observed on micrographs, comparing the craters obtained with a glow discharge and a spark (Fig. 28) [150]. The roughness of the burning crater can be measured with sensing probes; it imposes the ultimate limitation of the depth resolution obtainable when applying sputtering to study variations of sample composition with depth from the sample surface.

Achievable ablation rates depend on the sample composition, the discharge gas, and its pressure. As filler gas, a noble gas is normally used. With nitrogen or oxygen, chemical reactions at the sample surface occur and disturb the sputtering, as electrically insulating oxide or nitride layers are formed. Reactions with the ablated material produce molecular species which emit molecular band spectra in optical atomic spectrometry, or produce cluster ion signals in mass spectrometry. In both cases, severe spectral interference can hamper the measurement of the analytical signals. The relation between ablation rates and sample composition can be understood from impulse theory. In most cases argon $(m = 40)$ is used as sput-
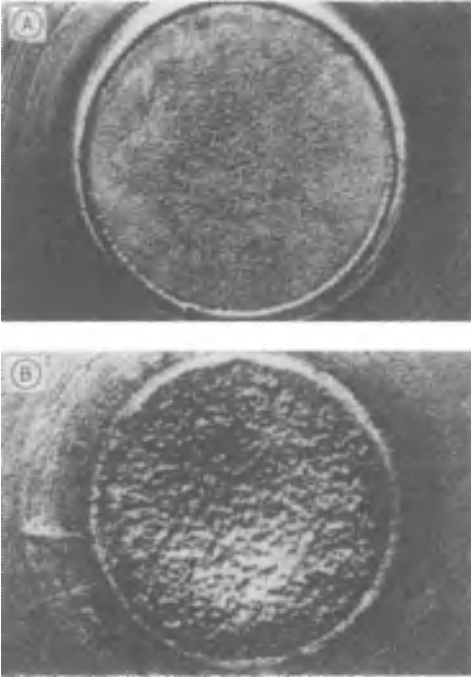
Figure 28. Burning spots obtained with a glow discharge (A) and a medium-voltage spark (B)
Sample: aluminum [150]

tering gas, and the sequence of the ablation rates follows the mass sequence: C < Al < Fe < steel < Cu < brass < Zn

Helium is not suitable, as its small mass renders its sputtering efficiency negligible. The sputtering rates further increase in the sequence Ne < Ar < Kr < Xe

The last two gases are rarely used because of their price. Neon is attractive because of its high ionization potential; argon may cause spectral interference. Gas pressure has a very important influence on electrical characteristics. At low gas pressure the burning voltage is high, as is the energy of the incident particles. At high pressure, the number density of potential charge carriers is higher and the voltage decreases. The number of collisions also increases, so the energy of the incident particles decreases. The resulting decrease of sputtering rate with gas pressure for a glow discharge with planar cathode and abnormal characteristic [151], can be described:

$$q = c/\sqrt{p} \qquad (150)$$

where c is a constant and p is the pressure.

Many studies have been performed on volatilization by cathodic sputtering in analytical glow discharges used as sources for atomic emission, atomic absorption, and mass spectrometry [152]. Also, studies on the trajectories of the ablated material have been performed, as described, e.g., for a pin glow discharge [153].

21.5. Atomic Absorption Spectrometry

21.5.1. Principles

As a method for elemental determinations, atomic absorption spectrometry (AAS) goes back to the work of WALSH in the mid-1950s [2]. In AAS, the absorption of resonant radiation by ground state atoms of the analyte is used as the analytical signal. This process is highly selective as well as very sensitive, and AAS is a powerful method of analysis, used in most analytical laboratories. Its methodological aspects and applications are treated in several textbooks [154]–[156].

A primary source is used which emits the element-specific radiation. In the beginning, continuous sources were used with a high-resolution spectrometer to isolate the primary radiation. However, due to the low radiant densities of these sources, detector noise limitations occurred, or the spectral band width was too large for sufficiently high sensitivity.

Generally, a source which emits discrete lines was found more suitable, with low-cost monochromators to isolate the radiation. It was also necessary in early work to use sources in which the physical widths of the emitted analyte lines are low. This is necessary to get high absorbances (Fig. 29). When the band width of the primary radiation is low with respect to the absorption profile of the line, a higher absorption results from a given amount of analyte as compared with a broad primary signal. Narrow lines in the primary radiation are obtained with low-pressure discharges, as in hollow cathode lamps or low-pressure r.f. discharges. Recently, the availability of narrow-band, tunable laser sources, such as the diode lasers, has opened new perspectives [157], [158]. Only the analytical line is present and no monochromator is necessary. If tuning is applied, it is possible to scan across the absorption profile, to increase the dynamic range or to perform back-
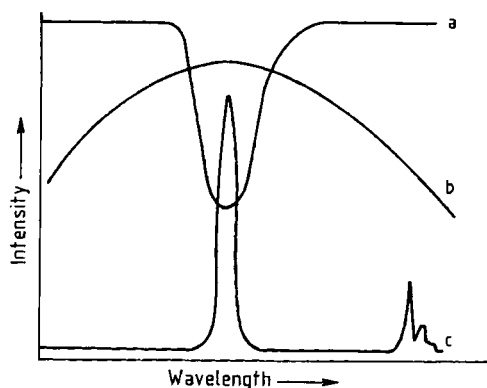
Figure 29. Importance of physical line widths in atomic absorption spectrometry
a) Absorption signal for elemental line; b) Spectral bandpass of monochromator; c) Emission of hollow cathode lamp

ground correction. Switching from one analyte element to another becomes feasible, certainly when several diode lasers are provided. However, laser sources are restricted by the limited spectral range which can be covered — at the time of writing, only down to the green.

The use of line sources generally leads to high analytical selectivity, sensitivity, and power of detection as well as a high dynamic range within which the Lambert – Beer law (Eq. 38) applies. A linear relationship between absorption and concentration can only apply when all radiation passing to the detector is absorbed to the same extent by the analyte atoms (ideal case). In the normal case, the curve of concentration and absorption falls off toward the concentration axis, as a result of the presence of nonabsorbed radiation. For a primary source emitting narrow lines (widths below one-fifth of those of the absorption lines) this consists mainly of contributions from nonabsorbed lines of the cathode material or of the filler gas, which fall within the spectral bandwidth of the monochromator. At high concentrations, a decrease in dissociation gives rise to lower absorbances, and suppression of ionization leads to higher absorbances.

For AAS, the analyte must be present as an atomic vapor, i.e., an atomizer is required. Both flames and furnaces are used, and the corresponding methodologies are known as flame AAS and graphite furnace AAS. Special methods of atomization are based on volatile compound formation, as with the hydride technique (Sections 21.4.3 and 21.5.5). AAS is generally used for the analysis of

liquids, so solids must first be dissolved. Wet chemical decomposition methods are of use, but involve all the care required in trace analysis to prevent losses or contamination, and the resultant systematic errors. For direct solid analysis, a few approaches exist and are discussed in Chapter 21.4.

21.5.2. Spectrometers

Atomic absorption spectrometers contain a primary source, an atomizer with its sample introduction system, and a monochromator with a suitable detection and data acquisition system (Fig. 30).

21.5.2.1. Spectrometer Details (Fig. 30)

Radiation from the primary source (a) is led through the absorption volume (b) and subsequently into the monochromator (c). As a rule, radiation densities are measured with a photomultiplier and processed electronically. Czerny – Turner or Ebert monochromators with low focal length (0.3 – 0.4 m) and moderate spectral bandpass (normally not below 0.1 nm) are frequently used.

In most instruments, the radiant flux is modulated periodically. This can be achieved by modulating the current of the primary source, or with the aid of a rotating mirror (d) in the radiation beam. Accordingly, it is easy to differentiate between the radiant density emitted by the primary source and that emitted by the flame by lock-in amplification. Both single- and dual-beam instruments are used. In the latter, part of the primary radiation is fed directly into the monochromator, while the rest passes first through the flame. In this way, fluctuations and drift can be compensated insofar as they originate from the primary radiation source or the measurement electronics.

The spectrometer can be equipped for quasi-simultaneous measurement of the line and background absorption [159]. Radiation from a second, continuous source, such as a deuterium arc lamp, is rapidly switched with the radiation of the primary source and fed through the absorption volume, by a semitransparent mirror. The radiant flux of the continuous source is not decreased significantly by atomic absorption, owing to the low spectral resolution of the monochromator; however, it will be weakened by broad-band absorption of molecules, or stray radiation. By this means, nonelement-specific background absorption can
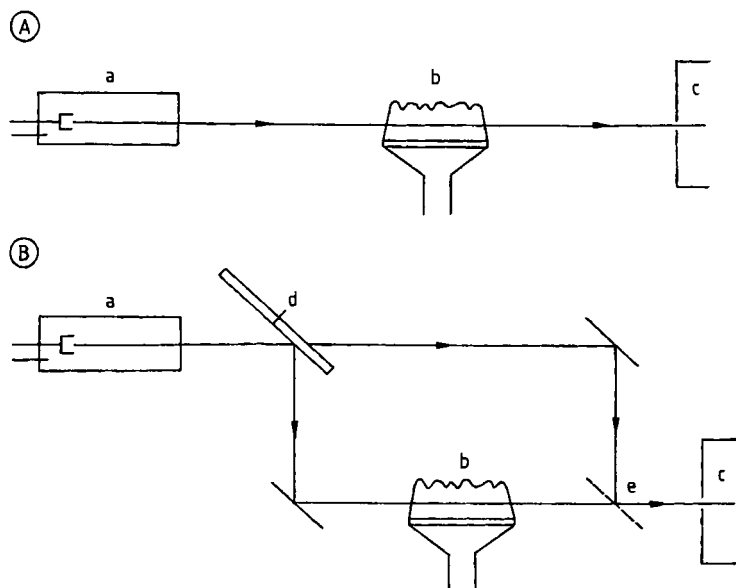
Figure 30. Flame atomic absorption spectrometer
A) Single beam; B) Dual-beam system
a) Source; b) Burner; c) Monochromator; d) Rotating mirror; e) Half-silvered mirror

be compensated. This and other approaches to background correction are discussed in detail in Section 21.5.6.

Dual-Channel Instruments. Apart from single-channel instruments, with which only measurements at a single wavelength in one channel can be performed, dual-channel instruments are also used. They contain two independent monochromators which enable simultaneous measurements at two wavelengths. They are of use for the simultaneous determination of two elements of which one element may be the analyte and the second a reference element. It is also possible to use two lines with widely different intensity to determine one element over a wide concentration range.

Multichannel spectrometers which allow the simultaneous determination of a large number of elements, as in atomic emission spectrometry have not encountered a breakthrough in AAS yet. However, over a number of years, work with high-intensity continuous sources and high-resolution echelle spectrometers for multielement AAS determinations has aroused some interest [160]. Fourier transform spectrometry and multichannel detection with photodiode arrays has opened new perspectives for the simultaneous detection of

larger numbers of spectral lines, and high-intensity sources have improved considerably.

21.5.2.2. Primary Sources

The primary radiation sources used in AAS have to fulfill several conditions:

1) They should emit the line spectrum of one or more analyte elements with line widths smaller than those of the absorption lines in the atomizer

2) They should have high spectral radiance density (radiant power per unit area, solid angle, and unit wavelength) in the center of the spectral line

3) Their optical conductance (radiant area multiplied by usable solid angle) must be high

4) The radiances of the analytical lines must be constant over a long time

These conditions are fulfilled by low-pressure discharge such as hollow cathode discharges and, for some elements, high-frequency discharges.

In a commercially available hollow cathode lamp (Fig. 31), the cathode (a) has the form of a hollow cylinder closed at one side. The lamp is sealed and contains a noble gas at a pressure of the order of 100 Pa. At discharge currents up to
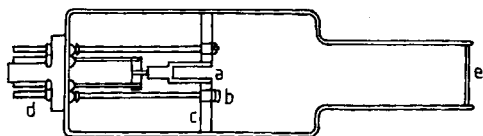
Figure 31. Hollow cathode source for atomic absorption spectrometry
a) Hollow cathode; b) Anode; c) Mica insulation; d) Current supply; e) Window (e.g., quartz)

10 mA (at ca. 500 V), a glow discharge is sustained between this cathode and a remote anode (b). The atomic vapor is produced by cathodic sputtering and excited in the negative glow contained within the cathode cavity. Lines of the discharge gas are also emitted, which may lead to interference in AAS. In most cases, high-purity argon or neon are used. For mechanical reasons, it may be necessary to manufacture the cathode mantle from a material other than that of the internal part of the hollow cathode, which is usually made of the analyte, so a further atomic spectrum may be emitted.

For a number of elements, lamps in which the hollow cathode consists of several elements may also be used. The number of elements contained in one lamp is limited because of the risk of spectral interference.

For a few elements, electrodeless discharge lamps are to be preferred over hollow cathode lamps. This applies for volatile elements such as As, Se, and Te. In the hollow cathode lamps of these relatively volatile elements self-absorption at low discharge currents may be considerable, and self-reversal may even occur. This is not the case with electrodeless discharge lamps. The latter consist of a quartz balloon in which the halide of the element is present. The analyte spectra are excited with the aid of a high-frequency (MHz) or microwave field (GHz), which may be supplied through an external antenna.

21.5.3. Flame Atomic Absorption

In flame AAS, the sample solutions are usually nebulized pneumatically in a spray chamber and the aerosol produced is transferred, together with a mixture of a combustion gas and an oxidant, into a suitable burner.

21.5.3.1. Flames and Burners

Flames may use propane, acetylene, or hydrogen as combustion gases, with air or nitrous oxide as oxidant. With these mixtures, temperatures of 2000 – 3000 K can be obtained. Temperatures of flames (in kelvin) used in AAS are given below [157]:

Air – propane	1930
Air – acetylene	2300
Air – hydrogen	2045
Nitrous oxide – acetylene	2750
Oxygen – acetylene	3100

The *air – acetylene flame* is most often used. Its temperature is high enough to obtain sufficient atomization for most elements which can be determined by AAS, but not so high that ionization interference becomes significant. The latter occur only with the alkali and alkaline earth metals. The analytical conditions can be optimized by changing the composition of the gas mixture. The sensitivity for the noble metals (Pt, Ir, Au, etc.) can be improved by increasing the oxidative properties of the flame (excess air), which also lowers interference. The use of a reducing flame is more advantageous for determination of the alkali metals.

For more than 30 elements, however, the temperature of this flame is too low. This applies for elements which have very stable oxides, such as V, Ti, Zr, Ta, and the lanthanides. For these elements, *nitrous oxide – acetylene flame*, as introduced in flame AAS by Amos and Willis in 1966 [161], is more suitable. However, this flame has a higher background emission, especially due to the radiation of the CN bands.

The *hydrogen – air flame* is of litte use. For the determination of Sn, however, its sensitivity is higher than that of the acetylene – air flame. Because of its higher transmission at short UV wavelengths the hydrogen – air flame is also used for the determination of As (193.7 nm) and Se (196.0 nm).

The *propane – air flame* has the lowest temperature and is used for the determination of easily atomized elements such as Cd, Cu, Pb, Ag, Zn, and especially the alkali metals.

In flame work, the burners have one or more parallel slits of length 5 – 10 cm, providing an absorption path for the primary radiation. In special laminar burners, the density of the analyte atoms at relatively high observation positions (up to more than 5 cm) can be kept high, which is ad-
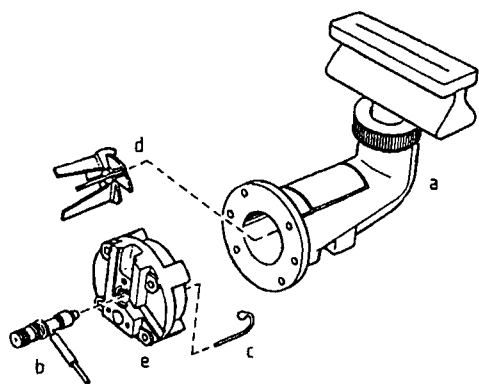
Figure 32. Burner – nebulizer assembly for flame atomic absorption spectrometry
a) Burner head with mixing chamber; b) Nebulizer; c) Impactor bead; d) Impact surfaces; e) Nebulizer socket

vantageous as measurements are then possible in these zones where more complete atomization occurs. Multislit burners may be useful to prevent deposits in solutions with high salt content. High-temperature flames require the use of special burners. Detailed information on flames is available in classical textbooks [162] – [164].

21.5.3.2. Nebulizers

The nebulizer is mounted in the spray-chamber burner assembly, as shown for the nitrous oxide or air – acetylene burner system (Fig. 32). Air or nitrous oxide are fed to the mixing chamber through the nebulizer. The combustion gas is fed directly to the burner. A self-aspirating concentric nebulizer is often used, but other types, including the Babington and cross-flow types (Section 21.4.1) are also applied. Nebulizers may be made of glass, corrosion-resistant metals such as Pt – Ir, or plastics such as Ryton. Sample solution consumption usually amounts to ca. 5 mL/min in the case of gas flows of 1 – 5 L/min.

In flame AAS, the nebulization system must be able to generate small droplets ($< 10 \mu m$), as only such droplets can be transported and completely vaporized in the flame. This is a prime condition for the realization of high sensitivity and low volatilization interference. Nebulizers for AAS generate not only small droplets, but also larger ones (up to 25 μm or more). For fragmentation of the latter, the nebulizer should be positioned in a mixing chamber provided with impact surfaces. In pneumatic nebulization, the sampling efficiency

remains restricted (a few percent) as fragmentation is limited, and all particles larger than 10 μm are led to waste. In flame AAS with pneumatic nebulization, the sample solution is usually aspirated continuously. After 5 – 7 s, a stable signal is obtained and signal fluctuations depending on concentration are of the order of 1 %.

Apart from continuous sample aspiration, flow injection and discrete sampling can be applied (see Section 21.4.1), which both deliver transient signals. In the latter, 10 – 50 μL aliquots can be injected manually or with a sample dispenser into the nebulization system [78], [79]. This approach is especially useful for preventing clogging in the case of sample solutions with high salt content, for the analysis of microsamples as required in serum analysis, or when aiming for maximum absolute power of detection in combined analytical procedures for trace analysis.

21.5.3.3. Figures of Merit

Power of Detection. The lowest elemental concentrations which can be determined by flame AAS are often given in terms of the so-called characteristic concentrations (μg/mL). For aqueous solutions of such concentration an absorption of 1 % corresponds to an extinction of 0.0044. The noise comprises contributions from the nebulizer as well as from the flame. Detector noise limitations may also occur. The latter can be minimized by operating the hollow cathode lamp at sufficiently high current.

In order to obtain maximum power of detection, the atomization efficiency should be as high as possible. Optimization of the geometry of the spray chamber and the nebulizer gas flow is required. The primary radiation should be well isolated by the monochromator and the amount of nonabsorbed radiation reaching the detector should be minimized by selection of the appropriate observation zone with the aid of a suitable illumination system.

The detection limits of flame AAS are especially low for rather volatile elements, which do not form thermally stable oxides or carbides, and have high excitation energies, such as Cd and Zn. Apart from these and some other elements such as Na and Li, the detection limits in flame AAS are higher than in ICP – AES(see Table9).

Analytical Precision. By integrating the extinction signals over 1 – 3 s, relative standard deviations down to 0.5 % may be achieved. When in-

jecting discrete samples into the nebulizer or applying flow injection analysis, slightly higher relative standard deviations are obtained. They increase rapidly on leaving the linear part of the calibration curve and applying linearization by software.

Interference. Interference in flame AAS comprises spectral, chemical, and physical interference.

Spectral Interference. Spectral interference of analyte lines with other atomic spectral lines is of minor importance compared with atomic emission work. It is unlikely that resonance lines emitted by the hollow cathode lamp coincide with an absorption line of another element present in the atomizer. However, it may be that several emission lines of the hollow cathode are within the spectral band width or that flame emission from bands or continuum occurs. Both contribute to the nonabsorbed radiation, and decrease the linear dynamic range. Nonelement-specific absorption (Section 21.5.6) is a further source of spectral interference.

Chemical Interference. Incomplete atomization of the analyte causes chemical interference, due to the fact that atomic absorption can only occur with free atoms. Reactions in the flame which lead to the formation of thermally stable species decrease the signal. This is responsible for the depression of calcium signals in serum analysis by proteins, as well as for the low sensitivities of metals which form thermally stable oxides or carbides (Al, B, V, etc.) in flame AAS. A further example of chemical interference is the suppression of the extinction of alkaline earth metals as a result of the presence of oxyanions (OX) such as aluminates or phosphates. This well-known "calcium phosphate" interference is caused by the reaction:

$$\text{MOX} \overset{\text{excess}}{\rightleftharpoons} \text{X} + \text{MO} \overset{\text{high }T}{\rightleftharpoons} \text{M} + \text{O} \overset{+\text{C}}{\rightleftharpoons} \text{M} + \text{CO}$$
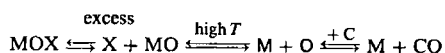
In hot flames, such as the carbon-rich flame, the equilibrium lies to the right, but with excess oxyanions it is shifted to the left, and no free M atoms are formed. This can be corrected by adding a metal R, the so-called "releasing agent", which forms yet more stable oxysalts and releases metal M again. La and Sr compounds can be used:

$$\text{MOX} + \text{RY} \overset{\text{excess}}{\rightleftharpoons} \text{ROX} + \text{MY}$$
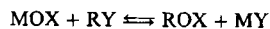
When $LaCl_3$ is added to the sample solutions, phosphate can be bound as $LaPO_4$.

For alkali metals, the free atom concentrations in the flame can decrease as a result of ionization, especially in hot flames. This leads to a decrease of extinction for alkali metals. It may also lead to false analysis results, as the ionization equilibrium for the analyte element is changed by changes in concentration of easily ionized elements. In order to suppress these effects, ionization buffers can be added. The addition of excess (low ionization potential) is most effective to suppress ionization changes of other elements, as it provides a high electron number density in the flame.

Physical Interference. Physical interference may arise from incomplete volatilization, and occurs especially in strongly reducing flames. In steel analysis, the depression of Cr and Mo signals as a result of excess Fe is well known. It can be reduced by adding NH_4Cl. Further interference is related to nebulization effects, and arises from the influence of the concentration of acids and salts on the viscosity, density, and surface tension of the analyte solutions. Changes in physical properties from one sample solution to another influence aerosol formation efficiencies and the aerosol droplet size distribution, as discussed in Section 21.4.1. However, related changes of nebulizer gas flow also influence the residence time of the particles in the flame.

21.5.4. Electrothermal Atomic Absorption

The use of furnaces as atomizers for quantitative AAS goes back to the work of L'vov, and led to the extension of AAS to very low absolute detection limits. In electrothermal AAS, graphite or metallic tube or cup furnaces are used and, with resistive heating, temperatures are achieved at which samples can be completely atomized. For volatile elements, this can be accomplished at 1000 K, whereas for more refractory elements the temperature should be up to 3000 K.

The high absolute power of detection of electrothermal AAS is due to the fact that the sample is completely atomized and vaporized, and the free atoms spend a long time in the atomizer. The signals obtained are transient, as discussed in Section 21.4.4.1.
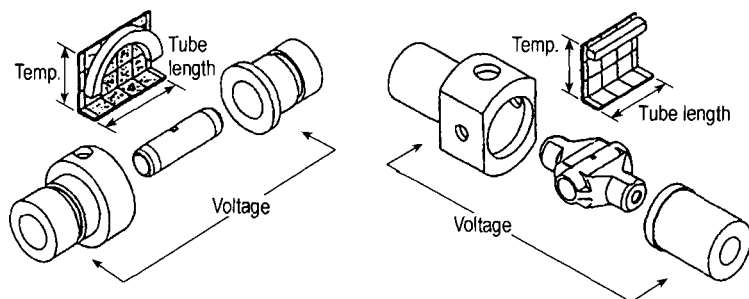
Figure 33. Temperature distribution in a longitudinally (left) and transversely (right) heated graphite furnace
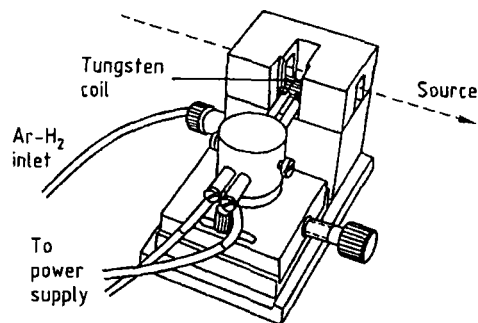
Figure 34. Tungsten filament atomizer [166]

21.5.4.1. Atomizers

Graphite furnaces, cups, and filaments are all used as atomizers in electrothermal AAS [165]. The models originally proposed by L'vov et al. [3] and by MASSMANN [4] are described in Section 21.4.4.2. In the most widely used type the optical beam is led centrally through a graphite tube, closed on both sides, with a quartz viewing port mounted in the cooled tube holders. Sample aliquots are introduced with the aid of a micropipette or a computer-controlled dispenser, through a sampling hole in the middle of the tube. The normal temperature profile of a graphite furnace leads to differences in spreading of the analyte over the graphite surface and may produce changes in volatilization from one sample to another. This effect can be avoided by using a transversely heated furnace, where the temperature is constant over the whole tube length (Fig. 33). Especially in electrothermal AAS, problems may arise from recombination of the atomized analyte with oxygen and other nonmetals, as the free atom concentration measured in AAS decreases. This may occur when the volatilized analyte enters a fairly cool plasma,

as is normally the case in a tube furnace. This can be prevented [109] by dispensing the sample on a low-mass graphite platform located in the tube furnace. The low heat content then allows very rapid heating and volatilization of the analyte. The analyte is surrounded by a furnace plasma of about the same temperature as the sample carrier, which lowers the risk of recombination. Tungsten furnaces (see Section 21.4.4.2) [112] are useful for the determination of refractory carbide-forming elements, but they are more difficult to use with real samples, and reduction of analyte oxides does not occur. BERNDT et al. [166] have shown that tungsten coils are suitable atomizers for dry solution aliquots, especially for matrix-free solutions obtained in combined analytical procedures involving separation of the analyte from the matrix elements. Owing to the relatively large coil surface area (Fig. 34), salt problems in real samples may be lower than with a metal tube furnace.

21.5.4.2. Thermochemistry

The dissociation equilibria between the analyte elements and their compounds are very important, as they determine the fraction of analyte available as free atoms for AAS. They affect both the analytical sensitivity, and the possibility of systematic errors. The thermochemical behavior of the sample is of prime importance for the height of the extinction signal as well as its form.

The acids present in the sample solution are normally removed during drying and matrix decomposition. Residues present during the absorption measurement lead to nonelement-specific absorption, especially in the case of acids with high boiling points, such as $HClO_4$, H_2SO_4, or H_3PO_4. The absorption stems from ClO, SO, SO_2, or PO molecular bands. These species may also be

produced by the dissociation of salts. Further problems may be caused by the oxides of the analytes, which result from salt dissociation. This nondissociated oxide is lost for AAS, and may give rise to nonelement-specific absorption. When salt dissociation and reduction of analyte oxides is not complete before the absorption measurement, several peaks may occur in the absorption signal. This can often be avoided by the platform technique, which produces a sharp rise in heating of the furnace and lowers the risks of analyte deposition at the cool parts of the furnace.

Thermochemistry is especially important reliable matrix destruction and removal of matrix elements without analyte loss. Thermochemical reagents such as quarternary ammonium salts $(R_4N^+Cl^-)$ mineralize organic samples at temperatures below 400 °C, and prevent losses of elements which are volatile or form volatile, perhaps organic, compounds. This may be helpful for Cd and Zn, which form volatile compounds in a number of organic matrices. Removal of NaCl, present in most biological samples, may be helpful to prevent matrix interference; this must be done at low temperature to prevent analyte loss, and can be achieved by the addition of NH_4NO_3:

$$NH_4NO_3 + NaCl \quad NH_4Cl + NaNO_3$$

Excess NH_4NO_3 dissociates at 350 °C, NH_4Cl sublimates and $NaNO_3$ decomposes below 400 °C, so NaCl is completely removed at temperatures below 400 °C. Without the addition of NH_4NO_3, this would be possible only at the volatilization temperature of NaCl (1400 °C), with inevitable analyte loss. In the graphite furnace, several elements (Ti, V, etc.) form thermally stable compounds such as carbides, producing negative errors some of the analyte is bound and does not contribute to the AAS signal. The use of pyrolytically coated graphite tubes is helpful, as diffusion of the analyte solution into the graphite decreases the risk of carbide formation. Flushing the furnace with nitrogen can be helpful. In the case of Ti, a nitride is formed which, in contrast to the carbide, can be easily dissociated. Other thermochemical means to decrease interference are known as matrix modification. Addition of Pd compounds, $Mg(NO_3)_2$, etc., has been successful for matrix-free vapor cloud formation [167]. The mechanisms involved also relate to surface effects in the furnace [168] and are an active field of research.

Development of the temperature program is vital in furnace AAS. Selection of the different temperature steps, and the use of all kinds of thermochemical effect are most important to minimize matrix influence without analyte loss.

Radiotracers are helpful for understanding and optimizing analyte volatilization in furnace AAS: element losses and their causes at all stages of the atomization process can be followed quantitatively. This has been studied in detail for elements such as As, Pb, Sb, and Sn by KRIVAN et al. [169].

21.5.4.3. Figures of Merit

The *analytical sensitivity* and *power of detection* of electrothermal AAS are orders of magnitude higher than those of flame AAS. This is due to the fact that, in the furnace, a higher concentration of atomic vapor can be maintained compared with flames. Dilution of the analyte by the solvent is avoided, as it has already evaporated before atomization and large volumes of combustion gases are not involved. For most elements, the characteristic mass, i.e., the absolute amount for which an extinction of 0.0044 (1 % absorption) is obtained, are orders of magnitude lower than in flame AAS [165].

Detection limits in flame AAS: 0.1 – 1 ng/mL (Mg, Cd, Li, Ag, Zn); 1 – 10 ng/mL (Ca, Cu, Mn, Cr, Fe, Ni, Co, Au, Ba, Tl); 10 – 100 ng/mL (Pb, Te, Bi, Al, Sb, Mo, Pt, V, Ti, Si, Se); > 100 ng/mL (Sn, As)

Detection limits in furnace AAS, (20 µL): 0.005 – 0.05 ng/mL (Zn, Cd, Mg, Ag); 0.05 – 0.5 ng/mL (Al, Cr, Mn, Co, Cu, Fe, Mo, Ba, Ca, Pb); 0.5 – 5 ng/mL (Bi, Au, Ni, Si, Te, Tl, Sb, As, Pt, Sn, V, Li, Se, Ti, Hg)

When sample aliquots of 20 – 50 µL are used, the concentration detection limits are often in the sub-ng/mL range. They are especially low for elements such as Cd, Zn, and Pb which have high volatility and high excitation energies, but also for Mg, Cu, Ag, Na, etc.

Interference in furnace AAS is much higher than in flame AAS. This applies both for physical and chemical interference.

Physical Interference. Physical interference may stem from differences in viscosity and surface tension, resulting in different wetting of the graphite tube surface with the sample solution, and changes in diffusion of the sample solution into

the graphite. The first may be suppressed by the addition of surfactants to the sample solutions, and the second can be avoided by using pyrolytically coated graphite tubes, where diffusion of the analyte into the tubes is lower.

The influences of concomitants on volatilization can be minimized by the use of platform techniques and of an isothermal furnace. Both measures are particularly helpful in direct solid sampling.

Undissociated molecules, which may be oxides formed by dissociation of ternary salts (MgO, ZnO, etc.) or radicals and molecular species from solvent residues (OH, PO, SO_2, etc.) may cause nonelement-specific absorption. Rayleigh scattering of primary radiation by nonevaporated solid particles may occur, which also leads to radiation losses. Both phenomena necessitate the application of suitable techniques for background correction for most analyses of real samples by furnace AAS (Section 21.5.6). Emission by the furnace itself may give continuous radiation, which may lead to systematic errors.

Chemical Interference. Chemical interference may be caused by losses of analyte during ashing, especially in the case of volatile elements such as As, Sb, Bi, Cd, but also for elements forming volatile compounds (e.g., halides). In this respect, removal of NaCl, which is present in most biological samples, is critical. Further element losses may occur from the formation of thermally stable compounds such as carbides or oxides which cannot be dissociated and hamper atomization of the analyte element; elements such as Ti or V are especially difficult to determine by furnace AAS. Some of these sources of systematic error can be reduced in furnace AAS by the use of thermochemical reactions.

Differences in the volatility of elements and their compounds may also be used for speciation work. It is possible to determine organolead compounds such as $Pb(CH_3)_4$ and $Pb(C_2H_5)_4$ directly by furnace AAS [170], [171] or after on-line coupling with chromatographic separation. This approach can be used for other species, but needs to be carefully worked out in each case.

Calibration in furnace AAS is usually done by standard addition, as many of the sources of interference change the slope of the calibration curve.

Recommendations to be followed to enable determinations to be as free from interferences as possible include: (i) isothermal operation; (ii) the use of a matrix (chemical) modifier; (iii) an integrated absorbance signal rather than peak height measurements; (iv) a rapid heating rate during atomization; (v) fast electronic circuits to follow the transient signal; (vi) the use of a powerful background correction system such as the Zeeman effect. Most or all of these recommendations are incorporated into virtually all analytical protocols nowadays and this, in conjunction with the transversely heated tubes, has decreased the interference effects observed considerably.

21.5.5. Special Techniques

21.5.5.1. Hydride and Cold Vapor Techniques

Determination of traces and ultra-traces of Hg, As, Se, Te, As, and Bi is often by formation of either volatile mercury vapor, or of the volatile hydrides. This gives a high sampling efficiency and power of detection. The absorption measurement is often performed in a quartz cuvette. Hg, for instance, can be transported after reduction by means of a carrier gas. For the absorption measurement no heating is required. The other elements are reduced to the volatile hydride, which is then transported with a carrier gas flow (argon or nitrogen) into a cuvette. In order to dissociate the hydrides thermally, the quartz cuvette must be heated to a temperature of $600-900\,^{\circ}C$.

In both cases, flow cells enable continuous formation of the volatile compound and subsequent separation of the reaction liquid from the gaseous compound [172]. This is very useful (Section 21.4.3) for routine work and has practically replaced earlier systems, where $NaBH_4$ solution was reacted with a given volume of acidified sample [173]. In any case, trapping of the analyte vapor released may increase the absolute power of detection. In the case of Hg, this may be achieved by passing the Hg vapor over a gold sponge and binding it as an amalgam [174]. After collection, Hg can be liberated completely by heating. In the case of the hydrides, cold trapping can be applied. Here, the hydrides can be condensed by freezing in a liquid nitrogen trap, and freed from the excess hydrogen formed during the reaction [175]. The isolated hydrides can be released by warming. This removes hydrogen and gives trapping enrichment factors well above 100.

Hot trapping can also be used [176] as the hydrides have decomposition temperatures lower than the boiling points of the elements. Trapping can thus be performed in the graphite furnace itself. It is only necessary to heat the furnace to

ca. 600–700 °C to decompose the volatile hydrides, and the elements condense in the furnace. Here they can be preenriched and volatilized at temperatures above 1200 °C during absorption measurement.

The cold vapor technique for Hg gives detectin limits < 1 ng with 50 mL of sample, and this can be further improved by trapping. The hydride technique gives detection limits below the ng/mL level for As, Se, Sb, Bi, Ge, Sn, etc., levels appropriate for analyses of drinking water. Hydride techniques, however, are prone to interference (Section 21.4.3). In AAS, interference may occur not only as a result of influences on the hydride formation reaction, but also as a result of influences of concomitants on the thermal dissociation of the hydride. Interference by other volatile hydride-forming elements may also occur [177].

21.5.5.2. Direct Solid Sampling

In both flame and furnace AAS, direct solid sampling is possible. In flames, sampling may involve a boat or cup, as introduced by DELVES [113]. New approaches such as the combustion of organic samples and the introduction of the vapors released into a flame are possible [178]. For volatile elements such as Pb, detection limits are in the μg/g range. Direct powder sampling in furnaces was introduced in atomic spectrometry by PREUSS in the 1940s, and in furnace AAS it can be easily done with a powder sampling syringe [116]. For direct powder analyses with flame and furnace AAS, however, it is preferable to work with slurries. Slurry atomization was introduced in flame AAS work by EBDON and CAVE [179] in 1982. It has been shown to be useful for flame work, and to improve powder sampling in furnace AAS. The latter has been successfully used for direct analysis of SiC powders [180]. Direct solid sampling, however, should be used very carefully. The amount of sample used is often of the order of a few milligrams, and sampling errors may occur as a result of sample inhomogeneity. Therefore, the use of furnaces in which larger powder amounts can be handled has been proposed. In direct powder analysis, calibration may be troublesome because the particle size affects analyte volatilization and transport.

In compacted samples, direct analysis by AAS is feasible. The use of cathodic sputtering combined with AAS of the atomic vapor cloud produced has been recently reviewed [181], with jet-enhanced sputtering to give high analyte number densities. The approach has been made commercially available [182]. Its feasibility for direct analyses of steels has been shown, especially for samples which are difficult to dissolve, in which refractory oxide-forming elements such as Zr need to be determined [183]. For both electrically conducting and insulating samples, laser ablation combined with AAS may be useful. AAS measurements can be performed directly in the laser plume. Measurement of nonelement-specific absorption is very important, because of the presence of particles, molecules, and radicals, and the emission of continuous radiation. The absorption measurement should be performed in the appropriate zone. When using laser ablation for direct solid sampling, the atomic vapor produced can be fed into a flame for AAS work, as described by KANTOR et al. [184].

21.5.5.3. Indirect Determinations

For the determination of elements of which the most sensitive lines are in the vacuum UV conventional AAS cannot be applied. For this, and for the determination of chemical compounds, indirect methods can be used. For instance, sulfates can be determined by precipitation as $BaSO_4$ and a determination of excess barium with AAS. A similar approach can be used for the determination of halides after precipitation with silver.

Further indirect determinations can be performed by making use of chemical interference. Calcium phosphate interference in flame AAS can be used to determine phosphate by measuring the decrease of the absorption signal for calcium. However, for real samples, indirect techniques should be used very carefully in view of possible systematic errors.

21.5.6. Background Correction

In AAS, systematic errors are often due to nonelement-specific absorption, which necessitates the use of background correction procedures. They can only be omitted in the analysis of sample solutions with low matrix concentrations by flame AAS, and are especially necessary in furnace AAS. Determination of nonelement-specific absorption can be performed in several ways.
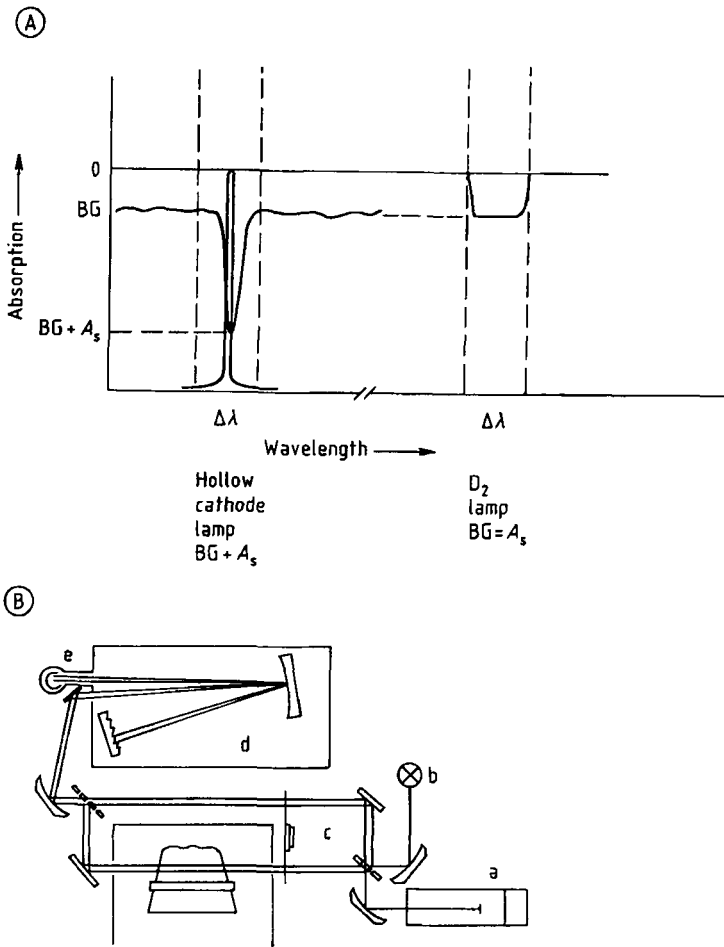
Figure 35. Background correction with the D_2 lamp technique
A) $\Delta\lambda$ = Spectral bandwidth of monochromator; B) Optical diagram of the spectrometer (Model 2380, Bodenseewerk Perkin-Elmer)
a) Hollow cathode lamp; b) D_2 lamp; c) Chopper; d) Monochromator; e) Photomultiplier
BG = Nonelement-specific background absorption; A_S = Element-specific absorption signal

21.5.6.1. Deuterium Lamp Technique

This approach was introduced by KOIRTHYO-
HANN and PICKETT in 1965 [159] and is now
provided in almost every AAS system. The total
absorption resulting from the presence of the ele-
ment and the background absorption are measured
with hollow cathode lamp radiation, but, in addi-
tion, a continuum source is used which measures
only the background absorption. This is possible
as the monochromators used in AAS have a large
spectral band width compared with the physical
width of the resonance line emitted by the hollow
cathode source, and the width of the absorption

line. Accordingly, all radiation emitted by the hol-
low cathode lies within the absorption profile of
the line, and elemental absorption does not con-
tribute significantly to the absorption of the con-
tinuous source (Fig. 35), which extends homoge-
neously over the whole spectral band with. As the
continuous source, an electrical discharge in
deuterium can be used, as it emits a rather smooth
continuum in the UV range up to 400 nm.

Alternative measurements of the absorption of
the hollow cathode lamp radiation and of the ra-
diation of the continuous source can be performed
in a rapidly alternating mode to give a quasi-si-
multaneous measurement of line and background

absorption, as required with transient signal. The latter can be realized by a rotating mirror and is rapid enough for high time resolution [185].

This type of background correction assumes that the background absorption is continuous within the spectral band width of the monochromator. This is not the case when the background absorption stems from molecular bands, which have a rotation–vibration hyperfine structure. They can arise from radicals produced by dissociation of the solvent (OH, SO_2, SO_3, N_2^+, CN, etc.) or from molecular oxides (MO). As such contributions occur especially with the complex matrices of real samples, background correction with a D_2 lamp often leads to systematic errors, which cannot be corrected by standard addition methods. Furthermore, the radiant density of the D_2 lamp over a large part of the spectrum is rather low. Accordingly, there are limits to the number of analytical lines which can be used and the number of elements which can be determined. As the spectral radiance of the D_2 lamp is lower than that of the hollow cathode, the latter must be operated at low current, which leads to detector noise limitations and poor detection limits. Finally, the two primary radiation sources are difficult to align, so they have to pass through the same zone of the atomizer, and this may lead to further systematic errors.

21.5.6.2. Zeeman Effect Technique

Zeeman AAS makes use of the splitting of atomic spectral lines by a magnetic field. When a magnetic field B (≤ 1 T) is applied, the shift in wavenumber (ΔT_m) of the so-called σ-components, where the π-components remain constant, is:

$$\Delta T_m = M g L B \qquad (151)$$

where M is the magnetic quantum number, L is the Lorentz unit ($e/(4\pi mc^2)$), and g the Landé factor, a function of the total quantum number J, the orbital momentum quantum number L, and the spin quantum number S. The intensities of the σ-components (for which $\Delta M = \pm 1$) and the π-components (for which $\Delta M = 0$) are a function of ΔJ (0, 1) and ΔM (0, ± 1) for the transitions. In the normal Zeeman effect, which occurs for singlet transitions (e.g., with alkaline earth metals and metals of group 12 such as Cd and Zn) $g = 1$, whereas in all other cases multiplets occur (anomalous Zeeman effect). For a transverse magnetic field (per-

pendicular to the observation direction) there is a π-component at the original wavelength. For this component, $\Delta M = 0$ and it is polarized parallel to the field. There are two σ-components (σ^+ and σ^-) for which $\Delta M = \pm 1$. They are polarized perpendicular to the field. For a longitudinal field (parallel to the observation direction) there is no π-component ($\Delta M = 0$) and the σ-components ($\Delta M = \pm 1$) are circularly polarized. In order to use the Zeeman effect for background correction, several approaches are possible (Fig. 36) [186]. A constant transverse field may be used, measuring the π- and the σ-components alternately with the aid of a polarizer and a rotating analyzer. Alternatively, an a.c. longitudinal field can be used, however, measuring σ-components only with the aid of a static polarizer, at zero and at maximum field strength.

In both cases, substracting one signal from the other eliminates the background absorbance. This assumes that both signals have constant intensities through the whole analytical system and that both have the same absorption coefficients for the background. The latter technique has the additional advantage that the magnetic field can be controlled since it is produced by an electromagnet. Accordingly, the splitting can be optimized with respect to the element determined and to the background structure. The magnetic field can be applied in the atomizer, which is possible both with permanent and electromagnets. In principle, the magnetic field can also be placed at the primary source, which is optimal in the case of a permanent magnet; both flames and furnaces can be used as atomizers and the exchange is easier. Moreover, a larger furnace can be used which is very useful for direct solid sampling.

Zeeman AAS has several analytical advantages. The accuracy of correction of a structured background is better than with the D_2 lamp technique. However, when the background structure stems from molecular bands, it must be remembered that molecular bands may also display a Zeeman effect. Systematic errors resulting from this fact may be larger when one line component is measured in an alternating field.

The *detection limits* in Zeeman AAS should be lower than in the case of background correction with a D_2 lamp. The system uses only one source; it can be operated at high intensity, so detector noise limitations are avoided. This advantage is most pronounced when one component is measured in an alternating field.

(A)

Pulsed hollow cathode   Rotating   Furnace/flame in   Monochromator
or electrodeless   polarizer   magnetic field   and signal
discharge lamp   processing

(B)

Magnetic discharge   Phase-squeezer for rotation   Phase
lamp (permanent   of polarization plane and   synchronization
magnet)   static analyzer

                     Furnace or   Monochromator
                     flame   and signal
                                         processing

(C)

Pulsed hollow cathode   Furnace in   Static   Monochromator
or electrodeless   a.c. magnetic   polarizer   and signal
discharge lamp   field   and   processing
                                                         filter

Figure 36. Zeeman atomic absorption
A) Rotating polarizer and permanent magnet applied to the atomizer; B) Permanent magnet around the primary source;
C) Longitudinal field of a.c. magnet applied to the atomizer [186]

Another consequence of the use of one primary source is improved stability. The analytical sensitivity in Zeeman AAS, however, is inferior to that of conventional AAS, is the linear dynamic range. The latter relates to the fact that a difference between two absorbances is involved, which may even lead to bending of the calibration curve. These effects are less pronounced when measuring one component in an alternating field. Zeeman AAS is widely used, for instance, for trace determinations in biological samples.

21.5.6.3. Smith – Hieftje Technique [187]

This technique of background correction makes use of the fact that resonant atomic spectral lines emitted by a hollow cathode lamp may display self-reversal when the lamp is operated at high discharge current. A hollow cathode lamp operates during the first part of the measurement cycle, at low current. Self-reversal does not occur and the resonant radiation is absorbed both by the analyte atoms and by background-producing species. In the second part of the measurement cycle, the current is briefly pulsed to > 500 mA, and a pronounced self-reversal occurs. The intensity at the analytical wavelength decreases, but the intensities in the side-wings remain high (mostly background absorption). By subtraction of the absorbances, the net atomic absorption signal is obtained.

Similar to Zeeman AAS, the Smith – Hieftje technique can be used for lines over the entire spectral range, and uses only one primary radiation source, so that alignment and stability are optimum. The technique is simple and much cheaper than Zeeman AAS. It does not suffer from limitations due to Zeeman splitting of molecular bands. However, as self-reversal is not complete, the technique can be used only for rather low background absorbance, sensitivity is decreased as the

self-absorption is 40% at most, and special provisions have to be made for pulsing the lamps at high currents.

21.5.6.4. Coherent Forward Scattering

The intensity of scattered radiation is particularly high when monochromatic radiation is used as primary radiation, the wavelength of which is equal to that of a resonance line of the scattering atoms. When the latter are subjected to a magnetic field, the scattered radiation becomes coherent in the direction of the primary beam and the scattering atomic vapor becomes optically active (magneto-optic effect). Depending on whether a transverse or longitudinal magnetic field is used, a Voigt or Faraday effect is observed. In coherent forward scattering, radiation from the primary source is led through the atomizer across which a magnetic field is applied. When the atomizer is placed between crossed polarizers, scattered signals for the atomic species occur against zero background. When a line source, such as a hollow cathode lamp or a laser, is used, determinations of the respective element can be performed. For a continuous source, such as a xenon lamp, and a multichannel spectrometer, simultaneous multielement determinations can be performed. The method is known as coherent forward scattering (CFS) atomic spectrometry [188] – [190].

The *intensities of the scattering signals* in both Voigt and Faraday effects can be calculated [188]. In the *Faraday effect*, two waves are polarized parallel to the magnetic field. If n^+ and n^- are the refractive indices, $n_m = (n^+ + n^-)/2$ and $\Delta n = (n^+ - n^-)/2$, the intensity $I_F(k)$ at wavenumber k is:

$$I_F(k) = I_0(k)F[\sin(k\,l\Delta n)]\exp(k\,l\,n_m) \qquad (152)$$

where $I_0(k)$ is the intensity of the primary radiation at wavenumber k, and l is the length of the atomizer. The sinusoidal term relates to the rotation of the polarization plane and the exponential term to the atomic absorption. As both n_m and Δn are functions of the density of the scattering atoms, $I_F(k)$ is proportional to the square of the density of scattering atoms N:

$$I_F(k) = I_0 N^2 l^2 \qquad (153)$$

In the *Voigt effect*, the scattered radiation has two components. One is polarized parallel to the magnetic field (normal component) and the other perpendicular (abnormal component). If n_0 and n_e are the respective refractive indices, their intensities can be calculated as for the Faraday effect.

Coherent forward scattering atomic spectrometry is a multielement method. The instrumentation required is simple and consists of the same components as a Zeeman AAS spectrometer. As the spectra contain only a few resonance lines, a spectrometer with low spectral resolution is sufficient. The detection limits depend considerably on the primary source and on the atomizer. With a xenon lamp as primary source, multielement determinations can be performed, but the power of detection is low as the spectral radiances are low compared with those of a hollow cathode lamp. By using high-intensity laser sources, the intensities of the signals and the power of detection can be considerably improved. When furnaces are used as atomizers, typical detection limits in the case of a xenon arc are: Cd 4 ng; pb 0.9; Tl 1.5; Fe 2.5; and Zn 50 ng [188]. These are considerably higher than in furnace AAS.

The sensitivity of CFS atomic spectrometry is high as the signals are proportional to the square of the atom number densities. The dynamic range is similar to atomic emission spectrometry, of the order of three decades. Information on matrix effects in real samples is still scarce. As scattering by molecules and undissociated species is expected to be low, background contributions may be low compared with AAS.

21.5.7. Fields of Application

The different methods of AAS are very powerful for analysis of solutions. The instruments are simple and easy to operate; they are in use in almost any analytical laboratory. Especially when one or only a few elements have to be determined in a large number of samples, as is the case in clinical or food analysis, AAS methods are of great importance compared with other methods of elemental analysis.

When solids need to be analyzed, samples have to be dissolved. Sample decomposition methods range from simple dissolution in aqueous solutions to treatment with strong and oxidizing acids. In all sample dissolution and pretreatment work for ASS, attention must be paid to all the problems which be set trace elemental analytical chemistry. This includes precautions for avoiding contamination from the reagents, the vessels used, and from the laboratory atmosphere ($\rightarrow$Trace Analysis). All

factors which may cause analyte loss as a result of adsorption or volatilization must be studied in detail [191].

Clinical chemistry, the metals industry, the analysis of high-purity chemicals and pure substances, environmental analysis, and the life sciences are important fields for the application of AAS. Applications in all these areas are regularly reviewed [192].

Clinical Chemistry. In clinical analysis, flame AAS is very useful for serum analysis. Here, Ca and Mg can be directly determined in serum samples after 1:50 dilution, even in micro-aliquots of $20-50$ µL [193]. In the case of Ca, La^{3+}, or Sr^{2+} is added to avoid phosphate interference. Na and K are mostly determined in the flame emission mode, which can be carried out with any flame AAS instrument. Here, the burner head is often turned to shorten the optical path, avoiding self-reversal. For the direct determination of Fe, Zn, and Cu, flame AAS can be used, but at lower sample dilution. For trace elements such as Al, Cr, Co, Mo, and V, determination with flame AAS often requires preenrichment, but in serum and other body fluids as well as in biological matrices, some elements can be directly determined with furnace AAS. This applies to toxic elements such as Ni, Cd, Pb, etc., in screening for workplace exposure.

For *metal samples*, flame and furnace AAS are important analysis methods. They find applications in the characterization of incoming material, and for product analysis. In combination with matrix removal, they are indispensable for the characterization of laboratory standards. The latter are used to calibrate direct methods such as X-ray spectrometry and spark emission spectrometry for production control. However, relevant elements such as B, the rare earths, Hf, Zr, etc., are much better determined by plasma emission spectrometry. For the determination of environmentally relevant elements in slags and ores, dissolution by fusion and subsequent AAS may be applied.

In the analysis of *high-purity substances*, matrix removal is often very important for preconcentration. For this, separation techniques such as ion exchange, liquid–liquid extraction of metal complexes with organic solvents, fractional crystallization, precipitation, coprecipitation, and electrochemical methods may be used [194].

For analytical problems in the *life sciences* and *environmental analyses*, AAS is widely used. For the analysis of wastewater, flame and furnace

AAS are complementary to plasma emission spectrometry. The latter enables determination of a large number of elements per sample, but is more expensive than flame AAS in terms of instrumentation, and less sensitive than furnace AAS. For analysis of drinking water, furnace AAS is very useful, but because of its single-element character, it is increasingly being replaced by plasma mass spectrometry, especially in large laboratories.

Atomic absorption spectrometry is a useful tool for *speciation* work. This applies for combinations of species-specific extraction procedures and subsequent determination of the respective species by AAS. Such a method has been worked out for Cr(III) and CR(VI) compounds [195]. On-line coupling of gas or liquid chromatography with AAS in increasingly used [196]. In furnace AAS, coupling with gas chromatography has been shown to be very useful for determination of organolead [170] and organotin compounds [197]. By combining the cold vapor technique with HPLC, a very sensitive determination of Hg species at the sub-nanogram level becomes possible [198]. For speciation, coupling of flow injection analysis and column chromatography with flame AAS, and direct coupling of HPLC with flame AAS (possible with high-pressure nebulization) are very powerful, as shown by the speciation of Cr(III) and Cr(VI) for wastewater analysis [96].

Manufacturers. The most important manufacturers of AAS equipment are:

Baird Atomic, USA: arc/spark AES, ICP-AES, AAS

Analytic Jena AG, Germany: AAS

Fisher Scientific, USA: AAS

GBC, Australia: TOF-ICP-MS, AAS

Hitachi, Japan: ICP-AES, AAS, ICP-MS

PerkinElmer Instruments (incorporating Princeton Applied Research, ORTEC, Signal Recovery and Berthold products), USA: AAS, ICP-OES, ICP-MS, GC, GC/MS

Shimadzu, Japan: arc/spark AES, ICP-AES, AAS

TJA Solutions (incorporating Thermo Jarrell Ash, VG Elemental and Unicam AA), USA: AAS, ICP-AES, ICP-MS, SF-ICP-MS, GD-AES, GD-MS

Thermoquest (incorporating Finnigan), USA: SF-ICP-MS, AAS

Varian Associates, Australia: ICP-AES, ICP-MS, AAS

21.6. Atomic Emission Spectrometry

21.6.1. Principles

Atomic emission spectrometry (AES) goes back to BUNSEN and KIRCHHOFF, who reported in 1860 investigations of the alkali and alkali-neearth metals with the aid of their spectroscope [1], which is still the heart of any dispersive spectrometer. The new elements cesium, rubidium, thorium, and indium were discovered from their atomic emission spectra. By the beginning of the twentieth century, HARTLEY, JANSSEN, and DE GRAMONT further developed spectrochemical analysis, and flames, arcs, and sparks became powerful analytical tools. To overcome the difficulties arising from unstable excitation conditions, GERLACH in 1925 introduced the principle of the internal standard [199], where the ratio of intensities of the analytes to that of the matrix or of an added element is used. This was the breakthrough in atomic spectroscopy for elemental analysis. In AES, a reproducible and representative amount of the sample is introduced into a radiation source. Its role in AES is similar to the dissolution procedures of wet chemical analysis; continuous endeavors are made to improve existing radiation sources, and to develop new sources, to perform analyses with enhanced figures of merit. The role of the radiation source in AES does not end at the atomization stage (as in AAS), but includes excitation of states from which emission or even ionization occur.

Qualitative analyses can be performed with AES; the unambiguous detection and identification of a single, interference-free atomic spectral line of an element is sufficient to confirm its presence. The most intense line under a given set of working conditions is known as the most sensitive line. Elemental lines for different elements are situated in widely different spectral regions and may differ from one radiation source to another, as a result of the excitation and ionization processes. Despite the fact that spectral lines are very narrow (a few picometers), line coincidence can occur. High-resolution spectrometers are necessary for the qualitative evaluation of spectra, and more than one line should be used to decide on the presence of an element. Qualitative emission spectral analysis is easy when the spectra are recorded photographically or with a scanning monochromator. For the first case, atlases are available, in which the spectra are reproduced with the most sensitive lines indicated. Spectral line tables are also very useful. They are available for arc and spark sources [200]–[202], and in much less complete form for newer radiation sources such as glow discharges [203] and inductively coupled plasmas [204]–[206].

Quantitative AES. In quantitative AES, the intensity of an elemental atomic or ionic line is used as analytical signal. As AES is a relative method, calibration has to be performed.

Calibration Function. The determination of the calibration function is important; it relates the intensity of a spectral line I to the concentration c of an element in the sample. SCHEIBE [207] and LOMAKIN [208], propose the following relation between absolute intensities and elemental concentrations:

$$I = a\,c^b \tag{154}$$

where a and b are constants. Absolute intensities are practical only in flame work or in plasma spectrometry, where very stable radiation sources are used. Normally, the intensity ratio of an analyte line to a reference signal is used, as proposed by GERLACH. This leads to calibration functions of the form:

$$I/I_R = a\,c^b \tag{155}$$

The inverse calibration function is the analytical evaluation function:

$$c = c_R(I/I_R)^\eta \tag{156}$$

or:

$$\log c = \log c_R + \eta \log (I/I_R) \tag{157}$$

In trace analyses, the slope of the analytical evaluation curve (in logarithmic form) is usually 1; at higher concentrations η may be > 1 as a result of self-reversal. In trace analysis by AES, the intensity of the spectral background I_U can be used as reference signal and c_R should be replaced by c_U, the concentration at which the line and background intensities are equal (background equivalent concentration or BEC value):

$$c_U = [(I_X/I_U)/c]^{-1} \tag{158}$$

The ratio (I_X/I_U) and c_U for a given element and line depend on the radiation source and its

working conditions, and on the spectral apparatus [209], [210]:

$$c_U = \frac{(dB_U/d\lambda)\,\Delta\lambda_L}{B_0}\,\frac{s_{eff}R_L}{s_0R_0} \cdot [(\pi/(\ln 16))]^{1/2}\left\{\left[(s_a/s_{eff})^4 + 11\right]\right\}^{1/4}$$

$$c_U = c_{U,\infty}A_1A_2 \tag{159}$$

where $c_{U,\infty}$ describes the influence of the radiation source, $(dB_U/d\lambda)$ reflects the spectral distribution of the background intensity, B_0 is the radiant density for an analytical line at $c = 1$, and $\Delta\lambda_L$ is the physical width of the analysis line. The term A_1 describes the influence of the spectral apparatus, and:

$$s_{eff} = \left(s_e^2 + s_0^2 + s_L^2 + s_z^2\right)^{1/2} \tag{160}$$

where s_L is the slit width corresponding to the physical width of the spectral line $\Delta\lambda_L$, and $R_L = \lambda/\Delta\lambda_L$, the resolution required to resolve the line; s_z is the slit width corresponding to the optical aberrations in the spectral apparatus:

$$s_z = \Delta\lambda_z\,d\tau/d\lambda \tag{161}$$

where $\Delta\lambda_z$ can be determined by measuring the practical resolution of the monochromator from deconvolution of the two components of the Hg 313.1 nm doublet and subtraction of the contributions of the diffraction slit width $s_0 = \lambda f/W$ and the entrance slit width s_e. The contribution of the natural width of the Hg lines can be neglected, as it is very low in the case of a hollow cathode lamp. In monochromators, s_z predomates, but at very high resolution it becomes less significant and $A_1 \longrightarrow 1$.

The factor A_2 describes the influences of the profile of the analysis line and the effective measurement slit. For photoelectric measurements, the latter is the exit slit width s_a of the spectrometer. This contribution is relevant only when $s_a/s_{eff} < 2$, which is practically the case only for photographic measurements where the line profile is scanned with a very narrow densitometer slit. For photoelectric measurements, $A_2 = s_a/s_{eff} >> 1$, as the exit slit width must be larger than the effective line width because of the thermal and mechanical stability of the system. The detection limit is a relevant figure of merit for an analytical method (Section 21.3.1). It can be written for AES:

$$c_L = c_U(I_X/I_U) \tag{162}$$

where I_X/I_U is the smallest line-to-background ratio which can be measured. When the definition is based on a 99.86% probability, it is three times the relative standard deviation of the background intensities.

For photographic or diode array registration, the intensities of the line and the spectral background are recorded simultaneously, and:

$$\sigma^2 = \left[\sigma_r^2\,(I_X + I_U) + \sigma_r^2\,(I_U)\right]^{1/2} \tag{163}$$

As the intensities of the line and the spectral background are almost equal near the detection limit, their standard deviations are equal and the relevant relative standard deviation reduces to:

$$\sigma^2 = \sigma_r(I_U)\,3\sqrt{2} \tag{164}$$

Its value can be measured directly, but in photographic measurements it can be calculated from the standard deviation of the blackening with the aid of Equation (115). The latter shows that for trace analyses, emulsions with a high γ must be used. In photoelectric measurements, σ is the relative standard deviation of the background intensity, measured at a wavelength adjacent to the spectral line or in a blank sample at the wavelength of the analytical line. It may comprise several contributions:

$$\sigma^2(I_U) = \sigma_P^2 + \sigma_D^2 + \sigma_A^2 + \sigma_V^2 \tag{165}$$

where σ_P is the photon noise of the source or the noise of the photoelectrons and $\sigma_P \approx \sqrt{n}$, where n is the number of photoelectrons; σ_D is the dark current noise of the photomultiplier and $\sigma_D \approx I_D$, where I_D is the dark current; σ_A is the flicker noise of the radiation source and $\sigma_A \approx \varphi$, where φ is the radiant density; and σ_V is the amplifier noise.

In most cases, σ_V can be neglected. For very constant sources, σ_A is small. Owing to the proportionality between σ_D and I_D it is advisable to select a photomultiplier with a low dark current, to avoid detector noise limitations. When these points are taken care of, the photon noise limits the power of detection.

When signals include a blank contribution, the latter also has fluctuations, which limit the power of detection. If I_X is the lowest measurable signal without blank contribution, I'_X the lowest analytical signal including a considerable blank value, I_U the background signal, and I_{Bl} the blank signal:

$$(I_X' + I_{Bl} + I_U) - (I_{Bl} + I_U) =$$

$$3 \sqrt{[\sigma^2(I_X' + I_{Bl} + I_U) + \sigma^2(I_{Bl} + I_U)]}$$

and

$$(I_X + I_U) - I_U = 3 \sqrt{[\sigma^2(I_X + I_U) + \sigma^2(I_U)]} \qquad (166)$$

This leads to:

$$(I_X'/I_U) = (I_X/I_U)(1 + I_{Bl}/I_U)[\sigma_r(I_{Bl} + I_U)/\sigma_r(I_U)] \qquad (167)$$

and the detection limit becomes [211]:

$$c_L' = c_U\{(I_X/I_U)(1 + I_{Bl}/I_U)[\sigma_r(I_{Bl} + I_U)/\sigma_r(I_U)]\}$$

or:

$$c_L[\sigma_r(I_{Bl} + I_U)/\sigma_r(I_U)](1 + I_{Bl}/I_U) \qquad (168)$$

To obtain maximum power of detection in AES, the standard deviations of the background, the blank contributions, and the c_U values must be minimized. The latter are a function of the radiation source, the elements, and the lines (reflected in $c_{U,\infty}$), and of the spectrometer (through A_1 and A_2). In order to keep A_1 and A_2 as close to unity as possible, it is necessary to use a spectrometer with high resolving power (high R_0), a narrow entrance slit ($s_e = 1.5 s_0$), and to take $s_a = s_{eff}$. On the other hand, the spectral background intensities must still be measurable. Indeed, the background intensities obtained are proportional to the entrance slit width, and detector noise limitations can arise when slit widths become too narrow. Also thermal and mechanical stabilities become limiting at narrow slit widths.

Apart from high power of detection, maximum analytical accuracy is very important. This relates to the freedom from interference. Whereas interference from influences of the sample constituents on sample introduction or volatilization and excitation in the radiation source differ widely from one source to another, most sources emit complex spectra and the risks of spectral interference in AES are much more severe than in absorption or fluorescence. Therefore, it is advisable to use high-resolution spectrometers. This is especially the case when trace determinations are performed in matrices emitting complex spectra. Knowledge of the atomic spectra is also very important so as to be able to select interference-free analysis lines for a given element in a well-defined matrix at a certain concentration level. This requirement is also much more stringent than in absorption or fluorescence.

21.6.2. Spectrometers

As all elements present in the radiation source emit their spectra at the same time, AES is by definition a multielement method. As well as simultaneous determinations, so-called sequential analyses are possible, provided the analytical signals are constant. Sequential and simultaneous multielement spectrometers have their own advantages and limitations.

Sequential spectrometers usually include Czerny–Turner or Ebert monochromators with which lines can be rapidly selected and measured one after another. Owing to the complex emission spectra, focal lengths are often up to 1 m and a grating with width up to 100 mm and spacing of 1/2400 mm is used. The instrument requires very accurate wavelength presetting, which is difficult for random access. The half-width of a spectral line corresponds to a grating angle of 0.0001°. Therefore, it is possible to work in a scanning mode and integrate stepwise at different locations across the line profile; alternatively, optical encoders and direct grating drives may be used. Systems employing fixed optics with a multislit exit mask, a movable photomultiplier, and fine wavelength adjustment by computer-controlled displacement of the entrance slit are also used. Sequential spectrometers are flexible because they can measure any elemental line and any background wavelength within the spectral range (usually 200–600 nm); however, they are slow when many elements need to be determined. They are especially of use in the case of stable radiation sources such as plasmas at atmospheric pressure or glow discharges operated at steady-state sputtering conditions.

Simultaneous Spectrometers. Modern simultaneous spectrometers use photoelectric radiation detection rather than the photographic plates formerly used for qualitative and semi-quantitative analyses. Simultaneous photoelectric spectrometers mostly have a Paschen–Runge optical mounting. They have many exit slits and are especially suitable for rapid determinations of many elements in a constant, well-characterized matrix; the achievable precision is high. However, the

analysis program is fixed and simultaneous spectrometers are not suitable for the analysis of random samples. Owing to stability requirements, larger exit slits are often used to overcome thermal drift, but spectral resolution is reduced. Especially for trace analysis, background correction may be required. This can be achieved by computer-controlled displacement of the entrance slit or by rotation of a quartz refractor plate behind the entrance slit.

Echelle spectrometers [36] are also often used. By a combination of an order-sorter and an echelle grating, in either parallel or crossed-dispersion mode, high practical resolution (up to 300 000) can be realized with an instrument of rather low focal length (down to 0.5 m). Therefore, stability and luminosity are high. By using an exit slit mask with a high number of preadjusted slits or a solid-state, two-dimensional array detector highly flexible and rapid multielement determinations are possible.

As the most sensitive lines of a number of elements such as S, P, C, As, are in the vacuum, it is desirable to work in this spectral range, which necessitates the elimination of oxygen from the entire optical path. This can be achieved by continuous evacuation of the spectrometer, purging with nitrogen or argon, or by the use of sealed gas-filled spectrometers. It is often helpful to measure low wavelengths in the second order, to increase spectral resolution and wavelength coverage (into the region where the alkali metals have their most sensitive lines). Solar blind photomultipliers can be used to filter out long-wavelength radiation of the first order.

21.6.3. Flame Emission

The oldest spectroscopic radiation sources operate at low temperature (Section 21.5.3.1), but have good spatial and temporal stability. They readily take up wet aerosols produced by pneumatic nebulization. Flame atomic emission spectrometry [162] is still a most sensitive technique for determination of the alkali metals, e.g., for serum analysis. With the aid of hot flames such as the nitrous oxide – acetylene flame, a number of elements can be effectively excited, but cannot be determined at low concentration. Interference arising from the formation of stable compounds is high.

21.6.4. Arcs and Sparks

Arc and spark AES have been widely used for analysis of solids and remain important methods for routine analysis.

Electrical arc sources for AES use currents between 5 and 30 A and burning voltages between the electrodes, for an arc gap of a few millimeters, of 20 – 60 V. Direct current arcs can be ignited by bringing the electrodes together, which causes ohmic heating. Under these conditions, d.c. arcs have a normal (decreasing) current – voltage characteristic owing to the high material ablation by thermal evaporation. A d.c. arc displays the cathode fall near the cathode (cathode layer) the positive column, and the anode fall area. The cathode glow delivers a high number of electrons carrying the greater part of the current. This electron current heats the anode to its sublimation or boiling temperature (up to 4200 K in the case of graphite) and the cathode reaches temperatures of ca. 3500 K as a result of atom and ion bombardment. An arc impinges on a limited number of small burning areas. The plasma temperature in the carbon arc is of the order of 6000 K and it can be assumed to be in local thermal equilibrium. Arcs are mostly used for survey trace analysis requiring maximum power of detection, but they may be hampered by poor precision (relative standard deviation $\geq 30\%$).

Direct Current Arc Spectrometry. The sample is presented to an electrode with suitable geometry and volatilizes as a result of the high electrode temperature. Cup-shaped electrodes made of graphite are often used. In the case of refractory matrices, anodic evaporation may be applied and the positive column or, for very sensitive methods, the cathode layer, where the volatilized sample ions are enriched (cathode layer effect) is selected as analytical zone.

Solid samples must be powdered and sample amounts up to 10 mg (powder or chips) are brought to the electrode after mixing with graphite powder and pelleting. Thermochemical reagents such as NH_4Cl may be added, to create similar volatilization conditions for all samples. Spectrochemical buffers such as CsCl may be added for controlling the arc temperature or the electron pressure. In most cases internal standards are used. The plasma is operated in noble gas atmospheres or in O_2 or CO_2 to create favorable dissociation and excitation conditions or to avoid intense bands

such as those of CN. Procedures for the analysis of oxide powders by d.c. arc analysis are well known [212]; d.c. arcs are also of use for the analysis of metals, as is still done for copper with the so-called globule arc. Here distillation of volatile elements from the less volatile matrix is applied.

Stabilized d.c. arcs may be obtained by sheathing gas flows or by wall stabilization [213]. Magnetic fields can be used to produce rotating arcs, which have higher precision.

Dry solution analysis can be performed with metal electrodes (e.g., copper) or with pretreated graphite electrodes.

Direct current arc spectrography is still a most powerful method for trace determinations in solids, even with difficult matrices such as U_3O_8. Detection limits for many elements are in the sub-µg/g range [214]. It is still in use for survey analysis.

Alternating Current Arc Spectrometry. Alternating current arcs have alternating polarity at the electrodes. Arc reignition is often by a high-frequency discharge applied at the a.c. discharge gap. Thermal effects are lower and the burning spot changes more frequently, so reproducibility is better than in d.c. arcs. Relative standard deviations of the order of 5 – 10 % may be obtained.

Spark Emission Spectrometry. Spark emission spectrometry is a standard method for direct analysis of metallic samples and is of great use for production as well as for product control in the metals industry [215], [216].

Sparks are short capacitor discharges between at least two electrodes. The spark current is delivered by a capacitor charged to voltage U_C by an external circuit (Fig. 37 A). When the voltage is delivered directly by an a.c. current, an a.c. spark is produced. The voltage can also be delivered after transforming to a high voltage and rectifying, to produce d.c. spark discharges. In both cases, the voltage decreases within a very short time from U_C to U_B, the burning voltage. During the formation of a spark, a spark channel is first built up and a vapor cloud is produced at the cathode. This plasma may have a temperature of 10 000 K or more, whereas in the spark channel even 40 000 K may be reached. This plasma is in local thermal equilibrium, but its lifetime is short and the afterglow is soon reached. Here the temperature and the background emission are low, but owing to the longer lifetime of atomic and ionic

excited states, line emission may still be considerable.

In a discharge cirucit with capacitance C, inductance L, and resistance R, the maximum discharge current is:

$$I_0 = (U_0 - U_B)\sqrt{(C/L)} \approx U_0\sqrt{(C/L)} \qquad (169)$$

The spark frequency within the spark train is:

$$f = 1/(2\pi)\sqrt{[1/LC - R^2/(4L^2)]} \qquad (170)$$

or, when R is small:

$$f = 1/(2\pi)\sqrt{(1/LC)} \qquad (171)$$

The capacitor may store energy:

$$Q = 1/2U_C^2 C \qquad (172)$$

of which 20 – 60 % is dissipated in the spark and the rest in the so-called auxiliary spark gaps in series with the analytical spark, or as heat. Depending on R, L, and C the spark has different properties; e.g., when L increases the spark becomes harder. As shown in Figure 37 B, there may be an oscillating (a), critically damped (b), or an overcritically damped spark (c).

High-voltage sparks with $10 < U_C < 20$ kV were frequently used in the past. Their time cycle was regulated mechanically or electrically. Medium-voltage sparks ($0.5 < U_C < 1.5$ kV) are now used for metal analysis. They need a high-frequency discharge for ignition and are mostly unipolar discharges in argon. Accordingly, band emission is low, the most sensitive lines of elements such as C, P, S, and B can be excited, and the vacuum UV radiation is not absorbed. Often, a high energy is used to prespark, by which the sample melts locally, is freed from structural effects, and becomes homogeneous. Working frequencies are often up to 500 Hz, so high energies can be dissipated and high material ablation occurs.

Analytical Features. Spark emission spectrometry enables rapid and simultaneous determination of many elements in metals, including C, S, B, P, and even gases such as N and O. Therefore, spark emission AES is complementary to X-ray spectrometry for metallurgical analyses [217]. Analysis times including presparking (5 s) may be down to 10 – 20 s and the method is of great use for production control. Special sampling devices are used for liquid metals, e.g., from blast
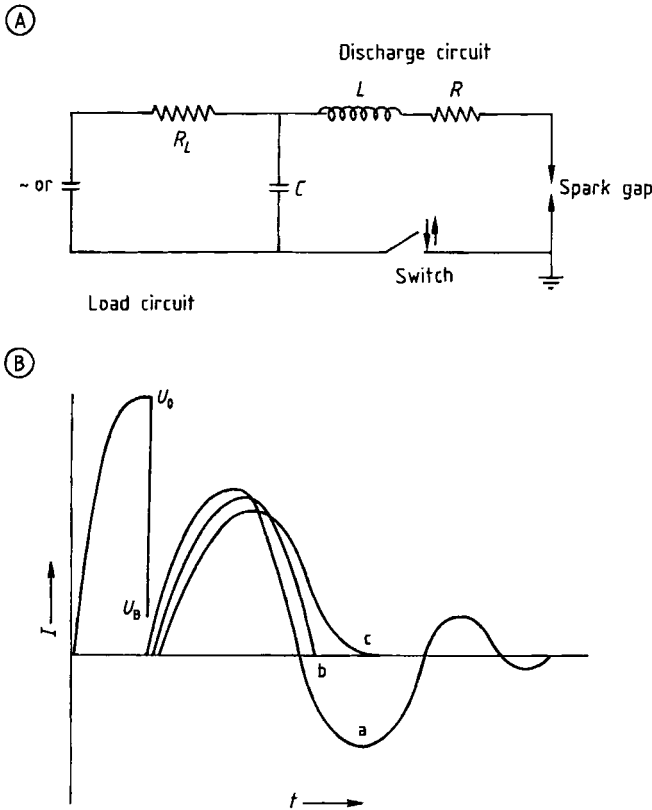
Figure 37. Circuitry in spark generator for optical emission spectrometry (A) and types of spark discharge (B)
a) Oscillating spark ($R < 2\sqrt{L/C}$); b) Critically damped spark ($R = 2\sqrt{L/C}$); c) Overcritically damped spark ($R > 2\sqrt{L/C}$)

furnaces or convertors, and special care must be taken that samples are homogeneous and do not include cavities. The detection limits obtained in metal analysis differ from one matrix to another, but for many elements they are in the μg/g range (Table 4). The power of detection of spark analysis is lower than that of d.c. arc analysis. Analytical precision is high. By using a matrix line as internal standard, relative standard deviations at the 1 % level and even lower are obtained.

Matrix interference is high and relates to the matrix composition and the metallurgical structure of the sample. Therefore, a wide range of standard samples and matrix correction procedures are used (Section 21.5.2.1). Better selection of the analytical zone of the spark is now possible with the aid of optical fibers and may decrease matrix interference.

Acquisition of the spectral information from each single spark is now possible with advanced measurement systems, using integration in a small capacitor and rapid computer-controlled readout combined with storage in a multichannel analyzer. With the aid of statistical analysis, sparks with outlier signals can be rejected and the precision improves accordingly. Pulse differential height analysis enables discrimination between dissolved metals and excreted oxides for a number of elements such as Al and Ti. The first give low signals and the latter high intensities as they stem from locally high concentrations [218]. By measurements in the afterglow of single sparks, or by the application of cross-excitation in spark AES [219], there is still room for improvement of the power of detection.

Spark AES has also been used for oil analysis. The oil is sampled by a rotating carbon electrode. Electrically insulating powders can be analyzed after mixing with graphite powder and pelleting.

Table 4. Detection limits (μg/g) for steels in spark and glow discharge AES and glow discharge MS

	Spark AES [a]	GD – AES [b]	GD – MS [c]
Al	0.5	0.1	
B	1	0.3	
Cr	3	0.05	
Cu	0.5	0.3	
Mg	2	0.9	
Mn	3	0.2	
Mo	1	0.8	
Nb	2	0.6	
Ni	3	0.1	
Pb	2		0.004
Si	3	0.4	
Ti	1	0.6	
V	1	1	
Zr	2	1.5	

[a] Spark discharge in argon with polychromator [215].
[b] Grimm-type glow discharge with 1 m Czerny – Turner monochromator [290].
[c] Grimm-type glow discharge and quadrupole mass spectrometer [334].

21.6.5. Plasma Sources

In order to overcome the disadvantages related to the low temperature of chemical flames, but to have sources with similarly good temporal stability and versatility with respect to analyte introduction, efforts were directed toward electrically generated plasmas. At atmospheric pressure, these discharges have a temperature of at least 5000 K and, provided their geometry is properly optimized, they allow efficient introduction of analyte aerosols. In most cases, they are operated in a noble gas to avoid chemical reactions between analyte and working gas.

Among these plasma sources, d.c. plasmajets, inductively coupled plasmas (ICP), and microwave discharges have become important sources for AES, mainly for the analysis of solutions [220], [221] .

21.6.5.1. Direct Current Plasmas

Direct current plasmajets (DCPs) were first described as useful devices for solution analysis by KOROLEW and VAINSHTEIN [222] and by MARGOSHES and SCRIBNER [223]. Sample liquids were converted to aerosols by pneumatic nebulization, and arc plasmas operated in argon at temperatures of ca. 5000 K.

Types of Plasmajets. It is customary to distinguish between so-called current-carrying and cur-

rent-free plasma jets. With *current-carrying plasmajets*, cooling as a result of changes in aerosol introduction has little influence. The current decrease resulting from a lower number density of current carriers can be compensated by an increase of the burning voltage, and this stabilizes the energy supply. The discharge, however, is very sensitive to changes in the electron number density, caused by varying concentrations of easily ionized elements. When the latter increase, the burning voltage energy supply and break down.

In *current-free* or *transferred plasmas*, the observation zone is situated outside the current-carrying zone. This can be achieved by the use of a supplementary gas flow directed perpendicular to the direction of the arc current, and observation of the tail flame. In this observation zone, there is no current. This type of plasma reacts strongly to cooling as no power can be delivered to compensate for the temperature drop. Therefore, it is rather insensitive to the addition of easily ionized elements. They do not cause a temperature drop, but shift the ionization equilibrium and give rise to so-called ambipolar diffusion. When the ionization in the observation zone increases, the pressure (proportional to the total particle number density — ions and electrons) increases, and ions as well as electrons move toward the outer zones of the plasma. Owing to the higher mobility of the electrons, a field is built up which amplifies the force on the positive ions directed to the outer plasma zones and thus increases the plasma volume. The d.c. plasmajet described by MARGOSHES and SCRIBNER [223] is a current-carrying plasma, whereas the plasma described by KRANZ [224] is a transferred plasma.

Three-Electrode Plasmajet. The three-electrode d.c. plasmajet developed by Spectrometrics (Fig. 38) [225] uses two graphite anodes protected by sheathing gas flows, each having a separate supply (2×10 A, 60 V). There is a common shielded cathode made of tungsten. The sample solution is usually nebulized by a cross-flow nebulizer, and the aerosol is aspirated by the gas flows around the anodes. The observation zone is small (0.5×0.5 mm^2), but very intense and located just below the point where the two positive columns meet. Optimally, this plasma is combined with a high-resolution echelle monochromator with crossed dispersion, as the latter has an entrance slit height ca. 0.11 mm. It has the advantages of a current-free plasma, as the observation zone is just outside the current-carrying part of the
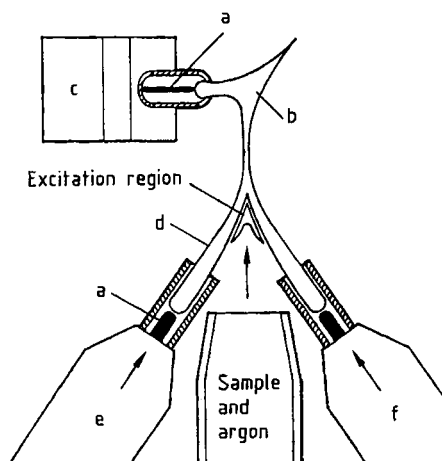
Figure 38. Three-electrode plasmajet [225]
a) Electrode; b) Plume; c) Cathode block; d) Plasma column (background) continuum; e) Argon anode block front; f) Argon anode block back

discharge, but still close enough to be compensated for strong cooling. The plasma temperature is of the order of 6000 K, $n_e = 10^{14}$ cm^3 [226], and the plasma is near local thermal equilibrium.

Detection limits for most elements are of the order of 5 – 100 ng/mL. For elements with very sensitive atomic lines such as As, B, and P, the detection limits are slightly lower than in ICP–AES [227]. The high power of detection also relates to the high resolution of the echelle spectrometer. Different concentrations of alkali metals, however, cause higher matrix effects than in ICP – AES and even may necessitate the use of spectrochemical buffers. The analytical precision achievable is high and relative standard deviations below 1 % can be reached. The system can cope with high salt content (>100 g/L), and has found considerable use in water analysis [228]. Its multielement capacity is limited, as the optimal excitation zones for most elements differ slightly; as a result of the small observation zones, considerable losses in power of detection occur. Owing to the high resolution of the echelle spectrometer, DCP–AES is very useful for the analysis of materials with complex emission spectra, such as high-alloy steels, tungsten, molybdenum, and other refractory metals.

21.6.5.2. Inductively Coupled Plasmas

Inductively coupled plasmas in their present form date back to the work of REED [229], who used these sources for crystal growth in the 1960s.

They were introduced in spectrochemical analysis by GREENFIELD et al. in 1964 [230] and WENDT and FASSEL [231], who used them in their present form as sources for AES.

In inductively coupled plasmas, the energy of a high-frequency generator (6 – 100 MHz) is transferred to a gas flow at atmospheric pressure (usually argon) in a quartz tube system, with the aid of a working coil (Fig. 39). The electrons take up energy and collide with atoms to form a plasma with a temperature of up to 6000 K. At a suitable gas flow, torch geometry, and frequency, a ring-shaped toroidal plasma occurs, where the sample aerosol passes centrally through the hot plasma. The burner consists of three concentric quartz tubes. The aerosol is fed, with its carrier-gas, through the central tube. Between the outer and intermediate tubes, a gas flow is introduced tangentially. It takes up the high-frequency energy, and prevents the torch from melting. The Fassel torch has a diameter of ca. 18 mm and can be operated at 0.6 – 2 kW with 10 – 20 L/min of argon. In so-called minitorches, gas consumption is ca. 6 L/min, by using a special gas inlet or by further reducing the torch dimensions [232]. GREENFIELD used a larger torch (outer diameter up to 25 mm), in which an ICP can be operated at higher power and where argon, oxygen, or air can be used as outer gas flows (up to 40 L/min) with an intermediate argon flow (ca. 8 L/min). With the Fassel torch, an intermediate gas flow (ca. 1 L/min) is useful when analyzing solutions of high salt content or organic solutions, to prevent salt or carbon deposition.

The ICP is near local thermal equilibrium. Excitation temperatures (from atomic line intensity ratios) are about 6000 K [233] and rotation temperatures (from the rotation lines in the OH bands) are 4000 – 6000 K [234], [235]. From the broadening of the H_β-line, an electron number density of 10^{16} cm^{-3} is obtained, whereas from the intensity ratio of ionic and atomic lines of the same element the electron number density is 10^{14} cm^{-3}. Measured line intensity ratios of ionic to atomic lines were a factor of 100 higher than those calculated for a temperature of 6000 K and an electron number density of 10^{16} cm^{-3}. This indicates the existence of overionization. The excitation processes taking place include:

1) Excitation through electrons or processes involving radiation:

$$Ar + e \longrightarrow Ar^*, \quad Ar^m \quad (\text{metastable levels}$$
$$E \approx 11.7 \text{ eV})$$
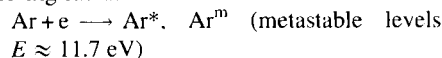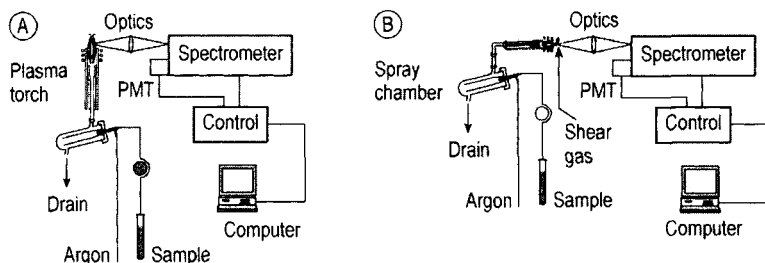
Figure 39. ICP-AES with (a) side and (b) axial viewing of the plasma

and

Ar^+ $(E_i = 15.7$ eV$)$

$Ar + hv \longrightarrow Ar^+$ (radiation trapping)

$M + e \longrightarrow M^*, M^{**} + e$ (electron impact)

$M^+ + e \longrightarrow M + hv$ (radiative recombination)
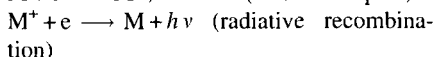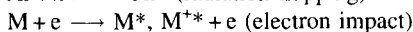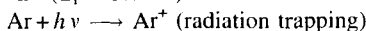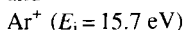
2) Excitation through argon species:

$Ar + M \longrightarrow Ar + M^*$

$Ar^m + M \longrightarrow Ar + M^+$ (Penning effect)

$Ar^+ + M \longrightarrow Ar + M^{+*}$ (charge transfer)
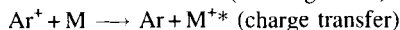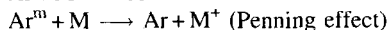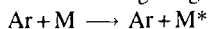
Accordingly, an overpopulation of the argon metastable levels would explain both the overionization as well as the high electron number density in the ICP. Indeed, it may be that argon metastables act as ionizers, but at the same time are easily ionized [236]. This could explain the rather low interference caused by easily ionized elements, and the fact that ionic lines are excited very efficiently, despite the fact that their standard temperatures are much higher than the plasma temperatures. However, discrepancies are not encountered to such a great extent when the temperatures from the Saha equation are used in the calculations [237]. Nevertheless, a number of processes may be considered to predominate in well-defined zones of the plasma, as indicated by spatially resolved measurements of various plasma parameters [238].

Instrumentation (Fig. 39). Inductively coupled plasma generators are $R - L - C$ circuits producing a high-frequency current at high power. As endstage a triode or transistor is used. The high-frequency energy can be led to the remote plasma stand by coaxial cable. This requires the use of a matching unit to tune the plasma with the generator before and after ignition. In other systems, the working coil may be an integral part of the oscillator. Generators with a fixed frequency and generators where the frequency may change with load

are used. In the first case, a change of load changes the ratio of the forward power and the power reflected to the generator, and requires retuning; in the second case, a small frequency shift occurs. Both types of generator may perform equally, provided the power supply to the plasma, usually ca. 50% of the energy dissipated in the generator, remains constant. Both the power stability and constant gas flows are essential to obtain maximum analytical precision.

In plasma spectrometry, the sample must be introduced into the plasma as a fine aerosol, usually by pneumatic nebulization. The aerosol should enter the hot plasma zones with low injection velocity for residence times in the millisecond range, required for efficient atomization and excitation. Therefore, the carrier gas flow should not exceed 1 – 2 L/min. With advanced pneumatic nebulizers for ICP spectrometry, aerosols with a mean droplet diameter ca. 5 µm can be obtained with an efficiency of a few percent [90], [91] for a gas flow of 1 – 2 L/min and a sample uptake of 1 – 2 mL/min. The concentric glass Meinhard nebulizer is self-aspirating, the cross-flow nebulizer may have capillaries made of different materials and is pump-fed, but less sensitive to clogging in the case of high-salt solutions. The Babington nebulizer can be used for slurry nebulization [179]. It has been used in ICP–AES for many types of powder analyses, including ceramic powders [125]. The fritted disk nebulizer has not achieved widespread use. New approaches for ICP – AES lie in the use of the jet-impact nebulizer [93], direct injection nebulization [94] (Section 21.4.1). In pneumatic nebulization for ICP – AES, continuous sample feeding requires a sample aspiration time of ca. 30 s, to give a stable signal, a measurement time of some 5 – 10 s, and a minimum rinsing time of 30 s. Discrete sampling is also possible with injection systems developed for flame AAS [78], [79], and by flow injection. Work

with sample aliquots down to 10 μL becomes possible — especially useful with microsamples [91], or for the analysis of high-salt solutions [239].

For ICP – AES, sequential, simultaneous, and combined instruments are used. In sequential spectrometers, special attention is given to the speed of wavelength access and, in simultaneous spectrometers, to the provision of background correction. In combined instruments, a number of frequently used channels are fixed and a moving detector or an integrated monochromator gives access to further lines. Echelle instruments allow maximum resolution at low focal length and can be combined with advanced detectors such as CCDs.

Analytical Performance. *Power of Detection.* The commonest configuration for an ICP-AES instrument is for the monochromator and detector to view the plasma side-on as shown in Figure 39 (a). This means that there is an optimum viewing height in the plasma which yields the maximum signal intensity, lowest background, and least interferences, and it is common practise to optimize this parameter to obtain the best performance from the instrument. Criteria of merit commonly used include the signal-to-noise ratio (SNR), signal-to-background ratio (SBR) and net signal intensity. Different elements, and different analytical lines of the same element, will have different optimum viewing heights so a set of compromise conditions is usually determined which gives satisfactory performance for a selected suite of analytical lines. An alternative configuration which is gaining popularity is end-on viewing of the plasma [Fig. 39 (b)]. This has the advantage that an optimum viewing height does not have to be selected because the analytical signal from the whole length of the central channel is integrated. However, any interferences present are also summed, so that interference effects are expected to be more severe with this configuration.

The nebulizer gas flow and power, and the observation height are also important parameters to optimize. Such multivariate optimizations can be done by single-factor studies or Simplex optimization [240], [241]. The nebulizer gas influences the aerosol droplet size, nebulizer efficiency, and plasma temperature. For each element and line there is a rather sharp optimum for the aerosol gas flow where the SBR is maximum. The power determines mainly the plasma volume and is optimum for soft lines (atomic and ionic lines of elements with low ionization potential) at rather

low values (1 – 1.2 kW), whereas for hard lines (lines of elements with high excitation potential and high-energy ionic lines) and for organic solutions, the optimum value is higher (1.5 – 1.7 kW). The optimum observation height is a compromise between the analyte number densities (highest in the lower zones) and completeness of atomization and excitation (a few millimeters above the working coil). Only elements such as the alkali metals are measured in the tail flame. As the analytical lines in ICP – AES have a physical width of 1 – 20 pm and the spectra are complex, a high-resolution spectrometer is required.

The detection limits for most elements are at the 0.05 ng/mL (Mg, Ca, etc.) to 10 ng/mL level (As, Th, etc.). Especially for elements which have refractory oxides (Cr, Nb, Ta, Be, rare earths, etc.), fairly low ionization potential, and sensitive ionic lines, the detection limits are much lower than in AAS. For P, S, N, O, and F, the most sensitive lines are at vacuum wavelengths.

Interference. Most interference in ICP – AES is additive and relates to coincidences of spectral background structures with the elemental lines used. Interference is minimized by the use of high-resolution spectrometers and background correction procedures. In the wavelength range 200 – 400 nm more than 200 000 atomic emission lines of the elements are listed [200] and molecular bands also occur. Accordingly, coincidences often occur within the spectral line widths themselves. In order to facilitate the selection of interference-free spectral lines in ICP – AES, BOUMANS worked out a system which gives the interference for about 900 prominent lines [205] for spectrometers with a given spectral slit width, in terms of critical concentration ratios. Further, line tables for ICP – AES [242] and atlases [204], [206] are available, although incomplete. In ICP – AES multiplicative interference also occurs. This may have several causes:

1) Nebulization effects: as discussed earlier, differences in physical properties of the sample solutions lead to variations in the aerosol droplet size and thus in the efficiency of nebulization and sample introduction. This effect is strongest in the case of free sample aspiration and relatively low nebulizer gas flow, and can be minimized (Section 21.4.1).
2) Easily ionized elements have a complex influence. First they may cause a decrease in excitation temperature, as energy is consumed for ionization. Further, they may shift the ion-

ization equilibrium for partially ionized elements as their easy ionization can influence the electron number densities. They may also cause changes in plasma volume as a result of ambipolar diffusion or they may even quench excited species such as metastables [243]. After correcting for the influence of easily ionized elements on the spectral background, signal enhancements or suppressions are negligible up to concentrations of 500 µg/mL for Na, K, Ca, or Mg. At larger differences in concentration matrix matching of the standards should be applied.

Analytical Precision. As sample introduction in the plasma is very stable, relative standard deviations in ICP–AES are ca. 1 % and the limiting noise is proportional, including flicker and nebulizer noise. Below 250 nm detector noise limitations may occur, especially with high-resolution spectrometers. At integration times beyond 1 s, the full precision is normally achieved.

With the aid of an internal standard, fluctuations in sample delivery and, to a certain extent, changes in excitation conditions can be compensated. To gain the full benefit of internal standardization, the intensities for the analytical and internal standard lines (and in trace determinations that of the spectral background near the analytical line) should be measured simultaneously. The selection of line pairs for ICP–AES has been investigated by FASSEL et al. [248] and the background correction by MEYERS and TRACY [249]. With internal standardization relative standard deviations below 0.1 % can be obtained.

Multielement Capacity. The multielement capacity of both sequential and ICP–AES is high, as is the linear dynamic range (up to five orders of magnitude). It is limited at the lower end by the limits of determination and at higher concentrations by the onset of self-reversal. This starts at analyte concentrations of ca. 20 000 µg/mL, as analyte atoms hardly occur at all in the cooler plasma zones.

Special Techniques. The analytical performance of ICP–AES can be considerably improved by using alternative techniques for sample introduction.

Organic solutions (methyl isobutyl ketone, xylene, kerosene) can be nebulized pneumatically and determinations in these solvents can be performed directly with ICP–AES [250]. The addition of oxygen or the use of higher power is helpful in avoiding carbon deposits. Oils can be analyzed after dilution 1:10 with xylene ($c_L = 0.5 - 1$ µ/g) and heavy metals can be extracted from wastewater as complexes with ammonium pyrrolidinedithiocarbonate (APDTC) and separated from more easily ionized elements [251].

Ultrasonic nebulization has been used right from the beginning of ICP–AES [99]. Nebulizer types where the sample liquid flows over the nebulizer transducer crystal and types where ultrasound at 1 MHz is focused through a membrane on the solution have both been used. With aerosol desolvation the power of detection of ICP–AES can be improved by a factor of ten by using ultrasonic nebulization. This applies especially to elements such as As, Cd, and Pb, which are of environmental interest. However, because of the limitations discussed in Section 21.4.2, the approach is of practical use only for dilute analytes, as in water analysis [98]. New approaches such as thermal spray nebulization [252] and high-pressure nebulization [94] increase analyte introduction efficiency and thus the power of detection, but again only when aerosol desolvation is applied. They are especially interesting for speciation by coupling of ICP–AES to HPLC.

Electrothermal evaporation (ETV) and direct sample insertion increase sampling efficiency so that work with microsamples becomes possible, and the power of detection can be further improved. ETC–ICP–AES has been applied with graphite cups [117], graphite furnaces [66], and tungsten filaments [253]. Direct sample insertion enables the direct analysis of waste oils [124], and of microamounts of biological samples [122], as well as the determination of volatile elements in refractory matrices, as shown for Al_2O_3 [121]. Difficulties lie in calibration and in signal acquisition. In trace analysis, the latter necessitates simultaneous, time-resolved measurement of the transient signals for the line and background intensities.

Hydride generation with flow cells [102] decreases the detection limits for As, Se, etc., by 1.5 orders of magnitude down to the ng/mL level — useful in water analysis. However, possible systematic errors from heavy metals and analytes in nonmineralized form remain.

Gas and liquid chromatography coupled to ICP–AES are useful for speciation work [196], [254], e.g., for Cr(III)/Cr(VI) speciation.

Direct solid sampling is of special interest, even though ICP–AES is of use mainly for the

analysis of liquids. In a number of cases the same precision and accuracy can be obtained as in solutions, but without the need for time-consuming sample dissolution, involving analyte dilution. For compacted solids, arc/spark ablation is a viable approach in the case of metals [255], [135]. Aerosols with particle sizes at the μm level [137] and detection limits at the μg/g level are obtained [256]. Owing to the separate ablation and excitation, matrix influences are especially low, as shown for aluminum [135] and steels [256]. Laser ablation can be used with similar figures of merit [257], even for electrically insulating samples with lateral resolution (up to ca. 10 μm).

High-frequency discharges at low power and with capacitive power coupling, e.g., stabilized capacitively coupled plasmas (SCP) [255] for element-specific detection in chromatography may be a useful alternative, especially when advanced sample introduction techniques are used.

Applications. Nowadays, ICP – AES is a routine analytical method, of special use when a large number of elements has to be determined in many samples (solutions or dissolved solids). In many cases it is complementary to AAS. ICP – AES is of special interest for the analysis of geological samples, environmental analysis, clinical analysis, food analysis, metals analysis, chemical analysis, and in the certification of reference materials [220], [221].

In the analysis of *geological samples* [243], sample dissolution is important with respect to both nebulization effects and ionization interference.

The analysis of *environmentally relevant samples* is a major field of application [243] and many standard methods have been developed, e.g., by bodies such as the US Environmental Protection Agency [244], [245]. This covers procedures for sample preparation and determination of 25 metals in waters, wastes, and biological samples. Details of quality assurance procedures, and interference correction are included.

In *clinical analysis*, Ca, Fe, Cu, Mg, Na, and K can be directly determined in serum samples even of microsize [211]. Electrothermal evaporation and direct sample insertion are particularly useful from this point of view. Generally, however, the power of detection of ICP – AES is too low for most analytical problems encountered in the life sciences.

In the field of industrial products, especially *metals analysis* deserves mention [246]. Apart

from solution analysis, direct metal analysis by spark ablation is very useful [256]. Analysis of refractory powders by slurry nebulization, even for ZrO_2 matrices [125] or analysis subsequent to sample decomposition, is very powerful. Laser ablation of solid samples is an important technique which has benefitted from the advent of solid-state detectors [247].

21.6.5.3. Microwave Plasmas

Microwave plasmas are operated at 1–5 GHz. They are mostly produced in a magnetron. Electrons emitted by a glowing cathode are led through a chamber with a series of radially arranged resonance chambers to which an ultrahigh-frequency field is applied. The currents produced, can be coupled to the anode and transported by coaxial cable. The microwave current is led to a waveguide or a cavity. Only when the latter have dimensions below certain critical values can microwave radiation be transported. The transport efficiency can be regulated with the aid of movable walls and screws. In a cavity, a standing wave is produced and the microwave energy is coupled to the resonator by a loop or antenna; $\lambda/2$, $\lambda/4$, etc., cavities have been proposed [267] and allow tuning with respect to a favorable ratio of forward to reflected power.

The role of microwave plasmas in optical spectrometry has recently been reviewed [268].

Capacitively-Coupled Microwave Plasmas. This type of single-electrode discharge goes back to the work of MAVRODINEANU and HUGHES [269]. The microwave radiation is led to a pin-shaped electrode, and the surrounding burner housing acts as counter-electrode. A bush-shaped discharge burns at the top of the electrode which enters the sample aerosol concentrically. The plasma, which can be operated with 2–4 L/min argon, helium or nitrogen (at 400–800 W), is not in local thermal equilibrium ($T_{ex} = 5000-8000$ K, $T_g = 3000-5700$ K, $n_e = 10^{14}$) and it is a current-carrying plasma. As the sample enters, with fairly low efficiency, into a hot zone where the background emission is also high, the detection limits are poor, especially for elements with thermally stable oxides [270], e.g., Na 589 nm, 0.05 μg/mL; Be 234.9 nm, 0.006 μg/mL; Mg 285.2 nm, 0.04 μg/mL; Al 396.1 nm, 0.16 μg/mL. Alkali and alkaline-earth metals cause high matrix effects; e.g., for 500 μg/mL they amount to 100% in the case of Mg. Therefore, spectrochemical
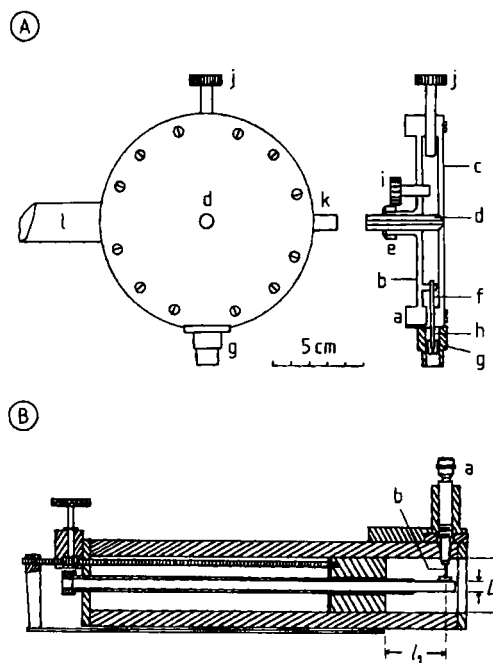
Figure 40. Microwave-induced plasma discharges
A) MIP in a TM_{010} resonator according to BEENAKKER [272]
a) Cylindrical wall; b) Fixed bottom; c) Removable lid;
d) Discharge tube; e) Holder; f) Coupling loop; g) PTFE
insulator; h) Microwave connector; i, j) Tuning screw;
k) Fine-tuning stub; l) Holder
B) Surfatron MIP, original version 20 mm $< l_1 < 250$ mm,
D_1 11 mm, D_2 40 mm, cut-off frequency 2 GHz [278]
a) Standard coaxial connector; b) Coupler

buffers such as CsCl are often used. Because it can work with He, the source is also useful for the speciation of widely diverse elements [271].

Microwave-Induced Plasmas. These electrode-less discharges are operated in a noble gas at rather low power (usually < 200 W). They became important as sources for atomic spectrometry when BEENAKKER et al. [272] succeeded in operating a stable plasma at atmospheric pressure with less than 1 L/min argon or helium and power below 100 W (Fig. 40). The so-called TM_{010} resonator is used. It has a cylindrical resonant cavity (diameter ca. 10 cm) where the power is introduced with a loop or antenna and the filament plasma is contained in a centrally located capillary. At higher power toroidal argon [273] and diffuse helium [274] discharges can be obtained. The plasma is not in local thermal equilibrium and metastable argon or helium species are thought to contribute

to the excitation. MOUSSANDA et al. [275] studied the spectroscopic properties of various microwave-induced plasma (MIP) discharges and reported rotational temperatures of ca. 2000 K. These plasmas are very useful for coupling with electrothermal evaporation, where detection limits in the upper picogram range can be obtained. This can be achieved with both graphite furnace [81] and tungsten filament atomizers [115]. Excitation of desolvated aerosols is also possible. Wet aerosols can only be taken up with the toroidal argon MIP. This allows element-specific detection in liquid chromatography [276]. Helium MIPs are excellent for element-specific detection in gas chromatography, and are commercially available [277]. Not only the halogens and other elements relevant in pesticide residue analysis, but also organolead and organotin compounds can be determined down to low concentrations. This makes MIP – AES very useful for speciation work [197]. The MIP is also useful for excitation of volatile hydride-forming elements after removal of excess hydrogen. With different trapping techniques detection limits down to the sub-nanogram level are easily obtained [104]. A further useful type of MIP is the so-called surfatron described by HUBERT et al. [278] (Fig. 40 B). Here, the power is introduced by a perpendicular antenna directed near the plasma capillary; a plasma is formed by microwave coupling through a slit and power propagates along the surface of the discharge. Compared with the MIP obtained in a TM_{010} resonator this plasma is more stable and allows more variability in sampling, as shown by direct comparative studies with the same tungsten filament atomizer [279]. The MIP torch described by QUEHAN JIN et al. [280] can operate both argon and helium discharges and make use of hydride generation, electrothermal evaporation, etc.

A strip-line microwave plasma has been described by Barnes and Reszke [281], [282]. A particular advantage of this design is that it is possible to sustain more than one plasma in the cavity at the same time. Such a plasma has compared favorably with other designs [282], though it has not yet been widely used.

21.6.6. Glow Discharges

Glow discharges have long been recognized as unique sources for AES [283]. Their special features relate to the possibility of analyte introduction by sputtering as well as to the advantages of
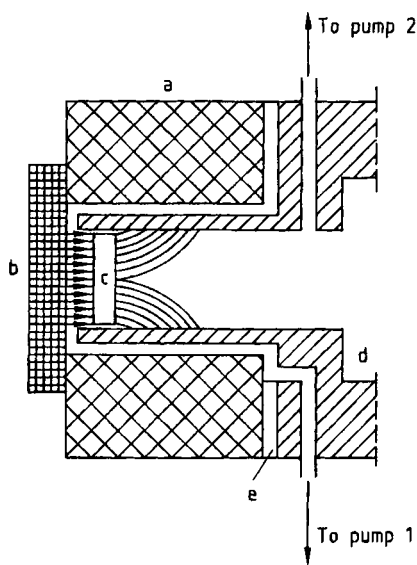
Figure 41.   Classical GRIMM glow discharge lamp [149]
a) Cathode body; b) Sample; c) Negative glow; d) Anode body; e) PFTE sheath (0.2 mm thick)

plasmas not in local thermal equilibrium, enabling the discrete excitation of terms, even those of high energy. These advantages are realized in the hollow cathode and related sources, in glow discharges with flat cathodes, with d.c. or r.f. powering, and in special sources such as gas-sampling glow discharges.

21.6.6.1. Hollow Cathodes

Here, the energetically important parts of the discharge (cathode layer, dark space, and negative glow) as well as the sample, are inside the cathode cavity. The volatilization results from cathodic sputtering and/or thermal evaporation and the analyte is excited in the negative glow. The characteristic depends on the discharge gas (at $100-500$ Pa Ar, $i=0.2-2$ A, $V=1-2$ kV; at $1-2$ kPa He, $i=0.2-2$ A, $V<1$ kV), and also on the cathode mounting. In a cooled cathode, the characteristic is abnormal and the analyte volatilizes by cathodic sputtering only, whereas in a hot hollow cathode, thermal evaporation also takes place, and the characteristic becomes normal especially at high currents.

Owing to the long residence times in the excitation zone and the high line-to-background ratios, resulting from the nonequilibrium character of the plasma (gas temperatures of $600-1000\,°C$

[19] and much higher electron temperatures), the hollow cathode is the radiation source with the lowest absolute detection limits. For this reason, it is widely used for dry solution residue analysis, which is useful subsequent to matrix separation in the analysis of high-purity substances [284]. The Doppler widths of the lines are low and this source can even be used for isotopic analysis (determination of $^{235}U/^{238}U$). It is still employed for the determination of volatile elements in a nonvolatile matrix, such as As, or Sb in high-temperature alloys [285], where detection limits below $1\ \mu g/g$ can be obtained. A treatment of the hollow cathode lamp as an AES source is given in [286].

21.6.6.2. Furnace Emission Spectrometry

Similar to hollow cathodes, electrically heated graphite furnaces in a low-pressure environment can be used for sample volatilization. The released analyte can then be excited in a discharge between the furnace and a remote anode. This source is known as furnace atomic nonresonance emission spectrometry (FANES), and was introduced by FALK et al. [287]. Owing to the separation of volatilization and excitation, its absolute detection limits are in the picogram range. For real samples, however, volatilization and excitation interference may be considerable.

21.6.6.3. Glow Discharges with Flat Cathodes

With this source, which is normally a d.c. discharge, compact electrically conducting samples can be analyzed. They must be flat, are used as the cathode, and ablated by cathodic sputtering. The sputtered material is excited in the negative glow of the discharge, which is usually operated at $100-500$ Pa of argon. As the sample is ablated layer-by-layer, both bulk and in-depth profiling analysis are possible.

Glow Discharge Sources. A useful device was first described by GRIMM in 1968 [288]. The anode is a tube (diameter $8-10$ mm) with the distance from its outer side to the cathode block and the sample less than the mean free path of the electrons (ca. 0.1 mm). The anode – cathode interspace is kept at a lower pressure than the discharge lamp. According to the Paschen curve, no discharge can take place at these locations, and the glow discharge is restricted to the cathode surface (Fig. 41). In later versions, a lamp with floating restrictor and remote anode is used, so that no

second vacuum is necessary. The Grimm lamp has an abnormal characteristic, the current is set at 40–200 mA and the burning voltage is usually below 1 kV. In both krypton and neon, the burning voltage is higher than in argon, but it also depends on the cathode material. The ablated material is deposited both at the edge of the burning crater, limiting the discharge time to a few minutes, and inside the anode tube, making regular rinsing necessary. In AES, the glow discharge is viewed end on.

Characteristics. After ignition of the discharge, a burn-in time is required to achieve sputtering equilibrium. First, the surface layers have to be removed, but once equilibrium is reached the discharge penetrates into the sample at constant velocity. Burn-in times are usually up to 30 s for metals (at 90 W in argon, zinc 6 s; brass 3–5 s; steel 20 s; Al 40 s), but depend on the sputtering conditions (shorter at high voltage, low pressure, etc.). The burning crater has a structure depending on the material structure or inclusions, and the electrical field may induce a small curvature, especially in the classical Grimm lamp [289]. Material volatilization is of the order of mg/min and increases in the sequence: C < Al < Fe < steel < Cu < brass < Zn. It also depends on the gas used, normally a noble gas. Helium is not used; because of its low mass sputtering is too limited and the sputtering rate increase in the sequence: neon < argon < krypton.

Sputtering is proportional to $1/\sqrt{p}$ [151], where p is the gas pressure. The sputtering rates of alloys can be predicted to a certain extent from the values for the pure metals [149].

For excitation, collision with high-energy electrons, but also other processes such as charge transfer or excitation by metastables are important. Electron temperatures are 5000 K (slow electrons for recombination) and > 10 000 K (high-energy electrons for excitation), and gas temperatures are below 1000 K, so line widths are low. Owing to the low spectral background intensities, the limits of detection arising from the source are low and, for metals, range down to μg/g in a classical Grimm lamp [290]. Elements such as B, P, S, C and As can also be determined. The plasma is optically thin, but for resonance lines of matrix elements, self-reversal may occur. However, the linear dynamic range is high compared with arc or spark sources.

Owing to the stable nature of the discharge, noise is low. There is hardly any flicker noise and only a few frequency-dependent components from the vacuum line are superimposed on a white noise background [44]. Relative standard deviations with no internal standard can be below 1 %. However, the low radiant densities may lead to shot noise limitations. Because of its stable nature, the glow discharge source can be coupled with Fourier transform spectrometry. As sample volatilization is due to cathodic sputtering only, matrix interference as a result of the thermochemical properties of the elements do not occur.

Applications. Glow discharge AES is widely used for *bulk analysis of metals.* This applies less to production control (the burn-in time required is long compared with sparks) than to product control. Here, the easy calibration due to the absence of volatilization interference and low spectral interference, as well as the high linear dynamic range are advantageous. For elements with high sputtering rates, such as copper, the method is especially useful. Samples have to be available as flat and vacuum-tight disks (thickness 10–30 mm), but with special holders threads and metal sheets can be analyzed. The samples must be polished and freed from oxide layers prior to analysis.

Depth-profile analysis is also possible, as the sample is ablated layer-by-layer with a penetration rate of 1–3 μm/min. Here, the intensities of the analyte lines are measured as a function of time. However, the sputtering rates of alloys with varying composition must be known to convert the time scale into a depth scale. The intensities must be related to concentrations, which can be done by using theoretical models and sputtering constants [291]. The power of detection may be quite good and depth resolution is of the order of 5 nm, when elemental concentrations > 0.1 % are monitored. Depth profile analysis is now a main field of application of glow discharge emission spectrometry in the metals industry (Fig. 42) [291].

Electrically insulating powders may be analyzed after mixing with Cu powder and pelleting [134]. Performance depends greatly on the particle size of the powders, but detection limits in the μg/g range are possible for the light elements (Al, Be, B, etc.).

21.6.6.4. New Developments

Since it was found that a great part of the analyte released by sputtering consists of ground-state atoms, cross-excitation by d.c. [292],
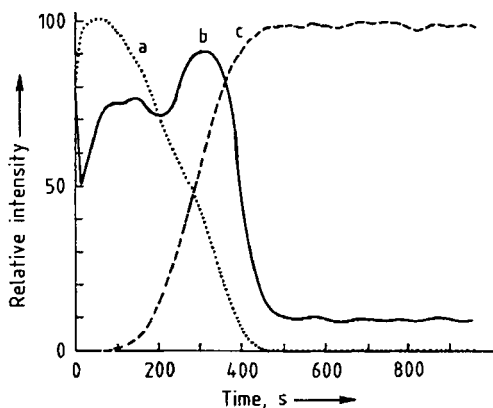
Figure 42. Depth profile through a galvanized steel sheet surface. The Zn coating (a) is ca. 18 μm thick, and the Al (b) concentration in the Fe bulk (c) is 0.049%, the voltage increases from 510 V in the Zn to 740 V in the bulk steel [291]

high-frequency [293], and microwave discharges [290] was investigated. The latter was found to improve the power of detection by up to a factor of ten, e.g., for steels. Further, it was found that simple types of d.c. discharges were capable of exciting vapors and even molecular species such as volatile hydrides [294].

Radio-frequency glow discharges are very useful sources for atomic spectrometry. By means of a bias potential in the vicinity of the sample, insulating samples such as ceramics can be directly ablated and analyzed by AES [295].

21.6.7. Laser Sources (→ Laser Analytical Spectroscopy)

High-power lasers have proved useful sources for the direct ablation of solids. In AES, ruby and Nd–YAG lasers have been used since the 1970s for solid ablation. When laser radiation interacts with a solid, a laser plume is formed. This is a dense plasma, containing both atomized material and small solid particles, evaporated from the sample. The processes occurring and the figures of merit in terms of ablation rate, crater diameter (ca. 10 μm), depth at various energies, and the expansion velocities of the plume are well known. The plume is optically thick and seems of little interest as an emission source. By applying cross-excitation with a spark or microwave energy, however, the power of detection is considerably improved [141], [142]. Absolute detection limits go down to 10^{-15} g, but relative standard devia-

tions are no better than 30%. The method, however, can be applied for identification of mineral inclusions and microdomains in metallurgical analysis, forensic samples, archeology, and remote analysis in the nuclear industry.

The method has evoked renewed interest, mainly due to the availability of improved Nd–YAG laser systems. Different types of detector, such as microchannel plates coupled to photodiodes and CCDs in combination with multichannel analyzers, make it possible to record an analytical line and an internal standard line simultaneously, so that analytical precision is considerably improved. By optimizing the ablation conditions and the spectral observations, detection limits obtained with the laser plume as source for AES are in the range 50–100 μg/g with standard deviations ca. 1% [142].

Manufacturers. Equipment for atomic emission spectrometry now is available from many different manufacturers. Several types of sources — arc, spark, ICP, MIP, glow discharge (GD), etc. — are available:

Agilent Technologies (formerly Hewlett Packard), USA: ICP-MS, MIP-AES
Baird Atomic, USA: arc/spark AES, ICP-AES, AAS
Analytic Jena AG, Germany: AAS
Hitachi, Japan: ICP-AES, AAS, ICP-MS
Instruments SA/Jobin Yvon Emission, France: ICP-AES, GD-AES
LECO Corporation, USA: TOF-ICP-MS, GD-AES
Leeman Labs Inc., USA: ICP-AES
PerkinElmer Instruments (incorporating Princeton Applied Research, ORTEC, Signal Recovery and Berthold products), USA: AAS, ICP-AES, ICP-MS, GC, GC/MS
Shimadzu, Japan: arc/spark AES, ICP-AES, AAS Spectro Analytical Instruments, Germany: ICP-AES, arc/spark AES, ICP-MS
TJA Solutions (incorporating Thermo Jarrell Ash, VG Elemental and Unicam AA), USA: AAS, ICP-AES, ICP-MS, SF-ICP-MS, GD-AES, GD-MS
Varian Associates, Australia: ICP-AES, ICP-MS, AAS
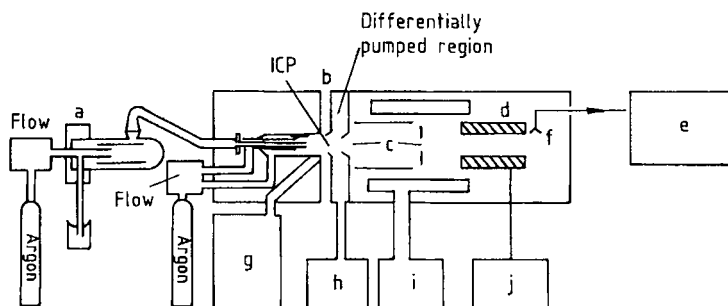
Figure 43. Instrumentation for low-resolution ICP–MS
a) Nebulizer; b) Sampler, skimmer; c) Ion lenses; d) Quadrupole mass filter; e) Computer; f) Electron multiplier; g) Radio-frequency supply; h) Mechanical pump; i) Cryosystem; f) Quadrupole supply

21.7. Plasma Mass Spectrometry

The plasma sources developed for AES have also proved very suitable ion sources for mass spectrometry. This is especially true for electrical discharges at pressures in the 100–500 Pa range and sources at atmospheric pressure with the advent of powerful vacuum systems with which to bridge the pressure difference between the mass spectrometer (of the order of 10^{-4} Pa) and the source. Elemental mass spectrometry, however, goes back to the use of high-vacuum arcs and sparks, for ultratrace and survey analyses of metal samples [50]. Spark source mass spectrography, with high-resolution sector field mass spectrometers, though still useful for characterization of electrically conducting solids down to the ng/g level, needs expensive instruments, and the analytical precision is low. In this chapter, only the newer mass spectrometric methods, which make use of plasma sources from optical atomic spectrometry are treated, namely, ICP mass spectrometry (ICP–MS) and glow discharge mass spectrometry (GD–MS). For a detailed treatment of mass spectrometric methods see →Mass Spectrometry.

21.7.1. Principles

Plasma mass spectrometry developed on the basis of the experience gathered from the extraction of ions from flames, as performed for diagnostic purposes in the mid-1960s [296]. In 1980 Gray, Houk, and coworkers [297] first showed the analytical possibilities of ICP–MS. The method has the same ease of sample introduction

as ICP–AES, but has much lower detection limits, covers more elements, and allows the determination of isotope concentrations; ICP–MS rapidly became an established analytical method.

The principles and practice of ICP–MS have been treated in a number of excellent texts [220], [298], [299].

21.7.1.1. Instrumentation   (Fig.43)

In ICP–MS, the ions formed in the ICP are extracted with the aid of a sampler into the first vacuum stage, where a pressure of the order of 100 Pa is maintained. Here, a supersonic beam is formed and a number of collision processes take place as well as an adiabatic expansion. These processes, together with their influence on the ion trajectories in the interface and behind the second aperture (skimmer), are very important for the transmission of ions and for related matrix interference [300]. Today's commercially available instrumentation falls into two groups: quadrupole and high-resolution instruments.

21.7.1.2. Analytical Features

The analytical features of ICP–MS are related to the highly sensitive ion detection and the nature of the mass spectra themselves.

ICP Mass Spectra. In low-resolution ICP–MS, resolution is at best 1 atomic mass units and, especially in the lower mass range, signals from cluster ions occur, which may cause spectral interference with analyte ions [301]. Cluster ions may be formed from different types of compounds:

1) Solvents and acids: H^+, OH^+, H_2O^+, NO^+, NO^{2+}, Cl^+, SO^+, SO_2^+, SO_3H^+ (when residual H_2SO_4 is present)
2) Radicals from gases in the atmosphere: O_2^+, CO^+, CO_2^+, N_2^+, NH^+, NO^+
3) Reaction products of the above-mentioned species with argon: ArO^+, $ArOH^+$, $ArCl^+$, Ar_2^+

Cluster ions give strong signals in the mass range up to 80 atomic mass units and may hamper the determination of light elements as a result of spectral interference. At higher masses, singly charged ions of the heavier elements occur, as well as doubly charged ions of the light elements, especially those with low ionization energies.

Signals from compounds of the analyte with various other species occur, such as MO^+, MCl^+, MOH^+, MOH_2^+. These ions may be formed by the dissociation of nitrates, sulfates, or phosphates in the plasma, or they may result from reactions between analyte ions and solvent residuals or oxygen in the plasma, or at the interface between the plasma and the mass spectrometer. Selection of the sampling zone in the plasma and the aerosol carrier gas flow may strongly influence the signal intensities of cluster ions and doubly charged analyte ions [302], [303].

For each element, signals from its different isotopes are obtained. Their intensity ratios correspond to the isotopic abundances in the sample. This fact can be used in calibration by isotopic dilution with stable isotopes, and in tracer experiments. Isotopic abundance is also useful for the recognition of spectral interference. The spectra in ICP–MS are simpler than in ICP–AES.

In ICP–MS, the background intensities are mostly low and are mainly due to the dark current of the detector, and to signals produced by ions scattered inside the mass spectrometer as a result of collisions with residual gas species or reflection. In order to shield the direct UV radiation, a beam stop is supplied often in the mass spectrometer. The plasma contributes little to the background, in contrast to ICP–AES where the background continuum stems from interactions between electrons and ions in the plasma, from molecular bands, from the wings of broad matrix lines, and from stray radiation.

A particular problem with quadrupole instruments is that they have insufficient resolution to resolve molecular ion interferences. Resolution is defined as:

$$R = \frac{M}{\Delta M} \tag{172a}$$

where: R is resolution; M is mass (strictly speaking m/z); and ΔM is peak width at 5% peak height. The resolution obtainable with quadrupoles is limited by the stabiliy and uniformity of the r.f./d.c. field, and by the spread in ion energies of the ions. Quadrupoles used in ICP–MS are typically operated at resolutions between 12 and 350, depending on m/z, which corresponds to peak widths of between 0.7 and 0.8. In comparison, sector field instruments (Fig. 44) are capable of resolution up to 10 000, resulting in peak widths of 0.008 at 80 m/z. For most applications the resolution provided by a quadrupole is sufficient, however, for applications when spectroscopic interferences cause a major problem, the resolution afforded by a magnetic sector may be desirable. Table 5 gives examples of some common spectroscopic interferences that may be encountered, and the resolution required to separate the element of interest from the interference. For example, a particular problem is the determination of arsenic in a matrix which contains chloride (a common component of most biological or environmental samples). Arsenic is monoisotopic (i.e., it only has one isotope) at m/z 75 and the chloride matrix gives rise to an interference at m/z 75 due to $^{40}Ar^{35}Cl^+$, so an alternative isotope is not available for analysis. A quadrupole has insufficient resolution to separate the two species but a magnetic sector could do so easily (Table 5). The ion transmission of sector field instruments is much greater, compared with a quadrupole, for comparable resolution, so sensitivity is much higher and detection limits are lower by three orders of magnitude or so (i.e., detection limits of $1-10$ fg ml^{-1} providing sufficiently clean reagents can be procured). Also, simultaneous isotopic detection is possible using an array of detectors, so extremely precise isotope ratio measurements can be made, allowing application of the technique for geochronology.

Recently, ICP time-of-flight (ICP–TOF) mass spectrometers have been commercialized (Figure 45). The TOF analyser has a number of advantages over other analyzers used for ICP–MS. The fact that all masses in the spectrum can be monitored simultaneously is the main attraction for many ICP–MS users. Even for simple multielement analysis, scanning analyzers suffer from a decrease in sensitivity as the number of ions to be monitored is increased. The fact that over 30 000 simultaneous mass spectra per second can

Table 5. Examples of spectroscopic interferences caused by molecular ions, and the resolution that would be necessary to separate the analyte and interference peaks in the mass spectrum

Analyte ion		Interfering ion		
Nominal isotope	Accurate m/z	Nominal isotope	Accurate m/z	Resolution required
^{51}V	50.9405	$^{16}O^{35}Cl$	50.9637	2580
^{56}Fe	55.9349	$^{40}Ar^{16}O$	55.9572	2510
^{63}Cu	62.9295	$^{40}Ar^{23}Na$	62.9521	2778
^{75}As	74.9216	$^{40}Ar^{35}Cl$	74.9312	7771
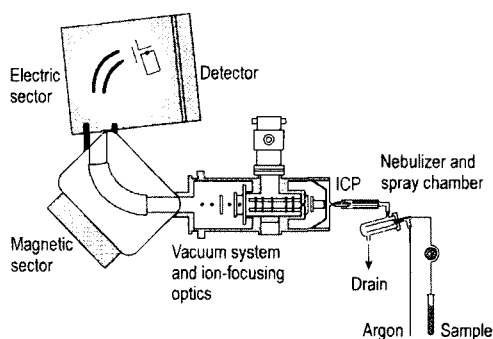
Figure 44. A sector-field ICP–MS instrument

be obtained by the TOF analyzer has many advantages for monitoring transient signals, such as those produced by laser ablation, chromatographic, capillary electrophoresis, and electrothermal vaporization sample introduction techniques. Also, improved isotope ratio measurements have been reported owing to the elimination of noise due to temporal fluctuations in the plasma and sample introduction systems.

Optimization. For the optimization of ICP–MS with respect to maximum power of detection, minimal spectral interference, signal enhancement or depression, and maximum precision, the most important parameters are the power of the ICP, its gas flows (especially the nebulizer gas), the burner geometry, the position of the sampler, and the ion optical parameters. These parameters determine the ion yield and the transmission, and thus the intensities of analyte and interference signals. At increasing nebulizer gas flow, the droplet size decreases (Section 21.4.1) and thus the analyte introduction efficiency goes up, but at the expense of the residence time in the plasma, the plasma temperature, and the ionization [304]. However, changes of the nebulizer gas flow also

influence the formation and breakdown of cluster ions as well as the ion energies (as shown for $^{63}Cu^+$ and ArO^+ [305]), and the geometry of the aerosol channel. Normally, the aerosol gas flow is 0.5–1.5 L/min. It must be optimized together with the power, which influences the plasma volume, the kinetics of the different processes taking place, and the position of the sampler. By changing the voltages of the different ion lenses, the transmission for a given ion can be optimized, enabling optimization of its detection limit and minimization of interference. In multielement determinations, a compromise must always be selected.

Power of Detection. For optimum power of detection, the analyte density in the plasma, the ionization, and the ion transmission must be maximized. The necessary power is 0.6–2 kW with the sample ca. 10–15 mm above the tip of the injector. The detection limits, obtained at single element optimum conditions, differ considerably from those at compromise conditions, but are still considerably lower than in ICP–AES (Table 6). For most elements they are in the same range, but for some they are limited by spectral interference. This applies to As ($^{75}As^+$ with $^{40}Ar^{35}Cl^+$), Se ($^{80}Se^+$ with $^{40}Ar^{40}Ar^+$), and Fe ($^{56}Fe^+$ with $^{40}Ar^{16}O^+$). The acids present in the measurement solution and the material of which the sampler is made (Ni, Cu, etc.) may have considerable influence on these sources of interference and the detection limits for a number of elements. The detection limits for elements with high ionization potential may be even lower when they are detected as negative ions (for Cl^+ $c_L = 5$ ng/mL and for Cl^- $c_L = 1$ ng/mL).

Precision and Memory Effects. The constancy of the nebulizer gas flow is of prime importance for the precision achievable in ICP–MS. After stabilizing the nebulizer gas flows, relative standard deviations may be below 1%. They can be further improved by internal standardization
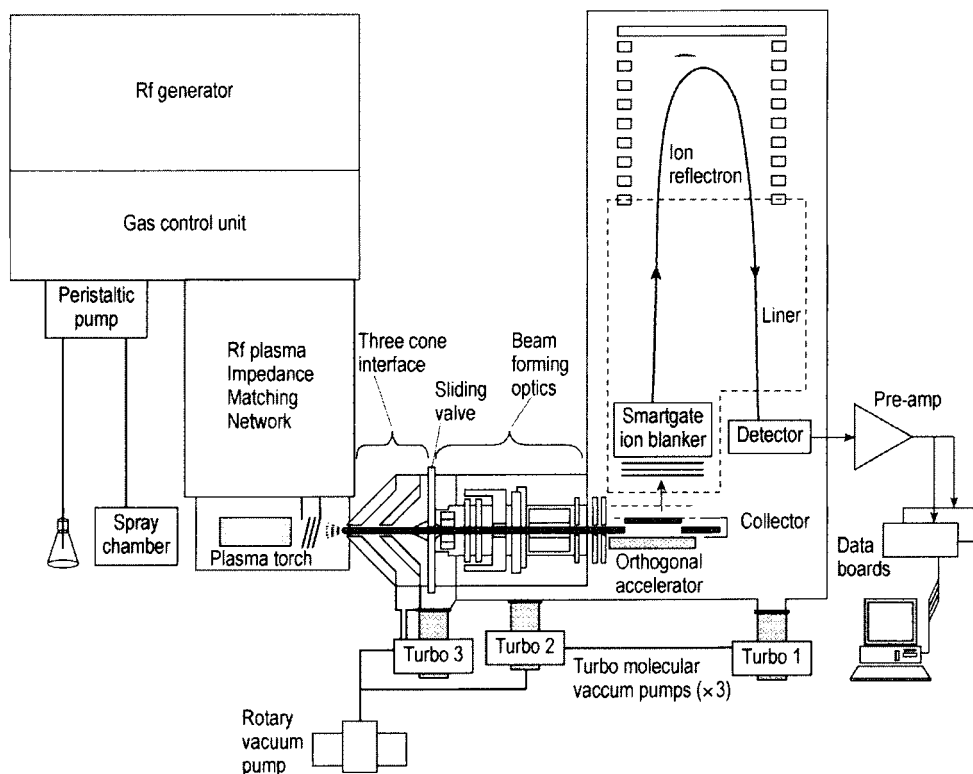
Figure 45. Orthogonal TOF-ICP-MS instrument

[306].The tolerable salt concentrations (1–5g/L) are much lower than in ICP–AES. Memory effects are low and 1–2 min rinsing times are appropriate.

Interference. ICP–MS suffers from two principal types of interference [301]. Spectroscopic interferences arise when an interfering species has the same nominal m/z as the analyte of interest. The interfering species can be either an isotope of another element (which are well documented and hence easily accounted for), or a molecular ion formed between elements in the sample matrix, plasma gas, water, and entrained atmospheric gases. The molecular ions are less easy to correct for since they will vary depending on the nature of the sample matrix. Some common molecular ion interferences are shown in Table 7. Many of these interferences can be overcome by choosing an alternative isotope of the analyte which is free from interference, though a sacrifice in sensitivity may result. If a "clean" isotope is not available then one recourse is to separate the analyte from the matrix before the analysis using

chemical extraction and chromatography, or use a magnetic sector instrument which is capable of resolving the interfering species from the analyte. Many of these interferences are thought to be formed in the interface due to a secondary discharge. This discharge can be eliminated by operating the plasma at low power, typically 600 W, and modifying the torch by inserting a grounded shield between it and the coil, or by using a center-tapped r.f. coil. Under these so-called "cool plasma" conditions interferences due to ArO^+, for example, can be eliminated, though the sensitivity for refractory elements and those with high ionization potentials may be reduced. An alternative method of reducing molecular ions is to use a collision cell prior to the quadrupole. A collision gas, such as helium, ammonia or water vapor, is introduced into the cell where it either reacts with, collisionally dissociates, or neutralizes the polyatomic species or precursors. It remains to be seen whether new interferences, formed from the matrix species and the gas, outweigh the benefits.

Non-spectroscopic interferences are caused by the sample matrix, and are manifested as an ap-

Table 6. Detection limits in ICP–AES and quadrupole ICP–MS

Element	Detection limit, ng/mL	
	ICP–AES [204]	ICP–MS [156]
Ag	7	0.005
Al	20	0.05
As	50	0.005
Au	20	0.005
B	5	0.05
Cd	3	0.005
Ce	50	0.001
Co	6	0.001
Cr	6	0.005
Fe	5	0.05
Ge	50	0.05
Hg	20	0.001
In	60	0.001
La	10	0.005
Li	80	0.001
Mg	0.1	0.05
Mn	1	0.001
Ni	10	0.005
Pb	40	0.001
Se	70	0.05
Te	40	0.05
Th	60	0.001
Ti	4	0.05
U	250	0.001
W	30	0.005
Zn	2	0.005

Table 7. Some commonly occurring spectroscopic interferences caused by molecular ions derived from plasma gases, air, and water, and the sample matrix

Molecular ion interference	Analyte ion interfered with	Nominal m/z
O_2^+	S^+	32
N_2^+	Si^+	28
NO^+	Si^+	30
NOH^+	P^+	31
Ar^+	Ca^+	40
ArH^+	K^+	41
ClO^+	V^+	51
CaO^+	Fe^+	56
ArO^+	Fe^+	56
ArN^+	Cr^+, Fe^+	54
$NaAr^+$	Cu^+	63
Ar_2^+	Se^+	80

parent enhancement or suppression in the analyte signal in the presence of a concentrated sample matrix. Such effects are thought to be caused primarily by space charge in the ion beam, whereby positive analyte ions are repelled from the ion beam by the high poisitive charge of the matrix ions, with low mass ions being relatively more affected than high mass ions. Such interferences are usually compensated for by using an internal standard, or by separating the matrix from the analyte before analysis. Because ICP–MS has such low detection limits it is also possible to dilute the sample to such an extent that the interference becomes negligible.

Isotope Ratio Analysis. Isotope ratio measurements are performed whenever the exact ratio, or abundance, of two or more isotopes of an element must be known. For example, the isotopic ratios of lead are known to vary around the world, so it is possible to determine the source of lead in paint, bullets, and petrol by knowing the isotopic abundances of the four lead isotopes 204, 206, 207, 208. Another example is the use of stable isotopes as metabolic tracers, where an animal is both fed and injected with an element having artificially

enriched isotopes and the fractional absorption of the element can be accurately determined. An important field is that of geological dating, where extremely precise isotope ratio measurements must be made.

In order to perform the isotope ratio experiment correctly it is necessary to compensate for a number of biases in the instrumentation. Mass spectrometers and their associated ion optics do not transmit ions of different mass equally. In other words, if an elemental solution composed of two isotopes with an exactly 1:1 molar ratio is analyzed using ICP–MS, then a 1:1 isotopic ratio will not necessarily be observed. This so-called mass bias will differ depending on mass, and even very small mass-biases can have deleterious effects on the accuracy of isotope ratio determinations, so a correction must always be made using an isotopic standard of known composition, as shown in Equation (173):

$$C = \frac{R_t}{R_o} \tag{173}$$

where C is the mass bias correction factor, R_t is the true isotopic ratio for the isotope pair, and R_o the observed isotope ratio for the isotope pair. For best results, such a correction should be applied to each individual isotopic pair that is to be ratioed, though this is not always possible in practice since a large number of isotopic standards would be needed to cover every eventuality. Also important is the effect of detector dead time. When ions are detected using a pulse counting (PC) detector the resultant electronic pulses are approximately 10 ns long. During and after each pulse there is a period of time during which the detector is effectively

"dead" (i.e., it cannot detect any ions). The dead time is made up of the time for each pulse and recovery time for the detector and associated electronics. Typical dead times vary between 20 and 100 ns. If dead time is not taken into account there will be an apparent reduction in the number of pulses at high count rates, which would cause an inaccuracy in the measurement of isotope ratios when abundances differ markedly. However, a correction can be applied as shown in Equation (174):

$$R_t = \frac{R_o}{1 - R_o \cdot D} \qquad (174)$$

where R_t is the true count rate and R_o is the observed count rate.

The relative abundances of the isotopes and operating conditions will affect the precision of any isotope ratio measurement. The rate at which the mass spectrometer scans or hops between masses the time spent monitoring each mass (the dwell time), and the total time spent acquiring data (total counting time) are important parameters which affect precision. The best precision achievable with a quadrupole instrument is ca. 0.1 % for a 1 : 1 isotope ratio at high count rate. This can be reduced by at least an order of magnitude if a magnetic sector instrument with an array of detectors is used to facilitate simultaneous detection. A long total counting time is desirable, because the precision is ultimately limited by counting statistics and the more ions that can be detected the better.

Isotope Dilution Analysis. A result of the ability to measure isotope ratios with ICP–MS is the technique known as isotope dilution analysis. This is done by spiking the sample to be analyzed with a known concentration of an enriched isotopic standard, and the isotope ratio is measured by mass spectrometry. The observed isotope ratio (R_m) of the two chosen isotopes can then be used in the isotope dilution equation (Eq. 174a) to calculate the concentration of the element in the sample:

$$C_x = \frac{C_s W_s M_x}{W_x M_s} \times \frac{A_s - R_m B_s}{R_m B_x - A_x} \qquad (174a)$$

where R_m is the observed isotope ratio of A to B; A_x and B_x the atom fractions of isotopes A and B, respectively, in the sample; A_s and B_s the atom fractions of isotopes A and B, respectively, in

the spike; W_x the weight of sample; W_s the weight of spike; C_x and C_s the concentration of element in sample and spike respectively; M_x and M_s the molar mass of element in sample and spike, respectively.

The best precision is obtained for isotope ratios near one, however, if the element to be determined is near the detection limit when the ratio of spike isotope to natural isotope should be between 3 and 10 so that noise contributes only to the uncertainty of natural isotope measurement. Errors also become large when the isotope ratio in the spiked sample approaches the ratio of the isotopes in the spike (overspiking), or the ratio of the isotopes in the sample (underspiking). The accuracy and precision of the isotope dilution analysis ultimately depends on the accuracy and precision of the isotope ratio measurement, so all the precautions that apply to isotope ratio analysis also apply in this case. Isotope dilution analysis is attractive because it can provide very accurate and precise results. The analyte acts as its own *de facto* internal standard. For instance, if the isotopic spike is added prior to any sample preparation then the spike will behave in exactly the same way as the analyte because it is chemically identical, providing it has been properly equilibrated by a series of oxidation/reductions. Hence, any losses from the sample can be accounted for because the analyte and spike will be equally affected.

Alternative Methods for Sample Introduction. Apart from continuous pneumatic nebulization, all sample introduction techniques known from ICP–AES have been used for ICP–MS. Similar to ICP–AES, the analysis of organic solutions is somewhat more difficult [308].

The use of *ultrasonic nebulization* as in ICP–AES increases sampling efficiency. The high water loading of the plasma has to be avoided by desolvation, not only to limit the cooling of the ICP, but also to minimize the formation of cluster ions and the related spectral interference and it should also be borne in mind that increased matrix loading in the plasma will occur. The use of high-pressure nebulization in ICP–MS has similar advantages [309], and is suitable for coupling ICP–MS to HPLC.

With the formation of *volatile hydrides*, the detection limits for elements such as As, Se, Sb, and Pb can be improved [310]. The improvement is due to enhanced analyte sampling efficiency and to the decrease of cluster ion formation.

Electrothermal evaporation (ETE) in addition to its advantages for microanalysis and improving sample transport , has the additional advantage in ICP – MS of introducing a dry analyte vapor into the plasma. Accordingly, it has been found useful for elements of which the detection limits are high as a result of spectral interference with cluster ions. In the case of ^{56}Fe, which is interfered by ^{40}ArO$^+$, PARK et al. [311] showed that the detection limit could be considerably improved by ETE. For similar reasons, the direct insertion of samples in ICP – MS maximizes the absolute power of detection [312], [313].

Although ICP – MS is mainly a method for the analysis of liquids and dissolved solids, techniques for *direct solid sampling* have been used [142]. They are especially needed when samples are difficult to dissolve or are electrically insulating and thus difficult to analyze with glow discharge or spark techniques or spatial analysis over the sample surface is required. In particular, the use of laser ablation for geological applications using UV or frequency doubled lasers is attracting considerable interest [314], [315]. For powders such as coal, slurry nebulization with a Babington nebulizer can be used in ICP – MS as well [316]. For the direct analysis of metals, spark ablation can be applied and the detection limits are in the ng/g range, as shown for steels [317]. For the analysis of metals [318] as well as of electrically insulating samples, laser ablation combined with ICP – MS is very useful. A Nd – YAG laser with a repetition rate of 1 – 10 pulses per second and an energy of ca. 0.1 J has been used. For ceramic materials such as SiC, the ablated sample amounts are of the order of 1 ng, and the detection limits are down to the 0.1 µg/g level.

21.7.1.3. Applications

Applications for ICP – MS are broadly similar to those for ICP-AES, though the better sensitivity of the former has resulted in a greater number of applications where ultra-low levels of detection are required. Sample preparation methods are similar to those used for FAAS and ICP-AES, however, nitric acid is favoured for sample digestion since the other mineral acids contain elements that cause spectroscopic interferences.

For geological applications ICP – MS is extremely well suited to the determination of the lanthanide elements because of its relatively simple spectrum. When coupled with UV laser ablation, it is an extremely powerful technique for spatial analysis and the the analysis of geological inclusions [319], especially with the extremely low limits of detection afforded by a sector field instrument. If simultaneous array detection is used then extremely precise isotope ratios can be measured, suitable for geochronology [320]. In environmental analysis the sensitivity of ICP – MS makes it ideal for applications such as the determination of heavy metals [243] and long-lived radionuclides [321] in the environment, and the determination of transition metals in seawater with on-line preconcentration [322]. Because of its capability for rapid multielement analysis ICP – MS is particularly suited to sample introduction methods which give rise to transient signals. For example, electrothermal vaporization, flow injection, and chromatographic methods for the speciation of organometallic and organohalide compounds [299]. A major attraction is the ability to perform isotope dilution analysis [320] which is capable of a high degree of accuracy and precision, and approaches the status of a primary method of analysis.

ICP – MS is the method of choice for routine analysis of samples containing ppb levels of analyte, and for speciation studies. Most quadrupole instruments are now benchtop-sized, and sector-field instruments are becoming easier to use and maintain, though high resolution operation still requires considerable skill and experience.

21.7.2. Glow Discharge Mass Spectrometry

Glow discharges [152] are known from their use as radiation sources for AES, and have been recognized as powerful ion sources for mass spectrometry. This development started from spark source mass spectrometry, where there was a continuous search for more stable sources to reduce the matrix dependence of the analyte signals [51].

In glow discharge mass spectrometry (GD – MS), the analyte is volatilized by sputtering and the ions are produced in the negative glow of the discharge, mainly as a result of collisions of the first kind with electrons and Penning ionization. Subsequently, an ion beam of composition representative of the sample is extracted. Between the glow discharge and the spectrometer a reduction in gas pressure is required. As the glow discharge is operated at a pressure around 100 Pa, a two-stage vacuum system is required. Cathodic extraction can be done at the cathode plume, by making
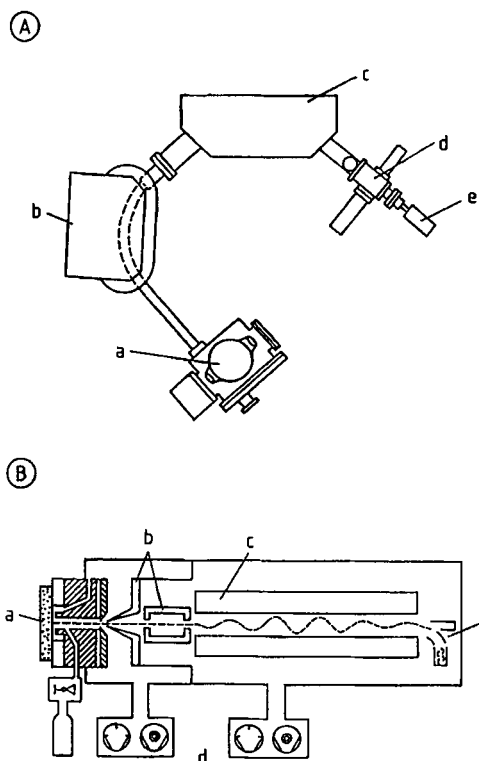
Figure 46. Glow discharge mass spectrometry
A) Sector field instrument VG 9000 [324]
 a) Source with sample introduction; b) Magnetic sector field; c) Electrostatic energy analyzer; d) Daly detector; e) Faraday cup
B) Quadrupole-based, low-resolution glow discharge mass spectrometer [325]
 a) Source; b) Ion optics; c) Quadrupole mass filter; d) Differential pumping system; e) Detector

the sample the cathode and drilling a hole in it. The aperture should reach the hot plasma center, and in most cases the extraction is done anodically.

Owing to impurities in the filler gas and the complexity of the processes taking place, not only analyte and filler gas ions, but many others may occur in the mass spectrum. Spectral interference may occur from several causes:

1) Isobaric interference from isotopes of different elements, e.g., $^{40}Ar^+$ with $^{40}Ca^+$, or $^{92}Zr^+$ with $^{92}Mo^+$

2) Interference of analyte ion signals with doubly charged ions, e.g., $^{56}Fe^{2+}$ with $^{28}Si^+$ This type of interferences is much rarer than in earlier work with high-vacuum sparks

3) Interference of analyte ions with cluster ions formed from analyte and gas species, e.g., $^{40}Ar^{16}O^+$ with $^{56}Fe^+$

4) Interference by signals from residual gas impurities, e.g., $^{14}N^{16}O^1H^+$ with $^{31}P^+$

Separation of the interfering signals depends on instrument resolution. In the case of quadrupoles, the mass resolution is at best unity, and interference can be corrected only by mathematical procedures. However, with high-resolution instruments with resolution of 5000 interference remains, especially for hydrides in the mass range 80 – 120.

21.7.2.1. Instrumentation

Glow discharges operated at pressures in the 10 – 100 Pa range can be coupled to mass spectrometers by using an aperture between both chambers with a size of ca. 1 mm. The glow discharge may use pin-shaped samples which can be introduced into the source without admitting air. BENTZ and HARRISON [323] coupled discharges with a pin cathode and hallow cathode plumes to a high-resolution sector feld mass spectrometer. Various electrode geometries were studied and diagnostic studies performed. This approach dates back to the VG 9000 spectrometer (Fig. 46 A), a double-focusing instrument in so-called reversed Nier – Johnson geometry, with a spectral resolution of ca. 10^4. The mass range to 280 can be covered with an acceleration voltage of 8 kV. As detectors, a Faraday cup and a so-called Daly multiplier are used, with the possibility of continuous switching. The vacuum of 10^{-4} Pa is maintained with the aid of oil diffusion or turbomolecular pumps. To reduce background, the source housing can be cooled to act as a cryopump.

Quadrupole mass spectrometers have also been used to detect the ions produced in a glow discharge. A high voltage is not required to achieve adequate ion extraction, and rapid scans can easily be performed, which enables quasi-simultaneous, multielement measurements. An electrostatic analyzer is often used in front of the spectrometer as an energy filter. JAKUBOWSKI et al. [325] described a combination of a Grimm-type glow discharge with a quadrupole mass spectrometer (Fig. 46 B) and studied the basic influence of the working and construction parameters. Glow discharge mass spectrometers using quadrupoles are commercially available.

21.7.2.2. Analytical Performance

The discharge is usually operated at about 100 Pa, a gas flow rate of a few mL/min, current ca. 2-5 mA, with a burning voltage of 500-1500 V. Normally the power is stabilized and pressure, current, voltage, or power are taken as optimization parameters. All electrically conducting samples, and even semiconductors, can be directly analyzed. After a burn-in phase, the discharge can be stable for hours. Detection limits down to the ng/g range are obtained for most elements.

Mass spectrometric methods are relative methods, and need to be calibrated with known reference samples; so-called relative sensitivity factors (RSF) are used:

$$RSF = (x_{el}/x_{ref})(c_{ref}/c_{el}) \qquad (175)$$

where x_{el} and x_{ref} are the isotope signals for the element to be determined and a reference element, and c are their concentrations. When matrix effects are absent RSF=1. In Gd-MS, RSFs are much closer to unity as compared for instance with spark source mass spectrometry [326] and secondary ion mass spectrometry.

Especially in quadrupole systems, spectral interference may limit the power of detection as well as the analytical accuracy. This is shown by a comparison of high- and low-resolution scans of a mass spectrum (Fig. 47) [326], [327]. In Figure 47 A the signals of $^{56}Fe^{2+}$, $^{28}Si^+$, $^{12}C^{16}O^+$, and $^2N_2^+$ are clearly separated, but not in the quadrupole mass spectra. In the latter (Fig. 47 B), interference can be recognized from comparisons of the isotopic abundances. Physical means, such as the use of neon as discharge gas [328], can be used to overcome the spectral interference problem.

21.7.2.3. Applications

Glow discharge mass spectrometry is of use for the direct determination of major, minor, and trace bulk concentrations in electrically conducting and semiconductor solids down to concentrations of 1 ng/g [329]. Because of limited ablation (10-1000 nm/s), the analysis times are long, especially when samples are inhomogeneous. The technique has been applied especially for the characterization of materials such as Al, Cd, Ga, In, Si, Te, GaAs, and CdTe. Both for high-purity substances [330], and semiconductor-grade materials [331], up to 16 elements can be deter-

mined in 5 N and 6 N samples. At high resolution, determination of Ca and K is possible [332]. In high-purity Ga, the detection limits are below the blank values of chemical analysis [333]; SiC analysis results agree with those of neutron activation analysis [331]. In Al samples, Th and U concentrations of 20 ng/g can be determined [330]. The capabilities of low-resolution GD-MS have been shown in steel analysis [333], where detection limits are down to 1 ng/g and as many as 30 elements can be determined. Electrically insulating powder samples can be analyzed after mixing with a metal powder such as Cu, Ag, or Au and pelleting, as known from glow discharge-optical emission spectrometry (GD-OES). However, degassing and oxygen release may necessitate long burn-in times and induce instabilities. Blanks arising from the metal powder may be considerable, e.g., for Pb, in the case of copper powder.

Depth profiles can be recorded just as in GD-OES, but with such high resolution that much thinner multilayers can be characterized [334]. Conversion of the intensity scale to a concentration scale is easier and the sensitivity higher. By the use of different working gases [335], the scope of the method can be continuously adapted.

Dry solution residue analysis can be performed with GD-MS. This has been shown with the aid of hollow cathodes, in which Sr, Ba, and Pb can be determined [336] and where in samples of 10-20 µL, up to 70 elements can be determined with detection limits in the picogram range [337]. With a Grimm-type discharge, solution residues can be formed on the cathode by evaporating microvolumes of analyte solutions with low total salt content. With noble metals [338], cementation can be used to fix the analyte on a copper target, so that it is preconcentrated and can be sputtered reproducibly. This has been found useful in fixing picogram amounts of Ir on a copper plate prior to analysis with GD-MS.

Manufacturers. Inductively coupled plasma (ICP) and glow discharge (GD) mass spectrometers with high (h) and low (l) resolution are available from several manufacturers:

Agilent Technologies (formerly Hewlett Packard), USA: ICP-MS, MIP-AES
Extrel, USA: GD-MS
GBC, Australia: TOF-ICP-MS, AAS
Hitachi, Japan: ICP-AES, AAS, ICP-MS
LECO Corporation, USA: TOF-ICP-MS
Micromass, UK: S-ICP-MS, ICP-MS
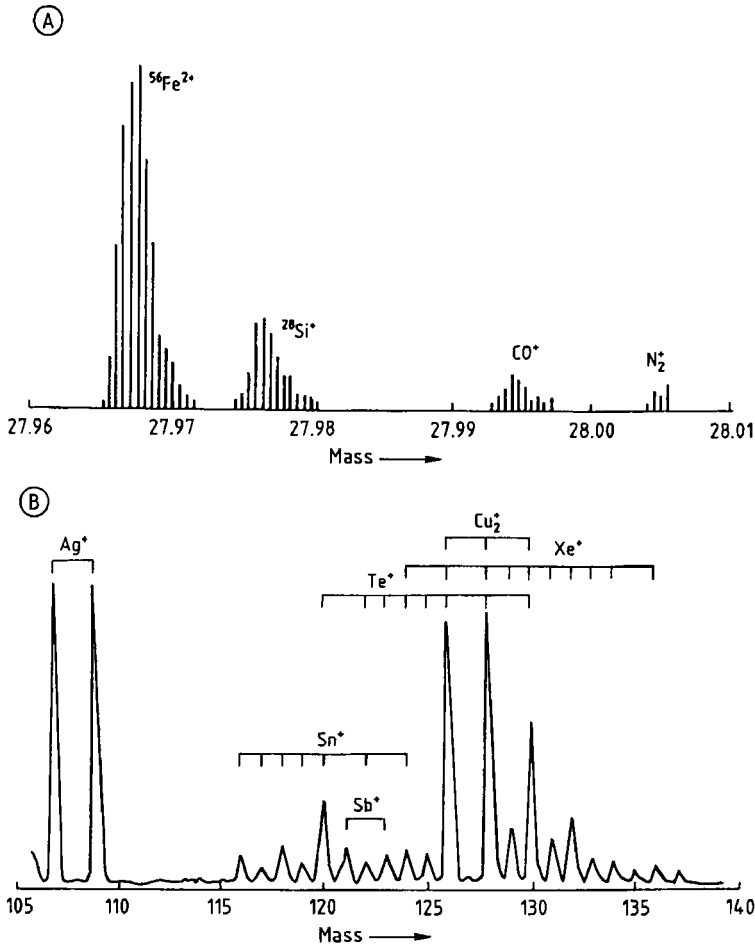
Figure 47. Glow discharge mass spectrum
A) High resolution [326]; B) Low resolution [327]

Nu Instruments Ltd., UK: SF-ICP-MS
PerkinElmer Instruments (incorporating Princeton Applied Research, ORTEC, Signal Recovery and Berthold products), USA: AAS, ICP-OES, ICP-MS, GC, GC/MS
Spectro Analytical Instruments, Germany: ICP-AES, arc/spark AES, ICP-MS
TJA Solutions (incorporating Thermo Jarrell Ash, VG Elemental and Unicam AA), USA: AAS, ICP-AES, ICP-MS, SF-ICP-MS, GD-AES, GD-MS
Thermoquest (incorporating Finnigan), USA: SF-ICP-MS, AAS
Varian Associates, Australia: ICP-AES, ICP-MS, AAS

21.8. Atomic Fluorescence Spectrometry

In atomic fluorescence spectrometry (AFS), the analyte is introduced into an atomizer (flame, plasma, glow discharge, furnace) and excited by monochromatic radiation emitted by a primary source. The latter can be a continuous source (xenon lamp) or a line source (hollow cathode lamp, electrodeless discharge lamp, or tuned laser). Subsequently, the fluorescence radiation, which may be of the same wavelength (resonance fluorescence) or of longer wavelength (nonresonance fluorescence), is measured.
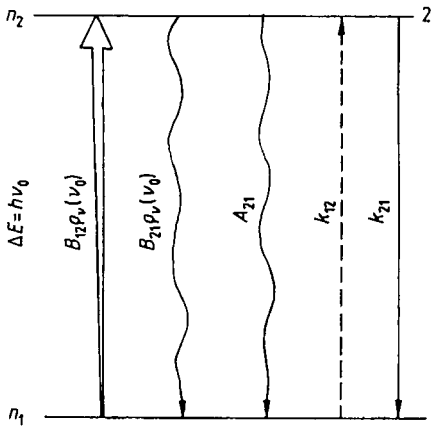
Figure 48. Transitions in a two-level system for atomic fluorescence spectrometry [339]

21.8.1. Principles [339]

For a two-level system (Fig. 48), with excitation only as a result of absorption of radiation with radiant density ϱ_v, the population of the excited level n_2 is:

$$\mathrm{d}n_2/\mathrm{d}t = n_1 B_{12}\varrho_v - n_2[A_{21} + k_{21} + B_{21}\varrho_v] \quad (176)$$

where B_{12} and B_{21} are the transition probabilities for absorption and stimulated emission, v the frequency, and n_1 the population of the lower level; A_{21} is the transition probability for spontaneous emission, and k_{21} is the coefficient for decay by collisions. The exciting radiation is supposed to be constant over the absorption profile. When the statistical weights of n_1 and n_2 are equal, $B_{21} = B_{12}$, and when $n_T = n_2 + n_1$, Equation (176) can be solved:

$$n_2 = n_T B \varrho_v t_r [1 - \exp(-t/t_r)] \quad (177)$$

where:

$$t_r = [2B\varrho_v + A_{21} + k_{21}]^{-1} \quad (178)$$

Accordingly, when the absorption of radiation increases to a certain value, A_{21} and k_{21} become negligible. The $n_2 = n_T/2$, and is independent of the radiant density. This state of saturation can be obtained when lasers are used as primary sources. As the fluorescence intensities are low, especially in the case of nonresonant AFS, special precautions must be taken for signal acquisition. Signal-to-noise ratios can be considerable im-

proved by using pulsed signals and phase-sensitive amplification. The noise is partly due to flickering (proportional to the product of measurement time t and the mean radiance $B_{\mathrm{ave}}^{\mathrm{s}}$ of the source and to background contributions (proportional to background mean intensity $B_{\mathrm{ave}}^{\mathrm{b}}$ and time t) and:

$$S/N = K B_{\mathrm{ave}}^{\mathrm{s}}/(B_{\mathrm{ave}}^{\mathrm{s}} + B_{\mathrm{ave}}^{\mathrm{b}})^{1/2} \quad (179)$$

By pulsing the source (with pulse duration t_{p}, frequency f, and pulse intensity $B_{\mathrm{p}}^{\mathrm{s}}$) as in the case of a laser, the signal-to-noise ratio is:

$$S/N = K B_{\mathrm{p}}^{\mathrm{s}} f t_{\mathrm{p}}/(B_{\mathrm{p}}^{\mathrm{s}} f t_{\mathrm{p}} + B_{\mathrm{ave}}^{\mathrm{b}})^{1/2} \quad (180)$$

but by phase-sensitive amplification, in the case of background noise limitations, it can be improved to:

$$S/N' = K B_{\mathrm{p}}^{\mathrm{s}} f t_{\mathrm{p}}/\left[f t_{\mathrm{p}}\left(B_{\mathrm{p}}^{\mathrm{s}} + B_{\mathrm{ave}}^{\mathrm{b}} \right) \right]^{1/2} = S/N(f t_{\mathrm{p}})^{1/2} \quad (181)$$

This improvement can be considerable when "boxcar" integrators with small time windows are used, to give $f t_{\mathrm{p}} \approx 10^{-8}$.

21.8.2. Instrumentation

An AFS system comprises a primary source, an atomizer, and a detection system for measuring the fluorescence radiation, which in some cases can be done without spectral dispersion.

Primary Source. As primary sources, continuous sources, such as tungsten halogenide or deuterium lamps, can be used. They have the advantage that multielement determinations are possible, but owing to the low radiant densities saturation is not obtained and the power of detection not fully exploited. With line sources, such as hollow cathode sources and electrodeless discharge lamps, much higher radiances can be obtained. Even ICPs into which a concentrated solution is introduced can be used as primary sources, so that multielement determinations become possible. With laser sources, saturation can be obtained. In most cases, dye lasers are used. They have a dye (mostly a substituted coumarin) as active medium. This liquid is continuously pumped through a cuvette and has a broad wavelength band (ca. 100 nm), within which a wavelength can be selected by a grating or prism. By using several dyes, the whole wave-
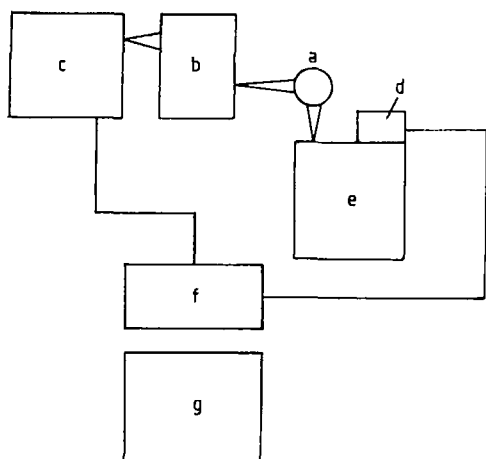
Figure 49. Experimental set-up for laser-induced atomic flourescence
a) Flame, ICP, etc.; b) Dye laser; c) Pumping laser;
d) Photomultiplier; e) Monochromator; f) Boxcar integrator;
g) Data treatment and display

length range down to 440 nm can be covered. The laser is often pumped by a second laser (e.g., an excimer laser) and operated in a pulsed mode. As the energy output is very high, it is possible to apply frequency doubling and Raman shifting, to access the wavelength range down to 200 nm. Accordingly, saturation can be flexibly achieved for almost any term. In the case of resonant transitions (with an oscillator strength: $0.1 - 1$) the energy required is $10 - 100$ mW, whereas for nonresonant transitions (oscillator strength 10^{-3}) it can reach 1 W. Tunable dye lasers can deliver frequencies up to ca. 10 Hz and the pulse length can be down to $2 - 10$ ns.

Atomizers. As fluorescence volumes, chemical flames are very useful, because of their good sample uptake and temporal stability. There has been interest in the low radiative background, low quenching, argon – hydrogen diffusion flame. The temperature of this flame is too low to prevent severe chemical interferences and therefore the argon separated air – acetylene flame has been most widely used. The hot nitrous oxide – acetylene flame (argon separated) has been used where atomization requirements make it essential. In all cases, circular flames, sometimes with mirrors around them, offer the preferred geometry. These problems do not occur when an ICP is used as atomizer [340]. In the ICP, measurements should

be made in the higher part of the tail flame, otherwise the ionized and excited analyte fraction is too high. To minimize the formation of oxides and sideways diffusion of analyte, a burner with an extended outer tube should be used for AFS. In a graphite furnace, the atomization efficiency is high and, owing to the limited expansion of the vapor cloud, the residence time of the analyte atoms is long; accordingly, graphite furnace atomization laser AFS has extremely low absolute detection limits. Low-pressure discharges can also be used as the fluorescence volume. They have the advantage that quenching of the fluorescent levels, which may limit the power of detection and cause matrix effects, is lower [341]. Laser-produced plasmas may be used as the fluorescence volume, [145], and are especially useful for direct trace determinations in electrically insulating solids.

Detection of fluorescence differs in resonant and nonresonant AFS. In *resonant AFS*, the radiation is measured perpendicular to the direction of the exciting radiation. The system suffers from stray radiation and emission of the flame. The latter can be eliminated by using pulsed primary sources and phase-sensitive detection. In *nonresonant fluorescence*, stray radiation problems are not encountered, but fluorescence intensities are lower, which necessitates the use of lasers as primary sources, and spectral apparatus to isolate the fluorescence radiation. A set-up for laser-excited AFS (Fig. 49) may make use of a pulsed dye laser pumped by an excimer laser. Here, pulses of $2 - 10$ ns occur at a frequency up to 25 Hz. A mirror can be positioned behind the source to pass the exciting radiation twice through the fluorescence volume and to lead the largest possible part of the fluorescence radiation into the monochromator. The photomultiplier signal is then amplified and measured with a box-car integrator with a nanosecond window synchronized with the pumping laser.

21.8.3. Analytical Performance

Detection limits in AFS are low, especially for elements with high excitation energies, such as Cd, Zn, As, Pb, Se, and Tl.

Elements such as As, Se, and Te can be determined by AFS with hydride sample introduction into a flame or heated cell followed by atomization of the hydride. Mercury has been determined

Table 8. Detection limits in laser atomic fluorescence and laser-enhanced ionization spectrometry

	Detection limit, ng/mL		
	Laser AFS ICP [340]	Laser AFS furnace [341]	LII flame [348]
Ag		0.003	0.07
Al	0.4		0.2
B	4		
Ba	0.7		0.2
Ca			0.03
Co		0.002	0.08
Cu		0.002	0.7
Eu		10	
Fe		0.003	0.08
Ga	1		0.04
In			
Ir		0.2	
K			0.1
Li			0.0003
Mn		0.006	0.02
Mo	5		10
Na		0.02	0.003
Pb	1	0.000025	0.09
Pt		4	
Rb			0.1
Si	1		40
Sn	3		0.3
Ti	1		1
Tl	7		0.008
V	3		0.9
Y	0.6		10
Zr	3		

by cold-vapor AFS. A non-dispersive system for the determination of Hg in liquid and gas samples using AFS has been developed commercially. Mercury ions in an aqueous solution are reduced to mercury using tin(II) chloride solution. The mercury vapour is continuously swept out of the solution by a carrier gas and fed to the fluorescence detector, where the fluorescence radiation is measured at 253.7 nm after excitation of the mercury vapor with a high intensity mercury lamp (detection limit 0.9 ng/L). Gaseous mercury in gas samples (e.g., air) can be measured directly or after preconcentration on an absorber comprising, for example, gold-coated sand. By heating the absorber, mercury is desorbed and transferred to the fluorescence detector. This preconcentration method can also be applied to improve the detection limits when aqueous solutions (or dissolved solid samples) are analyzed.

In *flame AFS*, elements which form thermally stable oxides, such as Al, Mg, Nb, Ta, Zr, and the rare earths, are hampered by insufficient atomization. This is not the case when an ICP is used as the fluorescence volume. Detection limits in the case of laser excitation and nonresonant fluorescence are lower than in ICP–AES (Table 8)[340]. For both atomic and ionic states, ICP–AFS can be performed [342].

In the case of a *graphite furnace* as fluorescence volume, and laser excitation, detection limits are very low; e.g., Ag 3×10^{-3} ng/mL; Co 2×10^{-3} ng/mL; Pb 2×10^{-3} ng/mL [339], [343].

Dry solution residue analysis with laser-excited AFS and a *hollow cathode discharge* atomizer offers very sensitive determinations of Mo [347]. BOLSHOV et al. [345] showed that very low levels of lead in Antarctic ice samples could be determined by laser-excited AFS of dry solution residues with graphite furnace atomization.

When performing laser-excited AFS at a *laser plume*, it seems useful to produce the laser plasma below atmospheric pressure, as then the ablation depends little on the matrix [146], [346].

In AFS, spectral interference is low as the fluorescence spectra are not very complex. The linear dynamic range of AFS is similar to that of OES, or even larger. There is no background limitation nor a serious limitation by self-reversal, both of which limit the linear dynamic range in OES, at the lower and at the higher end, respectively. It well may occur that the exciting radiation gets less intense when penetrating into the fluorescence volume, and that fluorescence radiation is absorbed by ground-state atoms before leaving the atomizer. These limitations, however, are less stringent than self-reversal in OES.

The analytical application of laser-excited AFS has so far, been seriously hampered by the complexity and cost of tunable lasers; this may change with the availability of less expensive but powerful and tunable diode lasers.

21.9. Laser-Enhanced Ionization Spectrometry (→Laser Analytical Spectroscopy)

Laser-enhanced ionization spectrometry (LEI) is based on the optogalvanic effect, known from early work on gas discharges. Here, an increase of the current occurs in a low-pressure discharge when irradiating with monochromatic radiation at a wavelength corresponding to a transition in the energy level diagram of the discharge gas. As an analytical method, LEI is based on measurement of the increase of analyte ionization in an
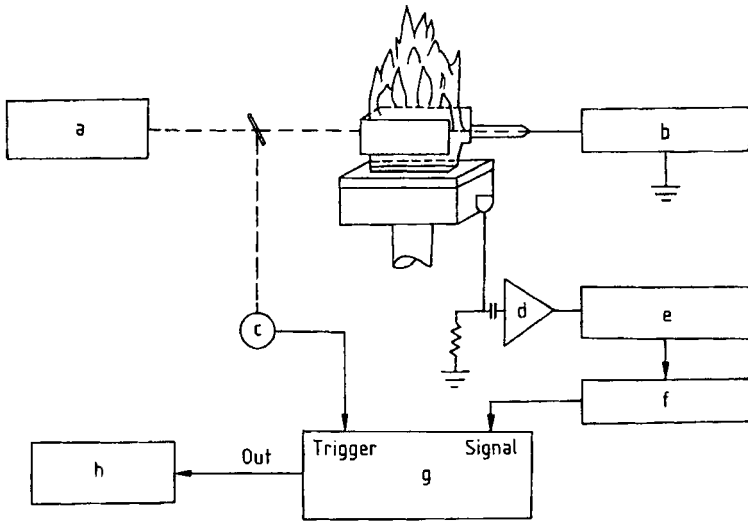
Figure 50. Flame laser-enhanced ionization spectrometry [0]
a) Flashlamp/dye laser; b) High voltage; c) Trigger photodiode; d) Preamplifier; e) Pulse amplifier; f) Active filter; g) Boxcar averager; h) Chart recorder

atomizer, resulting from selective excitation by laser radiation [347], [348].

Alternatively, two lasers can be used to both excite and photoionize the analyte. This is known as direct laser ionization (DLI) and the two methods (LEI and DLI) are collectively termed laser-induced ionization (LII).

Ionization of atoms in an atomizer may result from collisions with high-energy atoms of another species (Fig. 50) and the rate constant for this reaction can be written:

$$k_{1i} = n_X [8kT/(\pi m_r)]^{1/2} Q_{MX} \exp(-E_i/kT) \qquad (182)$$

where X refers to analyte atoms, Q_{MX} is the cross section for collisions between X and M (another element), k is Boltzmann's constant, E_i is the ionization energy, and T is the temperature in the atomizer. When the atoms X are in an excited state, being 1 eV above the original level k_{1i} changes to:

$$k_{2i} = \exp(1/kT)k_{1i} \qquad (183)$$

and the increase in ionization for a flame with $T = 2500$ K is $e^5 = 150$. In order to get sufficiently high population of the excited level, a laser must be used. For a closed system with two levels n_1 and n_2, the equilibrium is characterized by:

$$B_{12}\varrho_v n_1 = B_{21}\varrho_v n_2 + A_{21}n_2 + k_{21}n_2 \qquad (184)$$

where:

$$B_{21} = (g_1/g_2)B_{12}$$

and:

$$n_T = n_1 + n_2 + n_i$$

B_{12} and B_{21} are the transition probabilities for absorption and stimulated emission, and n_1, n_2, n_i, and n_T the analyte atom densities for the lowest state, the excited state, the ionized state, and the total number densities; g_1 and g_2 are the statistical weights, A_{21} is the transition probability for spontaneous emission and k_{21} the rate coefficient for collisional decay. Accordingly:

$$n_2 = \frac{B_{12}\varrho_v(n_T - n_i)}{[(1 + g_1/g_2)B_{12}\varrho_v + A_{21} + k_{21}]} \qquad (185)$$

and n_2 will increase until $A_{21} + k_{21}$ becomes negligible and $n_2/n_1 = g_2/g_1$. The required radiation densities are ca. $10-100$ mW/cm^3 for a resonant transition (with oscillator strength $0.1-1$) and 2 W/cm^3 for a nonresonant transition (with oscillator strength 0.001). The use of a continuous wave laser maximizes n_i, and the use of multistep excitation may be helpful, especially when high-energy terms (Rydberg states) have to be populated.

As an analytical technique, LEI requires a tunable laser, an atomizer, a galvanic detector and read-out electronics, including a boxcar integrator. It was first realized experimentally with a flame, with pneumatic nebulization for sample uptake and atomization [347].

A dye laser is used in most cases, as the selection of different dyes, frequency doubling, and Raman shifting allow coverage of the spectral range down to 200 nm. In the case of a flame (Fig. 50), the laser enters a few millimeters above the burner head, and a large collector electrode is situated a similar distance above the laser beam. This allows high-efficiency extraction of the ions formed. Apart from flames, furnaces and glow discharges are also very useful atomizers. The measurement unit includes phase-sensitive amplification to give high signal-to-noise ratios, but does not need any spectrometric equipment as only an ion current has to be measured.

The sensitivity in LEI depends on the properties of the laser system, the terms populated, the atomizer, and the detection system. It can be written:

$$S = c X_A \beta(A) \left[\exp \left(- E_i^* / kT \right) B_{12}(A) \varrho_v^p(A) \right] \tag{186}$$

where c is a constant, X_A is the number density of analyte atoms A in the lower level according to Boltzmann, β is the atomization efficiency in the flame, $\varrho_v^p(A)$ is the spectral radiation density per laser pulse, E_i^* is the ionization energy of the excited level, and S is the sensitivity in $nA \, ng^{-1} \, mL^{-1}$. The noise includes contributions from flicker noise, nebulizer noise, and detector noise, and with phase-sensitive amplification, detection limits are in the sub-ng/mL range (Table 8) [347], [349]. Because of the maximum atom number densities and residence times, detection limits are lowest for furnace atomizers. In pulsed lasers, the radiation densities favor sensitivity, whereas in continuous wave lasers the n_i values increase as a result of improved excitation efficiencies. In the thermionic diode, a heated collector electrode is placed in the vicinity of the laser beam and isolated from the vessel walls. No extraction voltage is applied and the ion current flowing as a result of decreasing negative space charge around the collector, caused by laser ionization, is taken as analytical signal. The whole system is usually operated at reduced pressure and can detect as few as 10^2 atoms [350].

In LEI, matrix effects are low. This applies especially for samples with complicated atomic spectra, as spectral interference is absent. Elements with low ionization energies may cause matrix effects, possibly due to a change of the ion current as a result of changes in the electron number density in the flame. This can be compensated by measuring alternately with and without laser irradiation. Ionization of the matrix by the laser may also falsify the ion currents measured. This can be eliminated by modulation of the frequency of the exciting laser radiation around the frequency v, which corresponds to $E_2 - E_1 = h v$ when E_2 is the intermediate level (Rydberg state) and E_1 the ground state. Ion production as a result of laser irradiation may influence the collection efficiency at the electrode as the electrical field around the electrode changes. This effect may be minimized by using a large electrode and by collecting all the ions produced. This sets limits to the pulsing frequency. For a distance of 5 mm between the collector and the burner, a field of 100 V/cm and a mobility of 20 cm^2 V^{-1} s^{-1} (typical for a flame in which $T = 2500$ K), the highest allowed pulsing rate is 4000 Hz. It has been found that for Na concentrations of 100 µg/mL, matrix effects are still negligible.

The linear dynamic range in flame LEI is of the order of three decades. However, the multielement capacity is limited as monochromatic laser radiation is required.

When using two lasers and applying two-photon spectroscopy, only those atoms which do not have a velocity component in the observation direction will undergo LEI. The absorption signals become very narrow (Doppler-free spectroscopy), enhancing the selectivity and the power of detection, and making isotope detection feasible.

The LEI technique has been applied successfully for certification purposes [351], [352], and represents a powerful approach for combinations of laser evaporation with all other atomization techniques. Apart from the galvanic detection of the ion currents, direct mass spectrometric detection of the ions can be applied, as in resonance ionization mass spectrometry [353]. Here, ionization can also be performed by multiphoton absorption, which requires very intense primary sources.

21.10. Comparison With Other Methods

Atomic absorption, optical emission, atomic fluorescence, plasma atomic spectrometry, and

Table 9. Figures of merit of analytical methods

Method	Power of detection	Matrix effects	Cost
Atomic absorption			
Flame	++	+	+
Furnace	+++	+++	++
Atomic emission			
d.c. Arc	++	+++	+++
Spark	+	++	++
d.c. Plasma	++	++	+
ICP	++	+	++
MIP	++	+++	+
Glow discharge	++	+	++
Laser	+	+	++
Atomic fluorescence			
(laser)	+++	++	+++
LEI	+++	+	+++
X-ray spectrometry			
XRF	++(+[a])	+	++
Electron microprobe	++	++	+++
PIXE	+++	++	+++
Auger electron			
spectroscopy	++	++	+++
Mass spectrometry			
Spark	+++	+++	+++
ICP	+++	++	+++
Glow discharge	+++	+	+++
SIMS	+++	++	+++
Laser	+++	++	+++
Activation analysis	+++	+	+++
Electrochemistry	+++	+++	+
Spectrophotometry	++	+++	+
Spectrofluorimetry	+++	+++	+
Chromatography			
Gas	+++	+++	++
Liquid	++	+++	++

+ Low; ++ medium; +++ very high
[a] TYXRF = Total reflection X-ray fluorescence.

newer approaches such as laser-enhanced ionization now represent powerful tools for elemental analysis, including speciation, and are to be found in many analytical laboratories. Their power of detection, reliability in terms of systematic errors, and cost are compared with those of other methods of analysis in Table 9.

21.10.1. Power of Detection

Atomic absorption spectrometry using furnaces remains a most sensitive method for easily volatilized elements with high excitation energies, and this also applies to AFS using laser excitation. The development of diode lasers should make it possible to eliminate the use of spectrometers in AAS. Dynamic range problems can be decreased by measuring in the wings of the absorption lines, and the background measurement can be per-

formed by tuning the laser [157]. This requires lasers with analytical wavelengths down to the vacuum UV. Laser-enhanced ionization is a further promising approach with respect to power of detection, especially when applying Doppler-free spectroscopy. However, these methods have limited multielement capacity. From this point of view, plasma atomic spectrometric methods offer the better possibilities; ICP – OES is likely to remain a powerful tool for multielement determinations in liquid samples and its power of detection is between that of flame and furnace AAS. It is the appropriate method for tasks such as wastewater analysis. Inductively coupled plasma – mass spectrometry offers further progress, provided the matrix loading of the sample solutions remains low. It may well have limitations in the low mass range (below 80), owing to interference by cluster ions. The use of sample introduction techniques for generating dry aerosols has great potential, e.g.,

for Fe, As, Se, etc. Further progress is possible by the use of high-resolution mass spectrometry with sector field instruments. It is likely to occur as now lower cost instruments make the technique available to a wider circle of users.

Whereas ICP atomic spectrometry is certainly a most useful method for multielement determinations in liquid samples at low concentrations, glow discharges are very useful for the analysis of solids. Glow discharge atomic emission has established itself, especially for depth profiling analysis, and is useful for bulk analysis, together with standard methods such as arc and spark emission spectrometry. Glow discharge mass spectrometry first found a role in the analysis of ultra-pure metals, as requird in microelectronics. However, its quantification is difficult and requires certified samples for calibration. In dry solution residue analysis, GD – MS has extremely low absolute detection limits. This approach is very useful for many applications in the life sciences, as it combines high power of detection with the possibilities of isotope dilution. In mass spectrometry, this approach offers new possibilities, as more classical thermionic techniques are only applicable to elements which form volatile compounds. Optical atomic and mass spectrometric methods can be used for the determination of light elements, which is an advantage over X-ray spectrometric methods. Apart from this restriction, total reflection X-ray fluorescence (TRXRF) offers high powers of detection, accuracy, and multielement capacity, with a minimum of sample preparation [354].

21.10.2. Analytical Accuracy

The analytical accuracy of methods can only be discussed in view of the complete analytical procedure applied. It is necessary to tune sample preparation and trace – matrix separations to the requirements of the analytical results in terms of accuracy, power of detection, precision, cost, number of elements, and, increasingly, the species to be determined. However, the intrinsic sensitivity of the different determination methods to matrix interference remains important. In optical emission and mass spectrometry, spectral interference remains an important limitation to the achievable analytical accuracy. In atomic emission, this applies especially to the heavier elements, as they have the more complex atomic spectra. Especially when they are present as the

matrix (as is often the case in metals analysis) the necessitity of trace matrix separation is obvious when trace analyses are to be performed. In order to overcome limitations due to spectral interference, high-resolution echelle spectrometers are increasingly used. They are compact and thus combine high resolution with excellent stability, and further enable multielement determinations, by using advanced detector technology, e.g., CCDs.

In mass spectrometry, spectral interference especially limits the accuracy in the low mass range. Progress may be expected from the use of sector field and time-of-flight mass spectrometry. However, in the first case transmission and in the second case dynamic range problems must be given attention. Signal depression and enhancement are a further cause of interference in ICP – MS. They can be taken care of by standard addition, as the sprectral background is low. Internal standardization may compensate for the effects of easily ionized element. This is more difficult in ICP – AES, where the spectral background (especially in trace analysis) is considerable and may be strongly influenced by changes in the concentrations of easily ionized elements.

In AAS and AFS, limitations in the analytical accuracy are mostly related to physical and chemical interference and less to spectral interference. Especially in furnace AAS, thermochemical processes limit the achievable accuracy and necessitate carefully worked out temperature programs. In AFS and LEI, it is necessary to control matrix influences relating to quenching when analyzing real samples.

21.10.3. Cost

The power of detection and accuracy of analytical methods cannot be discussed without considering the costs of instruments and their operation, including the costs of laboratory personnel. Methods allowing multielement analyses and high throughput are advantageous for routine analyses. The ICP spectrometric methods offer particularly good possibilities, despite the high instrument costs and the high consumption of power and gases. In many cases, however, the costs arising from sample preparation are decisive, and favor X-ray spectrometric methods, provided the limitations mentioned earlier are not encountered. Further progress is likely to depend on the availability of on-line sample treatment, e.g., with flow injection and on-line sample dissolution, which now be-

comes possible with microwave-assisted heating. In any case, the method selected has to be discussed in detail for any new analytical task.

21.11. References

[1] G. R. Kirchhoff, R. Bunsen, *Philos. Mag.* **20** (1860) 89–98.
[2] A. Walsh, *Spectrochim. Acta* **7** (1955) 108–117.
[3] B. V. L'vov, *Spectrochim. Acta* **17** (1961) 761–770.
[4] H. Maßmann, *Spectrochim. Acta Part B* **23 B** (1968) 215–226.
[5] W. Grotrian: *Graphische Darstellung der Spektren von Atomen mit ein, zwei und drei Valenzelektronen,* Springer Verlag, Berlin 1928.
[6] H. R. Griem: *Plasma Spectroscopy,* McGraw-Hill, New York 1964.
[7] L. de Galan, R. Smith, J. D. Winefordner, *Spectrochim. Acta Part B* **23 B** (1968) 521–525.
[8] W. J. Pearce, P. J. Dickerman (eds.): *Symposium on Optical Spectrometric Measurements of High Temperatures,* University of Chicago Press, Chicago 1961.
[9] P. W. J. M. Boumans: *Theory of Spectrochemical Excitation,* Hilger & Watts, London 1966.
[10] J. Junkes, E. W. Salpeter, *Ric. Spettrosc. Lab. Astrofis. Specola Vaticana* **2** (1961) 255.
[11] C. F. Bruce, P. Hannaford, *Spectrochim. Acta Part B* **26 B** (1971) 207–235.
[12] J. A. C. Broekaert, F. Leis, K. Laqua, *Spectrochim. Acta Part B* **34 B** (1979) 73–84.
[13] W. Lochte-Holtgreven (ed.): *Plasma Diagnostics,* North-Holland Publ., Amsterdam 1968.
[14] G. Elwert, *Z. Naturforsch. A: Astrophys. Phys. Phys. Chem.* **7 A** (1952) 703–711.
[15] J. M. Mermet in P. W. J. M. Boumans (ed.): *Inductively Coupled Plasma Emission Spectroscopy,* Part II, Wiley-Interscience, New York 1987, p. 353.
[16] R. Diermeier, H. Krempl, *Z. Phys.* **200** (1967) 239–248.
[17] G. Herzberg: *Molecular Spectra and Molecular Structure,* 2nd ed., Van Nostrand Reinhold, New York 1950.
[18] R. W. B. Pearce, A. G. Gaydon: *The Identification of Molecular Spectra,* Chapman and Hall, London 1953.
[19] J. A. C. Broekaert, *Bull. Soc. Chim. Belg.* **86** (1977) 895–906.
[20] B. Raeymaekers, J. A. C. Broekaert, F. Leis, *Spectrochim. Acta Part B* **43 B** (1988) 941–949.
[21] I. Ishii, A. Montaser, *Spectrochim. Acta Part B* **46 B** (1991) 1197–1206.
[22] N. P. Ferreira, H. G. C. Human, L. R. P. Butler, *Spectrochim. Acta Part B* **35 B** (1980) 287–295.
[23] K. Bockasten, *J. Opt. Soc. Am.* **51** (1961) 943–947.
[24] G. M. Hieftje, *Spectrochim. Acta Part B* **47 B** (1992) 3–25.
[25] J. A. C. Broekaert, *Anal. Chim. Acta* **196** (1987) 1–21.
[26] M. Borer, G. M. Hieftje, *Spectrochim. Acta Reviews* **14** (1991) 463–486.
[27] K. Doerffel: *Statistik in der analytischen Chemie,* Verlag Chemie, Weinheim 1984.
[28] L. Sachs: *Statistische Auswertungsmethoden,* Springer Verlag, Berlin 1969.
[29] V. V. Nalimov: *The Application of Mathematical Statistics to Chemical Analysis,* Pergamon Press, Oxford 1963.
[30] J. D. Ingle, Jr., S. R. Crouch: *Spectrochemical Analysis,* Prentice-Hall International, Englewood Cliffs 1988.
[31] G. M. Hieftje, *Anal. Chem.* **44** (1972) 81 A–88 A.
[32] H. Kaiser, *Spectrochim. Acta* **3** (1947) 40–67.
[33] H. Kaiser, H. Specker, *Fresenius Z. Anal. Chem.* **149** (1956) 46–66.
[34] E. L. Grove: *Analytical Emission Spectrometry,* vol. 1, Part I and II, Decker Inc., New York 1971.
[35] C. S. Williams, O. A. Becklund: *Optics,* Wiley-Interscience, New York 1972.
[36] P. N. Keliher, C. C. Wohlers, *Anal. Chem.* **48** (1976) 333 A–340 A.
[37] H. Kaiser, *Spectrochim. Acta* **3** (1948) 159–190.
[38] P. W. J. M. Boumans, *Colloq. Spectrosc. Int. Proc. 16th* **II** (1971) 247–253.
[39] H. Siedentopf, *Z. Phys.* **35** (1934) 454.
[40] V. K. Zworykin, E. G. Ramberg: *Photoelectricity and Its Application,* Wiley, New York 1949.
[41] Y. Talmi: "Multichannel Image Detectors," *ACS Symp. Ser.* **102** (1979).
[42] J. V. Sweedler, R. F. Jalkian, M. B. Denton, *Appl. Spectrosc.* **43** (1989) 954–962.
[43] J. A. C. Broekaert, K. R. Brushwyler, G. M. Hieftje, unpublished work.
[44] J. A. C. Broekaert, C. A. Monnig, K. R. Brushwyler, G. M. Hieftje, *Spectrochim. Acta Part B* **45 B** (1990) 769–778.
[45] A. Danielson, P. Lindblom, E. Södermann, *Chem. Scr.* **6** (1974) 5–9.
[46] R. B. Bilhorn, P. M. Epperson, J. V. Sweedler, M. B. Denton, *Appl. Spectrosc.* **41** (1987) 1114–1124.
[47] N. Furuta, K. R. Brushwyler, G. M. Hieftje, *Spectrochim. Acta Part B* **44 B** (1989) 349–358.
[48] P. J. Treado, M. D. Morris, *Anal. Chem.* **61** (1989) 723 A–734 A.
[49] L. M. Faires, *Anal. Chem.* **58** (1986) 1023 A–1043 A.
[50] F. Adams, R. Gijbels, R. Van Grieken (eds.): *Inorganic Mass Spectrometry,* Wiley, New York 1988.
[51] J. R. Bacon, A. Ure, *Analyst (London)* **109** (1984) 1229–1254.
[52] K. G. Heumann, F. Beer, H. Weiss, *Mikrochim. Acta I* 1983, 95–108.
[53] C. Brunnee, *Int. J. Mass Spectrom. Ion Processes* **76** (1987) 125–237.
[54] H. H. Willard, L. L. Merrit, A. J. Dean, A. F. Settle: *Instrumental Methods of Analysis,* 7th ed., Wadsworth, California 1998, Chap. 16.
[55] D. P. Myers, G. L. P. Yang, G. M. Hieftje, *J. Am. Soc. Mass Spectrom.* **5** (1994) 1008.
[56] D. P. Myers, G. Li, P. P. Mahoney, G. M. Hieftje, *J. Am. Soc. Mass Spectrom.* **6** (1995) 411.
[57] D. P. Myers, G. Li, P. P. Mahoney, G. M. Hieftje, *J. Am. Soc. Mass Spectrom.* **6** (1995) 400.
[58] J. R. Chapman: *Practical Organic Mass Spectrometry,* 2nd ed., John Wiley & Sons, Chichester 1993.
[59] A. Benninghoven, F. G. Rüdenauer, H. W. Werner: *Secondary Ion Mass Spectrometry,* Wiley, New York 1986.
[60] E. Denoyer, R. Van Grieken, F. Adams, D. F. S. Natusch, *Anal. Chem.* **54** (1982) 26 A–41 A.
[61] D. P. Meyers, G. M. Hieftje: "Use of a Time-of-Flight Mass Spectrometer for Plasma Mass Spectrometry," *1992 Pittsburgh Conference,* New Orleans, March 9–13, 1992, Paper 606.

[62] A. L. Gray, A. R. Date, *Analyst (London)* **108** (1983) 1033–1050.

[63] F. E. Lichte, A. L. Meier, J. G. Crock, *Anal. Chem.* **59** (1987) 1150–1157.

[64] F. Kohlrausch: *Praktische Physik.* 21st ed., vol. 1, Taubner Verlagsgesellschaft, Stuttgart 1961.

[65] F. Rosendahl, *Spectrochim. Acta* **10** (1957) 201–212.

[66] A. Aziz, J. A. C. Broekaert, F. Leis, *Spectrochim. Acta Part B* **37 B** (1982) 369–379.

[67] J. Sneddon (ed.): *Sample Introduction in Atomic Spectroscopy.* Elsevier, Amsterdam 1990.

[68] J. A. C. Broekaert: "Sample Introduction Techniques in ICP-AES," in P. W. J. M. Boumans (ed.): *Inductively Coupled Plasma Emission Spectroscopy.* vol. 1, Wiley, New York 1987, pp. 296–357.

[69] A. Montaser, D. W. Golightly (eds.): *Inductively Coupled Plasmas in Analytical Atomic Spectrometry,* 2nd ed., VCH Publishers, Weinheim 1992.

[70] R. F. Browner, A. W. Boorn, *Anal. Chem.* **56** (1984) 786 A–798 A.

[71] R. F. Browner, A. W. Boorn, *Anal. Chem.* **56** (1984) 875 A–888 A.

[72] J. A. C. Broekaert, G. Tölg, *Fresenius' J. Anal. Chem.* **326** (1987) 495–509.

[73] J. A. C. Broekaert, *Anal. Chim. Acta* **196** (1987) 1–21.

[74] J. E. Meinhard in R. M. Barnes (ed.): *Applications of Plasma Emission Spectroscopy,* Heyden, London 1979, p. 1.

[75] G. L. Gouy, *Ann. Chim. Phys.* **18** (1887) 5.

[76] E. L. Kimberley, G. W. Rice, V. A. Fassel, *Anal. Chem.* **56** (1984) 289–292.

[77] L. R. Layman, F. E. Lichte, *Anal. Chem.* **54** (1982) 638–642.

[78] E. Sebastiani, K. Ohls, G. Riemer, *Fresenius' Z. Anal. Chem.* **264** (1973) 105–109.

[79] H. Berndt, W. Slavin, *At. Absorpt. Newsl.* **17** (1978) 109.

[80] S. Greenfield, P. B. Smith, *Anal. Chim. Acta* **59** (1972) 341.

[81] A. Aziz, J. A. C. Broekaert, F. Leis, *Spectrochim. Acta Part B* **37 B** (1982) 381–389.

[82] J. Ruzicka, E. H. Hansen: *Flow Injection Analysis,* Wiley, New York 1988.

[83] S. D. Hartenstein, J. Ruzicka, G. D. Christian, *Anal. Chem.* **57** (1985) 21–25.

[84] Z. Fang: *Flow Injection Separation and Preconcentration,* VCH Verlagsgesellschaft, Weinheim 1993.

[85] P. Schramel, L. Xu, G. Knapp, M. Michaelis, *Fresenius J. Anal. Chem.* **345** (1993) 600–606.

[86] S. Nukuyama, Y. Tanasawa, *Nippon Kikai Cakkai Ronbunshu* **4** (1938) 86, **5** (1939) 68.

[87] R. Nießner, *Angew. Chem.* **103** (1991) 542–552.

[88] T. Allen: *Particle Size Measurement,* Chapman & Hall, London 1981.

[89] W. Van Borm et al., *Spectrochim. Acta Part B* **46 B** (1991) 1033–1049.

[90] S. D. Olsen, A. Strasheim, *Spectrochim. Acta Part B* **38 B** (1983) 973–975.

[91] A. W. Boorn, R. Browner, *Anal. Chem.* **54** (1982) 1402–1410.

[92] G. Kreuning, F. J. M. J. Maessen, *Spectrochim. Acta Part B* **42 B** (1987) 677–688.

[93] M. P. Doherty, G. M. Hieftje, *Appl. Spectrosc.* **38** (1984) 405–411.

[94] H. Berndt, *Fresenius' J. Anal. Chem.* **331** (1988) 321–323.

[95] J. Posta, H. Berndt, *Spectrochim. Acta Part B* **47 B** (1992) 993–999.

[96] J. Posta, H. Berndt, S. K. Luro, G. Schaldach, *Anal. Chem.* **65** (1993) 2590–2595.

[97] A. Syti, *CRC Crit. Rev. Anal. Chem.* **4** (1974) 155.

[98] V. A. Fassel, B. R. Bear, *Spectrochim. Acta Part B* **41 B** (1986) 1089–1113.

[99] P. W. J. M. Boumans, F. J. De Boer, *Spectrochim. Acta Part B* **30 B** (1975) 309–334.

[100] N. W. Barnett, L. S. Chen, G. F. Kirkbright, *Spectrochim. Acta Part B* **39 B** (1984) 1141–1147.

[101] D. C. Manning, *At. Absorpt. Newsl.* **10** (1971) 123.

[102] M. Thompson, B. Pahlavanpour, S. J. Walton, G. F. Kirkbright, *Analyst (London)* **103** (1978) 568–579.

[103] J. A. C. Broekaert, F. Leis, *Fresenius' Z. Anal. Chem.* **300** (1980) 22–27.

[104] E. Bulska, P. Tschöpel, J. A. C. Broekaert, G. Tölg, *Anal. Chim. Acta* **271** (1993) 171–181.

[105] H. Tao, A. Miyazaki, *Anal. Sci.* **7** (1991) 55–59.

[106] C. S. Stringer, M. Attrep, Jr., *Anal. Chem.* **51** (1979) 731–734.

[107] B. Welz, M. Melcher, *Analyst (London)* **108** (1983) 213–224.

[108] E. Preuß, *Z. Angew. Mineral.* **3** (1940) 1.

[109] B. V. L'vov, L. A. Pelieva, A. I. Sharnopolsky, *Zh. Prikl. Spektrosk.* **27** (1977) 395.

[110] J. McNally, J. A. Holcombe, *Anal. Chem.* **59** (1987) 1105–1112.

[111] T. S. West, X. K. Williams, *Anal. Chim. Acta* **45** (1969) 27.

[112] V. Sychra et al., *Anal. Chim. Acta* **105** (1979) 105.

[113] H. T. Delves, *Analyst (London)* **95** (1970) 431–438.

[114] H. Berndt, J. Messerschmidt, *Spectrochim. Acta Part B* **34 B** (1979) 241–256.

[115] E. I. Brooks, K. J. Timmins, *Analyst (London)* **110** (1985) 557–558.

[116] Z. Grobenski, R. Lehmann, R. Tamm, B. Welz, *Mikrochim. Acta I* 1982, 115–125.

[117] D. L. Millard, H. C. Shan, G. F. Kirkbright, *Analyst (London)* **105** (1980) 502–508.

[118] J. A. C. Broekaert, F. Leis, *Mikrochim. Acta II* 1985, 261–272.

[119] D. Sommer, K. Ohls, *Fresenius' Z. Anal. Chem.* **304** (1980) 97–103.

[120] E. D. Salin, G. Horlick, *Anal. Chem.* **51** (1979) 2284–2286.

[121] G. Zaray, J. A. C. Broekaert, F. Leis, *Spectrochim. Acta Part B* **43 B** (1988) 241–253.

[122] V. Karnassios, G. Horlick, *Spectrochim. Acta Part B* **44 B** (1989) 1345–1360.

[123] V. Karnassios, M. Abdullah, G. Horlick, *Spectrochim. Acta Part B* **45 B** (1990) 119–130.

[124] L. Blain, E. D. Salin, *Spectrochim. Acta Part B* **47 B** (1992) 205–217.

[125] R. Lobinski et al., *Fresenius' J. Anal. Chem.* **342** (1992) 563–568.

[126] B. Raeymaekers et al., *Spectrochim. Acta Part B* **43 B** (1988) 923–940.

[127] B. Raeymaekers, P. Van Espen, F. Adams, *Mikrochim. Acta II* 1984, 437–454.

[128] B. Docekal, V. Krivan, *J. Anal. At. Spectrom.* **7** (1992) 521–528.

[129] H. Min, S. Xi-En, *Spectrochim. Acta Part B* **44 B** (1989) 957–964.

[130] R. M. Barnes, S. Nikdel, *J. Appl. Phys.* **47** (1976) 3929.

[131] G. M. Hieftje, R. M. Miller, Y. Pak, E. P. Wittig, *Anal. Chem.* **59** (1987) 2861 – 2872.

[132] L. Ebdon, M. Foulkes, K. Sutton, *J. Anal. At. Spectrom.* **12** (1997) 213 – 229.

[133] J. L. Johnes, R. L. Dahlquist, R. E. Hoyt, *Appl. Spectrosc.* **25** (1971) 629 – 635.

[134] S. ElAlfy, K. Laqau, H. Maßmann, *Fresenius' Z. Anal. Chem.* **263** (1973) 1 – 14.

[135] A. Aziz, J. A. C. Broekaert, K. Laqua, F. Leis, *Spectrochim. Acta Part B* **39 B** (1984) 1091 – 1103.

[136] J. A. C. Broekaert, F. Leis, K. Laqua: "A Contribution to the Direct Analysis of Solid Samples by Spark Erosion Combined to ICP-OES," in B. Sansoni (ed.): *Instrumentelle Multielementanalyse,* Verlag Chemie, Weinheim 1985.

[137] B. Raeymaekers, P. Van Espen, F. Adams, J. A. C. Broekaert, *Appl. Spectrosc.* **42** (1988) 142 – 150.

[138] R. M. Barnes, H. V. Malmstadt, *Anal. Chem.* **46** (1974) 66 – 72.

[139] N. Bings, H. Alexi, J. A. C. Broekaert, Universität Dortmund, unpublished results.

[140] N. Omenetto: *Analytical Laser Spectroscopy,* Wiley, New York 1979.

[141] L. Moenke-Blankenburg: *Laser Micro Analysis,* Wiley, New York 1989.

[142] J. Sneddon, T. L. Thiem, Y. Lee: *Lasers in Analytical Spectrometry,* VCH, New York 1997.

[143] K. Laqua: "Analytical Spectroscopy Using Laser Atomizers," in S. Martellucci, A. N. Chester (eds.): *Analytical Laser Spectroscopy,* Plenum Publishing Corp., New York 1985, pp. 159 – 182.

[144] F. Leis, W. Sdorra, J. B. Ko, K. Niemax, *Mikrochim. Acta II* 1989, 185 – 199.

[145] W. Sdorra, A. Quentmeier, K. Niemax, *Mikrochim. Acta II* 1989, 201 – 218.

[146] J. Uebbing et al., *Appl. Spectrosc.* **45** (1991) 1419 – 1423.

[147] M. Kaminsky: *Atomic and Ionic Impact Phenomena on Metal Surfaces.* Springer Verlag, Berlin 1965.

[148] F. M. Penning: *Electrical Discharges in Gases,* Philips Technical Library, Eindhoven 1957, p. 41.

[149] P. W. J. M. Boumans, *Anal. Chem.* **44** (1973) 1219 – 1228.

[150] J. A. C. Broekaert, *J. Anal. At. Spectrom.* **2** (1987) 537 – 542.

[151] M. Dogan, K. Laqua, H. Maßmann, *Spectrochim. Acta Part B* **27 B** (1972) 631 – 649.

[152] R. Marcus (ed.): *Glow Discharge Spectrometries,* Plenum Press, New York 1993.

[153] M. Van Straaten, A. Vertes, R. Gijbels, *Spectrochim. Acta Part B* **46 B** (1991) 283 – 280.

[154] B. Welz: *Atomic Absorption Spectrometry.* Verlag Chemie, Weinheim 1985.

[155] W. J. Price: *Spectrochemical Analysis by Atomic Absorption,* Heyden, London 1979.

[156] S. J. Hill, A. Fisher: in L. Ebdon, E. H. Evans (eds.): *An Introduction to Analytical Atomic Spectrometry,* Wiley, Chichester 1998.

[157] A. Zybin, C. Schneurer-Patschan, M. A. Bolshov, K. Niemax, *Trends. Anal. Chem.* **7** (1998) 513.

[158] N. Omenetto, *J. Anal. At. Spectrom.* **13** (1998) 385 – 399.

[159] R. S. Koirthyohann, E. E. Pickett, *Anal. Chem.* **37** (1965) 601 – 603.

[160] G. P. Moulton, T. C. O'Haver, J. M. Harnly, *J. Anal. At. Spectrom.* **5** (1990) 145 – 150.

[161] M. D. Amos, J. B. Willis, *Spectrochim. Acta* **22** (1966) 1325 – 1343.

[162] R. Hermann, C. T. J. Alkemade: *Chemical Analysis by Flame Photometry,* 2nd ed., Interscience Publishers, New York 1983.

[163] C. T. J. Alkemade, T. Hollander, W. Snelleman, P. J. Th. Zeegers: *Metal Vapours in Flames,* Pergamon Press, Oxford 1982.

[164] J. A. Dean, T. C. Rains (eds.): *Flame Emission and Atomic Absorption Spectrometry.* vol. 1, Marcel Dekker, New York 1969.

[165] W. Slavin, *Anal. Chem.* **58** (1986) 589 A – 597 A.

[166] H. Berndt, G. Schaldach, *J. Anal. At. Spectrom.* **3** (1988) 709 – 713.

[167] G. Schlemmer, B. Welz, *Spectrochim. Acta Part B* **41 B** (1986) 1157 – 1166.

[168] B. Welz et al., *Spectrochim. Acta Part B* **41 B** (1986) 1175 – 1201.

[169] V. Krivan, *J. Anal. At. Spectrom.* **7** (1992) 155 – 164.

[170] W. R. A. De Jonghe, F. C. Adams, *Anal. Chim. Acta* **108** (1979) 21 – 30.

[171] D. Chakraborti et al., *Anal. Chem.* **56** (1984) 2692 – 2697.

[172] M. Yamamoto, M. Yasuda, Y. Yamamoto, *Anal. Chem.* **57** (1985) 1382 – 1385.

[173] E. Jackwerth, P. G. Willmer, R. Höhn, H. Berndt, *At. Absorpt. Newsl.* **18** (1978) 66.

[174] G. Kaiser, D. Götz, P. Schoch, G. Tölg, *Talanta* **22** (1975) 889 – 899.

[175] J. Piwonka, G. Kaiser, G. Tölg, *Fresenius' Z. Anal. Chem.* **321** (1985) 225 – 234.

[176] R. E. Sturgeon, S. N. Willie, S. S. Berman, *Anal. Chem.* **57** (1985) ff 2311.

[177] B. Welz, M. Melcher, *Anal. Chim. Acta* **131** (1981) ff 17.

[178] H. Berndt, *Spectrochim. Acta Part B* **39 B** (1984) 1121 – 1128.

[179] L. Ebdon, M. R. Cave, *Analyst (London)* **107** (1982) 172 – 178.

[180] B. Docekal, V. Krivan, *J. Anal. At. Spectrom.* **7** (1992) 521 – 528.

[181] D. S. Gough, *Spectrochim. Acta Part B* **54 B** (1999) 2067 – 2072.

[182] A. E. Bernhard, *Spectroscopy* **2** (1987) 118.

[183] K. Ohls, *Fresenius' Z. Anal. Chem.* **327** (1987) 111 – 118.

[184] T. Kantor, L. Polos, P. Fodor, E. Pungor, *Talanta* **23** (1976) 585 – 586.

[185] H. Berndt, D. Sopczak, *Fresenius' Z. Anal. Chem.* **329** (1987) 18 – 26.

[186] M. T. C. De Loos-Vollebregt, L. De Galan, *Prog. Anal. At. Spectrom.* **8** (1985) 47 – 81.

[187] S. B. Smith, Jr., G. M. Hieftje, *Appl. Spectrosc.* **37** (1983) 419 – 414.

[188] M. Yamamoto, S. Murayama, M. Ito, M. Yasuda, *Spectrochim. Acta Part B* **35 B** (1980) 43 – 50.

[189] P. Wirz, H. Debus, W. Hanle, A. Scharmann, *Spectrochim. Acta Part B* **37 B** (1982) 1013 – 1020.

[190] G. M. Hermann, *Anal. Chem.* **64** (1992) 571A – 579A.

[191] P. Tschöpel, G. Tölg, *J. Trace Microprobe Tech.* **1** (1982) 1 – 77.

[192] Analytical Spectrometry Updates, *J. Anal. At. Spectrom.*

[193] H. Berndt, E. Jackwerth, *J. Clin. Chem. Clin. Biochem.* **19** (1979) 71.

[194] A. Mizuike: *Enrichment Techniques for Inorganic Trace Analysis,* Springer Verlag, Berlin 1983.

[195] K. S. Subramanian, *Anal. Chem.* **60** (1988) 11 – 15.

[196] J. Szpunar, *Analyst* **125** (2000) 963 – 988.

[197] W. M. R. Dirkx, R. Lobinski, F. C. Adams, *Anal. Sci.* **9** (1993) 273 – 278.

[198] C. E. Oda, J. D. Ingle, *Anal. Chem.* **53** (1981) 2305 – 2309.

[199] W. Gerlach, E. Schweitzer: *Die chemische Emissions-spektralanalyse,* vol. 1, L. Voss, Leipzig 1930.

[200] G. R. Harrison: *Wavelength Tables,* The M.I.T. Press, Cambridge, Mass., 1991.

[201] A. N. Saidel, V. K. Prokofiev, S. M. Raiski: *Spektral-tabellen.* V.E.B. Verlag Technik, Berlin 1955.

[202] A. Gatterer, J. Junkes: *Atlas der Restlinien,* **vols. 3,** Specola Vaticana, Citta del Vaticano, 1945, 1947, 1959.

[203] E. W. Salpeter: *Spektren in der Glimmentladung von 150 bis 400 nm,* vols. 1 – 5, Specola Vaticana, Citta del Vaticano 1973.

[204] R. K. Winge, V. A. Fassel, V. J. Peterson, M. A. Floyd: *ICP Atomic Emission Spectroscopy.* Elsevier, Amsterdam 1985.

[205] P. W. J. M. Boumans: *Line Coincidence Tables for ICP-OES,* vols. 1 and 2, Pergamon Press, Oxford 1984.

[206] B. Huang et al.: *An Atlas of High Resolution Spectra for Rare Earth Elements for ICP – AES.* Royal Society of Chemistry, Cambridge 2000.

[207] G. Scheibe: *Chemische Spektralanalyse, physikalis-che Methoden der analytischen Chemie,* **vol. 1,** 1933.

[208] A. Lomakin, *Z. Anorg. Allg. Chem.* **75** (1930) 187.

[209] *Ullmann,* 4th ed., **5,** 441 – 500.

[210] K. Laqua, W.-D. Hagenah, H. Wächter, *Fresenius' Z. Anal. Chem.* **221** (1967) 142 – 174.

[211] A. Aziz, J. A. C. Broekaert, F. Leis, *Spectrochim. Acta Part B* **36 B** (1981) 251 – 260.

[212] N. W. H. Addink: *DC Arc Analysis,* MacMillan, London 1971.

[213] M. Z. Riemann, *Fresenius' Z. Anal. Chem.* **215** (1966) 407 – 424.

[214] R. Avni in E. L. Grove (ed.): *Applied Atomic Spec-troscopy,* vol. 1, Plenum Press, New York 1978.

[215] K. Slickers: *Die automatische Atom-Emissions-Spek-tralanalyse,* Brühlsche Universitätsdruckerei, Gießen 1992.

[216] K. Slickers: *Automatic Atomic Emission Spectrosco-py.* Brühlsche Universitätsdruckerei, Gießen 1993.

[217] K. H. Koch, *Spectrochim. Acta Part B* **39 B** (1984) 1067 – 1079.

[218] K. Tohyama, J. Ono, M. Onodera, M. Saeki: *Research and Development in Japan,* Okochi Memorial Foundation, 1978, pp. 31 – 35.

[219] D. M. Coleman, M. A. Sainz, H. T. Butler, *Anal. Chem.* **52** (1980) 746 – 753.

[220] A. Montaser: *Inductively Coupled Plasmas in Analyt-ical Atomic Spectrometry,* 2nd ed., Wiley-VCH, New York 1998.

[221] S. J. Hill (ed.): *Inductively Coupled Plasma Spec-trometry and its Applications,* Sheffield Academic Press, Sheffield 1999.

[222] V. V. Korolev, E. E. Vainshtein, *J. Anal. Chem. USSR (Engl. Transl.)* **14** (1959) 658 – 662.

[223] M. Margoshes, B. Scribner, *Spectrochim. Acta* **15** (1959) 138 – 145.

[224] E. Kranz in: *Emissionsspektroskopie.* Akademie-Verlag, Berlin 1964, p. 160.

[225] J. Reednick, *Am. Lab.* 1979, no. 5, 127 – 133.

[226] R. J. Decker, *Spectrochim. Acta Part B* **35 B** (1980) 19 – 35.

[227] F. Leis, J. A. C. Broekaert, H. Waechter, *Fresenius' Z. Anal. Chem.* **333** (1989) 2 – 5.

[228] G. W. Johnson, H. E. Taylor, R. K. Skogerboe, *Spec-trochim. Acta Part B* **34 B** (1979) 197 – 212.

[229] T. B. Reed, *J. Appl. Phys.* **32** (1961) 821 – 824.

[230] S. Greenfield, I. L. Jones, C. T. Berry, *Analyst (London)* **89** (1964) 713 – 720.

[231] R. H. Wendt, V. A. Fassel, *Anal. Chem.* **37** (1965) 920 – 922.

[232] R. Rezaaiyaan et al., *Appl. Spectrosc.* **36** (1982) 626 – 631.

[233] D. J. Kalnicki, V. A. Fassel, R. N. Kalnicky, *Appl. Spectrosc.* **31** (1977) 137 – 150.

[234] B. Raeymaekers, J. A. C. Broekaert, F. Leis, *Spectro-chim. Acta Part B* **43 B** (1988) 941 – 949.

[235] I. Ishii, A. Montaser, *Spectrochim. Acta Part B* **46 B** (1991) 1197 – 1206.

[236] P. W. J. M. Boumans, F. J. de Boer, *Spectrochim. Acta Part B* **32 B** (1977) 365 – 395.

[237] L. De Galan, *Spectrochim. Acta Part B* **39 B** (1984) 537 – 550.

[238] M. Huang, D. S. Hanselmann, P. Y. Yang, G. M. Hieftje, *Spectrochim. Acta Part B* **47 B** (1992) 765 – 786.

[239] J. A. C. Broekaert, F. Leis, *Anal. Chim. Acta* **109** (1979) 73 – 83.

[240] F. H. Walters, L. R. Parker, Jr., S. L. Morgan, S. N. Deming: *Sequential Simplex Optimization,* CRC Press, Inc., Boca Raton 1991.

[241] G. L. Moore, P. J. Humphries-Cuff, A. E. Watson, *Spectrochim. Acta Part B* **39 B** (1984) 915 – 929.

[242] M. L. Parsons, A. R. Forster, D. Anderson: *An Atlas of Spectral Interferences in ICP Spectroscopy,* Plenum Press, New York 1980.

[243] M. R. Cave, O. Butler, J. M. Cook, M. S. Cresser, L. M. Garden, D. L. Miles, *J. Anal. At. Spectrom.* **15** (2000) 181 – 235.

[244] U.S. Environmental Protection Agency: *Methods for the Determination of Metals in Environmental Sam-ples.* EPA600491010, Washington DC 1991.

[245] U.S. Environmental Protection Agency: *Methods for the Determination of Metals in Environmental Sam-ples,* Supplement 1, EPA600R94111, Washington DC 1994.

[246] B. Fairman, M. Hinds, S. M. Nelms, D. M. Penny, P. Goodall, *J. Anal. At. Spectrom.* **14** (1999) 1937 – 1969.

[247] R. E. Russo, X. L. Mao, O. V. Borisov, *Trends Anal. Chem.* **17** (1998) 461.

[248] W. B. Barnett, V. A. Fassel, R. N. Kniseley, *Spectro-chim. Acta Part B* **23 B** (1968) 643 – 664.

[249] S. A. Meyers, D. H. Tracy, *Spectrochim. Acta Part B* **38 B** (1983) 1227 – 1253.

[250] P. W. J. M. Boumans, M. C. Lux-Steiner, *Spectro-chim. Acta Part B* **37 B** (1982) 97 – 126.

[251] J. A. C. Broekaert, F. Leis, K. Laqua, *Talanta* **28** (1981) 745 – 752.

[252] K. Vermeiren, C. Vandecasteele, R. Dams, *Analyst (London)* **115** (1990) 17 – 22.

[253] K. Dittrich et al., *J. Anal. At. Spectrom.* **3** (1993) 1105 – 1110.

[254] A. G. Cox, L. G. Cook, C. W. McLeod, *Analyst (London)* **110** (1985) 331 – 333.

[255] H. G. C. Human, R. H. Scott, A. R. Oakes, C. D. West, *Analyst (London)* **101** (1976) 265 – 271.

[256] A. Lemarchand, G. Labarraque, P. Masson, J. A. C. Broekaert, *J. Anal. At. Spectrom.* **2** (1987) 481–484.

[257] T. Ishizuka, Y. Uwamino, *Spectrochim. Acta Part B* **38 B** (1983) 519–527.

[258] I. B. Brenner et al., *Spectrochim. Acta Part B* **36 B** (1981) 785–797.

[259] J. R. Garbarino, M. E. Taylor, *Appl. Spectrosc.* **33** (1979) 167–175.

[260] U.S. Environmental Protection Agency: Inductively Coupled Plasma-Atomic Emission Spectrometric Method for Trace Element Analysis of Water and Wastes, Washington, DC, 1979.

[261] DIN 38406,*Bestimmung der 24 Elemente Ag, Al, B, Ba, Ca, Cd, Co, Cr, Cu, Fe, K, Mg, Mn, Mo, Na, Ni, P, Pb, Sb, Sr, Ti, V, Zn und Zr durch Atomemissionsspektrometrie mit induktiv gekoppeltem Plasma (ICP – AES) (E 22), Beuth Verlag, Berlin* 1987.

[262] M. Thompson, B. Pahlavanpour, S. J. Walton, G. F. Kirkbright, *Analyst (London)* **103** (1978) 705–713.

[263] A. Miyazaki, A. Kimura, K. Bansho, Y. Umezaki, *Anal. Chim. Acta* **144** (1982) 213–221.

[264] H. Berndt, U. Harms, M. Sonneborn, *Fresenius' Z. Anal. Chem.* **322** (1985) 329–333.

[265] M. Hiraide et al., *Anal. Chem.* **52** (1980) 804–807.

[266] J. A. C. Broekaert, S. Gucer, F. Adams: *Metal Speciation in the Environment,* Springer Verlag, Berlin 1990.

[267] R. K. Skogerboe, G. N. Coleman, *Anal. Chem.* **48** (1976) 611 A–622 A.

[268] A. E. Croslyn, B. W. Smith, J. D. Winefordner, *Crit. Rev. Anal. Chem.* **27** (1997) 199–255.

[269] R. Mavrodineanu, R. C. Hughes, *Spectrochim. Acta* **19** (1963) 1309–1307.

[270] A. Disam, P. Tschöpel, G. Tölg, *Fresenius' Z. Anal. Chem.* **310** (1982) 131–143.

[271] S. Hanamura, B. W. Smith, J. D. Winefordner, *Anal. Chem.* **55** (1983) 2026–2032.

[272] C. I. M. Beenakker, *Spectrochim. Acta Part B* **32 B** (1977) 173–178.

[273] D. Kollotzek, P. Tschöpel, G. Tölg, *Spectrochim. Acta Part B* **37 B** (1982) 91–96.

[274] G. Heltai, J. A. C. Broekaert, F. Leis, G. Tölg, *Spectrochim. Acta Part B* **45 B** (1990) 301–311.

[275] P. S. Moussanda, P. Ranson, J. M. Mermet, *Spectrochim. Acta Part B* **40 B** (1985) 641–651.

[276] D. Kollotzek et al., *Fresenius' Z. Anal. Chem.* **318** (1984) 485–489.

[277] P. Uden: *Element-Specific Chromatographic Detection by Atomic Emission Spectroscopy,* American Chemical Society, Washington 1992.

[278] J. Hubert, M. Moisan, A. Ricard, *Spectrochim. Acta Part B* **33 B** (1979) 1–10.

[279] U. Richts, J. A. C. Broekaert, P. Tschöpel, G. Tölg, *Talanta* **38** (1991) 863–869.

[280] Q. Jin, C. Zhu, M. W. Borer, G. M. Hieftje, *Spectrochim. Acta Part B* **46 B** (1991) 417–430.

[281] R. M. Barnes, E. E. Reszke, *Anal. Chem.* **62** (1990) 2650.

[282] K. A. Forbes, E. E. Reszke, P. C. Uden, R. M. Barnes, *J. Anal. At. Spectrom.* **6** (1991) 57.

[283] R. K. Marcus (ed.): *Glow Discharge Spectroscopies.* Plenum Press, New York 1993.

[284] K. I. Zil'bershtein: *Spectrochemical Analysis of Pure Substances,* A. Hilger, Bristol 1977, pp. 173–227.

[285] B. Thelin, *Appl. Spectrosc.* **35** (1981) 302–307.

[286] S. Caroli: *Improved Hollow Cathode Lamps for Atomic Spectroscopy.* Ellis Horwood, Chichester 1985.

[287] H. Falk, E. Hoffmann, C. Lüdke, *Spectrochim. Acta Part B* **39 B** (1984) 283–294.

[288] W. Grimm, *Spectrochim. Acta Part B* **23 B** (1968) 443–454.

[289] J. B. Ko, *Spectrochim. Acta Part B* **39 B** (1984) 1405–1423.

[290] F. Leis, J. A. C. Broekaert, K. Laqua, *Spectrochim. Acta Part B* **42 B** (1987) 1169–1176.

[291] A. Bengtson, *Spectrochim. Acta Part B* **40 B** (1985) 631–639.

[292] R. M. Lowe, *Spectrochim. Acta Part B* **31 B** (1978) 257–261.

[293] N. P. Ferreira, J. A. Strauss, H. G. C. Human, *Spectrochim. Acta Part B* **38 B** (1983) 899–911.

[294] J. A. C. Broekaert, R. Pereiro, T. K. Starn, G. M. Hieftje, *Spectrochim. Acta Part B* **48 B** (1983) 1207–1220.

[295] M. R. Winchester, C. Lazik, R. K. Marcus, *Spectrochim. Acta Part B* **46 B** (1991) 483–499.

[296] T. M. Sugden in R. L. Reed (ed.): *Mass Spectrometry,* Academic Press, New York 1965, p. 347.

[297] R. S. Houk et al., *Anal. Chem.* **52** (1980) 2283–2289.

[298] A. Montaser (ed.): *Inductively Coupled Plasma Mass Spectrometry.* Wiley-VCH, New York 1998.

[299] E. H. Evans, J. J. Giglio, T. Castillano, J. A. Caruso: *Inductively Coupled and Microwave Induced Plasmas for Mass Spectrometry,* Royal Society of Chemistry, Cambridge 1995.

[300] H. Niu, R. S. Houk, *Spectrochim. Acta Part B* **51 B** (1996) 119–815.

[301] E. H. Evans, J. J. Giglio, *J. Anal. At. Spectrom.* **8** (1993) 1–8.

[302] A. L. Gray, R. S. Houk, J. G. Williams, *J. Anal. At. Spectrom.* **2** (1987) 13–20.

[303] M. A. Vaughan, G. Horlick, *Appl. Spectrosc.* **40** (1986) 434–445.

[304] G. Horlick, S. H. Tan, M. A. Vaughan, C. A. Rose, *Spectrochim. Acta Part B* **40 B** (1985) 1555–1572.

[305] N. Jakubowski, B. J. Raeymaekers, J. A. C. Broekaert, D. Stüwer, *Spectrochim. Acta Part B* **44 B** (1989) 219–228.

[306] C. Vandecasteele, M. Nagels, H. Vanhoe, R. Dams, *Anal. Chim. Acta* **211** (1988) 91–98.

[307] J. S. Crain, R. S. Houk, F. G. Smith, *Spectrochim. Acta Part B* **43 B** (1988) 1355–1364.

[308] D. Hausler, *Spectrochim. Acta Part B* **42 B** (1987) 63–73.

[309] N. Jakubowski, I. Feldman, D. Stüwer, H. Berndt, *Spectrochim. Acta Part B* **47 B** (1992) 119–129.

[310] X. Wang, M. Viczian, A. Lasztity, R. M. Barnes, *J. Anal. At. Spectrom.* **3** (1988) 821–828.

[311] C. J. Park, J. C. Van Loon, P. Arrowsmith, J. B. French, *Anal. Chem.* **59** (1987) 2191–2196.

[312] D. Boomer, M. Powell, R. L. A. Sing, E. D. Salin, *Anal. Chem.* **58** (1986) 975–976.

[313] G. E. M. Hall, J. C. Pelchat, D. W. Boomer, M. Powell, *J. Anal. At. Spectrom.* **3** (1988) 791–797.

[314] T. E. Jeffries, S. E. Jackson, H. P. Longerich, *J. Anal. At. Spectrom.* **13** (1998) 935–940.

[315] E. K. Shibuya, J. E. S. Sarkis, J. Enzweiler, A. P. S. Jorge, A. M. G. Figueiredo, *J. Anal. At. Spectrom.* **13** (1998) 941–944.

[316] J. G. Williams, A. L. Gray, P. Norman, L. Ebdon, *J. Anal. At. Spectrom.* **2** (1987) 469–472.

[317] N. Jakubowski, I. Feldman, B. Sack, D. Stüwer, *J. Anal. At. Spectrom.* **7** (1992) 121 – 125.

[318] P. Arrowsmith, *Anal. Chem.* **59** (1987) 1437 – 1444.

[319] D. L. Miles: "Geological Applications of Plasma Spectrometry," in S. J. Hill (ed.): *Inductively Coupled Plasmas in Analytical Atomic Spectrometry.* 2nd ed., Wiley-VCH, New York 1999.

[320] F. Vanhaecke, L. Moens, P. Taylor: "Use of ICP – MS for Isotope Ratio Measurements," in S. J. Hill (ed.): *Inductively Coupled Plasmas in Analytical Atomic Spectrometry,* 2nd ed., Wiley-VCH, New York 1999.

[321] J. B. Truscott, L. Bromley, P. J. Jones, E. H. Evans, J. Turner, B. Fairman, *J. Anal. At. Spectrom.* **14** (1999) 627 – 631.

[322] I.M. J. Bloxham, S. J. Hill, P. J. Worsfold, *J. Anal. At. Spectrom.* **9** (1994) 935 – 938.

[323] W. W. Harrison, B. L. Bentz, *Prog. Anal. At. Spectrosc.* **11** (1988) 53 – 110.

[324] D. J. Hall, K. Robinson, *Am. Lab.* 1987, no. 8, 14.

[325] N. Jakubowski, D. Stüwer, G. Tölg, *Int. J. Mass Spectrom. Ion. Processes* **71** (1986) 183 – 197.

[326] P. M. Charalambous, *Steel Res.* **5** (1987) 197.

[327] W. Vieth, J. C. Huneke, *Spectrochim. Acta Part B* **46 B** (1991) 137 – 154.

[328] N. Jakubowski, D. Stüwer, *Fresenius' Z. Anal. Chem.* **335** (1988) 680 – 686.

[329] A. Bogaerts, R. Gijbels, *Spectrochim. Acta Part B* **53** (1998) 1.

[330] VG Instruments, Technical Information GD 012, Manchester, U.K.

[331] VG Instruments, Application Note 02.681, Manchester, U.K.

[332] VG Instruments, Technical Information GD 701, Manchester, U.K.

[333] N. Jakubowski, D. Stüwer, W. Vieth, *Anal. Chem.* **59** (1987) 1825 – 1830.

[334] N. Jakubowski, D. Stüwer, W. Vieth, *Fresenius' Z. Anal. Chem.* **331** (1988) 145 – 149.

[335] M. Hecq, A. Hecq, M. Fontignies, *Anal. Chim. Acta* **155** (1983) 191 – 198.

[336] W. A. Mattson, B. L. Bentz, W. W. Harrison, *Anal. Chem.* **48** (1976) 489 – 491.

[337] G. O. Foss, H. J. Svec, R. J. Conzemius, *Anal. Chim. Acta* **147** (1983) 151 – 162.

[338] N. Jakubowski, D. Stüwer, G. Tölg, *Spectrochim. Acta Part B* **46 B** (1991) 155 – 163.

[339] N. Omenetto, J. D. Winfordner, *Prog. Anal. At. Spectrom.* **2** (1979) 1 – 183.

[340] N. Omenetto, H. G. C. Human, P. Cavalli, G. Rossi, *Spectrochim. Acta Part B* **39 B** (1984) 115 – 117.

[341] V. A. Bolshov, A. V. Zybin, II. Smirenkins, *Spectrochim. Acta Part B* **36 B** (1981) 1143 – 1152.

[342] N. Omenetto et al., *Spectrochim. Acta Part B* **40 B** (1985) 1411 – 1422.

[343] B. W. Smith, P. B. Farnsworth, P. Cavalli, N. Omenetto, *Spectrochim. Acta Part B* **45 B** (1990) 1369 – 1373.

[344] S. Grazhulene, V. Khvostikov, M. Sorokin, *Spectrochim. Acta Part B* **46 B** (1991) 459 – 465.

[345] M. A. Bolshov, C. F. Boutron, A. V. Zybin, *Anal. Chem.* **61** (1989) 1758 – 1762.

[346] W. Sdorra, K. Niemax, *Spectrochim. Acta Part B* **45 B** (1990) 917 – 926.

[347] G. C. Turk, J. C. Travis, J. R. DeVoe, *Anal. Chem.* **51** (1979) 1890 – 1896.

[348] R. B. Green, M. D. Seltzer: "Laser Induced Ionization Spectrometry," in J. Sneddon (ed.): *Advances in Atomic Spectroscopy,* vol. I, JAI Press, London 1992.

[349] G. J. Havrilla, S. J. Weeks, J. C. Travis, *Anal. Chem.* **54** (1982) 2566 – 2570.

[350] K. Niemax, *Appl. Phys. B* **B 38** (1985) 147 – 157.

[351] N. Omenetto, T. Berthoud, P. Cavalli, G. Rossi, *Anal. Chem.* **57** (1985) 1256 – 1261.

[352] G. C. Turk, M. DeMing, in W. F. Koch (ed.): *NBS Special Publication 260 – 106.* US Government Printing Office Washington DC 1986, pp. 30 – 33.

[353] J. P. Young, R. W. Shaw, D. H. Smith, *Anal. Chem.* **61** (1989) 1271 A – 1279 A.

[354] R. Klockenkämper, J. Knoth, A. Prange, H. Schwenke, *Anal. Chem.* **64** (1992) 1115 A – 1123 A.

22. Laser Analytical Spectroscopy

MICHAEL A. BOLSHOV, Institute of Spectroscopy, Russian Academy of Sciences, Troitzk, Russia; Institute of Spectrochemistry and Applied Spectroscopy, Dortmund, Germany

YURI A. KURITSYN, Institute of Spectroscopy, Russian Academy of Sciences, Troitzk, Russia

22.1. Introduction

Modern technologies in electronics, in high-purity materials and chemical reagents, in toxicology, biology, and environmental control and protection need extremely sensitive analytical techniques. Impurity concentrations at and below pg/mL (pg/g) levels are state-of-the-art for many analyses. In many cases the sample volume is very small; if an impurity concentration of the order of ng/g has to be detected in a sample of mass 1 µg, the analytical technique should have an absolute detection limit of ca. 1 fg (10^{-15} g). Even techniques such as atomic absorption spectrometry (AAS), optical emission spectrometry with inductively coupled plasma (ICP-OES), neutron activation spectroscopy (NAS), and X-ray fluorescence spectroscopy (XRFS) can not detect directly trace
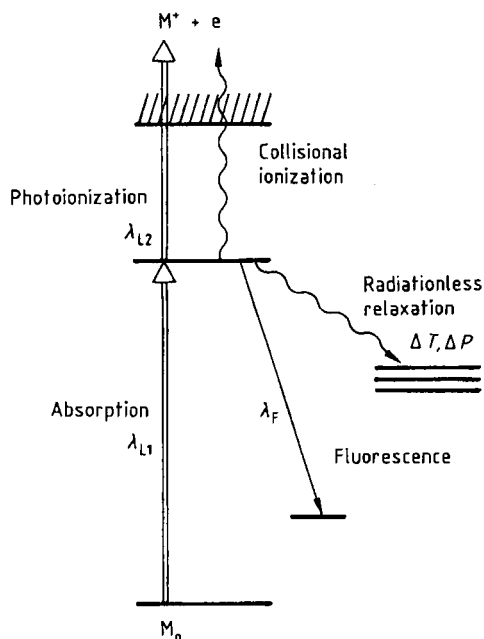
Figure 1. The basic laser analytical techniques

elements at such low levels of concentration and/or absolute mass. In most cases, these techniques need elaborate chemical pretreatment of samples. Contamination or loss of analyte are very probable during any chemical procedure at such low concentration.

During the last 30 years the development of laser technology has opened up new possibilities in analytical spectroscopy, both by improvements in traditional analytical techniques, such as absorption and fluorescence or optoacoustic spectroscopy, and by the introduction of new techniques such as multistep photoionization spectroscopy, thermal lens spectroscopy, and site-selection spectroscopy.

22.1.1. Properties of Laser Light

22.1.1.1. Tunability

Practically all laser analytical techniques are based on systems with tunable wavelength. By using different types of lasers and the methods of nonlinear optics, one can obtain intense tunable laser radiation over the spectral range of $0.2 - 40\ \mu m$ and in some regions of the vacuum ultraviolet and far infrared. Almost any strong

atomic or molecular transition is thus accessible to laser excitation.

22.1.1.2. Intensity

The intensities of lasers exceed by many orders of magnitude those of conventional light sources. At high intensity levels, nonlinear interaction of light with atomic and molecular systems becomes pronounced. New analytical techniques involving multistep and multiphoton excitation and ionization, optical saturation, and excitation of forbidden transitions become possible. High radiation intensity also increases the sensitivity of laser analytical techniques.

22.1.1.3. Time Resolution

Pulsed lasers with extremely short pulse durations (down to 10^{-15} s) are now available, enabling investigations of chemical reaction kinetics, energy and electron transfer in complex excited molecules, etc. The efficiency of excitation of higher states increases because the duration of the laser pulse is often shorter than the decay time of the intermediate state of atoms or molecules.

22.1.1.4. Temporal Coherence

The spectral linewidth of modern lasers is often less than the homogeneous linewidth of a given atom or molecule. This results in very high spectral intensity and thus very efficient excitation. It also allows highly selective excitation of a particular atom or molecule in a complex mixture. High selectivity is an important advantage of laser analytical techniques.

22.1.1.5. Spatial Coherence

The ability to focus a laser beam to a very small spot (a few μm diameter) provides high spatial resolution in sample analysis. High spatial coherence also enables collimated laser beams to be used for remote sensing.

22.1.2. Types of Laser Analytical Techniques

Laser analytical techniques can be classified on the basis of the physical process caused by the interaction of laser radiation with matter (Fig. 1).

The primary process is absorption of photons by atoms or molecules. If the absorption lines are strong and the number density of the absorbing species is sufficient, the difference between incident and transmitted energies may be measurable for a single pass of the laser beam through the sample. In this case the signal can be measured by a simple light detector, such as a photodiode. For weaker absorption lines and lower number densities of the absorbing species multipass schemes are used.

Another set of analytical techniques (optocalorimetric methods) is based on indirect measurement of the energy absorbed by the sample. Absorbed energy is transformed into kinetic energy of the sample particles followed by an increase in the local temperature of the sample. The resultant changes in pressure and refractive index can be measured by optoacoustic methods or by laser scattering techniques.

The absorption of laser photons leads to excitation of atoms or molecules. In many cases, radiative decay of the excited state is very efficient, i.e., quantum yields of luminescence or fluorescence are very high. Measurement of the intensity of emitted photons as a function of analyte concentration is a widespread analytical method with conventional light sources. The use of intense laser sources greatly increased the sensitivity and selectivity of traditional luminescence methods.

Excitation of atoms or molecules may be followed by collisional or multistep (multiphoton) ionization of the excited species. The charged particles so formed may be recorded by conventional methods. Several techniques involve a combination of optical excitation of a neutral particle and its subsequent ionization.

22.1.3. Peculiarities of Resonance Laser Radiation Absorption

The interaction of intense laser radiation with matter differs from that of conventional radiation sources. The high incident irradiances available with lasers can lead to a significant depletion of the initial level population. As a result, the optical absorption coefficient decreases as a function of the excitation energy and the absorbed energy tends towards a constant value. This so called optical saturation effect leads to a nonlinear dependence of the absorption signal on the light intensity. Because the amount of energy released by radiative and nonradiative relaxation processes

depends on the amount of energy absorbed, nonlinear intensity-dependent signals are also obtained in other spectrometric techniques, such as fluorescence, optoacoustic, thermal lens, etc.

In the simplest case for the interaction of a two-level system with cw (continuous wave) laser radiation (steady-state mode), the number of transitions per unit time from level 1 to level 2 (absorption) is equal to the number of transitions from level 2 to level 1 (stimulated and spontaneous emission):

$$\sigma(N_1 - N_2)I/h\nu = N_2/\tau$$

where σ_0 is the cross-section of the radiation (stimulated) transition, N_1 and N_2 are the number densities of particles at the levels 1 and 2, respectively, I is the intensity of laser radiation, $h\nu$ is the energy of the resonance photon, τ is the radiative decay time of upper state 2 (assuming that nonradiative transitions due to collisions are negligible). Both N_1 and N_2 are dependent on laser intensity, so that

$$N_2 = \frac{1}{2} N \frac{\dfrac{I}{I_{sat}}}{1 + \dfrac{I}{I_{sat}}}, \quad N_1 = N - N_2$$

where N is the total number density of particles, $I_{sat} = h\nu/2\sigma_0\tau$ is the saturation intensity, defined as the intensity for which the number density of the excited particles N_2 reaches one half of its maximum value $1/2\,N$.

If the intensity of laser radiation is much lower than I_{sat}, the absorption energy and the number density of the excited particles N_2 are directly proportional to the light intensity, and hence the analytical signal for different techniques (absorption, optocalorimetric, fluorescence, etc.) has a linear dependence on the intensity of the radiation. This linear regime is typical for conventional radiation sources used in elemental analysis: hollow cathode lamps (HCLs), electrodeless discharge lamps (EDLs) and for low intensity lasers.

In the opposite case, when the intensity of the resonance laser radiation is high enough ($I \gg I_{sat}$), the absorbed laser energy reaches its maximal value and the population of the upper level 2 approaches $1/2\,N$. The analytical signal for the absorption technique

$$\Delta I \approx \sigma_0(N_1 - N_2)Il = \frac{\sigma_0 N I l}{1 + I/I_{sat}}$$

(where l is the absorption pathlength) tends towards a constant value. Because the shot noise of radiation is directly proportional to $\sqrt{I}$, then optimum detection of the absorbing particles is achieved at $I = I_{sat}$.

If deep saturation is achieved ($I \gg I_{sat}$), two very important advantages of different laser analytical techniques (fluorescence, optocalorimetric, ionization) are gained: Maximum analytical signal and minimal influence of intensity fluctuations of the excitation source on the precision of the analysis. The ultimate sensitivity down to single atom/ molecule detection has been demonstrated under saturation conditions in many analytical applications with lasers.

In real analytes the scheme of the energy levels is rather complicated. Thus, there may be many channels of radiative and collisional transfer of the absorbed laser energy. The complex structure of the energy levels scheme offers different combinations of excitation and detection of a particular species, but nevertheless the simplest scheme described above gives correct qualitative presentation of the analytical signal behavior.

Details on the optimization of laser parameters for different analytical techniques can be found in [1], [2], and [3].

22.2. Tunable Lasers

In most analytical applications, tunable lasers are used. Tunability allows selective excitation of a particular analyte in a complex mixture and relatively simple "switching" from one analyte to another. For conventional sources with fixed spectral lines it is often necessary to use different lamps for different analytes.

For a few laser analytical techniques (molecular fluorimetry of liquids, spontaneous Raman scattering), fixed wavelength lasers are used. The following discussion is limited to the two most widely used types of tunable laser.

22.2.1. Dye Lasers [4]

For a long time the dye lasers were the most popular type for atomic and molecular analysis. The active medium of these lasers is an organic dye in an appropriate solvent. About 20 different dyes cover the spectral range from ca. 320 to ca. 900 nm.

Typically, dye lasers are pumped by an intense fixed-frequency laser, the most often used being nitrogen (337 nm), argon ion (488.0, 514.5 nm), XeCl excimer (308 nm), and frequency-doubled and frequency-tripled Nd-YAG (530, 353.3 nm).

Dye molecules have broad emission bands. The optical resonator of a dye laser includes a spectrally selective element (grating, prism, interferometer, Lyot filter, etc.) which narrows the dye laser spectral line. By appropriate adjustment (e.g., rotation of the grating) one can smoothly tune the laser wavelength within the luminescence band of a particular dye. The tuning range may be 30–40 nm. At the extremes of the tuning range the energy of a dye laser falls off.

The organic dyes used for lasers can be classified on the basis of molecular structure: terphenyls (spectral range: 312–400 nm); coumarines (420–580 nm); rhodamines (570–720 nm), oxazines (680–860 nm). Each class consists of several dyes with slightly modified chemical structure and, hence, slightly shifted absorption and luminescence bands. There are special dyes for UV and near IR.

The most utilized solvents are ethanol, methanol, cyclohexane, ethylene glycol, and dioxane. For different solvents, the tuning curves of the same dye may be shifted by 10–15 nm.

The dyes for the UV spectral range (below 380–400 nm) are not photochemically stable. In most analytical applications, intense UV radiation below 360 nm is obtained by nonlinear optics, i.e., the generation of harmonics and sum frequencies of the visible dye laser radiation in nonlinear crystals. The most widely employed crystals are KDP, ADP, LiF, LiIO$_3$, and BBO. For example, the second harmonic of a rhodamine dye laser (fundamental band near 640 nm) is a more practical way to obtain tunable UV radiation near 320 nm rather than direct lasing of an unstable UV dye laser.

The efficiency of energy conversion of the pumping laser radiation in commercially available lasers varies from a few percent to ca. 20%, depending on the type of dye and the spectral range.

22.2.2. Semiconductor Diode Lasers [5]

In the mid-1960s, tunable semiconductor diode lasers (DLs) were introduced into spectroscopic and analytical research. In conventional DLs the laser action can occur when a forward bias current is applied to a $p-n$ junction. Photons are generated in the process of electron–hole recombina-

tion. The wavelength of laser radiation is determined by the energy gap between conduction and valence bands, which, in turn, is determined by the chemical composition of the semiconductor. Wavelength tuning is possible by varying the injection current and by the temperature of the laser.

At first, DLs based on the IV – VI (Pb-salt) materials were used for analytical applications. These lasers operate in the $3 - 30 \, \mu m$ region, where the fundamental absorption bands of most molecules lie. Linewidths as small as $2 \times 10^{-6} \, cm^{-1}$ were observed for the DLs, but in the commercial systems used for analytical applications the DL linewidth is determined by the quality of temperature stabilization. Usually the temperature is stabilized at a level of $10^{-2} - 10^{-3} \, K$, which corresponds to a linewidth of $\sim 10^{-3} \, cm^{-1}$. This width is much smaller than the widths of molecular absorption lines at atmospheric pressure and even smaller than Doppler broadening for light molecules. The IV – VI DLs operate at cryogenic temperatures. In laboratory high-resolution spectrometers operating temperatures of $10 - 120 \, K$ are usually provided by closed-cycle Stirling microrefrigerators. Now lasers operating above 77 K are available at most wavelengths, and in analytical instruments temperature-controlled liquid-N_2 Dewars are the cooling method of choice. Typical emitting powers are $100 - 300 \, \mu W$, and cw lasers with powers of 1 mW can also be fabricated.

Recently, the characteristics of DLs operating in the visible/near-IR (NIR) region have been greatly improved. High quality lasers, made from the III-V group of semiconductor materials, are now commercially available in a number of discrete spectral ranges between 630 nm and 2.0 μm. In principle, lasers can be made at any wavelength, but because of the market demand the DL are commercialized mainly for specific spectral ranges. These lasers can work in cw mode at temperatures accessible with thermoelectric coolers, generally – 40 to 60 °C. The typical linewidth of a free-running DL is $20 - 100 \, MHz$. Commercially available edge-emitting Fabry – Perot (FP) DLs yield singe-mode radiation up to 150 mW in the $670 - 870 \, nm$ spectral range and up to 30 mW in the $630 - 670 \, nm$ range. DLs, working in the $630 - 780 \, nm$ region, have been shown to be very attractive for atomic absorption analyzers.

Compared with other lasers, the output radiation from DLs is highly divergent. The emission angles are typically 10° and 30° parallel and perpendicular to the laser junction, respectively, and may vary from sample to sample. These beam features can be corrected with appropriate collimating and anamorphic beam-conditioning optics.

The drawbacks of the FP-DLs are the frequency drift and mode jumping, therefore other types of lasers such as distributed feedback (DFB), distributed Bragg reflector (DBR), external cavity (ECDL), and vertical cavity surface emitting lasers (VCSELs) have been developed. In a typical DFB laser, an index grating is fabricated directly on a chip alongside the active layer, providing distributed feedback only for a selected wavelength specified by the grating. A DBR laser uses the same principle, but the grating is placed beyond the active region. DFB/DBR lasers provide single-mode operation, can be continuously tunable by current and temperature, and exhibit no mode hops. The drawbacks are limited tuning range and higher cost compared with simple FP lasers.

Wavelengths longer than 1 μm are attractive for molecular analysis. Commercially available DLs for this region include: InGaAsP/InP (1.3, 1.55 μm FP structures, 1.2 – 1.8 μm DFB and EC lasers). For wavelengths longer than 1.8 μm, DLs are made from compounds containing antimonides such as AlGaAsSb, InGaAsSb, and InAsSb. Lasing in the $2 - 2.4 \, \mu m$ region at room temperature is available, and the spectral range up to 3.7 μm at liquid N_2 is achieved.

Vertical cavity surface emitting lasers (VCSELs), in contrast to conventional FP lasers, emit from their top surface. A conventional VCSEL consists of a multi-quantum-well (MQW) active region placed in a cavity between two DBR mirrors. These mirrors control the wavelength and beam divergence. For typical VCSELs, the far field beam characteristic will be a near Gaussian shape, with a beam divergence less than 12°. The narrow beam divergence, circular symmetry, and lack of astigmatism greatly simplify the optical design with these type of DLs. Currently, VCSELs operating in cw mode at room temperature in the visible/NIR range up to 2.3 μm can be produced.

An alternative approach for tunable single-mode operation is to build an external cavity. In this case, a FP laser is placed into an optical cavity, and the wavelength is tuned by using a dispersive element, such as a grating. The wavelength can be tuned continuously over many wavenumbers $(20 - 50 \, cm^{-1})$ by rotating the grating. The linewidth depends on the construction of the external cavity and its stability and is usually less than 1 MHz.

To obtain generations in the regions of mode-hops for conventional FP lasers a simple design with a small external cavity was proposed. A glass plate mounted in front of the diode chip acts as the low-reflectivity mirror of an extended resonator and helps to select particular longitudinal modes within gain profile, even where laser action is not possible without the feedback.

Recently developed multi-electrode lasers are very promising devices for multielement sensors. These lasers separate the wavelength selection, gain, and phase control regions of the laser cavity and are available in a number of different configurations. A tuning range over 100 nm around 1.5 µm has been demonstrated. These tuning ranges are equivalent to EC devices, thereby simplifying present multi-laser sensor concepts by allowing comparable wavelength coverage with a single laser.

Frequently conversion is used to extend the available spectral range of DLs. Second harmonic generation (SHG) in bulk non-linear crystals, such as $LiIO_3$ and $KNbO_3$, can provide up to 0.1 µW of radiation in the 335–410 nm range and up to 1–3 µW in the 410–430 nm range. The generation of sum frequencies (SFG) in non-linear crystals can open the way to tunable radiation in the UV/visible part of the spectrum. Over the past several years, progress in the development of quasi-phase matching structures (QPMS) and waveguides, based on non-linear crystals, has been achieved. Conversion efficiencies in both types of such non-linear converters are much higher than in bulk crystals, thus single-mode powers up to some mW in the UV/blue range can be produced. Mid-IR radiation can be also produced by difference-frequency mixing of two NIR lasers. Difference-frequency generation in periodically poled $LiNbO_3$ and bulk $AgGaSe_2$ or $AgGaS_2$ has been demonstrated for generating tunable radiation between 3.2 and 8.7 µm.

In the mid-IR, very promising substitutes to lead-salt lasers are quantum-cascade lasers (QCLs), made using III-V compounds. In QCL the active region consists of several gain quantum well (QW) layers separated by the digitally graded electron-injecting regions. Electrons are injected into the high energy level of the first gain layer, fall down to the lower level with photon emission, then through the digitally graded region are injected into the high level of the next gain layer and the process repeats. Quantum efficiencies higher than 100 % are possible in such structures. It is important that such cascades can be organized

between quantized states in the conduction band only. The close spacing of these quantized states allows light to be emitted in the mid-IR range. Currently, room temperature lasing up to 8.2 µm in structures fabricated from AlInAs/InGaAs can be obtained. QCLs with wavelength as long as 11 µm at cryogenic temperatures have been demonstrated.

In 1999 blue DLs became commercially available. InGaN MQW DLs were demonstrated to have a lifetime of more than 10 000 h under conditions of room temperature cw operation. Single-mode emissions were produced in the region of 390–450 nm. Maximum output power is about 30 mW. With an external cavity the narrowband 1 MHz laser linewidth can be coarsely adjusted over a tuning range of 2.5 nm and then fine-tuned by a piezo stack over a more than 20 GHz range without mode hops.

22.3. Laser Techniques for Elemental Analysis

In this chapter we shall discuss laser analytical techniques for elemental analysis, i.e., for detection of impurity atoms of a particular element in a sample.

22.3.1. Laser-Excited Atomic Fluorescence Spectrometry [2], [3], [6]

22.3.1.1. Physical Principles

The principles of the laser-excited atomic fluorescence (LEAF) technique are very simple. A liquid or solid sample is atomized in an appropriate device. The atomic vapor is illuminated by laser radiation tuned to a strong resonance transition of an analyte atom. The excited analyte atoms spontaneously radiate fluorescence photons and a recording system registers the intensity of fluorescence (or total number of fluorescent photons). The extremely high spectral brightness of lasers makes it possible to saturate a resonance transition of an analyte atom. Therefore, the maximum fluorescence intensity of the free analyte atoms can be achieved while the effect of intensity fluctuations of the excitation source are minimized. Both factors provide the main advantage of LEAF— extremely high sensitivity. The best absolute detection limits achieved in direct analysis by LEAF
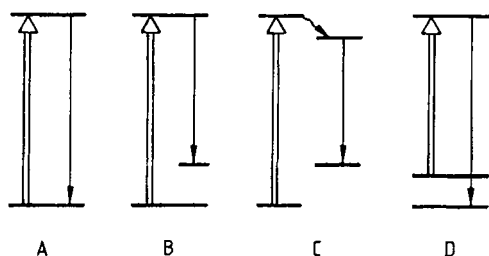
Figure 2. Classification of LEAF schemes
A) Resonance fluorescence; B) Direct-line fluorescence;
C) Collisionally assisted direct-line fluorescence; D) Anti-
Stokes direct-line fluorescence

lie in the femtogram-range, which corresponds to concentrations down to pg/mL.

Several different excitation–recording schemes used in LEAF are shown in Figure 2. The choice of a particular scheme is dictated by the structure of the atomic energy levels, the sources of background and noise, and spectral interference with the major sample components.

The resonance scheme (Fig. 2 A) has the great advantage of a possible cyclic interaction of analyte atoms with the laser radiation and, hence, a high yield of fluorescence photons. The highest sensitivity of LEAF, the detection of a single free atom, has been demonstrated with Na atomic beams and cw laser excitation. This scheme has a serious disadvantage—a high level of scattered laser light and, hence, high background. For a few elements (Na, Cd, etc.) the resonance scheme is the only possible one. For the most elements this scheme is not optimal in real analytical conditions when pulsed dye lasers are used for excitation.

The Stokes direct-line or collisionally assisted Stokes direct-line schemes (Fig. 2 B, C) are often employed (Tl, Pb, Au, Ag, Co, etc.). Their main advantage is that different wavelengths are used for excitation and fluorescence. Because of this, the laser scattered light can be easily separated from the weak fluorescence radiation by cut-off optical filters or monochromators. The analytical potential of collisionally assisted detection schemes depends strongly on the efficiency of collisional mixing between the neighboring levels. Typically this efficiency is high enough for energy gaps between the levels of ca. 1000 cm^{-1}. For some elements with low-lying excited states (Ga-type) the anti-Stokes direct-line scheme (Fig. 2 D) is very useful. An advantage of this scheme is a blue shift of the fluorescence and an appropriate reduction in background, caused by scattered laser

light. In some cases, double-resonance excitation by two lasers with different wavelengths is used.

The possibility to select the optimal excitation–detection scheme often simplifies the problem of spectral interference with the major components, providing the second advantage of LEAF—high selectivity.

22.3.1.2. Atomizers

Graphite electrothermal atomizers (GETA), flames, glow discharges, and laser plumes are the most widely used types of atomizers in modern LEAF analytical practice.

GETA. Graphite electrothermal atomizers are of two types: open and closed. The main unit is a graphite crucible (cup, profiled rod, boat, furnace) held between massive water-cooled electrodes. Typical sample volumes are 2–100 µL. The maximum temperature of a crucible depends on its mass, the electrical power supply, and the type of graphite used. In most applications, it is 2700–2900 °C. In open and closed atomizers an inert gas atmosphere is used to minimize quenching of the excited analyte atoms and sublimation of the hot graphite crucible. Graphite furnaces have proved preferable because of the high temperature of the analytical zone and, hence, less chemical interference. Low-pressure noble gas in a closed GETA has been used to minimize matrix interference.

Flames. The flames used in LEAF are very similar to those used in AAS. In most applications, cylindrically symmetric burners are used since the dimensions of the analytical zone in LEAF do not exceed 10 mm. Quenching of the excited atoms by components of the flame is a serious problem. As oxygen is one of the most effective quenching agents, flames with minimum oxygen concentration should be used (e.g., an oxygen–argon–hydrogen flame). The high temperature of the analytical zone increases not only the fluorescence signal, but also the background, caused by flame emission. The optimal position of the analytical zone on the flame axis maximizes the signal-to-noise ratio. The background radiaton is usually much higher in flames as compared with GETA because of fluctuations in the optical density of the analytical zone (Rayleigh scattering of the excited laser radiation), luminescence of organic compounds in the flame, Mie scattering by dust particles, etc. As a result detection limits are typically

much worse for flame-LEAF than for GETA-LEAF. On the other hand, the relative simplicity of design and the stationary analytical regime are definite advantages of flames.

Glow Discharges. Glow discharges (GD) have proved very useful atomizers for direct analysis of solid samples by LEAF. In GD the surface of a solid sample is bombarded by ions of a noble gas, which is continuously pumped through the atomizer at low pressure. Depending on the type of GD, the gas pressure ranges from 10^{-1} Pa to 10 kPa. At such pressures collisions between analyte or matrix atoms and atoms in the gas phase are relatively scarce, which substantially reduces the probability of gas phase reactions and fluorescence quenching. The sample to be analyzed is used as the cathode. The anode – cathode voltage is typically several hundred volts, the energy range of the bombarding ions is $10^{-17} - 10^{-15}$ J (0.1 – 10 keV). Both direct current and pulsed discharges are used. In the latter case the discharge is synchronized with the laser pulse, with an appropriate time delay, depending on the distance between the sample surface and the analytical zone and the mean energy of the sputtered atoms.

The process has substantial advantages over thermal atomization. The sputtering yields of different elements differ by no more than one order of magnitude, whereas their vapor pressures may differ by several orders of magnitude. Sputtering yields are much less sensitive to the chemical composition of the matrix as compared to thermal atomization efficiencies. Ion sputtering avoids the drawbacks of GETA (e.g., chemical interference, recombination of free atoms into molecules and clusters, carbide formation). The problem of surface contamination can be easily solved by preliminary cleaning in a discharge. After the surface layer is removed, sample sputtering stabilizes and the bulk composition of a homogeneous sample or the layer-by-layer composition of an inhomogeneous one can be analyzed.

22.3.1.3. Limits of Detection

Like most spectroscopic analytical techniques, LEAF needs a calibration procedure. In common with other modern analytical techniques, determination of limits of detection (LOD) poses a new problem. Commonly, LOD are defined as the concentration (or absolute amount) of analyte that gives an analytical signal three times the standard deviation of the blank. The analyte concentration in the purest solvent used for blank measurements can often be detected by modern analytical techniques. One should differentiate the LOD determined by the limited purity of the solvent from the limit determined by the intrinsic characteristics of the analytical technique (nature of background, electronic noise, recording sensitivity, etc.). The first LOD is measured with the laser tuned to the resonance line of the analyte (on-line LOD). The second is measured with the laser detuned (off-line LOD). The detection power of most laser analytical techniques is very high and most published LOD are off-line values.

The best LEAF LODs currently available are listed in Table 1 [7].

22.3.1.4. Real Sample Analysis

Trace amounts of Pb, Cd, and Bi at femtogram levels have been directly determined in ancient ice and recent snows in the Antarctic by LEAF – GETA. Toxic metal concentrations in deep Antarctic ice cores give unique information about the preindustrial tropospheric cycles of these metals and about recent anthropogenic sources of heavy metals. Concentrations of Pb and Cd in the samples analyzed were 0.3 – 30 pg/mL and 0.08 – 2 pg/mL, respectively. The values measured by LEAF – GETA were in good agreement with those measured by isotope dilution mass spectrometry (Pb) and graphite furnace AAS (Cd). Both traditional techniques required elaborate chemical pretreatment and large sample volumes (50 – 200 mL). The sample volume in LEAF experiments was 20 – 50 μL.

The most successful analysis of real samples by LEAF was done with sample atomization in a graphite furnace. The high detection power (absolute LODs at and below femtogram levels) and wide linear range (up to 5 – 7 orders of magnitude) of LEAF were combined with modern furnace technology:Fast heating,platform atomization, matrix modification, and direct solid sampling. As a result low LODs, reasonable accuracy and precision, less matrix interference, and simpler sample pretreatment (if necessary) were achieved (Table 2).

The potential of LEAF with glow discharge atomizers (LEAF-GD) have been demonstrated. The LODs for Pb and Ir as low as 0.1 and 6 ng/mL, respectively, were achieved with hollow cathode GD and aqueous standards. With a similar atomizer 40 ng/g LOD was obtained for direct analysis of Pb traces in Cu-based alloys. With

Table 1. Best detection limits for LEAF-GETA [7]

Element	λ_{11}, nm	λ_{12}*, nm	λ_0, nm	Limit of detection	
				fg	pg/mL**
Ag	328.1		338.3	8.0	0.4
Al	308.2		394	100	5.0
Au	242.8		312.3	10	0.5
Bi	306.8		472.2	800	40
Cd	228.8	643.8	361	18	0.9
	228.8		228.8	3.5	0.175
Co	308.4		345.4	4.0	0.2
Cu	324.7		510.5	60	3.0
Eu	287.9		536.1	3.6×10^5	1.8×10^4
Fe	296.7		373.5	70	3.5
Ga	403.3	641.4	250.0	1.0	0.05
In	410		451	20	1.0
Ir	295.1		322.1	6×10^3	300
Li	670.8		670.8	10	0.5
Mn	279.5		279.5	90	4.5
Mo	313.3		317.0	10^5	5×10^3
Na	589.6		589.6	30	1.5
Ni	322.2		361.9	10^3	50
Pb	283.3		405.7	0.2	0.01
Pt	265.9		270.2	10^3	50
Rb	780		780	20	1.0
Sb	212.7		259.8	10	0.5
Sn	286.3		317.5	30	1.5
Te	214.3		238	20	1.0
Tl	276.8		353	0.1	0.005
V	264.8		354.4	2.2×10^6	1.1×10^5
Yb	398.8	666.7	246.4	220	11

* λ_{12} is the wavelength of the second laser in a two-step excitation scheme.
** The pg/mL values are calculated assuming a sample volume of 20 µL.

the more efficient magnetron GD even lower LODs were achieved: 3 ng/g for Pb in Cu alloys, 0.8 ng/g and 2 ng/g for Si in In and Ga, respectively.

Analysis of pure gases by LEAF is much less developed. Nevertheless, some very impressive results have been demonstrated. Traces of Pb and Tl in the air of a normal laboratory and a class-100 clean room were detected by LEAF. Air from the room was pumped through a tantalum jet, fixed in a hole inside a graphite tube. The heavy particles impacted onto the tube wall opposite the jet. After pumping a certain volume of ambient air in a particular time interval the impacted metal traces were detected by graphite furnace LEAF. LODs of 0.1 pg/m^3 and 0.01 pg/m^3 for Pb and Tl, respectively, were achieved. The Pb LOD is two orders of magnitude better than the LOD for graphite furnace AAS. The measured concentrations of Pb and Tl in laboratory air were 1.25 ng/m^3 and 4.3 pg/m^3, respectively. The same values in clean-room air were about 0.3–0.6 ng/m^3 and 0.9–1.0 pg/m^3. The measured concentrations of the same metals inside a clean bench, in the clean room were 6.5 pg/m^3 and 0.035 pg/m^3.

Detection of atoms and molecules which have transitions in the vacuum UV region is a serious challenge for any spectroscopic technique. One way to solve the problem is excitation of a species, not from the ground state, but from a higher metastable state which can be efficiently populated by electron impact in a gas discharge. The population of a metastable level by electron impact followed by LEAF detection was successfully applied for the detection of Ne and N_2 traces in He, and for detection of NO and NO_2 traces in He, Ar, N_2, and air. LODs of ca. 10^{-7}–10^{-8} vol % were obtained. These LODs are 100–1000 times better as compared with any traditional analytical technique.

22.3.2. Laser-Enhanced Ionization [3]

22.3.2.1. Physical Principles

Laser-enhanced ionization (LEI) spectroscopy is based on the difference in collisional ionization rates of atoms in their ground and excited states. If

Table 2. Real sample analysis by LEAF

Analyte	Matrix[a]	Certified value	Measured value	LEAF LOD (aqueous standard)	Comments
Co	ASRMs	ng/g	ng/g	fg	digested, diluted samples
	wheat grains (SBMP)	60 ± 20	45 ± 2	60	
	dried potato (SBMK)	100 ± 30	90 ± 5		
	dried grass (SBMT)	60 ± 20	48 ± 2		
	Sn alloys			pg	vacuum atomizer
	OSCh		4.4 ± 1.5	4	
	OVCh		2.5 ± 0.8		
Mn	ASRMs	μg/g	μg/g	fg	slurried samples
	citrus leaves (SRM 1572)	23.0 ± 2	23.4 ± 0.7	100	
	bovine liver (SRM 1577 a)	9.9 ± 0.8	9.4 ± 0.5		
	milk powder (SRM 1549)	0.26 ± 0.06	0.29 ± 0.01		
Pb	Ni alloys	μg/g	μg/g	fg	direct solid sampling
	SRM 897	11.7 ± 0.8	12.6 ± 1.2	1	
	SRM 898	2.5 ± 0.6	2.2 ± 0.4		
	SRM 899	3.9 ± 0.1	3.7 ± 0.7		
	ASRMs				slurried samples
	citrus leaves (SRM 1572)	13.3 ± 2.4	14.1 ± 0.5		
Tl	ASRMs	ng/g[b]	ng/g	fg	dissolved samples
	pine needles (SRM 1575)	50	48 ± 2	3	
	tomato leaves (SRM 1573)	50	51 ± 4		
	bovine liver (SRM 1577 a)	3 ± 0.3	3.2 ± 0.2		
	Ni alloys	μg/g	μg/g	fg	direct solid sampling
	SRM 897	0.51 ± 0.03	0.56 ± 0.07	10	
	SRM 898	2.75 ± 0.02	2.5 ± 0.5		
	SRM 899	0.252 ± 0.003	0.24 ± 0.03		
Te	Ni alloys	μg/g	μg/g	fg	direct solid sampling
	SRM 897	1.05 ± 0.07	0.97 ± 0.15	20	
	SRM 898	0.54 ± 0.02	0.52 ± 0.05		
	SRM 899	5.9 ± 0.6	5.4 ± 0.7		
Sb	Ni alloys	μg/g[c]	μg/g	fg	dissolved samples
	PW 1 A	5	4.8 ± 0.5	10	
	PW 2 A	9	8.3 ± 0.5		
	PW 3 A	17	16.5 ± 0.8		
P	Ni alloys	μg/g	μg/g	pg	digested, diluted samples
	SRM 349	30 ± 10	23 ± 3	8	
	SRM 865	120 ± 30	100 ± 20		
	SRM 126 c	40[c]	37 ± 5		
	ASRMs	mg/g	mg/g		
	SRM 1577 a	11.1 ± 0.4	11.6 ± 0.9		
	SRM 1549	10.6 ± 0.2	10.9 ± 0.9		
	SRM 1575	1.2 ± 0.2	1.1 ± 0.2		

[a] ASRM = agricultural standard reference materials (Russian State Reference Materials). [b] Noncertified value. [c] Reference value.

a d.c. voltage is applied across the atom reservoir and analyte atoms are optically excited by laser radiation, the enhanced probability of collisional ionization can be detected as a change of current (or a change of voltage across the load resistor). The LEI signal is linearly proportional to the ionization rate of the excited analyte atom, which in turn depends on its ionization potential, the nature of its collision partners, and the temperature of the analytical zone. The maximum sensitivity of LEI occurs when both the probabilities of ionization and ion collection are maximal (close to unity). The ionization efficiency depends exponentially on the energy difference between the upper ex-

cited state and ionization potential of the analyte atom. The main advantage of LEI as compared with LEAF is the very high probability of ion collection (close to unity in the optimal scheme); the collection efficiency of spontaneously emitted fluorescence photons is much lower.

Different excitation schemes for LEI are shown in Figure 3. The simplest and most widely used is the one-step excitation scheme (Fig. 3 A). Elements with low ionization potential (mainly alkali metals) can be efficiently ionized by the one-step scheme with dye lasers in the visible range. For a number of elements (Cd, Mg, Pb, etc.) the one-step scheme can be efficient with
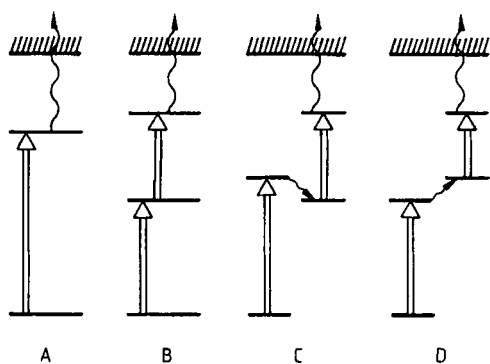
Figure 3. Classification of LEI schemes
A) One-step excitation; B) Two-step excitation; C, D) Collisionally assisted two-step excitation

excitation in the UV (harmonics of visible dye lasers). Excitation by UV photons below 300 nm may increase the background caused by nonselective one- or two-step photoionization of the thermally excited species in the analytical zone. The two-step excitation scheme (Fig. 3 B) is more complicated — thus more expensive — as it utilizes two tunable lasers, but it has significant advantages. Ionization efficiency increases up to three orders of magnitude compared with one-step excitation. The background does not change significantly if both lasers operate in the visible range. Selectivity increases dramatically as the selectivities of the single steps are multiplied in the two-step process. LEI is an efficient technique only if collision rates are high. Thus, the excitation – ionization scheme (Fig. 3 C), with noncoinciding levels of the first and second steps, might be efficient owing to effective collisional coupling of both levels.

22.3.2.2. Atomizers

The most commonly used atomizers for LEI are atmospheric flames. Flames have relatively high temperatures in the analytical zone and, hence, high efficiency of analyte atomization. Thermally excited molecular compounds in flames provide efficient ionization of the optically excited analyte atoms. All types of flames used in atomic absorption and emission spectrometry have been investigated in LEI. Relatively hot hydrocarbon flames such as C_2H_2/air or C_2H_2/N_2O are preferable for LEI because of the high degree of atomization of the elements. A large number of ele-

ments can be detected by LEI with C_2H_2/air flames; the hotter C_2H_2/N_2O flame is more suitable for the detection of refractory elements.

Few numerical values can be given for LEI in flames. In a flame at 2500 K, the rate of ionization of an element is enhanced ca. 100-fold for each electron volt (1.6×10^{-19} J) of laser excitation energy. If the energy gap between the excited level and the ionization continuum is less than 1 eV, the probability of ionization within 10 ns is close to unity. If cw laser excitation is used 100 % ionization can be achieved for an energy gap of ca. 3 eV (and less).

Other atomizers such as graphite furnaces, ICP, and GD have also been investigated for LEI. Argon buffer gas, typically used in graphite furnaces, is a much poorer collisional partner than the molecular constituents of flames, so collisional ionization is less efficient. On the other hand, good analytical results have been obtained with a combined flame – rod atomizer. A small volume of sample (positioned in a hole in a graphite rod) combined with the high temperature of the analytical zone (located above the rod in the surrounding flame) provided a low absolute detection limit in the picogram – femtogram range. ICP and GD proved to be electrically noisy for LEI.

22.3.2.3. Limits of Detection

The best LODs for one- and two-step LEI in flames are listed in Table 3.

22.3.2.4. Real Sample Analysis

The potential of LEI has been investigated by analysis of Standard Reference Materials (SRMs) at the National Institute of Standards and Technology (NIST). Traces of Co, Cu, Cd, Mn, Ni, and Pb were measured in different SRMs of water; Mn and Ni traces in bovine serum; Mn, Ni, and Fe traces in apple and peach leaves; Tl and Ni traces in river sediments; and Na, K, Ca, and Mg traces in rainwater.

The concentrations of the alkali metals Cs, Li, and Rb were detected by LEI in different types of reference rock samples. Rock samples were dissolved by standard methods and analyses performed without preconcentration, with pure aqueous standards for calibration. Concentrations were determined by both calibration curve and standard additions methods. The results obtained by the two techniques coincided within experimental errors. Concentrations varied in the range

Table 3. Best detection limits for one- and two-step flame LEI of aqueous solutions

Element	λ_1, nm	λ_2, nm	Flame*	Detection limit, ng/mL
Ag	328.1		AA	1
	328.1	421.1	AA	0.05
Al	309.3		AN	0.2
As	278.0		AA	3000
Au	267.7		AA	1.2
	242.8	479.3	AA	1
Ba	307.2		AA	0.2
Bi	227.7		AA	0.2
Ca	227.6		AA	0.006
	422.6	585.7	HA	0.03
Cd	228.8		AA	0.2
	228.8	466.2	AA	0.1
Co	240.8		AA	0.06
	252.1	591.7	AA	0.08
Cr	240.9		AA	0.2
	298.6	483.6	AA	0.3
Cs	455.5		AA	0.002
Cu	324.8		AA	3
	324.8	453.1	AA	0.07
Fe	302.1		AA	0.08
	271.9	468.1	AA	0.1
Ga	294.4		AA	0.01
In	271.0		AA	0.001
	451.1	571.0	AA	0.0009
K	766.5		AA	0.1
Li	670.8		AA	0.001
	670.8	610.4	AA	0.0002
Mg	285.2		AA	0.003
	285.2	470.3	AA	0.001
Mn	279.5		AA	0.04
	279.5	521.5	AA	0.02
Mo	319.4		AN	10
Na	285.3		AA	0.001
	589.0	568.8	AA	0.0006
Ni	229.0		AA	0.02
	282.1	501.4	AA	0.04
Pb	283.3		AA	0.2
	283.3	600.2	AA	0.09
Rb	780.0		HA	0.09
Sb	287.8		AA	50
Si	288.2		AN	40
Sn	284.0		AN	0.4
	284.0	597.0	HA	0.3
Sr	230.7		AA	0.003
	460.7	554.3	AA	0.2
Ti	320.0		AN	1
Tl	276.8		AA	0.006
	291.8	377.6	AA	0.008
V	318.5		AN	0.9
W	283.1		AA	300
Yb	267.2		AA	1.7
	555.6	581.2	AA	0.1
Zn	213.9	396.6	AA	1

* AA = acetylene – air; AN = acetylene – nitrous oxide; HA = hydrogen – air.

0.4 – 17.0 ppm and were in a good agreement with the certified values for the reference samples.

The LEI technique has been applied to the analysis of pure materials for the semiconductor industry. The concentrations of Cr, Fe, and Ni in Cr-doped GaAs disks were measured with the simplest pretreatment (dissolution in high-purity nitric acid and dilution with pure water). After dilution, a concentration of 0.1 ng/g in solution corresponded to 1 μg/g in the solid. Impurity concentrations were 4, 8, and 2.3 μg/g, respectively.

The rod – flame atomizer avoids the general disadvantage of flame-based analytical techniques — the necessity to dissolve the solid sample. The potential of two-step LEI with the rod –flame atomizer was investigated by analysis of Cu traces in Ge and In traces in CdHgTe. The LODs as low as 1 pg/g for In and 0.5 ng/g for Cu were achieved in direct analysis of 10-mg solid samples. Measured values in real samples varied in the range 30 – 70 ng/g for Cu and 2.6 – 3.4 ng/g for In.

Laser-enhanced ionization is a powerful, versatile and relatively simple technique which provides LODs in the ng/g – pg/g range with large linear dynamic range. A serious drawback of the technique is its susceptibility to interference: Thermal ionization of easily ionized matrix components (mainly alkali metals); photoionization of the thermally excited major components (especially if UV laser light is used for excitation); thermal ionization from levels excited by the spectral wing of an intense laser line; and ionization of the molecular components of the flame (especially NO). Interference may be greatly reduced or avoided by the appropriate choice of excitation wavelength, by the two-step excitation scheme, by specially constructed collecting electrodes; by optimal location of the analytical zone within the flame. To reduce interference one should investigate precisely the origins of noise and background for a particular analyte and matrix.

22.3.3. Resonance Ionization Spectroscopy [2]

22.3.3.1. Physical Principles

Resonance ionization spectroscopy (RIS) is based on resonance multistep excitation of highlying levels of free analyte atoms and their subsequent ionization. Ionization may be caused by the photons of the final laser step (photoioniza-
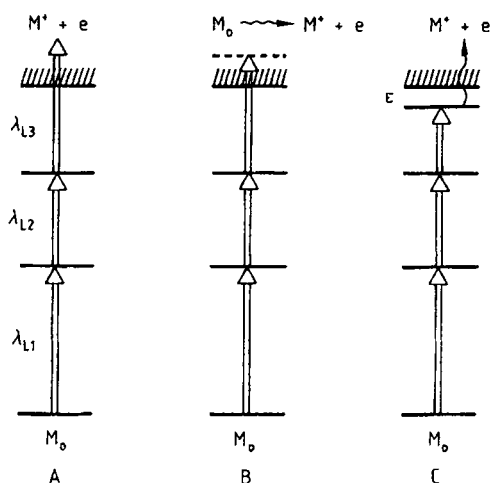
Figure 4. Classification of RIS schemes
A) Nonresonance photoionization to a continuum; B) Photoionization via an autoionization state; C) Electric field ionization of a Rydberg state

tion) or by an external electric field (field ionization). Sometimes the latter technique is referred to as field ionization laser spectroscopy (FILS). Multistep excitation of atoms may be effective if the rates of stimulated transitions to each intermediate level are much higher than the total decay rates (radiative and collisional) of any intermediate level. In contrast to LEI, RIS is efficient only if negligible collisions occur in the analytical zone. As in the case of LEI, the high sensitivity of RIS is based on the high efficiency of ion collection (close to unity). Thus, the overall efficiency of neutral atom detection by RIS depends on the efficiency of stepwise excitation and final ionization. Maximum sensitivity is achieved if all the intermediate transitions are saturated and the ionization efficiency is close to unity. In optimal conditions, a single free atom may be detected with a probability close to unity.

The three most widely used schemes for RIS are shown in Figure 4. They differ from each other in the final ionization step. Nonresonant photoionization to a continuum (Fig. 4 A) can be efficient if the energy flux of the ionizing laser is high enough $(0.01 - 1 \text{ J/cm}^2)$, as the nonresonant ionization cross sections are relatively low $(10^{-17} - 10^{-19} \text{ cm}^2)$. For comparison, typical cross-sections of resonance optical transitions are $10^5 - 10^6$ times higher.

The photoionization scheme via the so-called autoionization state (AIS) is shown in Figure 4 B.

The AISs are discrete states of the inner excited electrons with energies above the ionization potential of the atom. On the one hand, the cross section of a transition to AIS may be several orders of magnitude higher than direct photoionization to the continuum (Fig. 4 A); on the other hand, the AIS may decay to the continuum (i.e., the atom is ionized) with a typical lifetime of a few nanoseconds. Thus, in the field of an intense laser pulse with a duration of ca. 10^{-8} s, the probability of excited AIS depletion and ionization may be close to unity. Ionization via AIS is highly efficient for some elements (Gd, Lu).

Ionization through high-lying Rydberg states (Fig. 4 C) has been found to be very effective. Rydberg states close to the ionization limit can be efficiently ionized by an electric field pulse of moderate amplitude. The dependence of ionization yield of a particular Rydberg state on the electric field amplitude has a characteristic Z-profile with a pronounced threshold E_{crit}. For electric fields above E_{crit} the ionization yield is nearly constant and close to unity. Both the excitation cross section and E_{crit} depend strongly on the principal quantum number of the Rydberg state (n^*). For higher n^*, E_{crit} is lower (an advantage), but the excitation cross section is much lower (a disadvantage). Therefore, compromise values of n^* are used, which for most elements lie between 15 and 30. For such n^*, excitation cross sections are of the order $10^{-14} - 10^{-15} \text{ cm}^2$ and E_{crit} values are $1 - 10 \text{ kV/cm}$. For analytical applications, ionization via Rydberg states has proved to be an ideal scheme for RIS. Because of the relatively high excitation cross sections, maximum sensitivity can be realized with tunable lasers of moderate pulse duration and average power, and with reasonable electric fields. The effectiveness of AIS is limited by the absence of AIS near the ionization potential of many elements and the lack of appropriate spectroscopic information.

The fact that RIS measures charged particles is an important advantage. Highly efficient signal ion collection is possible by appropriate electromagnetic focusing. Second, discrimination of signal ions from background ions can be achieved by mass separation methods. These two factors improve both the sensitivity of RIS and its accuracy.

In most analytical RIS applications two- or three-step excitation is used. This provides extremely high selectivity (up to $10^{10} - 10^{12}$) as the overall selectivity of a multistep process is a product of one-step selectivities.

22.3.3.2. Atomizers

Vacuum or low-pressure atomizers are used in RIS to obtain collisionless interaction of the free analyte atoms with resonance laser radiation. Different types of GD and vacuum electrothermal atomizers (VETA) are most commonly used in RIS. In some specific analytical applications (detection of traces of rare radionuclides) RIS is used on-line with fast atomic beams.

In many applications of RIS – VETA the generated ions are detected by a secondary-emission multiplier incorporated in the atomizer. The vacuum should be better than 10^{-3} Pa throughout the analytical volume and the path of free ions from the interaction zone to the multiplier. This limits the speed of sample evaporation and dictates the efficiency of the pumping system. Both unified vacuum systems and systems with differential pumping of the analytical volume and free flight zones are used. In these respects the strategy of electrothermal atomization of a sample in RIS differs significantly from the traditional one used in graphite furnace AAS. In the latter technique fast atomization (1 – 10 s) gives the best analytical results. In RIS, very slow atomization (up to several minutes) is typical.

In some applications, a proportional counter operating near atmospheric pressure is used. In this case, the electrons created in the process of atomic resonance ionization are detected. By appropriate choice of gas mixture and accelerating voltage, an electron cascade with a magnification factor of about 10^4 can be achieved. This gain is high enough for single photoelectron detection.

For maximum sensitivity and isotope selectivity RIS is often combined with time-of-flight (RIS-TOF) or quadrupole mass spectrometry (RIMS).

22.3.3.3. Limits of Detection: Real Samples

The high sensitivity of RIS has been demonstrated in the analysis of pure semiconductor materials, biological samples, seawater, and environmental samples. In some cases, direct analysis was possible; for some matrices, sample pretreatment was used.

Traces of Na, Al, and B in high purity Ge were detected by RIS with direct electrothermal atomization of solid Ge chips in vacuum. For all these elements, two-step excitation of Rydberg states and electric field ionization were used. The meas-

ured concentrations of Na, Al, and B in real samples were 0.2, 1.0, and 0.2 ng/g while the LODs were 0.05, 0.01, and 0.05 ng/g, respectively. In the case of B, the strong ion background seriously limited the sensitivity and selectivity of traditional RIS. This background was caused by efficient two-step nonresonance photoionization of Ge atoms (the energy difference between the excited $2s^23s$ level of B and the $4s^24p5s$ level of Ge is only 20 cm^{-1}). Both analytical characteristics were greatly improved by a special technique, with two sequential electrical pulses of different duration, polarity, and voltage.

The distribution of ruthenium and rhodium traces in the ocean has been investigated by RIS – VETA. Both elements are rare and their distribution is of great geological and geochemical importance. The abundance of Ru and Rh in the Earth's crust is estimated to be 1 – 100 pg/g and can not be directly measured by current analytical techniques. The fire assay method and concentration by polymer sorbents were used prior to RIS analysis. Concentrations of Ru varied from 1.3 pg/mL in ocean water to 7.6 ng/g in Red Sea sediments. Concentrations of Rh in ocean water were 3 – 10 pg/mL.

The detection of ^{81}Kr, produced by cosmic rays, is of great importance for dating samples from the Earth's crust, groundwater, and polar ice cores. Measurement of ^{81}Kr in groundwater is a unique analytical problem as its concentration is ca. 1400 atoms per liter. The problem is further complicated by isotopic interference of stable krypton isotopes (^{82}Kr/^{81}Kr is about 10^{12}). Prior to final analysis by RIS – TOF [8] several consecutive steps of field sampling and isotope concentration were necessary. All steps required extreme care to avoid sample contamination. Water samples of 50 – 100 L were necessary to detect ^{81}Kr above the background. This value is more than 100 times less than for the routine technique of ^{39}Ar dating.

The environmentally important radionuclides ^{90}Sr and ^{210}Pb have been detected by RIMS – VETA. Two-step excitation of Rydberg states by two cw dye lasers, followed by nonresonant ionization by a cw infrared CO_2 laser, was used. The solid samples were dissolved by standard methods. The LODs were 60 ag (6×10^{-17} g) for ^{210}Pb and 30 ag for ^{90}Sr. Absolute ^{210}Pb concentrations of 0.12 – 0.14 fg were measured in brain tissue. Isotope selectivity of 10^9 for ^{90}Sr/^{89}Sr was achieved.

Recently the diode lasers were successfully used in a multiphoton RIMS scheme for isotopi-

cally selective detection of strontium. The two-step resonance excitation of ^{90}Sr was performed with two single-mode cw DLs ($\lambda_1 = 689.45$ nm, $\lambda_2 = 688.02$ nm). The excited ^{90}Sr atoms were photoionized by a powerful Ar$^+$ laser ($\lambda = 488.0$ nm). The ions were detected in a commercial quadrupole MS. Using graphite crucible atomization the absolute LOD of 0.8 fg and overall (optical + mass spectrometer) isotopic selectivity of $> 10^{10}$ in the presence of stable ^{88}Sr was achieved for ^{90}Sr. No other non-radiochemical technique can realize such sensitivity and selectivity of ^{90}Sr detection.

A similar scheme was used for the spectroscopic studies on hyperfine components and isotope shifts of rare stable isotopes of calcium and the radionuclide ^{41}Ca. Calcium atoms in an atomic beam were excited with single-frequency cw dye and titanium sapphire lasers and then photoionized with the 363.8 nm or 514.5 nm line of an Ar$^+$ laser.

The use of RIMS in chemical vapor deposition diagnostics has been investigated.

22.3.4. Diode Lasers in Elemental Analysis [6], [8]

22.3.4.1. Physical Principles of Absorption Spectrometry

Quantitative absorption measurements are based on the Lambert–Beer law:

$$P_t(\lambda) = P_0 \exp[-k(\lambda)l]$$

where P_t is the laser power at wavelength λ, transmitted through a layer with absorption coefficient $k(\lambda)$ and absorption pathlength l. The absorbed power ΔP is $P_0 - P_t = P_0[1 - \exp(-k(\lambda)l)]$. For small values of the absorption $k(\lambda)l \ll 1$, this is approximated with $\Delta P(\lambda) \approx P_0 k(\lambda)l$.

Theoretically, the minimum detectable absorption (MDA) is determined by the shot noise of detected radiation. In single pass measurements ($l = L$, where L is a sample absorption pathlength)

$$(kL)_{\min} = \Delta P_{\min}/P_0 = (2 Bh\nu/\eta P_0)^{1/2}$$

where η is the quantum efficiency of the photodetector, and B is a detection bandwidth. To obtain noise equivalent relative power variations, $\Delta P_{\min}/P_0$, smaller than 2×10^{-8} with a measurement time of 1 s, the laser photon flux should be at least 5×10^{15} photons/s. This corresponds to laser

powers of 1 mW at 1 μm and 100 μW at 10 μm. In practice there are additional noise factors affecting the detection sensitivity, such as intensity fluctuations of the laser (laser excess noise) and technical fluctuations. Thus, for typical DLs, the value of $\Delta P_{\min}/P_0 = 1 \times 10^{-8}$ can be considered as the lowest, fundamentally limiting one.

Usually, absorption measurements are not considered to be very sensitive since the signal $P_0 k(\lambda)l$ is a subtle variation of a large baseline P_0. In order to have zero baseline, sample or wavelength/frequency modulation (WM/FM) techniques can be used. In the case when λ is modulated with frequency f, a signal at the n-th harmonics of f is simply proportional to $P_0 k_{nf}(\lambda)l$, where k_{nf} is the amplitude of the corresponding harmonics of k. As $k(\lambda) = \sigma(\lambda)N$, where $\sigma(\lambda)$ is the absorption cross section, $\sigma(\lambda) = \sigma_0(\lambda)/(1 + I/I_{\text{sat}})$ (see Section 22.1.3), then the signal after the phase-sensitive detector is directly proportional to the concentration of absorbing particles N. Another important advantage of WM/FM is the detection of a signal at high frequency (> 10 kHz), where the low-frequency flicker noise is greatly reduced.

Besides reducing noise, improvements in the MDA can be achieved by enhancing the signal. This can be done by increasing the absorption pathlength using multipass cells. By an appropriate mirror arrangement a DL beam can be directed many times through the absorbing layer, increasing the overall absorption with an appropriate decrease in MDA. The number of passes is limited by the quality of the mirrors, which determines the cell transmission, or by interference effects.

22.3.4.2. Atomic Absorption Spectrometry with Diode Lasers [6], [8]

UV/visible/NIR semiconductor diode lasers (DLs) as sources of tunable narrowband resonance radiation have been discussed in Section 22.2.2. The practical and technical advantages of visible/NIR DLs are ease of operation, small size, room-temperature operation, and low price. In comparison with hollow cathode lamps (HCLs), they have such advantages as high spectral power, spatial coherence, and tunability. Owing to these features, detection limits, dynamic range, and selectivity of analysis are all significantly improved.

The presently available powers of DLs are several orders of magnitude higher than the powers of the best commercial HCLs. In contrast to an HCL, a DL emits a prominent single narrow line. This feature dramatically simplifies the construc-

tion of an analytical device because one does not need a monochromator, which is used in commercial AAS instruments for isolation of the analytical line from the spectral lines of the HCL buffer gas and unwanted lines of the cathode material. In the case when optimal experimental conditions are realized, more than 1 mW of laser radiation power can be obtained at the photodetector. It has been shown that a near fundamental shot-noise limit can be achieved and relative power variations $\Delta P_{min}/P_0$, which can be measured with DLs, can be as low as 10^{-8}.

The LODs of diode laser atomic absorption spectrometry (DLAAS) have been greatly improved using a WM technique. In conventional atomic absorption spectrometry with HCLs, low-frequency (10–60 Hz) amplitude modulation (mainly by mechanical choppers) is used. In contrast, the wavelength of the DL can be easily modulated with much higher frequencies, in DLAAS usually up to 10–100 kHz. By recording the analytical signal at the harmonics of the modulation frequency, usually the second one (2f-mode), one can dramatically suppress the 1/f noise, thus decreasing the LOD. Even in the first simple measurements with a DL power of ca. 100 µW, a modulation frequency of 5 kHz, and with 2f-mode of recording, absolute LODs in the femtogram range have been achieved experimentally. These LODs correspond to measured values of $\Delta P_{min}/P_0 \approx 10^{-4} - 10^{-5}$ (compared with $10^{-2} - 10^{-3}$ for conventional AAS).

Further improvement of WM-DLAAS has been demonstrated for the case when both the radiation source and the absorption are modulated. Such a situation was realized with a microwave induced plasma (MIP) or a direct current plasma (DCP) as the atom reservoir. The double-beam double-modulation approach with logarithmic signal processing (LSP) and 2f-mode recording permitted a complete elimination of the background, caused by non-selective absorption, low-frequency fluctuations, and drift of the baseline and laser intensity. A shot-noise limited MDA as low as 3×10^{-8} for a 1 s time constant was experimentally realized. The linear dynamic range can be also substantially extended with this technique, in the best cases up to six–seven orders of magnitude as compared with less than two orders of magnitude for conventional AAS.

Applications of DLs in analytical atomic spectrometry are collected in the following list (numbers in parentheses are analytical wavelengths in nanometers):

Diode laser atomic absorption spectrometry (DLAAS):
   Graphite furnace
      Al (396.15), Ba (791.13), Cs (894.35), K (769.90), La (657.85), Li (670.78), Rb (780.03), Sr (689.26)
   Flame
      Cs (425.44, 852.11), Cr (425.44), K (766.49), Li (670.78), Rb (780.03), Ti (399.97, 842.65)
   Discharge (direct current plasma (DCP) or microwave induced plasma (MIP))
      Cl (837.60), F (739.87), O (777.18)
Laser induced fluorescence (LIF):
   Li (670.78), Rb (780.03)
Isotope selective detection:
   Li, Pb, wavelength modulation diode laser atomic absorption spectrometry (WMDLAAS)
   ^{235}U/^{238}U in glow discharge, optogalvanic detection
   ^{235}U/^{238}U in solid samples, laser ablation + LIF with DL excitation
Physical vapor deposition process control:
   Y (668), external cavity diode laser (ECDL), high frequency modulation spectroscopy
   Al (394), second harmonic generation (SHG) of ECDL
Control of semiconductor films growth:
   Al (394), Ga (403), In (410), SHG in bulk crystals
Specific detection of haloform components:
   Cl (837.60), Br (827.24), MIP or DCP, GC-WMDLAAS, double-modulation
Environmental analysis:
   Cr^{III}/Cr^{VI} speciation in deionized and drinking water
      Cr (427.48), SHG in LiIo$_3$, flame, HPLC-WMDLAAS
   Cr^{VI} in tap water
      Cr (427.48), SHG in KNbO$_3$, flame, HPLC-WMDLAAS, double-beam, LSP
   Chlorinated hydrocarbons in oil from plastic material recycling
      Cl (837.60), MIP, GC-WMDLAAS
   Chlorophenols in plant extracts
      Cl (837.60), MIP, GC-WMDLAAS, double-beam, logarithmic signal processing (LSP)
   Speciation of methylcyclopentadienyl manganese carbonyl (MMT) in gasoline, human urine and tap water
      Mn (403.1), flame, SHG in LiIO$_3$, HPLC-WMDLAAS, 4f, LSP

Analysis of chlorine in polymers

    Cl (837.60), MIP, laser ablation + WMDLAAS, double-modulation, LSP

Resonance ionization mass spectrometry (RIMS):

    La (657.85), three-step RIMS, DL – first step

    Sr (689.45, 688.02), double-resonance excitation by two DLs

The detection capabilities for various species differ considerably depending on the type of atomizer, oscillator strength of a transition, detection technique used. Multielement analysis with DLs has also been investigated. Several laser beams may be collimated and directed simultaneously through an absorbing layer in a conventional atomizer, either by a system of mirrors, or by optical fibers. The monochromator – photomultiplier system can be replaced by a simple semiconductor photodiode. Appropriate modulation of a wavelength of a particular DL, followed by Fourier transform or lock-in amplification analysis of the photodiode signal, makes it possible to process analytical signals of individual analytes. A six-element analysis with DLs and a commercial graphite tube atomizer has been demonstrated.

22.3.4.3. Element Specific Detection by DLAAS

Diode laser based analytical devices can be used as element selective detectors for gas/liquid chromatography (GC/LC). The atomic constituents of different molecular species in a sample can be measured in an atomizer, such as a flame or low pressure plasmas, e.g., an MIP or DCP. The GC-MIP/DCP in combination with DLAAS enables the detection of non-metals, such as H, O, S, noble gases, and halogens. Most of these elements have metastable levels, which are efficiently populated in a plasma, and which have strong absorption transitions in the red – NIR spectral region.

Recently it has been possible to observe a definite trend in elemental analysis, from detection of the total content of a heavy metal in a sample to the detection of its different chemical compounds or oxidation states (speciation). This is because the essentiality or toxicity of the metals should not be attributed to an element itself but to element compounds with distinct biological, physical, and chemical properties. A compact and relatively cheap DL based instrument could be a reasonable alternative to the powerful but expensive ICP-MS, presently used for speciation. The potential of

DLAAS as an element selective detector for speciation has been demonstrated by its coupling with high performance liquid chromatography (HPLC). This system has been used for speciation of chromium (III)/(VI) and organo-manganese compounds. An LOD of 30 pg/mL for Cr^{III} and Cr^{VI} was obtained by use of 3 µW power frequency-doubled laser light, a double-beam configuration, and logarithmic processing of the signals. In the case of chromium, the LOD of HPLC-DLAAS is several times lower, than the LOD for HPLC-ICP-MS.

22.3.4.4. Isotope Selective Detection with Diode Lasers

The narrow spectral line of a DL enables isotope selective analysis. For light and heavy elements (such as Li and U) the isotope shifts in spectral lines are often larger than the Doppler widths of the lines, in this case isotopically selective measurements are possible using simple Doppler-limited spectroscopy – DLAAS or laser induced fluorescence (LIF). For example, ^{235}U and ^{238}U ratios have been measured by Doppler-limited optogalvanic spectroscopy in a hollow cathode discharge. DLAAS and LIF techniques have been combined with laser ablation for the selective detection of uranium isotopes in solid samples. This approach can be fruitful for development of a compact analytical instrument for rapid monitoring of nuclear wastes.

22.3.4.5. Commercial System

The DLAAS devices are now commercially available from Atomica Instruments (Munich, Germany). The first system was prepared for the measurement of ultra-trace levels of Al in the semiconductor industry. The commercial laser diode module includes the DL with a heat sink for temperature tuning, the microoptics, and the non-linear crystal for SHG. The module has the size of a HCL. Since delivery, the module has successfully worked routinely under the conditions of an industrial analytical laboratory without any repair or maintenance. The semiconductor industry in particular is interested in the measurements of light elements, such as Al, K, Ca, Na, in ultra pure water or chemicals. For example, the module designed for the detection of K provides an LOD of 0.5 pg/mL (10 µL aliquot). Similar compact and sensitive devices with electrothermal atomizers, flames or micro-plasma are currently

available for a number of elements, such as Mn, Ni, Pt, Pd, Ca, Cu, Fe, Ga, the alkali elements, the halogens, C, O, and H. In forthcoming years, one can also expect the commercialization of specific element selective DLAAS detectors for GC or LC.

22.4. Laser Techniques for Molecular Analysis [6], [8]

22.4.1. Molecular Absorption Spectroscopy with Diode Lasers

22.4.1.1. Detection of Molecular Gas Impurities with Lead-Salt Diode Lasers [2], [9]

The Pb-salt-based DLs (Section 22.2.2) efficiently generate tunable laser radiation in the region $3-30$ µm, where the great majority of molecules have intense vibrational rotational absorption lines. As the Doppler-broadened absorption linewidths of molecules are of the order of $(1-3) \times 10^{-3}$ cm^{-1}, DLs provide the best selectivity for linear absorption techniques.

Both cw and pulsed (sweep) modes of DL generation have been widely tested in analytical applications. Many well-developed modulation techniques have been used with consequent improvement of analytical sensitivity. The best results were achieved with a high FM technique ($f \sim 100-1000$ MHz). The minimal shot-noise-limited relative power variations $\Delta P_{min}/P_0$ which can be experimentally detected with FM DLs are about 10^{-7}. Typical values of $\Delta P_{min}/P_0 \approx 10^{-5}$ (1 Hz detection bandwidth) are mainly determined by instrumental factors (in most cases by interference effects). For this minimal detectable value of $\Delta P_{min}/P_0$ and for molecular species at atmospheric pressure the corresponding concentration LODs are ca. $0.1-10$ ppb m ($0.1-10$ parts per billion by volume oner an optical path of 1 m). Evidently, if 10 m absorption length could be realized, the minimal detectable concentration of absorbing molecules would be 10 times lower.

The most important analytical applications of mulecular absorption spectroscopy with mid-IR DLs are environmental monitoring and atmospheric studies [9]. Two different approaches are used: (1) measurements in sample cells at reduced pressure, when the absorption linewidths are of ca. $(1-3) \times 10^{-3}$ cm^{-1}, and (2) open-path measurements at atmospheric pressure, when the pressure broadening leads to ca. 10^{-1} cm^{-1} linewidths.

In a cell, optimal gas pressure and appropriate dilution of primer gas by a buffer can be used. A number of mulecules have been measured with lead-salt DLs: CO, NO, NO_2, N_2O, HNO_3, NH_3, SO_2, O_3, HCl, H_2CO, CH_4, C_2H_4, C_2H_6, CF_2Cl_2, etc. The effective absorption pathlength of $20-300$ m was realized with multipass cells with an appropriate decrease of the LOD down to a record value of 10 ppt (5 min integration time). Several Pb-salt DL-based gas monitors have been used to perform *in situ* trace gas measurements in the troposphere and lower stratosphere from airborne or balloon-borne platforms. The systems were designed to measure CH_4, CO, N_2O, NO_2, N_2O, HCl and H_2O with LODs at sub-ppb concentration levels ($3-30$ s integration time). A typical system includes a liquid-N_2 Dewar for diode lasers and detectors, a multipass cell, and can simultaneously measure up to four gases.

Owing to pressure broadening, ground-based remote detection of trace gases with DLs was not as effective as measurements in sample cells. Long-path atmospheric gas analyzers have been exploited mainly to monitor carbon monoxide concentrations. Mirror retroreflectors have been used to reflect the DL probing beam back to the detector providing a round-trip absorption pathlength of up to 1 km. Sensitivities achieved are in the range from 0.06 to 40 ppm · m.

Besides atmospheric trace measurements, Pb-salt DLs have successfully been used for detection of impurities in the process control of semiconductor materials, analysis of high purity gases, sensing of toxic gases, explosive detection, isotope analysis of gases, and medical diagnosis of various diseases by analysis of expired air. The use of a DL provides high time resolution ($0.1-1$ µs), facilitating sensitive control of transient processes (e.g., intermediates in chemical reactions, discharges, molecular beams). So far the high price, complexity, and limited reliability of the Pb-salt DL spectrometers limit their applicability to the research laboratories when utmost sensitivity and selectivity have to be realized.

22.4.1.2. Near IR Gas Sensors

Recent advances in room-temperature NIR lasers has led to a new generation of gas sensors. A number of DL-based instruments have been developed and are now commercially available.

These sensors have rather simple constructions, may be very compact, have low cost, are reliable, and allow completely autonomous operation for a long time.

Peculiarities of developing NIR-DL sensors are defined by characteristics of molecular absorption in this spectral region. There exists a large quantity of absorption lines that arise from overtones and combination bands. Typical lines-trengths of these transitions are two to three orders weaker than those for the fundamental bands in the mid-IR. The strongest absorptions in the $1.3-1.6 \mu m$ region are provided by the H_2O, HF, HCl, and C_2H_2 molecules. The LODs which are achieved for these molecules in pure gas can be at the ppt level. However, if the absorption of the matrix gas is larger than that of the detected impurity, the LOD will be determined by the interfering absorption of the matrix gas molecules.

There are a number of quite successful applications of NIR-DL sensors for monitoring of gas molecules, especially H_2O. A few types of hygrometers based on a $1.4 \mu m$ InGaAsP-laser are commercially available. A very small LOD of 65 ppt for water in air was achieved in laboratory test measurements with a hygrometer, designed to measure moisture in semiconductor feed gas. Another DL sensor, capable of measuring low levels of water vapor in the presence of methane, is designed to be attached to a sampling station where pipeline gas is extracted at near-ambient pressures $(1-2 atm)$.

Flight versions of NIR-DL instruments for trace water vapor measurements have been demonstrated on research aircraft with a sensitivity of the order of 10 ppm. Examples of such systems flown on NASA aircraft include both open-path and closed-path configurations designed for fast-response measurements of H_2O from the Earth's surface to an altitude of 20 km.

Similar instruments developed for monitoring O_2, NH_3, HCl, HF, H_2S, CO, CO_2 are suited to industrial stack monitoring, process and environmental control, where high temperatures, continuous vibration, thermal cycling, or high-speed air flow are present. The automated sensors provide fast and accurate ppm-level measurements and are used for optimizing incineration, sewage and waste treatment, power generation, and so on. Fiber optics are often preferred for their simplicity and ability to place the instrument at distances hundreds of meters from the hazardous zone.

22.4.1.3. Advanced Developments

Rapid improvements of the DL characteristics make these devices very attractive for absorption spectrometry not only of gas species (atomic or molecular) but also of liquid and condensed-phase samples. Much attention is now focused on using NIR-DLs for molecular fluorescence spectrometry, Raman spectroscopy, thermal lens spectrometry, medical diagnostics, development of fiber sensors, and so on.

It is unlikely that the NIR-DLs will permit measurement of all molecular trace constituents in the atmosphere with sensitivities sufficient for environmental control. New, room temperature operated lasers have been tested for trace measurements in mid-IR region. Potentially the most significant technology is quantum-cascade lasers (QCLs, see Section 22.2.2). Spectroscopic characteristics and noise properties of these structures enable measurements of trace impurities with LODs achieved using cryogenic Pb-salt lasers.

New absorption methods, like intracavity spectroscopy, cavity-ring-down and cavity-enhanced spectroscopy, have demonstrated very high sensitivities in laboratory measurements with DLs. An ultrasensitive technique that combines external cavity enhancement and FM spectroscopy has been developed recently. This method, which has been called NICE-OHMS, or noise-immune cavity-enhanced optical heterodyne molecular spectroscopy, is based on frequency modulation of the laser at the cavity free-spectral-range frequency or its multiple. The MDA of 5×10^{-13} $(k_{min} = 1 \times 10^{-14} cm^{-1})$ in the detection of narrow saturated absorption spectra of C_2H_2 and C_2HD has been reported, with a cavity finesse of 10^5. This result is one of the most sensitive absorption measurement ever reported.

22.4.2. Laser Optoacoustic Spectroscopy
[1], [2], [10]

The optoacoustic effect, based on partial non-radiative dissipation of the absorbed light energy, followed by generation of an acoustic wave, was discovered about a hundred years ago. The effect is very weak and, before the advent of lasers, it was applied mainly for spectroscopic investigations of gases. Lasers and improvements in acoustic vibration detection have led to the development of laser optoacoustic spectroscopy (LOAS). Several physical processes are involved in LOAS: (1)

absorption of laser radiation by one or more compounds in a medium; (2) radiationless conversion of a part of the absorbed energy into local heating of the sample; (3) formation of an acoustic wave due to a temperature gradient; and (4) detection of the acoustic vibrations.

The main advantages of LOAS are high sensitivity, applicability to samples of any state of aggregation, and relatively simple instrumentation. Its drawback is poor selectivity, due to spectral interference of the molecular compounds in a complex mixture. To improve selectivity, LOAS is often combined with a preselecting technique such as gas or liquid chromatography. This discussion is limited to the applications of LOAS for sensitive analysis of simple mixtures.

The LOAS instrument consists of the following main units: A tunable laser, a sample cell with an acoustic transducer (sometimes called a spectrophone or SP), an amplifier, and a recording system. The acoustic response of the medium can be stimulated by amplitude or frequency modulation of the laser light or by Stark or Zeeman modulation of the absorption line of the analyte. Both pulsed and cw lasers with mechanical or electro-optical amplitude modulation are often employed in LOAS analytical applications.

The key component of an LOAS instrument is the spectrophone. Its construction depends critically on the state of aggregation of the sample. The SP for gas samples consists of a gas cell with input and output windows for laser radiation and a microphone, fixed on an inside or outside wall of the cell. As in the case of MAS with SDLs, the sensitivity can be improved by the use of multipass SPs. Plane capacitor or electronic microphones are the most widely used. The sensitive element of these microphones is a thin ($1-10$ μm) elastic membrane made of Mylar, Teflon, or metallic foil. The membrane serves as one electrode of a dielectric capacitor (membranes made of dielectric materials have metallic coatings). Thus, the acoustic vibrations of a gas mixture in the cell can be directly converted to an electrical signal. Commercial microphones with sensitivities of $5-50$ mV/Pa are utilized in routine applications of LOAS; specially constructed and optimized ones are used for ultrasensitive analyses.

Resonant and nonresonant SPs can be used in LOAS. In the former type, the acoustic resonances are formed inside the cell. Resonant SPs provide higher sensitivity, but they are less flexible, as the acoustic resonances depend on the physical dimensions of the cell, the type of gas mixture, its

temperature, and pressure. The modulation frequency of cw laser radiation should be matched to the acoustic resonance frequency (typically of the order of several kilohertz). The dimensions of nonresonant cells are commonly $5-10$ mm diameter and $5-20$ cm length, while some resonant cells have a diameter up to 10 cm.

For the analysis of liquid and solid samples, more complicated SPs are used. For liquid samples, the simplest and most sensitive detection of acoustic waves is by direct contact of the microphone and a liquid. The acoustic vibrations of irradiated solid samples can be transduced to a microphone via a buffer gas or a "pure" liquid. Piezoelectric microphones are the most widely used type for solids and liquids.

Some important molecular impurities in pure gases have been successfully detected by LOAS. Typical gas pressures range from $1-100$ kPa; sample volumes $0.5-3$ cm^3; routine accuracy $5-15$ %; linear dynamic range up to 10^5. The most impressive LODs are: 0.1 ppb for SO_2 (multipass, resonant SP); 10 ppb for NO_2 (resonant SP); 1 ppm for HF (nonresonant SP, pulse laser, time resolution); 100 ppb for CH_4 (nonresonant SP); about 100 ppb for CO (resonant SP); 0.1 ppb for NO (nonresonant cell, six microphones); 10 ppb for SF_6 (resonant SP).

The best sensitivity of LOAS for liquid samples (up to 10^{-6} cm^{-1}) was achieved with pulsed lasers of ca. 1 mJ pulse energy. With modulated cw lasers of ca. 1 W average power, a minimum absorption of $10^{-5}-10^{-6}$ cm^{-1} was detected. A few examples of real analyses include the LODs for β-carotene, Se, and Cd in chloroform of 0.08, 15, and 0.02 ng/mL, respectively; the LODs for bacteriochlorophyll, chlorophyll b, and hematoporphyrin in ethanol of 1, 0.3, and 0.3 ng/mL, respectively; the LODs for cytochrome c, vitamin B_{12}, uranium(IV), and uranium(VI) in water of 30 ng/mL, 4 ng/mL, 8×10^{-7} mol/L and 10^{-6} mol/L, respectively. Typically, the LODs for direct analysis of liquids are worse as compared with "pure" gases because of higher background absorption of the solvent. To reduce this background, frequency modulation of lasers and differential two-beam detection schemes are used.

LOAS has been successfully applied for the detection of thin films and layers deposited on a solid substrate. Both direct measurement by a piezoelectric microphone connected to the substrate and detection of acoustic waves via an intermediate gas or liquid were used. For example, the absorption of thin Al_2O_3 layers (up to monolayer)
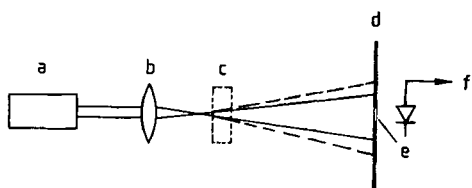
Figure 5. Schematic diagram of a single-beam thermal lens spectrometer
a) Laser; b) Lens; c) Sample; d) Screen; e) Pinhole; f) Detector
The dashed lines show the beam divergence due to the thermal lens effect

on an aluminum substrate was measured with a cw CO_2 laser, modulated by a mechanical chopper. Measurements of thin layers of liquids and powders as well as aerosols, suspensions, colloids, and small dispersed particles are important applications of LOAS.

22.4.3. Thermal Lens Spectroscopy [1], [3]

Thermal lens spectroscopy (TLS) is another laser-based technique, associated with dissipation of absorbed laser energy followed by the formation of temperature gradients in a medium. It was observed for the first time in 1964 when a 1-cm Brewster cell containing benzene was placed inside the resonator of a cw He–Ne laser. Contrary to the expected behavior, the laser intensity began to fluctuate and in a few seconds, when the steady state was reached, the beam diameter at the resonator mirrors had increased. The benzene cell acted as a spherical negative lens with a focal length of ca. 1 m.

The physical principles of TLS are rather simple. If a laser beam with a Gaussian-like intensity distribution propagates in a weakly absorbing medium the local temperature will increase and the transverse temperature profile will match the intensity profile of the laser beam. Since most liquids have a negative temperature coefficient of their refractive index ($dn/dT < 0$) the induced transverse profile of the refractive index acts as a negative (diverging) lens. The focal length of the lens depends on the parameters of the laser beam (power, diameter) and the characteristics of the medium (absorption coefficient, thermal conductivity, thermal diffusivity, dn/dT, cell length). Calculations show that in benzene a 0.01-W laser beam of 0.1 cm radius induces a negative thermal lens with focal length -1.14×10^4 cm. If a beam of

the same power is focused to radius 0.01 cm, the focal length is -2.27×10^2 cm [2].

The absorption coefficient of a medium and the concentration of the absorbing species can be evaluated from the experimentally measured focal length value of an induced thermal lens. The simplest way to determine focal length is to measure the "far field" diameter of a laser beam with and without the sample. Since the optical strength of a thermal lens is not very high, the distance between the sample and the plane of measurement should be several meters. The maximum TLS signal can be detected optimally when the laser beam is focused just before the sample cell. This distance (the confocal length) is defined by the parameters of the laser beam and the focal length of the lens.

An optimal TLS scheme detects the fraction of laser beam power P passed through a pinhole in a screen, located in the farfield, being detected by a photomultiplier (or vacuum photodiode) and lock-in amplifier (Fig. 5). The TLS signal can be expressed as a normalized fractional change in the detected laser power:

$$[P(t = 0) - P(t = \infty)]/P(t = \infty)$$

where $P(t=0)$ is the detected part of a laser beam not affected by the thermal lens effect, and $P(t=\infty)$ is the equilibrium value of P in the presence of the thermal lens. The absorption coefficient of a sample can be directly evaluated from this ratio and the parameters of the sample and laser beam. Both single- and two-beam versions of TLS are widely used in analytical applications. In the latter version, one beam of appropriate power is used as the heating beam while the second, low intensity, stable probe beam is used for signal detection.

From the introduction of TLS, cw lasers with discrete wavelengths (e.g., He–Ne, He–Cd, Ar^+, Kr^+, CO_2) are the most widely used. These lasers have the best power stability and beam profile. In most cases, the absorption bands of the trace species of interest do not coincide with a discrete wavelength of the heating laser. Chemical pretreatment is necessary for formation of an analyte derivative with an appropriate absorption band. As in the case of LOAS, TLS has poor selectivity because of spectral interference of sample components. Selectivity depends mainly on the appropriate choice of reagent for analyte derivative formation.

The signal also depends on the solvent. Carbon tetrachloride, benzene, and acetone are preferred.

Table 4. Best detection limits for pure solutions and some examples of real sample analysis by TLS

Analyte	Solvent	Sample*	Laser	Detection limit, wt%
Co^{2+}	$CHCl_3$	PS	He – Ne	1.2×10^{-6}
Cu^{2+}	water	PS	He – Ne	3.8×10^{-6}
Fe^{2+}	$CHCl_3$	PS	Ar^+	1.7×10^{-10}
Nd^{3+}	D_2O	PS	dye	3.0×10^{-5}
Pr^{3+}	water	PS	dye	7.2×10^{-3}
Tb^{3+}	butanol	PS	Ar^+	7.0×10^{-6}
UO_2^{2+}	water	NW	dye	4.7×10^{-7}
U^{6+}	$HClO_4$	NW	dye	7.0×10^{-6}
Am^{3+}	$HClO_4$	PS	dye	7.0×10^{-7}
Pu^{4+}	HCO_3^-	PS	dye	2.5×10^{-6}
Po_4^{3-}	water – acetone	PS	dye	5.0×10^{-10}
AsO_3^{3-}	2-butanol	NW	He – Ne	5.0×10^{-8}
SO_3^{2-}	water	PS	Ar^+	1.3×10^{-7}
S^{2-}	2-butanol	water, gas	He – Ne	5.0×10^{-8}
I_2	CCl_4	PS	Ar^+	1.0×10^{-16}**
Formaldehyde	water	PS	dye	4.5×10^{-8}
Nitroaniline	methanol – water	PS	He – Cd	1.2×10^{-9}**
Trimethylpentane	CCl_4	PS	He – Ne	8.0×10^{-9}
Azulene	liquid CO_2	PS	Kr^+	3.0×10^{-10}
NO_2	air	air	Ar^+	5.0×10^{-7}
Methanol	N_2	PS	CO_2	1.2×10^{-6}
CCl_2Fe_2	Ar	PS	CO_2	1.0×10^{-7}
CS_2	Ar	PS	CO_2	1.0×10^{-7}

* PS = pure solution; NW = natural water.
** Absolute detection limit, g.

For example, the TLS signal of a given analyte in water is 45 times smaller than in carbon tetrachloride. Suspended dust and air bubbles in a sample may perturb the laser beam, affecting the signal. To minimize these perturbations, liquids should be filtered and allowed to equilibrate in the cell before beginning the experiment. Absorption by the cell windows followed by a heating of the sample should also be minimized, as well as the formation of interference fringes due to the reflections of the laser beam by the optical components of the apparatus.

Originally, TLS was expected to be an absolute (standardless) technique because of the direct relationship between absorption coefficient and the experimental parameters. It was later recognized that the accuracy of the experimental and molecular parameters is not sufficient for absolute quantitative evaluation of the trace concentration of an analyte; calibration by standard reference solutions is necessary for quantitative analysis.

The minimum absorption coefficients measured by TLS are of the order of $10^{-6} - 10^{-7}$ cm^{-1}, which corresponds, for strong absorbers ($\varepsilon \geq 10^5$ M^{-1} cm^{-1}), to minimal concentrations of ca. $10^{-11} - 10^{-12}$ M. Some recent data from real sample analysis by TLS are listed in Table 4.

22.4.4. Fluorescence Analysis of Organic Molecules in Solid Solutions [2]

Fluorescence is one of the most sensitive techniques for the detection of complex organic molecules such as aromatic hydrocarbons, heterocyclic compounds, porphyrins (including chlorophyll and its derivatives), organic dyes, etc. Its high sensitivity depends on the relatively high fluorescence quantum yields for these classes of molecules and well-developed methods of photon detection. If a pure, single-component solution is analyzed, extremely low LODs can be achieved, e.g., highly luminescent dyes in aqueous solution can be easily detected with LODs down to 10^{-13} M. For many organic compounds in liquid or solid solution form, single molecules can be detected by modern experimental techniques. In many cases, the main analytical problem is not sensitivity, but selectivity. At room temperature, the absorption and luminescence spectra of organic molecules consist typically of one or more bands whose spectral width is hundreds or thousands of wavenumbers; severe spectral interference is a key problem in the analysis of multicomponent mixtures.

The Shpol'skii method is a powerful improvement. The spectra of many organic molecules embedded in a crystalline matrix of short-chain *n*-

paraffins consist, at helium temperature, of well-defined quasi-lines with linewidth $2-20\ cm^{-1}$. In spite of many successful analytical applications the Shpol'skii technique is not universally applicable because many organic molecules have low solubility in n-paraffins and, moreover, many compounds in n-paraffin matrices have broad-band spectra even at helium temperature.

The discovery that in many cases the spectra of organic compounds at low temperature are inhomogeneously broadened was the next and very important step in the understanding of the spectroscopy of complex molecules. In most cases, the broad spectra have an intrinsic line structure, which can be revealed by selective laser excitation. This new technique, laser fine-structure selective spectroscopy of complex molecules in frozen solutions is often called site-selection spectroscopy (SSS). A physically more realistic term is energy selection spectroscopy. By now the SSS technique is so far developed that inhomogeneous broadening of organic molecules at low temperature can be eliminated by selective laser excitation [2].

The fundamentals of SSS are based on the theory of impurity centers in a crystal. The optical spectrum of an organic molecule embedded in a matrix is defined by electron – vibrational interaction with intramolecular vibrations (vibronic coupling) and interaction with vibrations of the solvent (electron – phonon coupling). Each vibronic band consists of a narrow zero-phonon line (ZPL) and a relatively broad phonon wing (PW). ZPL corresponds to a molecular transition with no change in the number of phonons in the matrix (an optical analogy of the resonance γ-line in the Mössbauer effect). PW is determined by a transition which is accompanied by creation or annihilation of matrix phonons. The relative distribution of the integrated intensity of a band between ZPL and PW is characterized by the Debye – Waller factor:

$$\alpha = I_{ZPL}/(I_{ZPL} + I_{PW})$$

and depends critically on the temperature and the strength of electron – phonon coupling. At liquid helium temperature, if this coupling is not too strong, the vibronic bands of an isolated impurity molecule consist of a strong, narrow ZPL and a broad, weak PW. The observed broad band of an assembly of organic molecules in solution at low temperature is an envelope of a great number of narrow ZPLs (accompanied by PWs) of individual

molecules which are surrounded by different local fields (mainly because of small variations of the relative orientation of an analyte molecule and the solvent molecules). Statistical variations cause shifts in ZPL positions and, hence, inhomogeneous broadening of a spectral band. A narrow-band laser excites only a small fraction of the molecules with an appropriate ZPL position. Thus, well-separated, narrow ZPLs, accompanied by PWs, are observed in the fluorescence spectra of a multicomponent cryogenic solid mixture. This effect is often called fluorescence line narrowing (FLN). Informative spectra may be obtained for many types of molecule (polar and nonpolar), in different media (glasses, polymers, and crystals). The efficiency of FLN for a particular compound is well defined by the factor α. The higher the value of α, the better is excitation selectivity. For most systems, α varies within the interval $0.2-0.8$. For compounds with $\alpha \ll 1$, selective excitation is inefficient. Examples of such molecules are ionic dyes (proflavines and rhodamines in ethanol).

Fluorescence line narrowing depends strongly on temperature. For most organic molecules, ZPL disappears completely at temperatures above $40-50$ K. This is the main disadvantage of FLN for analytical applications; the sample temperature should be below $15-30$ K.

The main components of FLN instruments for analysis are a cryogenic system, a tunable laser, a medium resolution spectrometer, and a recording system. The cryogenic system consists of a liquid He optical cryostat with a He circulation system. In some cases, commercially available closed-cycle cryostats can be used. The optical system consists of a sample holder, optics for the exciting beam input, and a high aperture collector of the fluorescent radiation.

The most intense $0-0$ ZPLs of organic molecules of interest occupy the spectral range $650-250$ nm. Thus, lasers for FLN should be tunable within this range. The laser linewidth should match the FLN vibronic linewidth of the analytes, typically of the order of $1-3\ cm^{-1}$. To avoid sample heating, saturation effects, and/or stimulated two-photon photoreactions, the average laser power should not exceed $1-10$ mW. The cw dye lasers, pumped by Ar^+ or Kr^+ ion lasers are relatively cheap and reliable sources for the visible spectrum. For UV, frequency-doubled dye lasers pumped by copper or quasi-continuous Nd-YAG lasers (typical repetition rates $1-10$ kHz) may be preferable.

As the distance between the wavelengths of excitation and fluorescence in FLN is ca. $100-200$ cm^{-1}, background scattering of the exciting laser radiation may be a serious problem. Therefore, high-quality double- or even triple-grating spectrometers are used in FLN analysis. New multichannel recording systems (photodiode arrays, image intensifiers with linear CCDs) are potentially very useful detectors for FLN.

Polycyclic aromatic hydrocarbons (PAH) and heterocyclic aromatic compounds are the most widely analyzed organic compounds by FLN techniques. Both types of compound are widespread in nature. Many have strong carcinogenic and mutagenic properties and these properties depend on the isomeric structure of the compound. Analysis of PAH in a glassy matrix by FLN has many advantages compared with other techniques; simple sample preparation, high optical quality of the frozen matrices, etc., provide higher sensitivity and selectivity. LODs down to $10^{-12} - 10^{-13}$ g/cm^3 have been achieved for single-component solutions. For example, a concentration of perylene in ethanol of 10^{-12} g/mL (ca. 10^5 molecules in the sample) was easily detected in a frozen solution by FLN. The calibration curve was linear over the concentration range $10^{-6} - 10^{-9}$ g/cm^3. Well-characterized fine-structure FLN spectra were detected and identified in an artificial mixture of 14 PAHs in a water–glycerol solution. In another experiment, a mixture of 7 PAHs with concentrations of individual components ranging from 10^{-7} to 10^{-9} g/cm^3 was quantitatively analyzed and all compounds were correctly identified.

In some cases, direct analysis of raw products (without sample pretreatment or dissolution) is possible. To avoid self-absorption and energy transfer effects, low analyte concentrations and low optical density of the sample are necessary. Several PAH were successfully detected in gasoline when a frozen sample ($T = 4.2$ K) was selectively excited by laser radiation of appropriate wavelength. The same sample excited by a conventional Hg-lamp exhibited a smooth, unresolved spectrum with no possibility of identifying different compounds. As an example, the active carcinogen 3,4-benzpyrene was easily detected in gasoline by FLN. The concentration of 10^{-7} g/mL was measured by the standard additions method. The estimated LOD depended on the background luminescence of a gasoline sample which is mainly determined by its composition. The LODs varied within $10^{-10} - 10^{-11}$ g/mL for different gasoline samples. By the same technique, perylene at concentration 5×10^{-8} g/mL was detected in solid paraffin.

22.5. Laser Ablation [3], [6], [11]

Laser ablation (LA) is a technique in which focused laser radiation is used to release material from a sample. The advantages of LA are now well known: reduced or no sample preparation is needed; conducting or non-conducting samples may be analyzed; spatial resolution up to a few micrometers can be obtained.

Different analytical techniques are used for detection of the elemental composition of the ablated material. The simplest one is direct detection of the emission of a plasma plume formed above a sample surface. This technique is generally referred to as LIBS (laser induced breakdown spectroscopy). Strong continuous background radiation of the hot plasma plume does not allow detection of atomic and ionic lines of specific elements during the first few hundred nanoseconds of plasma evolution. One can achieve reasonable signal-to-noise ratio for the measurement of atomic and ionic spectral line intensities by optimization of experimental parameters: type and pressure of a buffer gas, time delay, and distance from the sample surface for detection of line intensities. The LODs in (mg–µg)/g range have been realized for direct analysis of metals, glasses, and ceramics by LIBS techniques for a variety of trace elements. Better LODs were demonstrated when LA was followed by detection of the ablated atoms or ions by laser atomic absorption or fluorescence spectroscopy (LAAS, LIF). Even though matrix-effects are significant and detection limits are not as good as for LA-ICP-MS, LIBS has one major advantage. It allows fast and on-line analysis in industrial environment, including the nuclear industry. When fibers are used both for excitation and emission collection, a very robust system can be built that allows on-line quality control.

The intrinsic drawback of LIBS is a short duration (less than a few hundred microseconds) and significant non-stationary conditions of a laser plume. Much higher sensitivity was realized by transport of the ablated material into a secondary atomic reservoir such as an MIP. Owing to the much longer residence time of ablated atoms and ions in a stationary MIP (typically several ms as compared with at most a hundred microseconds in a laser plume) and due to additional excitation of

the radiating upper levels in the low pressure plasma, the line intensities of atoms and ions are greatly enhanced. Because of both factors the LODs of the LA-MIP technique have been improved by 1 – 2 orders of magnitude as compared with LIBS.

Recently, analytical capabilities of LIBS and LA-MIP-OES were noticeably improved by use of an advanced detection scheme based on an Echelle spectrometer combined with a high sensitivity ICCDs (intensified charge-coupled devices) detector. This detection scheme enables simultaneous detection of a large spectral range from the UV to NIR in a single laser shot. It allows estimation of temperature of a laser plume by constructing Boltzman plots and correction for plasma temperature variations. The advantages of this technique are: complete sample analysis in a single laser shot, improved accuracy and precision, the possibility to detect sample inhomogeneities.

The sensitivity, accuracy, and precision of solid sample analysis were greatly improved by coupling of LA with ICP-OES/MS. The ablated species are transported with a carrier gas (usually argon) into the plasma torch. Additional atomization, excitation and ionization of the ablated species in a stationary hot plasma provide a dramatic increase in the sensitivity of emission detection (LA-ICP-OES) or detection of ions (LA-ICP-MS). The efficiency of the transport of ablated species into an ICP strongly depends on the size of the particles. The optimal conditions for ablation in the case of LA-ICP differ significantly from the optimal conditions for LIBS because the efficient transport of the ablated matter to an ICP requires a fine aerosol (with solid particle diameters less than a few micrometers), whereas direct optical emission spectroscopy of the laser plume needs excited atoms and ions.

The LODs in the sub-μg/g – ng/g range were realized for different materials (metals, glasses, polymers) by LA-ICP-OES. Even higher sensitivity was achieved for LA-ICP-MS due to the high efficiency of ion collection and detection. Under optimal conditions of ablation (choice of gas, diameter of a crater, flow rate of a carrier gas) the LODs in the ng/g – pg/g range are now routinely realized for most elements of the periodic table. Today the LA-ICP-MS technique is accepted as a most powerful technique for direct analysis of solid samples.

As a result of its attractive characteristics, LA-ICP-MS is currently being used for a large variety of applications, such as the *in situ* trace element analysis of glasses, geological samples, metals, ceramic materials, polymers, and atmospheric particulates. Fingerprinting (characterization of the trace element composition) of diamonds and gold and glass and steel using LA-ICP-MS has been reported to be very useful for provenance determination and for forensic purposes, respectively.

So far the pulsed lasers with a pulse duration of some nanoseconds have been most widely used for LA – Nd: YAG (1064, 532, 354.7, 266 nm), excimer lasers XeCl (308 nm), KrF (248 nm). For these lasers the mechanisms responsible for the material removal are thermal melting and evaporation or some kind of explosive evaporation. As a result, so called fractional evaporation could occur — the elements with different melting and boiling temperatures evaporate from a melt at different rates. Because of this composition of a laser plume does not match the bulk composition. This was a serious problem as inadequate probing of a sample could not be corrected for by any sensitive detection scheme. The optimal conditions for laser ablation and advanced methods of data processing were found to avoid the problem of fractionation for different classes of samples in the case of nanosecond pulses.

The use of the lasers with femtosecond duration opens the way to overcome the problem of fractionation in the most radical way. The first generation of all-solid-state femtosecond lasers is now commercially available, although expensive. For such a fast energy deposition the non-thermal mechanisms of material removal (e.g., ion repulsions due to Coulomb forces) can be significant. The non-thermal character of material removal with fs pulses enables the depth profiling of a solid sample with a few nm resolution, which was impossible with ns pulses because of melting and mixing of different layers.

22.6. References

[1] D. S. Kliger (ed.): *Ultrasensitive Laser Spectroscopy*, Academic Press, New York 1983.

[2] V. S. Letokhov (ed.): *Laser Analytical Spectrochemistry*, Adam Hilger, Bristol 1986.

[3] J. Sneddon, T. L. Thiem, Y. Lee (eds.): *Lasers in Analytical Atomic Spectrometry*, John Wiley & Sons, New York 1997.

[4] F. P. Schäfer (ed.): *Dye Lasers*, Springer Verlag, Heidelberg 1973.

[5] M. Ch. Amann, J. Buus: *Tunable Laser Diodes*, Artech House, Boston 1998.

[6] B. Smith: "Lasers in Analytical Chemistry," *Trends Anal. Chem.* **17** (1998) nos. 8 + 9.

[7] S. Sjöström, *Spectrochimica Acta Rev.* **13** (1990) 407 – 465.

[8] K. Niemax: "Diode Laser Spectroscopy," *Spectrochim. Acta Rev.* **15** (1993) no. 5.

[9] H. I. Schiff, G. I. Mackay, J. Bechara, in M. W. Sigrist (ed.): *Air Monitoring by Spectroscopic Techniques,* John Wiley & Sons, New York 1994.

[10] S. Martellucci, A. N. Chester (eds.): *Optoelectronics for Environmental Science,* Plenum Press, London 1991.

[11] R. Russo: "Laser Ablation," *Appl. Spectr.* **49** (1995) no. 9, 14 A – 28 A.

23. X-Ray Fluorescence Spectrometry

RON JENKINS, International Centre for Diffraction Data, Newtown Square, Pennsylvania, United States

23.1. Introduction

X-ray spectrometric techniques provided important information for the theoretical physicist in the first half of the 19th century, and since the early 1950s they have found increasing use in the field of materials characterization. While most of the early work in X-ray spectrometry was carried out with electron excitation [1], today most stand-alone X-ray spectrometers use X-ray excitation sources rather than electron excitation. X-ray fluorescence spectrometry typically uses a polychromatic beam of short wavelength X-radiation to excite longer wavelength characteristic lines from the sample to be analyzed. Modern X-ray spectrometers use either the diffracting power of a single crystal to isolate narrow wavelength bands, or a proportional detector to isolate narrow energy bands, from the polychromatic radiation (including characteristic radiation) excited in the sample. The first of these methods is called wavelength dispersive spectrometry and the second, energy dispersive spectrometry. Because the relationship between emission wavelength and atomic number is known, isolation of individual characteristic lines allows the unique identification of an element, and elemental concentrations can be estimated from characteristic line intensities.

23.1.1. Properties of X Rays

X rays are a short wavelength form of electromagnetic radiation discovered by WILHELM ROENTGEN in the late 19th century [2]. When a high-energy electron beam is incident upon a specimen, one of the products of the interaction is the emission of a broad-wavelength band called the continuum, also referred to as white radiation or

bremsstrahlung. This white radiation is produced as the atomic electrons of the elements making up the specimen decelerate the impinging high-energy electrons. Characteristic X-ray photons are produced following the ejection of an inner orbital electron from an excited atom and subsequent transition of atomic orbital electrons from states of high to low energy.

A beam of characteristic X rays passing through matter is subject to three processes, absorption, scatter, and fluorescence. The absorption of X rays involves mainly inner orbital atomic electrons and varies as the third power of the atomic number of the absorber. X rays are scattered mainly by the loosely bound outer electrons of an atom. Scattering of X rays may be coherent (same wavelength) or incoherent (longer wavelength). Coherently scattered photons may undergo subsequent mutual interference, leading in turn to the generation of diffraction maxima. Fluorescence occurs when the primary X-ray photons are energetic enough to create electron vacancies in the specimen, leading in turn to the generation of secondary (fluorescence) radiation produced from the specimen. This secondary radiation is characteristic of the elements making up the specimen. The technique used to isolate and measure individual characteristic wavelengths following excitation by primary X-radiation is called X-ray fluorescence spectrometry.

23.1.2. Scattering and Diffraction

Scattering occurs when an X-ray photon interacts with the electrons of the target element. Where this interaction is elastic, i.e., no energy is lost in the collision process, the scattering is referred to as coherent (Rayleigh) scattering. Since no energy change is involved, the coherently scattered radiation retains exactly the same wavelength as that of the incident beam. It can also happen that the scattered photon gives up a small part of its energy during the collision. In this instance, the scatter is referred to as incoherent (Compton scattering).

X-ray diffraction is a combination of two phenomena — coherent scatter and interference. At any point where two or more waves cross one another, they are said to interfere. Under certain geometric conditions, wavelengths that are exactly in phase may add to one another, and those that are exactly out of phase may cancel each other. Under such conditions, coherently scattered

photons may constructively interfere with each other, giving diffraction maxima. Since a crystal lattice consists of a regular arrangement of atoms, with layers of high atomic density throughout the crystal structure, a crystal can be used to diffract X-ray photons. Since scattering occurs between impinging X-ray photons and the loosely bound outer orbital atomic electrons, when a monochromatic beam of radiation falls onto the high atomic density layers, scattering will occur. In order to satisfy the requirement for constructive interference, it is necessary that the scattered waves originating from the individual atoms, that is, the scattering points, be in phase with one another. The geometric conditions for this are that:

$$n \lambda = 2d \sin\theta \tag{1}$$

where λ is the wavelength, d the interplanar spacing of the crystal, and n is an integer. Equation (1) is a statement of Bragg's law. The diffraction phenomenon is the basis of wavelength dispersive spectrometry, since by using a crystal of fixed $2d$, each wavelength is diffracted at a unique diffraction (Bragg) angle θ. Hence, by measuring the diffraction angle, knowledge of the d-spacing of the analyzing crystal allows the determination of the wavelength.

23.1.3. Absorption of X Rays

When a beam of X-ray photons of intensity $I_0(\lambda)$ falls on a specimen, a fraction of the beam passes through, this fraction being:

$$I(\lambda) = I_0(\lambda) \exp(\mu \varrho x) \tag{2}$$

where μ is the mass attenuation coefficient of absorber for wavelength λ, ϱ is the density of the specimen, and x the distance traveled by the photons through the specimen. It can be seen from Equation (2) that a number $(I_0 - I)$ of photons has been lost in the absorption process. Although a significant fraction of this loss may be due to scatter, by far the greater loss is due to the photoelectric effect. Photoelectric absorption occurs at each of the energy levels of the atom; thus the total photoelectric absorption is determined by the sum of each individual absorption effect within a specific shell. Where the absorber is made up of a number of different elements, as is usually the case, the total absorption is the sum of the products of the individual elemental mass attenuation coefficients and the weight fractions of the respec-

tive elements. This product is referred to as the total matrix absorption. The value of the mass attenuation in Equation (2) is a function of both the photoelectric absorption and the scatter. However, the photoelectric absorption influence is usually large in comparison with the scatter, and to all intents and purposes the mass absorption coefficient is equivalent to the photoelectric absorption. A plot of the mass attenuation coefficient as a function of wavelength contains a number of discontinuities, called absorption edges, at wavelengths corresponding to the binding energies of the electrons in the various subshells. Between absorption edges, as the wavelength of the incident X-ray photons increases, the absorption increases. This effect is very important in quantitative X-ray spectrometry, because the intensity of a beam of characteristic photons leaving a specimen is dependent on the relative absorption effects of the different atoms making up the specimen. This is called a matrix effect and is one of the reasons why a curve of characteristic line intensity as a function of element concentration may not be a straight line.

23.1.4. X-ray Fluorescence

X-ray fluorescence spectrometry provides the means for the identification of an element by measurement of its characteristic X-ray emission wavelength or energy. The method allows the quantitative estimation of a given element by first measuring the emitted characteristic line intensity and then relating this intensity to elemental concentration. The inherent simplicity of characteristic X-ray spectra make the process of allocating atomic numbers to the emission lines relatively easy, and the chance of making a gross error is rather small. Within the range of the conventional spectrometer, each element gives, on average, only half a dozen lines. A further benefit of using the X-ray emission spectrum for qualitative analysis is that, because transitions arise from inner orbitals, the effect of chemical combination, or valence state is almost negligible.

23.2. Historical Development of X-ray Spectrometry

While the roots of the X-ray spectrometric method go back to the early part of this century,

it is only during the last thirty years or so that the technique has gained major significance as a routine means of elemental analysis. The first use of the X-ray spectrometric method dates back to the classic work of HENRY MOSELEY [3]. One of the first papers on the use of X-ray spectroscopy for real chemical analysis appeared in 1922, when HADDING [4] described its use for the analysis of minerals. The use of X rays, rather than electrons, to excite characteristic X-radiation [5] avoids the problem of heating of the specimen. Use of X-rays rather than electrons represented the beginnings of X-ray fluorescence as we know it today. The fluorescence method was first employed on a practical basis in 1928 by GLOCKER and SCHREIBER [6]. However, widespread use of the technique had to wait until the mid-1940s when X-ray fluorescence was rediscovered by FRIEDMAN and BIRKS [7]. The basis of their spectrometer was a diffractometer, originally designed for the orientation of quartz oscillator plates. A Geiger counter was used to measure the intensities of the diffracted characteristic lines, and quite reasonable sensitivity was obtained for a very large part of the atomic number range. The first commercial X-ray spectrometers became available in the early 1950s and, although these operated only with an air path, they were able to provide qualitative and quantitative information on all elements above atomic number 22 (titanium). Later versions allowed use of helium or vacuum paths that extended the lower atomic number cut-off.

23.3. Relationship Between Wavelength and Atomic Number

23.3.1. Characteristic Radiation

If a high-energy particle, such as an electron, strikes a bound atomic electron, and the energy of the particle is greater than the binding energy of the atomic electron, it is possible that the atomic electron will be ejected from its atomic position. The electron departs from the atom with a kinetic energy $E - \varphi$, equivalent to the difference between the energy E of the initial particle and the binding energy φ of the atomic electron. When the exciting particles are X-ray photons, the ejected electron is called a photoelectron, and the interaction between primary X-ray photons and atomic elec-

Table 1. Atomic structures of the first three principal shells

Shell (electrons)	n	m	l	s	Orbitals	J
$K(2)$	1	0	0	$\pm\frac{1}{2}$	$1s$	$\frac{1}{2}$
$L(8)$	2	0	0	$\pm\frac{1}{2}$	$2s$	$\frac{1}{2}$
	2	0	0	$\pm\frac{1}{2}$	$2p$	$\frac{1}{2}$
	2	-1	1	$\pm\frac{1}{2}$	$2p$	$\frac{1}{2},\frac{3}{2}$
	2	$+1$	1	$\pm\frac{1}{2}$	$2p$	$\frac{1}{2},\frac{3}{2}$
$M(18)$	3	0	0	$\pm\frac{1}{2}$	$3s$	$\frac{1}{2}$
	3	1	0	$\pm\frac{1}{2}$	$3p$	$\frac{1}{2}$
	3	-1	1	$\pm\frac{1}{2}$	$3p$	$\frac{1}{2},\frac{3}{2}$
	3	$+1$	1	$\pm\frac{1}{2}$	$3p$	$\frac{1}{2},\frac{3}{2}$
	3	0	0	$\pm\frac{1}{2}$	$3d$	$\frac{1}{2}$
	3	-1	1	$\pm\frac{1}{2}$	$3d$	$\frac{1}{2},\frac{3}{2}$
	3	$+1$	1	$\pm\frac{1}{2}$	$3d$	$\frac{1}{2},\frac{3}{2}$
	3	-2	2	$\pm\frac{1}{2}$	$3d$	$\frac{3}{2},\frac{5}{2}$
	3	$+2$	2	$\pm\frac{1}{2}$	$3d$	$\frac{3}{2},\frac{5}{2}$

Figure 1. Observed lines in the K series (asterisks indicate ionized states)

trons is called the photoelectric effect. As long as the vacancy in the shell exists, the atom is in an unstable state, and there are two processes by which it can revert to its original state. The first of these involves a rearrangement that does not result in the emission of X-ray photons, but in the emission of other photoelectrons from the atom. This is known as the Auger effect [8], and the emitted photoelectrons are called Auger electrons.

The second process by which the excited atom can regain stability is by transfer of an electron from one of the outer orbitals to fill the vacancy. The energy difference between the initial and final states of the transferred electron may be given off in the form of an X-ray photon. Since all emitted X-ray photons have energies proportional to the differences in the energy states of atomic electrons, the lines from a given element are characteristic of that element. The relationship between the wavelength of a characteristic X-ray photon and the atomic number Z of the element was first established by MOSELEY. Moseley's law is written:

$$1/\lambda = K(Z - \sigma)^2 \qquad (3)$$

in which K is a constant that takes different values for each spectral series; σ is the shielding constant with a value just less than unity. The wavelength of the X-ray photon λ in Å is inversely related to the energy E of the photon in keV according to the relationship:

$$\lambda = 12.4/E \qquad (4)$$

Since there are two competing routes by which an atom can return to its initial state, and since only

one of these processes produces a characteristic X-ray photon, the intensity of an emitted characteristic X-ray beam depends on the relative effectiveness of the two processes within a given atom. As an example, the number of quanta of K series radiation emitted per ionized atom is a fixed ratio for a given atomic number, this ratio being called the fluorescent yield. The fluorescent yield varies as the fourth power of atomic number and approaches unity for higher atomic numbers. Fluorescent yield values are several orders of magnitude less for very low atomic numbers.

23.3.2. Selection Rules

An excited atom can revert to its ground state by transferring an electron from an outer atomic level to fill the vacancy in the inner shell. An X-ray photon is emitted from the atom as part of this de-excitation step, the emitted photon having an energy equal to the energy difference between the initial and final states of the transferred electron. The selection rules for the production of normal (diagram) lines require that the principal quantum number n must change by at least one, the angular quantum number l must change by only one, and the J quantum number [the total momentum J of an electron is given by the vector sum of $(l+s)$ where s is the spin quantum number] must change by zero or one (Table 1). Transition groups can now be constructed, based on the appropriate number of transition levels. Figure 1 shows the lines that are observed in the K series. Three groups of lines are indicated with the normal lines being shown on the left-hand side of the
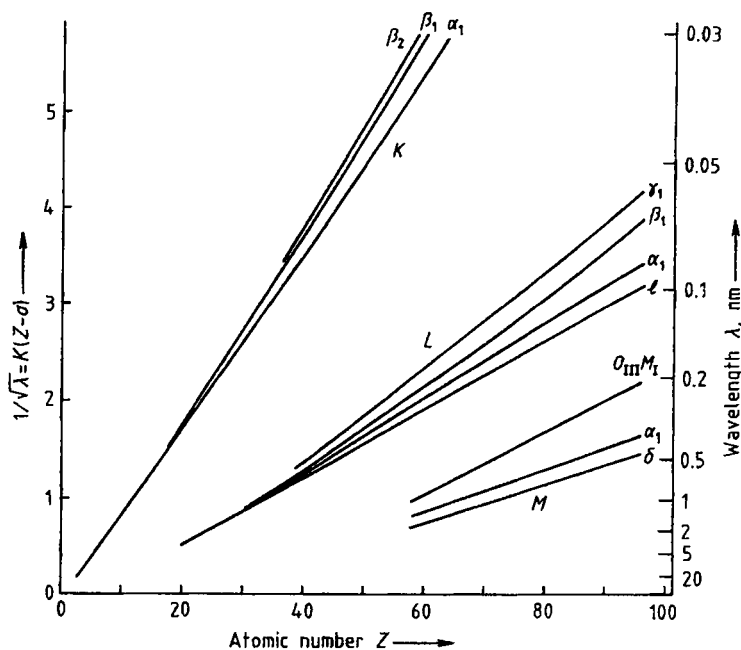
Figure 2. Moseley diagrams for K, L, and M series

figure. While most of the observed fluorescent lines are normal, certain lines may also occur in X-ray spectra that at first sight do not fit the basic selection rules. These lines are called forbidden lines and are shown in the center portion of the figure. Forbidden lines typically arise from outer orbital levels where there is no sharp energy distinction between orbitals. A third type of line may also occur, namely, satellite lines, which arise from dual ionizations. Neither forbidden transitions nor satellite lines have great analytical significance; they may cause confusion in qualitative interpretation of spectra, and may even be misinterpreted as coming from trace elements.

23.3.3. Nomenclature for X-ray Wavelengths

The classical nomenclature system for the observed lines is that proposed by SIEGBAHN in the 1920s. Figure 2 shows plots of the reciprocal of the square root of the wavelength as a function of atomic number for the K, L, and M series. As indicated by Moseley's law (Eq. 3), such plots should be linear. A scale directly in wavelength is also shown, to indicate the range of wavelengths over which a given series occurs. In practice, the

number of lines observed from a given element depends on the atomic number of the element, the excitation conditions, and the wavelength range of the spectrometer. Generally, commercial spectrometers cover the K series, the L series, corresponding to transitions to K, L, and M levels respectively. There are a much larger number of lines in the higher series and for a detailed list of all of the reported wavelengths the reader is referred to the work of BEARDEN [9]. In X-ray spectrometry most analytical work is carried out using either the K or the L series wavelengths. The M series may also be useful, especially in the measurement of higher atomic numbers.

23.4. Instrumentation

23.4.1. Sources

Several different types of source have been employed for the excitation of characteristic X-radiation, including those based on electrons, X rays, γ rays, protons, and synchrotron radiation. Sometimes a bremsstrahlung X-ray source is used to generate specific X-radiation from an intermediate pure element sample, called a secondary flu-

orescer. By far the most common source today is the X-ray photon source. This source is used in primary mode in the wavelength and primary energy dispersive systems, and in secondary fluorescer mode in secondary target energy dispersive spectrometers. A γ-source is typically a radioisotope that is used either directly, or in a mode equivalent to the secondary fluorescer mode in energy dispersive spectrometry. Most conventional wavelength dispersive X-ray spectrometers use a high-power (2 – 4 kW) X-ray bremsstrahlung source. Energy dispersive spectrometers use either a high- or low-power (0.5 – 1.0 kW) primary source, depending on whether the spectrometer is used in the secondary or primary mode. In all cases, the primary source unit consists of a very stable high-voltage generator, capable of providing a potential of typically 40 – 100 kV. The current from the generator is fed to the filament of the X-ray tube, which is typically a coil of tungsten wire. The applied current causes the filament to glow and emit electrons. A portion of this electron cloud is accelerated to the anode of the X-ray tube, which is typically a water-cooled block of copper with the required anode material plated or cemented to its surface. The impinging electrons produce X-radiation, a significant portion of which passes through a thin beryllium window to the specimen.

23.4.2. Detectors

An X-ray detector is a transducer for converting X-ray photon energy into voltage pulses. Detectors work by a process of photoionization, in which interaction between the incoming X-ray photon and the active detector material produces a number of electrons. The current produced by these electrons is converted to a voltage pulse by a capacitor and resistor, such that one digital voltage pulse is produced for each entering X-ray photon. In addition to being sensitive to the appropriate photon energies, i.e., being applicable to a given range of wavelengths or energies, there are two other important properties that an ideal detector should possess: proportionality and linearity. Each X-ray photon entering the detector produces a voltage pulse, and if the size of the voltage pulse is proportional to the photon energy, the detector is said to be proportional. Proportionality is needed where the technique of pulse height selection is to be used. Pulse height selection is a means of electronically rejecting pulses of voltage levels

other than those corresponding to the characteristic line being measured. X-ray photons enter the detector at a certain rate and, if the output pulses are produced at the same rate, the detector is said to be linear. Linearity is important where the various count rates produced by the detector are to be used as measures of the photon intensity for each measured line.

In the case of wavelength dispersive spectrometers, the gas flow proportional counter is generally employed for the measurement of longer wavelengths, and the scintillation counter for shorter wavelengths. Neither of these detectors has sufficient resolution to separate multiple wavelengths on its own, and so they have to be employed along with an analyzing crystal. However, in the case of energy dispersive spectrometry, where no dispersing crystal is employed, a detector of higher resolution must be used, generally the Si(Li) detector.

The Si(Li) detector consists of a small cylinder (ca. 1 cm diameter and 3 mm thick) of p-type silicon, compensated by lithium to increase its electrical resistivity. A Schottky barrier contact on the front of the silicon disk produces a p-i-n-type diode. To inhibit the mobility of the lithium ions and to reduce electronic noise, the diode and its preamplifier are cooled to liquid-nitrogen temperature. Incident X-ray photons interact to produce a specific number of electron – hole pairs. The charge produced is swept from the diode by the bias voltage to a charge-sensitive preamplifier. A charge loop integrates the charge on a capacitor to produce an output pulse. Special energy dispersive detectors have been used in miniaturized instrumentation for in vivo applications [10] and for in situ analysis [11].

23.4.3. Types of Spectrometer

The basic function of the spectrometer is to separate the polychromatic beam of radiation coming from the specimen in order that the intensity of each individual characteristic line can be measured. A spectrometer should provide sufficient resolution of lines to allow such data to be taken, at the same time providing a sufficiently large response above background to make the measurements statistically significant, especially at low analyte concentrations. It is also necessary that the spectrometer allow measurements over the wavelength range to be covered. Thus, in the selection of a set of spectrometer operating vari-

ables, four factors are important: resolution, response, background level, and range. Owing to many factors, optimum selection of some of these characteristics may be mutually exclusive; as an example, attempts to improve resolution invariably cause lowering of absolute peak intensities. There is a wide variety of instrumentation available today for the application of X-ray fluorescence techniques and it is useful to break the instrument types down into three main categories: wavelength dispersive spectrometers (sequential and simultaneous); energy dispersive spectrometers (primary or secondary); and special spectrometers (including total reflection, synchrotron source, and proton-induced). The wavelength dispersive system was introduced commercially in the early 1950s and probably around 25 000 such instruments have been supplied commercially. Energy dispersive spectrometers became commercially available in the early 1970s and today there are around 15 000 of these units in use. There are far fewer of the specialized spectrometers in use and these generally incorporate an energy dispersive rather than a wavelength dispersive spectrometer.

X-ray spectrometers differ in the number of elements measurable at one time, and the speed at which they collect data. All of the instruments are, in principle at least, capable of measuring all elements from $Z = 9$ (fluorine) upwards, and most modern wavelength dispersive spectrometers can do useful measurements down to $Z = 6$ (carbon). Most systems can be equipped with multisample handling facilities and can be automated by use of minicomputers. Typical spectrometer systems are capable of precision on the order of a few tenths of a percent, with sensitivities down to the low ppm level. Single-channel wavelength dispersive spectrometers are typically employed for routine and nonroutine analysis of a wide range of products, including ferrous and nonferrous alloys, oils, slags and sinters, ores and minerals, thin films, etc. These systems are very flexible, but somewhat slower than multichannel spectrometers. Multichannel wavelength dispersive instruments are used almost exclusively for routine, high-throughput analyses where the great need is for fast, accurate analysis, but where flexibility is of no importance. Energy dispersive spectrometers have the great advantage of being able to display information on all elements at the same time. They lack somewhat in resolution compared with wavelength dispersive spectrometers, but the ability to reveal elements absent, as well as elements present, makes the energy dispersive spectrometer ideal for general trouble-shooting problems. They have been particularly effective in the fields of scrap alloy sorting, in forensic science, and in the provision of elemental data to supplement X-ray powder diffraction data.

23.4.4. Wavelength Dispersive Systems

A wavelength dispersive spectrometer may be a single-channel instrument in which a single crystal and a single detector are used for the sequential measurement of a series of wavelengths; or a multichannel spectrometer in which many crystal – detector sets are used to measure elements simultaneously. Of these two basic types, the sequential systems are the most common. A typical sequential spectrometer system consists of the X-ray tube, a specimen holder, a primary collimator, an analyzing crystal, and a detector. A portion of the characteristic fluorescence radiation from the specimen is passed via a collimator or slit onto the surface of an analyzing crystal, where individual wavelengths are diffracted to the detector in accordance with the Bragg law. A goniometer is used to maintain the required $\theta - 2\theta$ relationship between crystal and detector. Typically, about six different analyzing crystals and two different collimators are provided in this type of spectrometer, giving the operator a wide choice of dispersion conditions. The separating power of a crystal spectrometer depends on the divergence allowed by the collimators, which to a first approximation determines the width of the diffracted lines, and the angular dispersion of the crystal [12]. While the maximum wavelength covered by traditional spectrometer designs is about 2 nm, recent developments allow extension of the wavelength range significantly beyond this value.

The output from a wavelength dispersive spectrometer may be either analog (a) or digital (d). For qualitative work, an analog output is traditionally used; the digital output from the detector amplifier is fed through a d/a converter, called a rate meter, to an x/t recorder synchronously coupled with the goniometer scan speed. The recorder thus records an intensity – time diagram, in terms of an intensity – 2θ diagram. It is generally more convenient to employ digital counting for quantitative work, and a timer – scaler combination is provided which allows pulses to be integrated over a period of several tens of seconds and then displayed as count or count rate. In more modern spectrom-

eters, a scaler – timer may take the place of the rate meter, using step-scanning. Some automated wavelength dispersive spectrometers provide the user with software programs for the interpretation and labeling of peaks [13].

23.4.5. Energy Dispersive Systems

The energy dispersive spectrometer consists of the excitation source and the spectrometer/detection system. The spectrometer/detector is typically a Si(Li) detector. A multichannel analyzer is used to collect, integrate, and display the resolved pulses. While similar properties are sought from the energy dispersive system as with the wavelength dispersive system, the means of selecting these optimum conditions are very different. Since the resolution of the energy dispersive system equates directly to the resolution of the detector, this feature is of paramount importance. The output from an energy dispersive spectrometer is generally displayed on a cathode ray tube, and the operator is able to dynamically display the contents of the various channels as an energy spectrum. Provision is generally made to allow zooming in on portions of the spectrum of special interest, to overlay spectra, to subtract background, etc. Generally, some form of minicomputer is available for spectral stripping, peak identification, quantitative analysis, and other useful functions.

All of the earlier energy dispersive spectrometers were operated in what is called the primary mode and consisted simply of the excitation source, typically a closely coupled low-power, end-window X-ray tube, and the detection system. In practice there is a limit to the maximum count rate that the spectrometer can handle and this led, in the mid-1970s, to the development of the secondary mode of operation. In the secondary mode, a carefully selected pure element standard is interposed between primary source and specimen, along with absorption filters, such that a selectable energy range of secondary photons is incident on the sample. This allows selective excitation of certain portions of the energy range, thus increasing the ratio of useful to unwanted photons entering the detector. While this configuration does not completely eliminate the count rate and resolution limitations of the primary system, it certainly does reduce them.

23.5. Accuracy

23.5.1. Counting Statistical Errors

The production of X rays is a random process that can be described by a Gaussian distribution. Since the number of photons counted is nearly always large, typically thousands or hundreds of thousands, rather than a few hundred, the properties of the Gaussian distribution can be used to predict the probable error for a given count measurement. The random error $\sigma(N)$ associated with a measured value of N, is equal to $\sqrt{N}$. For example, if 10^6 counts are taken, the standard deviation σ is $\sqrt{10^6} = 10^3$, or 0.1 %. The measured parameter in wavelength dispersive X-ray spectrometry is generally the counting rate R, and the magnitude of the random counting error associated with a given datum can be expressed as:

$$\sigma(\%) = 100/\sqrt{N} = 100/\sqrt{(Rt)} \qquad (5)$$

Care must be exercised in relating the counting error (or indeed any intensity-related error) with an estimate of the error in terms of concentration. Provided that the sensitivity of the spectrometer in count rate (c/s) per percent analyte is linear, the count error can be directly related to a concentration error. However, where the sensitivity of the spectrometer changes over the range of measured response, a given fractional count error may be much greater when expressed in terms of concentration. In many analytical situations, the peak lies above a significant background and this adds a further complication to the counting statistical error.

23.5.2. Matrix Effects

In the conversion of net line intensity to analyte concentration, it may be necessary to correct for absorption and/or enhancement effects. Absorption effects include both primary and secondary absorption. Primary absorption occurs because all atoms of the specimen matrix absorb photons from the primary source. Since there is competition for these primary photons by the atoms making up the specimen, the intensity — wavelength distribution of the photons available for excitation of a given analyte element may be modified by other matrix elements. Secondary absorption refers to the absorption of characteristic analyte radiation by the specimen matrix. As characteristic

radiation passes out from the specimen in which it was generated, it is absorbed by all matrix elements by amounts relative to their mass attenuation coefficients. The total absorption of a specimen depends on both primary and secondary absorption. Enhancement effects occur when a nonanalyte matrix element emits a characteristic line, of energy just in excess of the absorption edge of the analyte element. This means that the nonanalyte element is able to excite the analyte, giving characteristic photons in addition to those produced by the primary continuum. This gives an enhanced signal from the analyte.

23.6. Quantitative Analysis

The simplest quantitative analysis situation to handle is the determination of a single element in a known matrix. In this instance, a simple calibration curve of analyte concentration versus line intensity is sufficient for quantitative determination. A slightly more difficult case might be the determination of a single element where the matrix is unknown. Three basic methods are commonly employed in this situation: use of internal standards, addition of standards, and use of a scattered line from the X-ray source. The most complex case is the analysis of all, or most, of the elements in a sample, about which little or nothing is known. In this case a full qualitative analysis would be required before any attempt is made to quantitate the matrix elements. Once the qualitative composition of the sample is known, one of three general techniques is typically applied: type standardization, influence coefficient methods, or fundamental parameter techniques. Both the influence coefficient and fundamental parameter techniques require a computer for their application.

In principle, an empirical correction procedure can be described as the correction of an analyte element intensity for the influence of interfering elements, using the product of the intensity from the interfering element line and a constant factor, as the correction term [14]. This constant factor is today generally referred to as an influence coefficient, since it is assumed to represent the influence of the interfering element on the analyte. Commonly employed influence coefficient methods may use either the intensity or the concentration of the interfering element as the correction term. These methods are referred to as intensity correc-

tion and concentration correction methods, respectively. Intensity correction models give a series of linear equations which do not require much computation, but they are generally not applicable to wide ranges of analyte concentration. Various versions of the intensity correction models found initial application in the analysis of nonferrous metals where correction constants were applied as look-up tables. Later versions [15] were supplied on commercially available computer-controlled spectrometers, and were used for a wider range of application. A concentration model (e.g., [16]) requires the solving of a series of simultaneous equations by regression analysis or matrix inversion techniques. This approach is more rigorous than the intensity models, and so it became popular in the early 1970s as suitable low-cost minicomputers became available.

23.6.1. Internal Standards

One of the most useful techniques for the determination of a single analyte element in a known or unknown matrix is to use an internal standard. The technique is one of the oldest methods of quantitative analysis and is based on the addition of a known concentration of an element that gives a wavelength close to that of the analyte. The assumption is made that the effect of the matrix on the internal standard is essentially the same as the effect of the matrix on the analyte element. Internal standards are best suited to the measurement of analyte concentrations below about 10 %. Care must also be taken to ensure that the particle sizes of specimen and internal standard are about the same, and that the two components are adequately mixed. Where an appropriate internal standard cannot be found, it may be possible to use the analyte itself as an internal standard. This method is a special case of standard addition, and it is generally referred to as spiking.

23.6.2. Type Standardization

Provided that the total specimen absorption does not vary significantly over a range of analyte concentrations, and provided that enhancement effects are absent, and that the specimen is homogeneous, a linear relationship will be obtained between analyte concentration and measured characteristic line intensity. Where these provisos are met, type standardization techniques can be employed. Type standardization is probably the old-

est of the quantitative analytical methods employed, and the method is usually evaluated by taking data from a well-characterized set of standards and, by inspection, establishing whether a linear relationship is indeed observed. Where this is not the case, the analyte concentration range may be further restricted. The analyst of today is fortunate in that many hundreds of good reference standards are commercially available. While the type standardization method is not without its pitfalls, it is nevertheless extremely useful, especially for quality control applications where a finished product is compared with a desired product.

23.6.3. Influence Correction Methods

Influence coefficient correction procedures can be divided into three basic types: fundamental, derived, and regression. Fundamental models are those which require starting with concentrations, then calculating the intensities. Derived models are based on some simplification of a fundamental method but still allow concentrations to be calculated from intensities. Regression models are semiempirical in nature and allow the determination of influence coefficients by regression analysis of data sets obtained from standards. All regression models have essentially the same form and consist of: a weight fraction term W (or concentration C); an intensity (or intensity ratio) term I; an instrument-dependent term which essentially defines the sensitivity of the spectrometer for the analyte in question; and a correction term, which corrects the instrument sensitivity for the effect of the matrix. The general form is:

$$W = I(\text{instrument})[1 + \{\text{model}\}] \qquad (6)$$

Equation (6) shows that the weight fraction W of analyte is proportional to the product of the measured intensity I of the analyte, corrected for instrumental effects, and a matrix correction term. Different models vary only in the form of this correction term.

The major advantage of influence coefficient methods is that a wide range of concentration ranges can be covered by using a relatively inexpensive computer for the calculations. A major disadvantage is that a large number of well-analyzed standards may be required for the initial determination of the coefficients. However, where adequate precautions have been taken to ensure correct separation of instrument- and matrix-dependent terms, the correction constants are transportable from one spectrometer to another and, in principle, need only be determined once.

23.6.4. Fundamental Methods

Since the early work of SHERMAN, there has been a growing interest in the provision of an intensity – concentration algorithm which would allow the calculation of the concentration values without recourse to the use of standards. Sherman's work was improved upon, first by the Japanese team of SHIRAIWA and FUJINO [17] and later, by the Americans, CRISS and BIRKS [18], [19] with their program NRLXRF. The same group also solved the problem of describing the intensity distribution from the X-ray tube [20]. The problem for the average analyst in the late 1960s and early 1970s, however, remained that of finding sufficient computational power to apply these methods. In the early 1970s, DE JONG suggested an elegant solution [21], in which he proposed the use of a large main-frame computer for the calculation of the influence coefficients, then use of a minicomputer for their actual application using a concentration correction influence model.

23.7. Trace Analysis

One of the problems with any X-ray spectrometer system is that the absolute sensitivity (i.e., the measured c/s per % of analyte element) decreases significantly as the lower atomic number region is approached. The three main reasons for this are: first, the fluorescence yield decreases with atomic number; second, the absolute number of useful long-wavelength X-ray photons from a bremsstrahlung source decreases with increasing wavelength; and third, absorption effects generally become more severe with increasing wavelength of the analyte line. The first two of these problems are inherent to the X-ray excitation process and to constraints in the basic design of conventional X-ray tubes. The third, however, is a factor that depends very much on the instrument design and, in particular, upon the characteristics of the detector.

23.7.1. Analysis of Low Concentrations

The X-ray fluorescence method is particularly applicable to the qualitative and quantitative analysis of low concentrations of elements in a wide range of samples, as well as allowing the analysis of elements at higher concentrations in limited quantities of materials. A measurement of a line at peak position gives a counting rate which, in those cases where the background is insignificant, can be used as a measure of the analyte concentration. However, where the background is significant, the measured value of the analyte line at the peak position now includes a count-rate contribution from the background. The analyte concentration in this case is related to the net counting rate. Since both peak and background count rates are subject to statistical counting errors, the question now arises as to the point at which the net peak signal is statistically significant. The generally accepted definition for the lower limit of detection is that concentration is equivalent to two standard deviations of the background counting rate. A formula for the lower limit of detection LLD can now be derived [22]:

$$LLD = 3/m\sqrt{R_b/t_b} \qquad (7)$$

where R_b is the count rate at the background angle, t_b the time spent counting the background, and m the sensitivity of the spectrometer.

It is important to note that not only does the sensitivity of the spectrometer vary significantly over the wavelength range of the spectrometer, but so does the background counting rate. In general, the background varies by about two orders of magnitude over the range of the spectrometer. From Equation (7), the detection limit is optimal when the sensitivity is high and the background is low. Both the spectrometer sensitivity and the measured background vary with the average atomic number of the sample. While detection limits over most of the atomic number range lie in the low ppm range, the sensitivity of the X-ray spectrometer falls off quite dramatically towards the long-wavelength limit of the spectrometer due mainly to low fluorescence yields and the increased influence of absorption. As a result, poorer detection limits are found at the long-wavelength extreme of the spectrometer, which corresponds to lower atomic numbers. Thus, the detection limits for elements such as fluorine and sodium are at the levels of hundredths of one percent rather than parts per million. The detection limits for very low atomic number elements such as carbon ($Z = 6$) and oxygen ($Z = 8$) are very poor, typically on the order of 3 – 5 %.

23.7.2. Analysis of Small Amounts of Sample

Conventional X-ray fluorescence spectrometers are generally designed to handle rather large specimens with surface areas on the order of several square centimeters, and problems occur if the sample to be analyzed is limited in size. The modern wavelength dispersive system is especially inflexible in the area of sample handling, mainly because of geometric constraints arising from the need for close coupling of the sample to the X-ray tube and the need to use an airlock to bring the sample into the working vacuum. The sample to be analyzed is typically placed inside a cup of fixed external dimensions that is, in turn, placed in the carousel. This presentation system places constraints not only on the maximum dimensions of the sample cup, but also on the size and shape of samples that can be placed into the cup itself. Primary source energy dispersive systems do not require the same high degree of focusing, and to this extent are more easily applicable to any sample shape or size, provided that the specimen fits into the radiation-protected chamber. In some instances, the spectrometer can even be brought to the object to be analyzed. Because of this flexibility, analyses of odd-shaped specimens have been almost exclusively the purview of the energy dispersive systems.

In the case of secondary target energy dispersive systems, while the geometric constraints are still less severe than those of the wavelength system, they are much more critical than in the case of primary systems. Where practicable, the best solution for the handling of limited amounts of material is invariably found in one of the specialized spectrometer systems.

23.8. New developments in Instrumentation and Techniques

23.8.1. Total Reflection Spectrometry

One of the major problems that inhibits the obtaining of good detection limits in small samples is the high background due to scatter from the

sample substrate support material. The suggestion to overcome this problem by using total reflection of the primary beam was made as long ago as 1971 [23], but the absence of suitable instrumentation prevented real progress being made until the late 1970s [24], [25]. The TRXRF method is essentially an energy dispersive technique in which the Si(Li) detector is placed close to (ca. 5 mm) and directly above the sample. Primary radiation enters the sample at a glancing angle of a few seconds of arc. The sample itself is typically presented as a thin film on the optically flat surface of a quartz plate. A beam of radiation from a sealed X-ray tube passes through a fixed aperture onto a pair of reflectors placed very close to each other. Scattered radiation passes through the first aperture to impinge on the sample at a very low glancing angle. Because the primary radiation enters the sample at an angle barely less than the critical angle for total reflection, this radiation barely penetrates the substrate medium; thus scatter and fluorescence from the substrate are minimal. Because the background is so low, picogram amounts can be measured or concentrations in the range of a few tenths of a ppb. can be obtained without recourse to pre-concentration [26]. Ideally, the radiation falling onto the specimen surface should be monochromatic. Unfortunately, in earlier executions of TXRF spectrometers the use of monochromatic source radiation was not possible because of the high intensity loss. This problem was solved with he advent of multilayer monochromators, since these exhibit high reflectivity and produce an intense almost parallel beam [27]. A TXRF spectrometer has been designed to operate with a standard X-ray tube and a multilayer monochromator, and this was found to give a two- to threefold increase in intensity over the normal geometry [28].

23.8.2. Spectrometers with Capillary Optics

Curved glass capillaries with channel diameters on the order of a few microns can efficiently scatter X rays through large angles by multiple grazing incidence reflection. This technology generates a very high specific intensity, yielding an increase in intensity of two to three orders of magnitude [29]. The use of capillary optics has opened up tremendous possibilities in energy dispersive X-ray fluorescence. A new type of X-ray spectrometer has now evolved, broadly categorized as microscopic X-ray fluorescence (MXRF). Earlier versions of MXRF systems simply used an aperture X-ray guide to control the size of the irradiated spot. Use of a monolithic polycapillary optic in place of the original aperture was found to provide more than two orders of magnitude increase in intensity [30]. There are two basic types of MXRF based on capillary optics: the first is based on classical sealed X-ray tube sources, and the second on the use of the synchrotron. As an example, one low-power, portable XRF analyzer, based on this technology [31], allowed the analysis of specimens as small as 30 mm in diameter with detection limits down to 5×10^{-5} %. GORMLEY [32] has described a series of experiments, based on the use of a CCD camera, to better quantify the focusing and transmission characteristics of these capillaries. Another instrument was designed with an 80 mm^2 detector [33] and allowed the analysis of small (5 mm) areas of bulky objects. Each of these spectrometers was designed to work with low-power X-ray tubes. At the other end of the source power scale, VEKEMANS et al. described an instrument based on a rotating anode X-ray tube. This high-efficiency, microscopic X-ray fluorescence system has been utilized for the analysis of 16-20th century brass statues originating from Northern India and Nepal and for the imaging analysis of enamel paints on 17th century Chinese porcelain [34].

23.8.3. Applications of the Synchrotron

The availability of intense, linearly polarized synchrotron radiation beams [35] has prompted workers in the fields of X-ray fluorescence [36] to explore what the source has to offer over more conventional excitation media. In the synchrotron, electrons with kinetic energies on the order of several billion electron volts (typically 3 GeV at this time) orbit in a high-vacuum tube between the poles of a strong (ca. 1 T) magnet. A vertical field accelerates the electrons horizontally, causing the emission of synchrotron radiation. Thus synchrotron source radiation can be considered as magnetic bremsstrahlung as opposed to the normal electronic bremsstrahlung that is produced when primary electrons are decelerated by the electrons of an atom. It has been found that because the primary source of radiation is so intense, it is possible to use a high degree of monochromatization between source and specimen, giving a source that is wavelength- and, therefore, energy-tunable,

as well as being highly monochromatic. There are several different excitation modes that can be used for synchrotron-source X-ray fluorescence spectroscopy (SSXRF) including direct excitation with continuum, excitation with absorber-modified continuum, excitation with source crystal monochromatized continuum, excitation with source radiation scattered by a mirror, and reflection and transmission modes.

Most SSXRF is carried out in the energy dispersive mode. Since all solid-state detectors are count-rate limited, the tremendous radiation flux available from the synchrotron is, in general, not directly useful. The real power of the synchrotron comes from the use of highly conditioned beams, which provide a highly focused, monochromatic beam at the sample. A newer development in the area of beam conditioning for the synchrotron is the use of laterally graded multilayers [37]. The use of X-ray optical multilayers in X-ray instrumentation became widespread in the 1990s, and newer technology based on pulsed laser deposition is resulting in the production of multilayers of uniform thickness with a minimum of columnar thin-film growth. The intensity of the synchrotron beam is probably four to five orders of magnitude greater than the conventional bremsstrahlung source sealed X-ray tubes. This, in combination with its energy tunability and polarization in the plane of the synchrotron ring, allows very rapid bulk analyses to be obtained on small areas. Because the synchrotron beam has such a high intensity and small divergence, it is possible to use it as a microprobe of high spacial resolution (about 10 μm). Absolute limits of detection around 10^{-14} μg have been reported for such an arrangement [38]. Use of monolithic polycapillary X-ray lenses has added an even wider dimension to synchrotron microbeam analysis, and beam sizes down to 4–10 μm have been reported [34]. Such systems have been successfully applied in trace element determination [39].

Synchrotron source X-ray fluorescence has also been used in combination with TRXRF. The small divergence of the continuous beam can be caused to strike the sample at less than the critical angle. Very high signal-to-background ratios have been obtained on employing this arrangement for the analysis of small quantities of aqueous solutions dried on the reflector, with detection limits of < 1 ppb, i.e., 1 pg. [40]. Additional advantages accrue because synchrotron radiation is highly polarized, and background due to scatter can be greatly reduced by placing the detector at 90° to

the path of the incident beam and in the plane of polarization. A disadvantage of the SSXRF technique is that the source intensity decreases with time, but this can be overcome by bracketing analytical measurements between source standards and/or by continuously monitoring the primary beam.

23.9. References

[1] G. v. Hevesey: *Chemical Analysis by X-Rays and its Application,* McGraw-Hill, New York 1932.
[2] W. C. Roentgen, *Ann. Phys. Chem.* **64** (1898) 1.
[3] H. G. I. Moseley, *Philos. Mag.* **26** (1912) 1024;**27** (1913) 703.
[4] A. Hadding, *Z. Anorg. Allg. Chem.* **122** (1922) 195.
[5] D. Coster, J. Nishina, *Chem. News J. Ind. Sci.* **130** (1925) 149.
[6] R. Glocker, H. Schreiber, *Ann. Phys. (Leipzig)* **85** (1928) 1085.
[7] L. S. Birks: *History of X-Ray Spectrochemical Analysis,* American Chemical Society Centennial Volume, ACS, Washington, D.C., 1976.
[8] P. Auger, *Compt. R.* **180** (1925) 65;*Journal de Physique* **6,** 205.
[9] J. A. Bearden: *X-Ray Wavelengths,* U.S. Atomic Energy Commission Report NYO-10586, 1964, 533 pp.
[10] J. Iwanczyk, B. E. Patt, *Adv. X-ray Anal.,* **41** (1997) 951–957.
[11] W. T. Elam, et.al., *Adv. X-ray Anal.,* **42** (1999) 137–145.
[12] R. Jenkins: *An Introduction to X-Ray Spectrometry,* Chap. 4, Wiley/Heyden, London 1974.
[13] M. F. Garbauskas, R. P. Goehner, *Adv. X-Ray Anal.* **26** (1983) 345.
[14] M. J. Beattie, R. M. Brissey, *Anal. Chem.* **26** (1954) 980.
[15] H. J. Lucas-Tooth, C. Pyne, *Adv. X-Ray Anal.* **7** (1964) 523.
[16] G. R. Lachance, R. J. Traill, *Can. Spectrosc.* **11** (1966) 43.
[17] T. Shiraiwa, N. Fujino, *Bull. Chem. Soc. Jpn.* **40** (1967) 1080.
[18] J. W. Criss, L. S. Birks, *Anal. Chem.* **40** (1968) 1080.
[19] J. W. Criss, *Adv. X-Ray Anal.* **23** (1980) 93.
[20] J. V. Gilfich, L. S. Birks, *Anal. Chem.* **40** (1968) 1077.
[21] W. K. de Jongh, *X-Ray Spectrom.* **2** (1973) 151.
[22] R. Jenkins, J. L. de Vries: *Practical X-Ray Spectrometry,* 2nd ed., Springer Verlag, New York 1970.
[23] Y. Yoneda, T. Horiuchi, *Rev. Sci. Instrum.* **42** (1971) 1069.
[24] J. Knoth, H. Schwenke, *Fresenius Z. Anal. Chem.* **301** (1978) 200.
[25] H. Schwenke, J. Knoth, *Nucl. Instrum. Methods Phys. Res.* **193** (1982) 239.
[26] H. Aiginger, P. Wobrauschek, *Adv. X-Ray Anal.* **28** (1985) 1.
[27] E. Zeigler, O. Hignette, M. Lingham, A. Souvorov, *SPIE,* vol. 2856/61 (1996).
[28] R. Schwaiger, P. Wobrauschek, C. Streli, *Adv. X-ray Anal.,* **41** (1997) 804–811.

[29] M. A. Kumakhov, *Nucl. Instr. Methods,* **B48** (1990) 283.

[30] G. Worley, G. Havrilla, N. Gao, Q. F. Xiao, *Adv. X-ray Anal.* **42** (1998) 26 – 35.

[31] M. A. Kumakhov, *Adv. X-ray Anal.* **41** (1997) 214 – 226.

[32] J. Gormley, *Adv. X-ray Anal,* **41** (1997) 239 – 242.

[33] J. A. Nicolosi, E. Scrugge, *Adv. X-ray Anal.* **41** (1997) 227 – 233.

[34] *Adv. X-ray Anal.* **41** (1997) 278 – 290.

[35] C. J. Sparks Jr. in H. Winnick, S. Doniach (eds.): *Synchrotron Radiation Research,* Plenum Press, New York 1980, p. 459.

[36] J. V. Gilfrich et al., *Anal. Chem.* **55** (1983) 187.

[37] T. Holz et al., *Adv. X-ray Anal.* **41** (1997) 346 – 355.

[38] W. Petersen, P. Ketelsen, A. Knoechel, R. Pausch, *Nucl. Instrum. Methods Phys. Res. Sect. A* **246** (1986) no. 1 – 3, 731.

[39] K. W. Jones, B. M. Gordon, *Anal. Chem.* **61** (1989) 341A.

[40] A. Iida, Y. Gohshi, *Adv. X-ray Anal.* **28** (1985) 61 – 68.

24. Activation Analysis

Richard Dams, Gent University, Laboratory Analytical Chemistry, Gent, Belgium

Karel Strijckmans, Gent University, Laboratory Analytical Chemistry, Gent, Belgium

In addition to the standard symbols defined in the front matter of this volume the following symbols and abbreviations are used.

A	activity (count rate, decay rate)
AA	activation analysis
B	background
C, D	correction terms
CPAA	charged-particle activation analysis
E	element
E	energy
ENAA	epithermal neutron activation analysis
FNAA	fast-neutron activation analysis
FWHM	full width at half maximum
I	beam intensity
I_0	resonance integral
INAA	instrumental neutron activation analysis
IPAA	instrumental photon activation analysis
l	length
MCA	multichannel analyzer
n	particle density
N	number of nuclides, counts
NAA	neutron activation analysis
PAA	photon activation analysis
R	reaction rate; range
RBS	Rutherford backscattering
S	saturation function; signal; stopping power
SCA	single-channel analyzer
TNAA	thermal neutron activation analysis
γ	relative intensity of gamma radiation
ε	detection efficiency
Θ	abundance of isotope
κ	gamma quantum; particle
λ	disintegration constant
σ	cross section
τ	dead time of detector

φ	flux density of radiation (fluency rate)
Φ	flux (fluency)

24.1. Introduction

Activation analysis has, in a few decades, become one of the most important methods for determination of minor (0.1 – 10 mg/g), trace (1 – 100 µg/g), and ultratrace (1 – 1000 ng/g) elements in solid samples. The main advantages of activation analysis are its accuracy and sensitivity (detection limit). Moreover, it is an independent method, i.e., not subject to the same systematic errors as other, more commonly used analytical methods. When a radiochemical separation is applied (after irradiation and prior to measurement) it is not subject to systematic errors due to reagent blanks, sample contamination (after irradiation), and non-quantitative (and even non-reproducible) yield. Precision goes down to a few percent in most favorable cases. The method is applied in the semiconductor industry, medicine, biology, criminology, archaeology, geochemistry, and en-

vironmental studies or control. Activation analysis is based on the exposure of a sample to a flux of activating particles or radiation. Irradiation is performed mainly with thermal neutrons (90 %), but fast neutrons (3 %), charged particles (4 %), or photons (3 %) can also be used (figures based on publications in Chemical Abstracts over the last two decades). This induces a nuclear reaction, whereby an excited intermediate is formed, which deexcites in 10^{-14} s by emission of prompt gamma rays, which can be measured, or the prompt emission of another particle [neutron(s), proton, alpha] i.e., nuclear reaction analysis (NRA). However, in activation analysis (AA) the resulting radioactive nuclide is monitored by means of the emitted alpha, beta, gamma or X-ray quanta, or delayed neutrons. The radiation emitted during the decay is measured for the identification and quantification of the target elements.

24.2. Neutron Activation Analysis

For a number of reasons, neutrons have found the widest applications as bombarding particles in activation analysis. First, the cross sections for reactions of nuclides with neutrons, yielding primarily (n, γ) reactions, but also (n, α); (n, p), $(n, 2n)$, and (n, n') reactions, are often large. The type of reaction depends on the energy of the neutron and is chosen individually for each case, on the basis of the following considerations:

1) A suitable reaction product with not too long a half-life (preferably below one year) and with decay parameters compatible with the available counting equipment
2) A sufficiently high cross section
3) The absence of important interfering reactions yielding the same radionuclide
4) Easy treatment of the sample after irradiation (limited matrix activity)

Second, intense neutron sources (nuclear reactors, neutron generators, isotopic sources) were generally available. Third, the high penetration depth of neutrons allows a nearly homogeneous flux density of the entire sample to be analyzed.

24.2.1. Basic Principles

If a target with n^A nuclides of A per cubic centimeter is placed in a neutron beam φ (number of neutrons $cm^{-2} s^{-1}$) then the number of interactions to form nuclei B is given by the reaction rate R $(cm^{-3} s^{-1})$

$$R = \varphi \sigma n^A \tag{1}$$

where σ is the cross section.

The general nuclear reaction can be written as follows:

$$A(n, \kappa) \lambda \to C$$
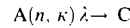
where n is a neutron; κ the promptly emitted gamma quantum or particle; λ is the disintegration constant of the radioactive nuclide B produced (equal to $\ln 2/t_{1/2}$; $t_{1/2}$ denotes the half-life of the radionuclide B); and C is the nuclide produced after radioactive decay of B.

From Equation (1) the definition of the cross section σ for this reaction is obtained:

$$\sigma = \frac{\text{Number of interactions}}{\varphi n^A} \tag{2}$$

The cross section is expressed as an area in square centimeters or, more practically, in barn $(1 \text{ b} = 10^{-24} cm^2)$.

The number N^A of target nuclei is assumed to remain constant, as is the flux density φ throughout the target. Since the nuclides B decay at a rate equal to $-\lambda N^B$,

$$\frac{dN^B}{dt} = \varphi \sigma N^A - \lambda N^B \tag{3}$$

If no nuclides B were present before irradiation $(N^B = 0$ for $t = 0)$, integration gives, after an irradiation time t,

$$N^B = \frac{\varphi \sigma N^A}{\lambda} \left(1 - e^{-\lambda t}\right) \tag{4}$$

The absolute decay rate A_0, in terms of the number of decays per second, at the end of the irradiation is given by

$$A_0 = N^B \lambda = \varphi \sigma N^A \left(1 - e^{-\lambda t}\right) \tag{5}$$

The expression

$$S = 1 - e^{-\lambda t} = 1 - e^{\ln(2) t/t_{1/2}} \tag{6}$$

is termed the saturation function. After an irradiation time equal to one half-life of B, a saturation of 50 % is obtained. For irradiation times that are

Table 1. Typical neutron fluxes at irradiation facilities in research reactors

Irradiation site	Neutron flux, $cm^{-2}\,s^{-1}$		
	Thermal	Epithermal	Fast
Large reactor			
Reactor core	2×10^{14}	3×10^{13}	4×10^{13}
Surrounding moderator	$1-2\times10^{14}$	$0.5-2\times10^{13}$	$0.2-2\times10^{13}$
Small reactor			
Reactor core	5×10^{12}	4×10^{11}	3×10^{12}
Reflector	$0.3-2\times10^{12}$	$0.2-8\times10^{10}$	$0.3-30\times10^{10}$

much lower than the half-life, the saturation function can be considered approximately linear:

$$S \approx \ln(2)\,t/t_{12} \quad (\text{for } t < 0.1t_{1/2}) \tag{7}$$

A correction term $(D = e^{-\lambda t_d})$ must be applied for the decay during the time t_d between the end of irradiation and the measuring time.

When the counting (measuring) time is not very short $(t_m > 0.3\,t_{1/2})$ compared with the half-life, a correction C for decay during the counting period t_m can also be applied. It equals

$$C = \frac{1 - e^{-\lambda t_m}}{\lambda t_m} \tag{8}$$

Finally the counting equipment has a detection efficiency ε, which is usually less than unity. Thus the number of counts N measured after an irradiation time t, a waiting time t_d, and during a counting time t_m becomes

$$N = \varphi\sigma N^A \left(1 - e^{-\lambda t}\right) e^{-\lambda t_d} \frac{1 - e^{-\lambda t_m}}{\lambda}\,\varepsilon \tag{9}$$

$$= \frac{\varphi\,\sigma\,m\,\theta N_A\,S\,D\,C\varepsilon}{A_r}\,t_m \tag{10}$$

where m is the mass of the element A; A_r is its relative atomic mass; θ is the abundance of the isotope yielding the radioactive isotope; and N_A is the Avogadro constant.

24.2.2. Methods of Neutron Activation

Reactor Neutrons. Nuclear reactors are the largest and most often used neutron sources. Neutrons formed in the reactor have a continuous energy spectrum extending from nearly 0 to 15 MeV. Conventionally the neutron spectrum is divided in three components:

1) *Fast or fission neutrons*, produced by the fission of ^{235}U with energies ranging from 0.1 to 15 MeV. The mean energy is ca. 2 MeV. These neutrons give rise to so-called threshold reactions, e.g., (n, p), (n, α), and $(n, 2\,n)$.

2) *Epithermal or resonance neutrons* have energies varying from 0.1 MeV down to about 0.5 eV. In the ideal case the epithermal flux Φ_{epi} is inversely proportional to the neutron energy E

$$\Phi_{epi}(E) = \frac{\Phi_{epi}}{E} \tag{11}$$

The highest resonance neutron flux is in the reactor core.

3) *Thermal or moderated neutrons* are in thermal equilibrium with the atoms of the moderator, and their velocities of 2200 m/s at 20 °C correspond to an energy of 0.0253 eV.

In some research reactors, pulses of very high neutron flux may be generated by removing the control rods. The neutron flux increases immediately to several times $10^{16}\,cm^{-2}\,s^{-1}$ and falls after ca. 10 ms to a very low value.

Table 1 shows that at most irradiation facilities the highest fluxes belong to thermal neutrons. Fortunately for activation analysis, they are the most useful ones. As shown in Figure 1 for the lower-energy neutrons (up to several hundred electron-volts) the decrease in cross section with increasing energy is roughly proportional to $1/v$ (where v is the velocity of the neutrons). Sometimes superimposed on this smoothly dropping curve are sharp resonance peaks in the epithermal region just above 1 eV. The total cross section for capture of epithermal neutrons is, therefore, the sum of the $1/v$ probability and that associated with these superimposed resonance peaks, and is called the resonance integral I_0. In neutron activation analysis, thermal and epithermal flux are separated by the so-called cadmium cutoff energy E_{Cd} of 0.55 eV. Cadmium in fact strongly absorbs neutrons with energies of less than 0.55 eV.
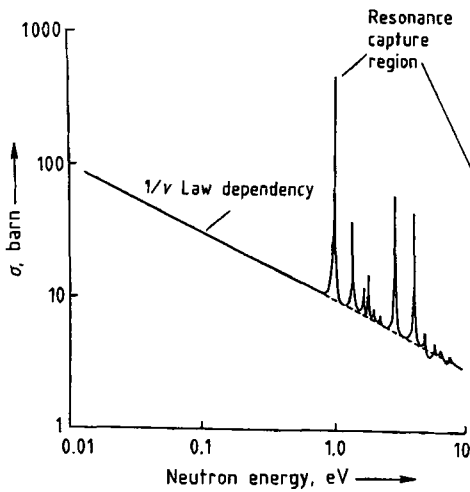
Figure 1. Typical excitation function for neutron-induced reaction, illustrating $1/v$ dependency and resonance peaks in epithermal region

In Equation (1) and following, the reaction rate can thus be split up according to the Høgdahl convention [29]

$$R = N\left(\Phi_{\text{th}}\,\sigma_0 + \Phi_{\text{epi}}\,I_0\right) \qquad (12)$$

where σ_0 is the tabulated thermal neutron cross section for the corresponding energy (0.0253 eV). The conventional thermal flux is, with a thermal neutron density n

$$\varphi_{\text{th}} = n\,v_0$$

The second term in Equation (12) is the somewhat more complicated solution for the integration of the reaction probability in the epithermal region. The resonance integrals I_0 have been tabulated:

$$I_0 = \int\limits_{E_{\text{Cd}}}^{\lambda} \sigma_{\text{epi}}(E)\,\frac{dE}{E} \qquad (13)$$

Values of φ_{th} and φ_{epi} can be found by irradiating a nuclide (monitor) with known σ_0 and I_0 in a cylindrical cadmium box with 1-mm wall thickness. Often, gold is used as a monitor.

The cross sections for slow neutron reactions are often on the order of several barns and can in some cases be as large as 10^4 barn. The predominant reaction for slow neutrons is the (n, γ) reaction.

Equation (10) allows calculation, to a first approximation, of the detection limit attainable by neutron activation analysis. If a count rate of $A = 1\ \text{s}^{-1}$ can be detected and $A_r = 100$, $S = 0.5$, $\sigma = 1$ barn, $\varphi = 3\times10^{12}\ \text{cm}^{-2}\,\text{s}^{-1}$, $\varepsilon = 0.33$, with θ, D, and C close to unity, a mass of 10^{-9} g of a typical element can be detected; for many elements with larger cross sections or when irradiating at high neutron fluxes, much lower masses can be detected.

When determining a specific element, the *half-life* of the radionuclide to be measured is a very important parameter. For the production of a radioisotope with a half-life of only a few hours or less, the irradiation time should be limited to a saturation of ca. 0.5–0.75, to avoid production of much longer-lived radionuclides. The short-lived radionuclides must be measured soon after the end of irradiation. Also, the counting period should not greatly exceed the half-life of the radionuclide. This however, results in poor counting statistics. For really short-lived radionuclides, this drawback can to some extent be overcome by *cyclic activation analysis* which is based on the use of repetitive short irradiations and counting periods, and summing of the γ-ray spectra. Also *pulsed neutron irradiation* enhances the detection efficiency of radionuclides with $t_{1/2} < 20$ s. In contrast, to determine elements producing longer-lived radionuclides (e.g., $t_{1/2} > 5$ d), much longer irradiation times are required although activation to a high saturation factor is not meaningful. Treatment of the sample and counting should be postponed several days to allow for decay of the shorter-lived radionuclides.

An additional selectivity factor is the neutron energy. By covering the sample with cadmium or boron carbide, the thermal neutrons are filtered out and the sample is activated largely with epithermal neutrons. (Boron has a nearly pure $1/v$ excitation function and is therefore also chosen to filter out thermal low-energy resonance neutrons.) A distinction is made between *thermal* (*TNAA*) and *epithermal neutron activation analysis* (*ENAA*). In the latter, selective activation of a number of elements relative to other interfering elements can be enhanced. ENAA is often advantageous because most of the matrix elements in geological and environmental samples (e.g., Na, Al, P, K, Ca, Sc, Cr, Mn, and Fe) are activated primarily with thermal neutrons (σ_0), whereas a number of trace or ultratrace elements (e.g., As, Sb, Se, Sn, Co, Ni, Ga, Mo, Ag, Pd, In, Cd, Br, I, Rb, Sr, Cs, Ta, W, Th, U, Pt, and Au) have large resonance integrals

(I_0). Not only is the relative activation of many trace elements enhanced, but the overall activity is also generally reduced so that the sample can be measured much closer to the detector, which improves the counting geometry and thus the precision. Since fast neutrons induce threshold reactions, *fast-neutron activation analysis* (*FNAA*) is used primarily for the determination of light elements (O, N, F, P, S, Cl, Si, Mg, Al, Fe), several of which cannot (or can barely) be determined by thermal neutron activation. Although the reactor neutron spectrum covers a wide range of fast-neutron energies, neutron generators are much more appropriate for FNAA.

Neutron Generators. Neutrons can also be obtained by bombarding an appropriate target with ions accelerated to high kinetic energies. In cyclotrons, a typical neutron flux of $2 \times 10^{11}\,\mathrm{s}^{-1}$ is obtained with a 10-µA beam of 30-MeV deuterons on a thick beryllium target. Muchmore important, however, are the Cockroft–Walton accelerators, called 14-MeV neutron generators, in which neutrons are produced by the $^3\mathrm{H}\,(\mathrm{d}, \mathrm{n})\,^4\mathrm{He}$ reaction. The neutrons obtained have an energy of 14.7 MeV and are emitted nearly isotropically from the target. The total neutron yield varies from 10^8 to $10^{12}\,\mathrm{s}^{-1}$, which gives a usable flux density for irradiation ca. 20–100 times lower. Since neutron generators are rather compact, are not too expensive, and do not need very large shielding, they are used in many research and industrial laboratories. Cross sections, especially for light elements, often range from 10 mb to a few hundred millibarn for (n, p) and (n, α) reactions on light elements, and even up to 2–3 b for some (n, 2 n) reactions on heavy elements. With a neutron flux of $10^9\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ typical detection limits of 10^{-4} to 10^{-6} g, and for a few elements (F, Si, P, Cu, Br, Ag, Sb) ca. 10^{-7}, can be reached. Disadvantages of FNAA with neutron generators are the many possibilities for producing the same indicator radionuclide from different elements and the large flux gradients.

The determination of oxygen by the reaction $^{16}\mathrm{O}\,(\mathrm{n}, \mathrm{p})\,^{16}\mathrm{N}$ is most important. The half-life of the indicator radionuclide $^{16}\mathrm{N}$ is only 7.1 s, and it emits high-energy gamma rays (>6 MeV), which can be measured without the interference of other radionuclides. A detection limit of about 0.005 % can be obtained by using large samples (10–100 g).

Isotopic Neutron Sources [30]. Neutrons can also be produced from isotopic sources. They have the unique advantage of being small, portable, and completely reliable, and of having a constant neutron flux and low cost. There is no need for complicated controlling mechanisms, and the shielding can be very limited. Isotopic neutron sources offer interesting analytical perspectives for accurate determination of minor and major constituents.

Their operation is based on releasing neutrons from a target nucleus either by the action of an α particle or a γ quantum, or during spontaneous fission of a transuranic element. Beryllium is mostly used as target material. As an α source, historically $^{226}\mathrm{Ra}$ was used; it has now been replaced by $^{227}\mathrm{Ac}$, $^{241}\mathrm{Am}$, or $^{238,239}\mathrm{Pu}$, and in the future possibly $^{224}\mathrm{Cm}$. As a γ source, $^{124}\mathrm{Sb}$ is employed ($t_{1/2} = 60.2$ d). The neutrons emitted have a broad energy spectrum that peaks at ca. 3–6 MeV. Paraffin wax is often used as a moderator. The neutron outputs range from 10^6 to $1.5 \times 10^7\,\mathrm{s}^{-1}\,\mathrm{Ci}^{-1}$. In practice, sources of 10^8 neutrons/s are available commercially. The isotope $^{252}\mathrm{Cf}$ disintegrates not only by alpha decay, but also by spontaneous fission, releasing several neutrons in the process. It has a number of advantages over (α, n) sources: the small size of the source, and the close similarity of the neutron spectrum to that obtained from the fission of uranium. The main disadvantage is obviously its relatively short effective half-life of 2.65 a.

Since the obtainable neutron flux density is generally less than $2 \times 10^6\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, elements with cross sections > 10 barn and half-lifes of less than 5 d have typically been determined. Examples are In, Dy, Rh, V, Ag, La, Eu, Mn, I, Br, Co, Ga, Cu, Ir, and Au. They are determined preferentially in matrices that do not activate very much so that nondestructive analyses are possible. Attainable precisions are better than 1 %, sometimes as good as 0.1 %.

24.2.3. Standardization

Standardization (i.e., calibration) is performed by Equation (10), which shows that m, the mass of the element to be determined, can be calculated from the measured activity. In practice, this is rarely done because although the values for θ, N_A, A_r, S, D, $C(\lambda)$, and t_m are known with good accuracy for all radionuclides considered, this is not the case for φ (φ_th and φ_epi), σ (σ_0 and I_0), and ε. Therefore a *relative or comparator method* is

normally applied, although more recently absolute methods have also been used. The relative method is essentially very simple. Each sample is coirradiated with a standard containing known amounts of the elements to be determined. When samples and standards are counted under identical conditions, all these factors may be assumed to cancel out, leaving only the following very simple equation:

$$\frac{m_x}{m_{st}} = \frac{A_x}{A_{st}} \cdot \frac{D_{st}}{D_x} = \frac{A_x}{A_{st}} \frac{(e^{-\lambda t_d})_{st}}{(e^{-\lambda t_d})_x} \tag{14}$$

where the subscripts x and st denote the unknown and the standard, respectively. This simple relationship is valid only when all other factors are the same or are held constant. For short-lived radionuclides, the sample and the standard are irradiated sequentially. Since the neutron flux cannot be relied on to remain entirely constant with time, the activities measured must be corrected for any flux variation by coirradiating flux monitors. Second, the sample and the standard, even when irradiated simultaneously, are not always exposed to the same neutron flux because vertical or lateral flux gradients may occur. This can be corrected by means of flux-gradient monitors. The standards containing known amounts of the elements to be determined can be mixtures of pure elements or compounds of known stoichiometry. Since in NAA the samples to be irradiated are mostly solid, the mixed standard solutions are spotted on clean paper filters or ultrapure graphite, dried, and finally pressed into pellets with similar geometry to the sample. To avoid problems associated with the preparation of these standards, reference materials or other "in-house standard materials" are often used in NAA.

k_0-Method. The k_0-method has been developed to overcome the labor-intensive and time-consuming work of preparing such multielement standards when routine multielement or panoramic analyses are required. Single comparators have been used for a long time, but they are applicable only to constant experimental parameters of activation and counting. Therefore, the k_0-method is being implemented increasingly in NAA laboratories. It is intended to be an absolute technique in which uncertain nuclear data are replaced by a compound nuclear constant, the k_0-factor, which has been determined experimentally for each radionuclide. This k_0 is given by

$$k_0 = \frac{A_r^* \, \theta \, \sigma_0 \, \gamma}{A_r \, \theta^* \, \sigma_0^* \, \gamma^*} \tag{15}$$

where γ is the relative intensity of the gamma radiation in the decay scheme of the radionuclide and * denotes the comparator.

For determination of the k_0-factor, gold was used as comparator, which was coirradiated as dilute Au – Al wire. This Au – Al wire should be coirradiated with each sample. It can, however, be converted to any comparator that is found suitable for coirradiation and that has been coirradiated with gold before. To determine the epithermal-to-thermal flux ratio, coirradiation of a zirconium monitor is suitable. The k_0-factors for 68 elements and their relevant gamma lines of 135 analytically interesting radionuclides have been determined and published by DE CORTE and coworkers. For nearly all of them, the uncertainly is < 2 %. An overview is given in Table 2 [31] – [33].

24.2.4. Sources of Errors

Besides typical analytical errors in sample preparation (contamination, inhomogeneity, inaccurate standards) and errors during counting (pulse pileup, instability, spectrometric interference, electronic failure, etc.), some systematic bias may also occur during the irradiation step, such as neutron shielding and nuclear interference. By taking the proper measures, these can be avoided or corrected for. *Flux gradients* can be determined experimentally and corrected for. For accurate work, vertical and horizontal gradients should be measured for the different types of flux (e.g., thermal, epithermal, and fast). When relatively large samples or standards with large absorption cross sections or resonance integrals are irradiated, nonnegligible *neutron shadowing* or *self-shielding* may also occur.

In addition, *flux hardening* (preferential absorption of low-energy neutrons) and *self-moderation* (unmoderated neutrons are further moderated inside the sample) may occur. Since the correction is difficult to apply for samples with mixed composition and irregular shape, the effect is often avoided by using as small a sample and a standard as possible or by diluting the sample with a material having a low absorption cross section (graphite, cellulose). An internal standard can also be applied.

In thermal neutron activation, the determination of an element $_Z^M E$ is often based on an (n, γ)

Table 2. Elements and their analyte radionuclides that can be determined by k_0 INAA

F	^{20}F
Na	^{24}Na
Mg	^{27}Mg
Al	^{28}Al
Si	^{31}Si
S	^{37}S
Cl	^{38}Cl
K	^{42}K
Ca	^{47}Ca
	^{47}Ca $\rightarrow$ ^{47}Sc
	^{49}Ca
Sc	^{46}Sc
Ti	^{51}Ti
V	^{52}V
Cr	^{51}Cr
Mn	^{56}Mn
Fe	^{59}Fe
Co	^{60}Co
Ni	^{65}Ni
Cu	^{64}Cu
	^{66}Cu
Zn	^{65}Zn
	^{69m}Zn
	^{71}Zn
Ga	^{72}Ga
Ge	^{75m}Ge
	^{75m}Ge $\rightarrow$ ^{75}Ge
	^{77m}Ge
	^{77m}Ge $\rightarrow$ ^{77}Ge
As	^{76}As
Se	^{75}Se
	^{77m}Se
Br	^{80m}Br
	^{80m}Br $\rightarrow$ ^{80}Br
	^{82m}Br $\rightarrow$ ^{82}Br
Rb	^{86m}Rb $\rightarrow$ ^{86}Rb
	^{88}Rb
Sr	^{85m}Sr
	^{85m}Sr $\rightarrow$ ^{85}Sr
	^{87m}Sr
Y	^{90m}Y
Zr	^{95}Zr
	^{95}Zr $\rightarrow$ ^{95m}Nb
	^{95}Zr $\rightarrow$ ^{95m}Nb $\rightarrow$ ^{95}Nb
	^{97}Zr
	^{97}Zr $\rightarrow$ ^{97m}Nb
	^{97}Zr $\rightarrow$ ^{97m}Nb $\rightarrow$ ^{97}Nb
Nb	^{94m}Nb
Mo	^{99}Mo
	^{99}Mo $\rightarrow$ ^{99m}Tc
Mo	^{101}Mo
	^{101}Mo $\rightarrow$ ^{101}Tc
Ru	^{97}Ru
	^{103}Ru
	^{105}Ru
	^{105}Ru $\rightarrow$ ^{105m}Ru
	^{105}Ru $\rightarrow$ ^{105m}Ru $\rightarrow$ ^{105}Ru
Rh	^{104m}Rh
	^{104m}Rh $\rightarrow$ ^{104}Rh
Pd	^{109m}Pd
	^{109m}Pd $\rightarrow$ ^{109}Pd
	^{109m}Pd $\rightarrow$ ^{109}Pd $\rightarrow$ ^{109m}Ag
	^{111m}Pd
Ag	^{108}Ag

Table 2. continued

	^{110m}Ag
	^{110}Ag
Cd	^{115}Cd
	^{115}Cd $\rightarrow$ ^{115m}In
In	^{114m}In
	^{116m}In
Sn	^{113m}Sn $\rightarrow$ ^{113}Sn $\rightarrow$ ^{113m}In
	^{117m}Sn
	^{123m}Sn
	^{125m}Sn
	^{125m}Sn $\rightarrow$ ^{125}Sn
	^{125m}Sn $\rightarrow$ ^{125}Sn $\rightarrow$ ^{125}Sb
Sb	^{122m}Sb $\rightarrow$ ^{122}Sb
	^{124m}Sb $\rightarrow$ ^{124m1}Sb $\rightarrow$ ^{124}Sb
	^{124m2}Sb $\rightarrow$ ^{124m1}Sb
Te	^{131m}Te $\rightarrow$ ^{131}Te $\rightarrow$ ^{131}I
I	^{128}I
Cs	^{134m}Cs
	^{134m}Cs $\rightarrow$ ^{134}Cs
Ba	^{131m}Ba $\rightarrow$ ^{131}Ba
	^{133m}Ba
	^{139}Ba
La	^{140}La
Ce	^{141}Ce
	^{143}Ce
Pr	^{142m}Pr $\rightarrow$ ^{142}Pr
Nd	^{147}Nd
	^{149}Nd
	^{149}Nd $\rightarrow$ ^{149}Pm
	^{151}Nd
	^{151}Nd $\rightarrow$ ^{151}Pm
Sm	^{153}Sm
	^{155}Sm
Eu	^{152m}Eu
	^{152m}Eu $\rightarrow$ ^{152}Eu
	^{154m}Eu $\rightarrow$ ^{154}Eu
Gd	^{153}Gd
	^{159}Gd
	^{161}Gd
Tb	^{160}Tb
Dy	^{165m}Dy
	^{165m}Dy $\rightarrow$ ^{165}Dy
Ho	^{166}Ho
Er	^{171}Er
Tm	^{170}Tm
Yb	^{169}Yb
	^{175}Yb
	^{177}Yb
Lu	^{177}Lu
	^{176m}Lu
Hf	^{175}Hf
	^{179m}Hf
	^{180m}Hf
	^{181}Hf
Ta	^{182m}Ta $\rightarrow$ ^{182}Ta
W	^{187}W
Re	^{186}Re
	^{188m}Re
	^{188m}Re $\rightarrow$ ^{188}Re
Os	^{185}Os
	^{191m}Os $\rightarrow$ ^{191}Os
	^{193}Os
Ir	^{194}Ir
Pt	^{199m}Pt $\rightarrow$ ^{199}Pt $\rightarrow$ ^{199}Au

Table 2. continued

Au	^{198}Au
Hg	^{197m}Hg
	^{203}Hg
	^{205}Hg
Th	^{233}Th $\rightarrow$ ^{233}Pa
U	^{239}U
	^{239}U $\rightarrow$ ^{239}Np

reaction producing the radionuclide $^{M+1}_{Z}$E. An (n, p) reaction on nuclide $^{M+1}_{Z+1}$E' or an (n, α) reaction on nuclide $^{M+3}_{Z+2}$E" will produce the same indicator radionuclide, resulting in positive errors. This *nuclear interference* depends on the concentration ratios of the elements in the sample, the fast-to-thermal flux ratio, and the cross sections involved, and can be estimated by a simple calculation. The cross sections for (n, p) and (n, α) reactions are usually much smaller than the (n, γ) cross sections. Well-known examples are ^{27}Al (n, γ) ^{28}Al interference by ^{28}Si (n, p) ^{28}Al, and ^{55}Mn (n, γ) ^{56}Mn interference by ^{56}Fe (n, p) ^{56}Mn.

If the interference is not too important it can be corrected for. If uranium or other fissionable material is present in the sample, the fission products with high fission yields (Sr, Mo, Zr, Ce, Ba) can induce important positive errors. The interference can easily be estimated when the uranium content is known. In FNAA, *secondary interference reactions* may occur when fast neutrons interact with other elements and produce particles that induce a nuclear reaction that forms the same indicator nuclide. These particles are usually protons ejected by fast neutrons from a matrix with a high hydrogen content. Examples are:
^{19}F (n, 2 n) ^{18}F interference by: ^{17}O (p, γ) ^{18}F and ^{18}O (p, n) ^{18}F
^{14}N (n, 2 n) ^{13}N interference by: ^{13}C (p, n) ^{13}N

24.2.5. Measuring Equipment

For identification and quantification of the radionuclides produced during the activation step a radiation detector is needed. The major way to select a specific radionuclide is of course by *radiochemical separation*. At the end of the irradiation a nonradioactive carrier of the element to be measured can be added prior to separation. A second parameter used for selection is the *half-life* of the radionuclide. Selection on the basis of the *energy of the emitted radiation* is, however, the

most powerful technique. Obviously, detection of the γ and X rays is much more suitable for identification and selective for quantification purposes than β-ray detection. Alpha emitters are almost never produced in activation analysis.

Gas-filled detectors, such as ionization chambers, proportional counters, and Geiger–Müller counters are mainly sensitive to β-radiation. They rarely allow any selection on the basis of energy. Gamma counting with a selection for energy is called gamma spectrometry and is performed by means of *scintillation* or *semiconductor detectors.* Currently the gamma spectrometric measurements in neutron activation analysis are only performed with semiconductor detectors. Scintillation counters will therefore be discussed only briefly.

Interaction of γ or X rays with matter leads to three effects: (1) A photoelectric effect in which the entire energy is absorbed by the detector and the result is a photopeak in the spectrum, with a pulse height proportional to the full energy of the gamma radiation. (2) Compton scattering results in a partial absorption of the γ-ray energy. In the spectrum a continuous background is seen, which can interfere with the measurement of other lower-energy γ or X rays. (3) Pair production may result in the escape of one or two annihilation photons, each with an energy of 0.511 MeV. The *efficiency* ε of a detector refers to the ratio of the number of radiations actually detected to the number emitted by the source. In addition a photopeak-efficiency ε_p may be defined as the number of counts recorded under the photopeak. The *energy resolution R* is the ability of the detector to discriminate between two γ or X rays of different energy. This resolution is defined as the full width ΔE of the peak at half-maximum (FWHM). For scintillation detectors, it is expressed mostly as a percentage of the energy E corresponding to the centroid of the peak. For semiconductor detectors, it is expressed in energy units (kiloelectronvolts or electronvolts). Finally, the *dead time* τ of the detector is the amount of time required before the detector can recover from an incoming radiation and respond to the next event. In gamma spectrometry the main cause of dead time is not the detector itself but the electronic equipment and, more specifically, the amplifier (i.e. pulse pile up) and the multichannel analyzer (MCA).

Scintillation Detectors. In scintillation detecors the incoming radiation interacts with the material by ionization or excitation. The excited atoms or
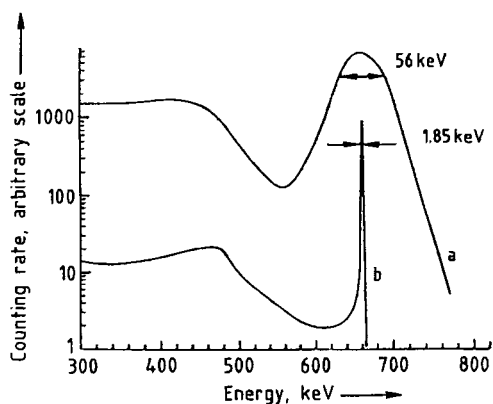
Figure 2. Comparison of gamma spectrum of ^{137}Cs (gamma energy 661.6 keV) obtained with 7.5×7.5-cm NaI(Tl) scintillation detector (a) and with a 50-cm^3 Ge(Li) semiconductor detector (b)

molecules undergo deexcitation by the emission of a photon of light. This light passes through the optically transparent scintillator material and is guided through a light pipe to a photomultiplier, which converts the light to an electrical signal. The signal is then amplified and fed to a counting or storage system. This can be a single-channel analyzer (SCA) or a multichannel analyzer (MCA). The detection efficiency is excellent. The major drawback of NaI(Tl) scintillators is their poor energy resolution, which is typically 7 % for the 661.6-keV line of ^{137}Cs (Fig. 2). Dead times are important at high count rates. More selective counting can be performed with NaI(Tl) detectors when coincidence or anticoincidence spectrometry is applied by using two or even more detectors.

In charged-particle activation analysis (Chapter 24.4) pure positron emitters have to be measured by their annihilation photons, emitted simultaneously in a 180° geometry. The sample is then introduced between two NaI detectors, coupled to an SCA to select the photopeak of the 511-keV annihilation photons. The fast SCA output is fed into a coincidence circuit, which selects only coincident signals (typically within 40 ns). The poor energy resolution of the NaI detector is no drawback, while the high detection efficiency (as compared with semiconductor detecors) is an advantage. The NaI detectors are much cheaper than semiconductor detectors (for the same detection efficiency) and do not require liquid nitrogen cooling. The signal-to-noise ratio is much higher for the NaI coincidence set-up, than for a Ge spec-

trometer, mainly because the noise (background) is much lower.

Semiconductor Detectors. These have a $p-i-n$ diode structure in which the intrinsic region is formed by depletion of charge carriers by applying a reverse bias across the diode. The energy of an electron in a semiconductor material is confined to the conduction and the valence bands, both states separated by forbidden energies. The passage of radiation through the semiconductor may inject sufficient energy to raise an electron from the valence band to the conduction band, thus creating an electron – hole pair. In order for the material to be used as a radiation detector, this electrical charge must be collected. Therefore, several thousand volts are applied to the detector. Owing to the small amount of energy needed to create an electron – hole pair, a large number of pairs (>300 000 pairs/MeV) is created, with small relative statistical fluctuations. Since the energy resolution of the detector depends on this fluctuation, which can be predicted by Poisson statistics, much better energy resolution is obtained than with scintillation counters. For a highly efficient detector, a large, nearly intrinsic or depletion region must be present. Therefore, the $p-n$ junction is created within a single p or n crystal. Ways to achieve this are the surface barrier detector (used only for detection of charged particles because the depletion region is too small), the lithium-drifted detector [Ge (Li) and Si(Li)], and at present the high-purity or intrinsic germanium detector (HP-Ge). *Ge (Li) detectors* must be cooled constantly at liquid nitrogen temperature, they are still in use but are no longer available. They are more efficient for γ-ray detection than Si (Li) detectors, owing to their higher atomic number Z. *Si (Li) detectors* are sufficiently efficient for detection of X rays and low-energy γ rays and must be cooled only during operation to ensure optimum energy resolution. Typical resolutions of Si(Li) detectors are 140 – 160 eV for the 5.9-keV Mn K line from a ^{55}Fe source.

HP-Ge detectors have the advantage that they can be stored at room temperature and are cooled only during operation to reduce the problem of thermal excitation of electrons. Coaxial p-type detectors and n-type detectors are available. In the *n-type detector* the inner contact is made by diffused lithium, while the outer contact is achieved by ion implantation, which ensures a very thin entrance window (0.3 μm). Thin *planar germanium detectors* (5 – 20 mm) can also be used

to measure low-energy radiation and have the advantage of not stopping the more energetic γ rays, thereby reducing interference effects such as Compton radiation. *Well-type detectors* are also available with well sizes up to 15 mm in diameter and a depth of 40 mm. For cooling, a cold finger and a large cryostat are needed, which makes semiconductor detectors bulky and difficult to maneuver.

Germanium detectors are characterized by three parameters: resolution, peak-to-Compton ratio, and efficiency. The *resolution* is typically given for the 1332-keV ^{60}Co line and varies from 1.8 keV for the very best to 2.3 keV for the very large detectors. The *peak-to-Compton ratio* is measured as the ratio of the number of counts in the 1332-keV peak to the number of counts in a region of the Compton continuum. Values vary from 30 to 90 for the most expensive model. The efficiency is expressed as a relative efficiency compared with the 7.5×7.5-cm NaI(Tl) scintillation detector. Relative efficiencies of HP-Ge detectors vary from 10 % up to 150 %. The dead time of semiconductor detectors is low, so the count rate is limited largely by the electronic circuit.

Gamma spectrometry is performed mainly by using MCAs, and the relation between the obtained pulse height or corresponding channel number and gamma energy must be established. When an absolute method of standardization is applied, the peak detection efficiency ε_p of the detector with surrounding material must also be measured.

For *peak area calculation*, two procedures can be applied in principle: integration or fitting. In *integration*, the number of counts recorded in the channels in which about 90 % of the peak is located is corrected for Compton contribution from other peaks by subtracting the content of an equal number of channels before and after the peak. The method fails, of course, when other peaks are located in the immediate neighborhood or in the case of superimposed peaks. *Fitting procedures* assume a defined shape, either Gaussian or with non-Gaussian distortions (determined from the shape of a well-defined peak in the spectrum).

The overall procedure for the evaluation of γ-ray spectra as part of activation analysis consists of several steps as defined by ERDTMANN und PETRI [34]:

1) Search for the peaks
2) Calculation of γ-ray energy
3) Calculation of peak areas

4) Calculation of γ-ray emission rates
5) Identification of radionuclides
6) Calculation of decay rates
7) Calculation of amounts of elements

The amount of data to be handled during this evaluation procedure requires the use of computers. At present, the multichannel analyzer not only carries out pulse height analysis but also does most of the data processing. The manufacturers of computer-based MCAs usually provide built-in or software programs for the different aforementioned steps.

Obviously the use of a personal computer is an important step toward automation of the analysis. The computer can control the automatic sample changer, the timing of the analysis, counting, storing of the spectra, and finally, data processing.

24.2.6. Procedures for Activation Analysis

Sample Preparation. As in any other analytical technique, the quality of the final results depends strongly on the care taken during sampling (→ Sampling). Since sample sizes can often be small (2 mg up to 20 g), owing to the high sensitivity of NAA, a well-homogenized material is required. In a number of cases, some preirradiation treatment will be necessary, such as: (1) the removal of water by drying, freeze drying, or ashing for biological materials, or evaporation for liquid samples, to avoid pressure buildup due to evaporation or radiolysis during irradiation; (2) the removal of surface contamination; or (3) perhaps even preconcentration. All of these treatments should be performed in clean environments.

For the irradiation itself, the sample is packed in polyethylene vials for total neutron doses of less than 10^{17} cm^{-2} and in high-purity quartz ampoules for longer irradiations at higher fluxes. The packed samples are transferred to the reactor in containers (rabbits) of aluminum, polyamide, graphite, or high-pressure polyehtylene for low-dose irradiation.

The major advantage of activation analysis, however, is that after irradiation, contamination with nonradioactive material does not introduce more errors. On the contrary, when radiochemical separations are required, an amount of nonradioactive carrier of the element to be separated is added, after irradiation.

Instrumental Neutron Activation Analysis (INAA).

A purely instrumental approach is applicable for the determination of many elements (i.e., after irradiation, the sample can be unpacked and the induced radioactivity measured without further treatment). Such a technique has a number of advantages. After decay of the induced activities, the sample is essentially unchanged (nondestructive activation analysis), and is therefore available, in principle, for further investigation. In addition, short-lived radionuclides can be used for the analysis. Tedious chemical separations are not required, which implies reduction of analysis time and avoidance of errors associated with separation. INAA becomes feasible when the activity induced in the sample matrix is not prohibitively high and no single major activity is produced that overshadows that of the other radionuclides. Of course, the differences in half-lifes of various radionuclides can also be exploited by limiting the saturation factor S of interfering long-lived activities or by allowing the short-lived matrix activity to decay before counting. Virtually all instrumental multielement activation analysis is based on high-resolution gamma spectrometry using semiconductor detectors after thermal or epithermal activation of the sample in a reactor or a neutron generator.

INAA is favored especially for many organic, biological, and environmental samples, because the elements O, H, C, N, S, and P are not activated or do not produce gamma emitters. The most abundant minor or trace elements such as Si, Fe, Al, Ca, Mg, and Cl have relatively small cross sections or short half-lifes, the nuclides with the highest activity often being ^{24}Na($t_{1/2} = 15.0$ h) or ^{82}Br ($t_{1/2} = 36$ h). Also, a number of metallic matrices can be analyzed purely instrumentally with thermal neutrons. Examples of these metals are Be, Mg, Al, Ca, Ti, V, Ni, Y, Zr, Nb, Rh, Pb, and Bi, some of them after decay of the short-lived matrix activity.

In multielement analysis, choosing the optimal irradiation, cooling, and counting conditions for each element is obviously not feasible, so a compromise must be found. The balance between working up a reasonable number of samples and the desired quality of the analytical results will dictate the mode of the irradiation–counting scheme. A typical scheme has been described for atmospheric aerosols (Fig. 3) [35]. For the *short-lived radionuclides*, irradiation of a 2- to 25-mg sample for a few minutes, together with a flux monitor, is employed. Typically, 10–15 elements are determined from this irradiation. *Longer irradiation* of a number of samples, together with a multielement standard or a flux monitor, is followed by two to three counts after decay times varying from 1 to 30 d. In total, the scheme allows detection of up to 45 elements in favorable cases although, more realistically, about 30 elements can be determined with precisions varying from 2 to 10%. Shorter and less complicated schemes can be applied at the expense of precision and the number of elements determined. The selectivity for the determination of some elements can be enhanced by applying epithermal, pulsed, or cyclic activation and by counting with low-energy photon detectors or anticoincidence spectrometers.

Radiochemical Neutron Activation Analysis.

The ultimate sensitivity and the lowest detection limits are generally obtained when radiochemical separations are performed after irradiation. Chemical treatment of the samples and the standards is aimed at separating the radionuclides formed. The separated fractions are then used in a chemical and physical form suitable for counting. Because the actual amount of radionuclides is very small, a chemical carrier is added before the decomposition step to ensure chemical equilibration with the radionuclide. The next stage is chemical separation using classical methods such as precipitation, distillation, solvent extraction, chromatography, ion exchange, or electrodeposition. Since the amount of carrier added is known and greatly exceeds the amount of the element originally present, the yield of the chemical separation can be determined. Correction of the activity measured is possible.

If clean separation of one radionuclide is achieved, scintillation counting or even β-counting can be performed in principle. Nevertheless, high resolution HP-Ge spectrometry is applied to check the absence of any interfering radionuclide. For most types of samples, such as geological, environmental, and biological materials, typical group separation schemes have been developed and can be found in the literature. For biological samples, the emphasis is on the removal of major activities from ^{32}P, ^{24}Na, and ^{82}Br. In geological samples, rare earth elements or noble metals are often determined after separation. Sometimes carriers or so-called scavengers such as hydroxides of Fe^{3+}, Mn^{4+}, and Sn^{4+} are used to remove interfering activities or to separate the analyte radionuclides.

Sample + Ti flux monitor

Irradiation
t_{irr} = 5 min
φ = 2.6 × 10^{12} neutrons cm^{-2} s^{-2}

^{28}Al
^{52}V
^{66}Cu
^{51}Ti
^{37}S
^{49}Ca

Count of sample
t_d = 3 min
t_c = 6 min

Count of Ti flux monitor
t_d = 12 min
t_c = 30 s

^{27}Mg
^{80}Br
^{38}Cl
^{128}I
^{56}Mn
^{24}Na
^{116m}In
^{165}Dy

Count of sample
t_d = 15 min
t_c = 20 min

Sample + standard

Irradiation
t_{irr} = 7–21 h
φ = 1.0 × 10^{12} neutrons cm^{-2} s^{-1}

Count of sample + standard
t_d = 20–30 h
t_c = 30 min

^{42}K
^{64}Cu
^{69m}Zn
^{82}Br
^{76}As
^{72}Ga
^{122}Sb
^{140}La
^{153}Sm
^{152m}Eu
^{187}W
^{198}Au

Count of sample + standard
t_d = 5–10 d
t_c = 1 h

^{82}Br
^{122}Sb
^{140}La
^{155}Sm
^{198}Au

Count of sample + standard
t_d = 20–30 d
t_c = 1–2 h

^{46}Sc
^{51}Cr
^{59}Fe
^{60}Co
^{58}Co
^{65}Zn
^{75}Se
^{110m}Ag
^{124}Sb
^{134}Cs
^{131}Ba
^{177m}Lu
^{141}Ce
^{152}Eu
^{203}Hg
^{233}Pa

Figure 3. Irradiation – counting scheme for INAA of air particulates collected on filter paper [35]
φ = Thermal neutron flux; t_{irr} = Irradiation time; t_d = Decay time; t_c = Counting time

24.2.7. Sensitivity and Detection Limits

The smallest amount of an element that can be determined depends on the specific activity produced and the minimum activity measurable with sufficient precision. From Equation (10) the activity produced and measured per gram of an element can be calculated. However, the minimum activity that can be measured depends not only on the decay properties of the radionuclide and on the counting equipment, but also on the background of the detector or the Compton continuum on which the photopeak must be detected in gamma spectrometry. Thus the detection limit of an element for specific irradiation – counting conditions is not immutable since it depends on the presence of other radionuclides.

For measurement results giving a Poisson distribution, as in radioactive decay, the equation used to calculate the standard deviation σ is simply the square root of the number of counts recorded when the number is large. The net signal S is obtained as the difference between the measured signal $(S+B)$ minus the background B (e.g., Compton continuum in gamma spectrometry). Thus the statistically derived standard deviation is given as

$$\sigma_s = (S + B + B)^{1/2} \tag{16}$$

The limit of detection depends on whether the background is well known or is not well known as in gamma spectrometry. CURRIE [36] defines-

Table 3. Detection limits (units: 10^{-12} g) for INAA under ideal conditions of gamma spectrometry [135 cm^3 Ge(Li) detector, lead shielded] [a] [35]

Na	0.1	Ge	30	Sb	0.02	Tm	2
Si	900 000	As	0.03	Te	3	Y	0.06
Cl*	10 000	Se	3	I*	1500	Lu	0.03
Ar*	30	Br	0.05	Cs	0.5	Hf	0.4
K	5	Rb	8	Ba	40	Ta	0.4
Ca	1 500	Sr*	50	La	0.02	W	0.04
Sc	0.04	Y	1 500	Ce	1.5	Re	0.015
Ti	600	Zr	15	Pr	0.5	Os	0.8
Cr	3	Nb	15 000	Nd	5	Ir	0.006
Mn*	0.4	Mo	0.5	Sm	0.002	Pt	1
Fe	300	Ru	2	Eu	0.003	Au	0.001
Co	0.8	Pd	40	Gd	0.9	Hg	1
Ni	200	Ag	2	Tb	0.15	Tl	3 000
Cu	2	Cd	0.6	Dy*	2	Pb	30 000
Zn	6	In*	4	Ho	0.3	Th	0.2
Ga	0.06	Sn	100	Er	0.5	U	0.06

[a] For detection of radionuclides with $t_{1/2} > 1$ h, irradiation: 48 h at 2×10^{14} cm^{-2} s^{-1} (cadmium ratio $= 11$, i.e., ratio obtained by irradiation without and with cadmium cover); decay time 10 h; counting time 64 h. For detection of radionuclides with $t_{1/2} < 1$ h (marked with *), irradiation: 2 h at 3×10^{11} cm^{-2} s^{-1}; decay time 1 h; counting time 3 h.

three specific quality levels for a measurement ($\rightarrow$ Chemometrics):

1) *Decision limit* L_c is the net signal above which an "observed" signal can be accepted (95 % probability) as detected; L_c equals $1.64\,\sigma_B$ or $2.33\,\sigma_B$; for well-known or not well-known background, respectively

2) *Detection limit* L_D is the "true" net signal that a priori may be expected to be detected; L_D equals $3.29\,\sigma_B$ or $4.65\,\sigma_B$

3) *Determination limit* L_Q is the net signal on which quantitive determination with a precision of 10 % can be performed; L_Q equals $10\,\sigma_B$ or $14.1\,\sigma_B$

Because the sources of errors in NAA are generally known, they can be kept under control. When good laboratory practice is applied, the reproducibility and precision can be nearly as good as the counting statistics, in favorable cases as low as $1-2$ %.

Detection limits obtainable by multielement INAA were determined by VERHEYKE and RATH [37] for compromise irradiation–counting conditions but for ideal gamma spectrometry conditions, namely, calculated from a background (Compton continuum) spectrum obtained from very pure irradiated graphite (Table 3). For several elements (Ar, Mn, Sr, Y, Dy) a better sensitivity can be calculated for a shorter decay time. For some elements (F, Mg, Al, V, and Rh) giving rise to radionuclides with half-lifes of less than 10 min no limits are given, although they can be determined with low detection limits after a short irra-

diation and a fast count. Elements such as S, P, and Bi produce only β-emitters and are therefore absent from the table.

24.3. Photon Activation Analysis [23], [24]

Photons — often obtained as bremsstrahlung produced in an electron accelerator — with energies varying from 10 to 45 MeV can overcome the threshold of (γ, n), (γ, p), $(\gamma, 2n)$, and (γ, α) reactions. Inelastic scattering with low-energy photons leads to short-lived metastable isotopes by (γ, γ') reaction. The light elements C, N, O, and F are determined by (γ, n) reactions, but the reaction leaves the nuclides proton rich, so that a positron emitter is formed. Radiochemical separation or decay curve analysis is needed. Detection limits of < 1 µg are obtainable. *Instrumental photon activation analysis* (IPAA) is possible for heavier elements in the presence of an excess of low atomic number material, because when a bremsstrahlung spectrum is used, the yields of the reactions increase monotonically with increasing atomic number. Typical detection limits range from 0.01 to 1 µg. IPAA has been applied to the analysis of many environmental samples. Typically, about 15 elements can be determined precisely. Photon activation analysis was considered a valuable complement to NAA for the determination of light elements and some heavy elements before charged-particle activation analysis

(CPAA) was developed. The major drawback is the limited availability of high-energy electron beams with sufficient intensity.

24.4. Charged-Particle Activation Analysis

Although NAA and PAA have a few characteristics in common with charged-particle activation analysis (CPAA), there are fundamental differences, owing to the electric charge of the bombarding particle.

In NAA there are endo-ergic reactions [e.g., (n, 2n) or (n, α) in FNAA] as well as exo-ergic reactions [e.g., (n, γ) in TNAA] and this is also the case for CPAA. However, owing to the Coulombic repulsion between a charged particle (CP) and the target nucleus, CPs always require a minimum energy, even for endo-ergic reactions. For CPAA, mono-energetic particles with energies between 5 and 50 MeV are utilized, leading to more nuclear reactions, and more possible nuclear interferences. For a proton energy below 10 MeV, e.g, in general (p, n) and (p, α) reactions are possible. Below 20 MeV also (p, 2n) (p, d) (p, t), (p, ^{3}He), (p, αn) ... reactions have to be considered. For TNAA only (n, γ) reactions have to be considered, nuclear interference from a (n, p) reaction is fairly infrequent.

CPs are slowed down in matter and are all stopped at depth, called the range, being 0.1 to 1 mm. As a consequence the beam intensity is hardly reduced, in contrast to neutron shadowing. Moreover, the induced activity is not related to the mass of the analyte, but to its mass fraction. So standardization (calibration) is very different from NAA.

The *stopping power (S)* is defined as the energy loss (– dE) per unit pathlength (d*l*) for a particular CP in a particular target:

$$S = \frac{-\,dE}{dl} \tag{17}$$

It depends on the CP energy. As a result the CPs are stopped at a depth, called *range R*. The range is characteristic for a CP, a target, and the CP energy, being mono-energetic.

The equivalent of $\varphi\,\sigma\,N_A$ in Equations (3)–(10), i.e., the reaction rate in s^{-1}, can be written for an infinitively thin target

$$I\,\sigma\,n\,dl \tag{18}$$

with I = beam intensity (number of CPs per s), n = number of target nuclides per cm^3, and d*l* = thickness (cm) of the infinitively thin target. For a target thicker than the range R, equation (18) is integrated

$$In \int_0^R \sigma_I dl = In \int_{E_i}^0 \frac{\sigma_E dE}{dE/dl} = In \int_{E_i}^{E_t} \frac{\sigma_E dE}{S_E} \tag{19}$$

with E_i the incident CP energy.

The sample is supposed to be homogenous. The CP beam intensity is nearly unchanged when penetrating the target, while the CP energy decreases, and consequently the reaction cross section. As the reaction cross section σ is only known as a function of the energy, the stopping power S is used. For a target for which the thickness D is less than the range R, Equation (19) can be rewritten

$$In \int_0^D \sigma_I dl = In \int_{E_i}^{E_0} \frac{\sigma_E dE}{dE/dl} = In \int_{E_0}^{E_t} \frac{\sigma_E dE}{S_E} \tag{20}$$

with E_0 = outgoing energy.

It is clear that the induced activity is proportional to the mass fraction ($\sim n$ in cm^{-3}) of the analyte, in contrast to NAA where it is proportional to the mass of the analyte ($\sim N^A$ dimensionless) in Equations (3)–(10). This difference in dimensions may be suprising for two different modes of activation analysis. However, it is related to (1) the analysed depth, being limited to the range (cm) and (2) the use of beam intensity (in s^{-1}) versus fluency rate (in cm^{-2} s^{-1}) in NAA.

As relative standardization is applied in CPAA, there is no need for absolute activity and beam intensity measurements, nor for absolute data of the reaction cross section σ and stopping power S. If the (relative) cross section is not accurately known, approximations have been developed. The stopping power of any elemental matter can be calculated accurately; for mixtures or compounds the major matrix should be known. If not, an internal standardization method has been developed.

CPAA has been applied successfully for the determination of (trace) elements in the bulk of solid samples, such as pure metals and alloys, semiconductors, geological and solid environmental materials. The method has proven its unique capabilities and outstanding performance in the

determination of light elements such as B, C, N, and O. Activation analysis has the unique possibility of removing surface contamination after irradiation and prior to measurement. CPAA has been a milestone, e.g., in the assessment of the bulk oxygen concentration in aluminum.

Recently CPAA has been extended to surface characterization. If it is possible to determine a trace element A in a matrix B at the μg/g level in the bulk of a sample, i.e., in a 1 mm thick layer, then it is also possible to determine the thickness of a mono-elemental layer A on a substrate B (that does not contain A) down to 1 nm. Alternatively, one can determine the composition of such a layer if the thickness is known, and consequently the stoichiometry without any foreknowledge of the thickness.

The principles of CPAA for bulk analysis are summarized in [25] and [26]; for surface characterization in [27] and in [28]. Applications of CPAA for bulk analyses are reviewed in [38]; the feasibility study on surface characterization in [39].

24.5. Applications

24.5.1. High-Purity Materials

Since the 1960s, activation analysis has been used intensively for determining trace elements in high-purity materials. Analysis of semiconductor materials, as well as of high-purity metals, obtained by zone refining, and of materials used in nuclear technology, has been one of the major fields of interest in activation analysis. Impurities are known to affect the properties of these materials, even when present in concentrations so low that they can barely be detected by other methods.

INAA is especially suited for the multielement analysis of matrices that, on neutron irradiation, do not produce intense gamma radiation or produce only short-lived ones. Examples of these matrices are graphite, silicon, and aluminum, or their oxides. A typical irradiation counting scheme allows the determination of 55 elements at submicrogram-per-gram concentrations in wafers of silicon [37]. Also, other semiconductor or alloying materials such as Se, Ge, Sb, Te, and ultrapure metals such as Cu, Fe, Ti, Mo, Nb, Ni, Ga Zn, Sn, Zr, Bi, Pb, and noble metals have been analyzed. Often a postirradiation separation of the matrix activity is required, or a radiochemical separation scheme must be applied based on ion ex-

change, extraction, or precipitation. Matrix activation can sometimes be avoided by irradiation in a well-thermalized neutron flux or under the cover of cadmium (ENAA).

Photon activation (PAA) is used for elements not activated by (n, γ) reaction, with the emphasis on carbon, nitrogen, and oxygen. Detection limits of the order of 20 μg can be obtained.

Even lower limits of detection can be obtained for determining the light elements in semiconductor materials or ultrapure metals by *CPAA*.

The major advantage of activation analysis over nearly all other techniques of analysis — namely, the absence of contamination problems after irradiation — is extremely valuable in this field of application. Preirradiation treatments should be carried out in clean environments, and to further avoid surface contamination, postirradiation washing or etching is highly recommended. Assessing the impurity levels of these typical industrial products by activation analysis is often used in the quality control of manufacturing companies and has therefore been entirely automated and computerized by using a single comparator for calibration.

24.5.2. Environmental Materials

Neutron activation analysis is one of the major techniques for the determination of many minor and trace elements in a large variety of solid environmental and pollution samples, such as atmospheric aerosols, particulate emissions, fly ash, coal, incineration ash, and sewage sludge. *Instrumental neutron activation analysis* of total, inhalable, or respirable airborne particulate matter collected on a cellulose or membrane filter, or in a cascade impactor on some organic substrate, allows the determination of up to 45 elements by an irradiation–counting scheme similar to the one given in Figure 3. *Radiochemical NAA* is applied only when extremely low limits of determination are required. *Instrumental photon activation analysis* is also complementary to INAA.

The problem in charged-particle activation analysis related to a matrix of unknown composition and degradation (votalization) of the organic matrix during irradiation, has been solved by the development of an internal standardization method.

When applied to water samples, NAA suffers from some severe drawbacks compared with other multielement techniques. Irradiation of water re-

sults in radiolysis. Evaporation, freeze drying, adsorption on charcoal, and ion exchange can be used to remove water. A final problem is the interference of intense radiation induced in matrix elements such as Na, K, Cl, Br, and P. Saline water in particular, yields difficult samples for INAA. The preconcentration step may include separation from these elements. A wealth of literature exists describing these environmental applications [40], [41].

24.5.3. Biological Materials

Activation analysis, particularly with neutrons, is a very effective method for elemental trace analysis of biological materials and plays an important role in studies of trace elements related to health. Because of the limited sensitivity of classical analytical methods, the trace and ultratrace elements that play an important role in biological systems can rarely be determined by the usual methods. Radiochemical NAA has contributed significantly to establishing the normal levels of these elements in blood (total blood, serum, packed cells, and erythrocytes), liver, kidney, lung, muscles, bones, teeth, nerves, and hair and in studies involving disease [42]. It is the method of choice for the determination of Sb, As, Cs, Co, Mn, Mo, Rb, Se, Ag, and V, and is very reliable for Br, Cr, Cu, Hg, I, and Zn, but it lacks sensitivity for Cd, Ni, and Sn [43]. Since radiochemical separations are generally required, a number of group separations using ion exchange or distillation procedures have been developed to remove at least the interfering activities from P, Na, and Br. For a limited number of elements (Co, Cs, Fe, Rb, Se, and Zn), INAA is possible only after a long decay period [44]. Activation analysis offers many advantages for the analysis of foods and plants. INAA of vegetables, meat, fish, and poultry allows the detection of about 20 trace elements [43].

In the last decade speciation (the determination of the species of an element, i.e., its oxidation state or its compounds) has become more important than trace element determination. Speciation analysis is obtained by hyphenated techniques, i.e., coupling of, for example, a chromatographic separation technique to a very sensitive elemental detector. As NAA cannot measure on-line, the role of NAA becomes less important. However, for the development of separation and preconcentration techniques, the use of radiotracers with very high specific activity is an outstanding tool, as these techniques can be developed at realistic concentration levels. Radiotracers produced by charged-particle activation, provide just such a specific activity, as the radionuclide formed is not an (radio)isotope of the element irradiated.

24.5.4. Geo- and Cosmochemistry

Activation analysis has probably most often been applied in the analysis of geological samples. The NAA results on rare earth elements in meteorites, rocks, and sediments were a significant contribution to the development of modern geo- and cosmochemistry. Also in geological surveys, hundreds of analyses are performed on soil and stream sediments by this technique. Silicate rock, nonsilicate rock, meteorites, minerals, and marble can often be analyzed purely instrumentally. *Epithermal activation* enhances the detectability of a number of interesting elements. When, however, up to 40 elements are to be determined — including the rare earths and the platinum-group elements — a preconcentration or some group separations based on ion exchange and liquid – liquid extraction are applied. A *14-MeV activation* can be applied for the determination of some major elements (O, F, Mg, Al, Si, P, Fe, Cu, Zn). Radioanalysis in geochemistry has been described in [45].

24.5.5. Art and Archaeology

Archaeologists and museum directors hope that small samples of art objects can be analyzed with high sensitivity, preferentially in a nondestructive way. INAA can usually achieve this. The wealth of information on the concentration of trace elements, obtained by the technique, can often serve as so-called fingerprints for identification or classification purposes. The concentrations may give clues to the provenance of the artifact, the methods of treatment, or the cultural or technological context in which it was made. The artifacts of archaeological interest analyzed range from metals (coins, medals, statuettes, utensils) to nonmetallic artifacts (clays, pottery, marble, obsidian, and paintings). An overview is given in [46].

24.6. Evaluation of Activation Analysis

When activation analysis is compared with other analytical techniques, it is most suited for the multielemental analysis of minor, trace, and ultratrace elements in solid samples. Often a purely instrumental approach is possible, making the technique nondestructive and leaving the sample intact for other investigations. About 65 % of the elements of the Periodic Table can be determined by *INAA* at concentrations lower than micrograms per gram and, in favorable cases, down to nanograms per gram. The detection of another 20 % is less sensitive, and about 10 – 15 % (mostly light elements) can rarely be detected at all. These can, however, be detected by *FNAA* (F) or *CPAA* (B, C, N, O). The lowest detection limits are always obtained after radiochemical separation. These low detection limits are possible because the technique can be made highly selective. Adjustable parameters are irradiation and waiting time before counting, the nature and energy of the bombarding particle or radiation, and above all the high resolving power of the measurement. A *major benefit* of activation analysis is that it can provide very accurate results for trace concentrations. Contamination from reagents or the laboratory environment is excluded after irradiation. Radiochemical separations can be performed after the addition of inactive carrier, avoiding the need for working with trace amounts. The absorption effects of neutrons during irradiation and of gamma rays during counting are small and generally negligible or can be corrected for. The information obtained is independent of matrix or chemical form. Other sources of error, such as interfering reactions, can generally be calculated and thus accounted for or avoided. Therefore the reproducibility and the precison can be as good as the counting statistics when good laboratory practice is applied. In favorable cases and for homogeneous samples, it can be as good as 1 – 2 %. For the detection of major contents, a precision of > 0.5 % is rarely achievable. The accuracy depends to a large extent on the calibration procedure. When appropriate standards are used, very accurate results can be obtained, making this a technique preeminently suited for certification purposes. The *major drawbacks* are probably that it is expensive, requires access to nuclear facilities such as a reactor, and is generally slow. Rapid analysis based on short-lived isotopes is possible

only in exceptional cases. The cost of the counting equipment is of the order of or much less than most other instruments for trace analysis [atomic absorption spectrometry (AAS), inductively coupled plasma optical emission spectrometry (ICP-OES) and mass spectrometry (ICP-MS), X-ray fluoroescence (XRF), etc.]. To work with radio-isotopes, a laboratory must meet legal requirements for radiological safety. An additional disadvantage is the lack of information on the chemical form of the element, which makes speciation studies difficult. Finally, analysis of strongly activated matrices renders the instrumental approach impossible and necessitates the application of important shielding material for the radiochemical separations.

24.7. References

General References

[1] W. D. Ehmann, D. E. Vance, J. D. Winefordner, I. M. Kolthoff (eds.): *Radiochemistry and Nuclear Methods of Analysis*, Wiley-Interscience, New York 1991.
[2] G. Erdtmann, H. Petri in I. M. Kolthoff, P. J. Elving, V. Krivan (eds.) *Treatise on Analytical Chemistry*, 2nd ed., part I, vol. 14, Wiley-Interscience, New York 1986, pp. 419 – 643.
[3] J. Hoste et al. in I. M. Kolthoff, P. J. Elving, V. Krivan (eds.): *Treatise on Analytical Chemistry*, 2nd ed., part I, vol. 14, Wiley-Interscience, New York 1986, pp. 645 – 775.
[4] S. J. Parry in J. D. Winefordner, I. M. Kolthoff (eds.): *Activation Spectrometry in Chemical Analysis*, Wiley-Interscience, New York 1991.
[5] Z. B. Alfassi (ed.): *Activation Analysis*, vols. I and II, CRC-Press, Boca Raton 1989.
[6] J. Tölgyessy, M. Kyrš: *Radioanalytical Chemistry*, vols. 1 and 2, Ellis Horwood, Chichester 1989.
[7] D. De Soete et al. (eds.): *Neutron Activation Analysis*, Wiley-Interscience, London 1972.
[8] Z. B. Alfassi (ed.): *Determination of Trace Elements*, VCH Verlagsgesellschaft, Weinheim 1994.
[9] Z. B. Alfassi (ed.): *Chemical Analysis by Nuclear Methods*, John Wiley, Chichester 1994.
[10] M. D. Glascock: Activation Analysis, in: Z. B. Alfassi (ed.): *Instrumental Multi-element Chemical Analysis*, Kluwer, Dordrecht 1999.
[11] O. Navrátil et al. (eds.): *Nuclear Chemistry*, Ellis Horwood, New York 1992.
[12] W. Seelmann-Eggebert, G. Pfennig, H. Münzel, H. Klewe-Nebenius: *Karlsruher Nuklidkarte – Charts of the Nuclides*, 5th ed., Kernforschungszentrum, Karlsruhe 1981.
[13] J. R. Parrington, H. D. Knox, S. L. Breneman, E. M. Baum, F. Feiner: *Chart of the Nuclides*, 15th ed., KAPL, Knolls Atomic Power Lab., New York 1996.
[14] R. B. Firestone, V. S. Shirley, C. M. Baglin, S. Y. F. Chu, J. Zipkin: *Table of Isotopes*, 8th ed. (book and CD-ROM + yearly update), John Wiley, Chichester 1996.

[15] M. D. Glascock: *Tables for Neutron Activation Analysis*. 4th ed., University of Missouri, Columbia 1996.

[16] http://nucleardata.nuclear.lu.se/database/se

[17] http://www.nndc.bnl.gov/nndc/nudat

[18] http://www.nndc.bnl.gov/nndcscr/pc_prog/   *Only* NDTxx_16.EXE *and* RADTIONS.ZIP *are required.*

[19] http://www.nndc.bnl.gov/nndc/exfor/

[20] http://www-nds.iaea.or.at/exfor/

[21] http://www.nea.fr/html/dbdata/x4/welcome.html   *Only for users registered* at http://www.nea.fr/html/signon.html

[22] http://nucleardata.nuclear.lu.se./database/toi/

[23] C. Segebade, H. P. Weise, G. L. Lutz: *Photon Activation Analysis*. De Gruyter, Berlin 1987.

[24] A. P. Kushelevsky: "Photon Activation Analysis", in Z. B. Alfassi (ed.): *Activation Analysis*, **vol. II**, CRC Press, Boca Raton 1990, pp. 219–2237.

[25] K. Strijckmans: "Charged Particle Activation Analysis", in Z. B. Alfassi (ed.): *Chemical Analysis by Nuclear Methods*, Chap. 10, John Wiley, Chichester 1994, pp. 215–252.

[26] K. Strijckmans: "Charged Particle Activation Analysis", in A. Townshend, R. Macrae, S. J. Haswell, M. Lederer, I. D. Wilson, P. Worsfold (eds.): *Encyclopedia of Analytical Science*, Academic Press, London 1995, pp. 16–25.

[27] K. Strijckmans: "Charged Particle Activation Analysis", in D. Brune, R. Hellborg, H. J. Whitlow, O. Hunderi (eds.): *Surface Characterisation: a User's Sourcebook*. Wiley-VCH, Weinheim 1997, pp. 169–175.

[28] K. Strijckmans: "Charged Particle Activation Analysis", in R. A. Meyers (ed.): *Nuclear Methods – Theory and Instrumentation of the Encyclopedia of Analytical Chemistry (EAC), Part 2: Instrumentation and Applications*, Section: Nuclear Methods, John Wiley, New York 2000.

Specific References

[29] O. T. Høgdahl in IAEA (ed.): *Radiochem. Methods Anal. Proc. Symp. 1964* I (1965) 23.

[30] J. Hoste: "Isotopic Neutron Sources for Neutron Activation Analysis," *IAEA-TECDOC* 1988, 465.

[31] F. De Corte, F. Simonits, A. De Wispelaere, J. Hoste, *J. Radioanal. Nucl. Chem.* **113** (1987) 145–161.

[32] F. De Corte et al., *J. Radioanal. Nucl. Chem.* **169** (1993) 125–158.

[33] S. Van Lierde, F. De Corte, D. Bossus, R. Van Sluijs, S. S. Pommé, *Nucl. Instr. Meth.* **A422** (1999)874–879.

[34] G. Erdtmann, H. Petri in I. M. Kolthoff, P. J. Elving, V. Krivan (eds.): *Treatise on Analytical Chemistry*, 2nd ed., **part I, vol. 14.** J. Wiley & Sons, New York 1986, p. 446.

[35] R. Dams, J. A. Robbins, K. A. Rahn, J. W. Winchester, *Anal. Chem.* **42** (1970) 861–866.

[36] L. A. Curie, *Anal. Chem.* **40** (1968) 586–593.

[37] M. L. Verheyke, H. J. Rath in M. Grasserbauer, H. W. Werner (eds.): *Analysis of Microelectronic Materials and Devices*, J. Wiley and Sons, New York 1991, pp. 1–39.

[38] G. Blondiaux, J. L. Debrun, C. J. Maggiore: "Charged Particle Activation Analysis", in J. R. Temer, M. Nastasi, J. C. Barbour, C. J. Maggiore, J. M. Mayer (eds.): *Handbook of Modern Ion Beam Materials Analysis*, Material Research Society, Pittsburgh, PA 1995.

[39] K. De Neve, K. Strijckmans, R. Dams: "Feasibility Study on the Characterization of Thin Layers by Charged-Particle Activation Analysis." *Anal. Chem.* **72** (2000) 2814–2820.

[40] J. Tolgyessy, E. H. Klehr: *Nuclear Environmental Chemical Analysis*, Ellis Horwood, Chichester 1987.

[41] R. Dams in I. M. Kolthoff, P. J. Elving, V. Krivan (eds.): *Treatise on Analytical Chemistry*, 2nd ed., part I, vol. 14, J. Wiley & Sons, New York 1986, p. 685.

[42] J. Versieck, R. Cornelis: *Trace Elements in Human Plasma or Serum*. CRC Press, Boca Raton 1989.

[43] W. C. Cunningham, W. B. Stroube, *Sci. Total Environ.* **63** (1987) 29–43.

[44] R. Cesareo (ed.): *Nuclear Analytical Techniques in Medicine*, Elsevier, Amsterdam 1988.

[45] H. A. Das, A. Faanhof, H. A. van der Sloot: *Radioanalysis in Geochemistry*, Elsevier, Amsterdam 1989.

[46] J. Op de Beeck in I. M. Kolthoff, P. J. Elving, V. Krivan (eds.): *Treatise on Analytical Chemistry*, 2nd ed., part I, vol. 14, J. Wiley & Sons, New York 1986, p. 729.

25. Analytical Voltammetry and Polarography

GÜNTER HENZE, Institut für Anorganische und Analytische Chemie der Technischen Universität Clausthal, Clausthal-Zellerfeld, Germany

25.1. Introduction

The term *voltammetry* is used to classify that group of electoanalytical techniques in which the current (ampere) that flows through an electrochemical cell is measured as the potential (volt) applied to the electrodes in the cell is varied. The term is derived from the units of the electrical parameters measured — *volt-am* (pere)-*metry*.

The essential difference between voltammetric and other potentiodynamic techniques, such as constant current coulometry, is that in voltammetry *an electrode with a small surface area* ($< 10 \, mm^2$) is used to monitor the current produced by the species in solution reacting at this electrode in response to the potential applied. Because the electrode used in voltammetry is so small, the amount of material reacting at the electrode can be ignored. This is in contrast to the case in coulometry where large area electrodes are used so that all of a species in the cell may be oxidized or reduced.

When mercury is used as the electrode in the form of small drops falling slowly from a fine capillary tube in the test solution, the technique has the special name *polarography*. This name is derived from the fact that the electrode can be polarized. An electrode is said to be polarized when no direct current flows across its interface

with the solution even though there is a potential difference across this interface. In his work published in 1922 on "Electrolysis with a dropping mercury electrode", Jaroslav Heyrovsky referred to this phenomenon and called the recorded current – potential polarization curves *polarograms* [1].

The recommendation of IUPAC is that voltammetry is the general term to be used when current – potential relationships are being investigated and that only when a flowing conducting liquid electrode (such as a dropping mercury electrode) is used as the working electrode should the term polarography be used. Polarography, the original technique, is thus a special case of voltammetry.

The *voltammetric cell* is a multi-phase system, in the simplest form designed with two electronic conductors called *electrodes* immersed in an electrolytic conductor (ionic cell solution). The electrolytic conductor consists of the sample solution with the electrochemically active *analyte* and an excess of an inert *supporting electrolyte*. The analyte is an inorganic or organic species and can be present as a cation, an anion or a molecule.

The application of a voltage to the electrodes produces, as the result of an electrode process, an electrical response — a current — from the analyte in the cell solution. The nature and magnitude of this response may be used to both identify and

quantify the analyte. Most polarographic and voltammetric procedures utilize electrode processes in which electrons are exchanged between the two phases. Such a process is referred to as a *charge transfer reaction*, because of the flow of charge, i.e., a current, through the electrode.

The small electrode used to monitor the response of the analyte is known as the *working electrode* (WE). Even though only a negligible amount of material is involved in the processes occurring at the working electrode, its small size ensures that a high current density develops at its surface. Consequently, it is the processes which occur at this small electrode that control the current flow through the cell. The WE may be constructed from a wide variety of conduction materials, preferably of mercury and graphite, or alternatively of gold and platinum. For polarographic experiments the dropping mercury electrode (DME) is important and for voltammetric investigations the thin mercury film electrode (TMFE), the hanging mercury drop electrode (HMDE), the glassy carbon electrode (GCE), the carbon paste electrode (CPE), the rotating platinum electrode (RPE), and chemically modified electrodes (CME's) may be used.

The second electrode in the simple voltammetric cell, called the *counter electrode* (CE), serves two purposes. It is used to control the potential applied to the working electrode and to complete the circuit for carrying the current generated at the WE. In the former role it must act as a reference electrode.

In modern measuring systems the current carrying role of the counter electrode is separated from its potential control role by introducing the *auxiliary electrode* (AE) as a third electrode of the cell. The addition of the auxiliary electrode means that the counter electrode is now used only to control the potential of the working electrode and so becomes a true *reference electrode* (RE). Two electrodes which are commonly used as reference electrodes for the precise control of the working electrode potential in aqueous media are the silver – silver chloride electrode in a solution of fixed chloride concentration and the saturated calomel electrode or SCE (a mercury – mercurous chloride electrode in a saturated KCl solution). These electrodes are robust, easily constructed and maintain a constant potential. The three electrode system will be discussed in more detail in Chapter 25.3.

When the voltage applied to the electrodes (WE and AE) is such that no charge transfer reactions are occuring, the working electrode is said to be "polarized". At higher voltages, the polarization of the working electrode disappears as soon as the conditions exist for a charge transfer reaction, $Ox + ne^- \rightleftarrows Red$, to occur at the interface between the electrolyte solution and the WE. The current flow that results from the oxidation or reduction of the analyte is known as a *Faradaic current*. Its magnitude depends on the concentration of the analyte in the sample solution and on the kinetics of all steps in the associated electrode process. Because of the small surface area of the working electrode ($1 - 10$ mm^2) the current flow is normally in the nA to µA range.

The temporal decrease of the analyte concentration in the interfacial region of the electrode, as a result of reduction or oxidation, is balanced chiefly by diffusion of the analyte from the bulk of the solution. The transport of charged analytes by migration is insignificant in the presence of an excess of supporting electrolyte in the test solution. Normally migration can be neglected if the supporting electrolyte concentration is at least 10^3 times higher than the analyte concentration. However, mass transport to the interface is assisted by convection, facilitated by movement of the solution relative to the electrode, e.g., by stirring or by the action of the mercury dropping from the capillary of the DME.

According to Nernst, diffusion occurs in a stable layer of thickness δ at the interface of the working electrode, i.e., from the electrode surface to some distance into the solution (Fig. 1). Within the Nernst diffusion layer, the decrease of analyte concentration c_a in the sample solution to c_s at the electrode surface is linear and voltage dependent. Convection within the layer is negligible.

For evaluating the diffusion-limited electron transfer current, the following equations are important:

1) The general equation for the cell current as a function of the number of moles converted:

$$i = \frac{dN}{dt} n F A \tag{1}$$

where N is the number of moles, t the time, n the number of electrons involved in the electrode reaction, F the Faraday constant, and A the surface area of the working electrode;

2) Fick's first law

$$\frac{dN}{dt} = D\left(\frac{dc}{dx}\right) \tag{2}$$
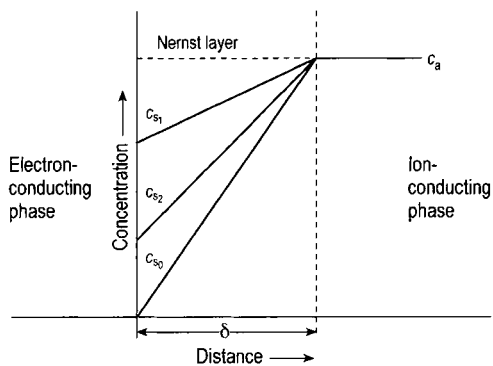
Figure 1. Variation of concentration in the Nernst diffusion layer

where D is the diffusion coefficient, and dc represents the change in concentration over the distance dx from the electrode surface. According to the ideas behind the Nernst diffusion layer, dc/dx can be replaced by $(c_a - c_s)/\delta$ (Fig. 1). From Equations (1) and (2):

$$i(t) = DnFA\frac{c_a - c_s}{\delta} \tag{3}$$

The concentration gradient increases with increasing voltage (see c_{s_1} and c_{s_2} in Fig. 1) and reaches its highest value when all active species reaching the electrode by diffusion immediately participate in the electron transfer reaction. Then $c_s = c_{s_0} = 0$ and Equation (3) becomes

$$i(t) = DnFA\frac{c_a}{\delta} \tag{4}$$

In this case the cell current is also called the *limiting diffusion current* i_d, whose value remains constant with further increase of voltage; $i_{\lim}$ is accordingly the maximum value of i and, being proportional to the concentration, is important for analytical voltammetry. However, these relations apply only if the thickness of the diffusion layer remains constant during the cell reaction (stationary state). This is true if the electrode is moved at constant velocity (rotating disk or ring electrode) or if the sample solution is stirred or flows past the electrode at constant velocity (flow cell). In such cases and under ideal experimental conditions, when i is plotted against the voltage E, a line parallel to the E axis is obtained on reaching the limiting diffusion current (Fig. 2 A). The relationships change if both electrode phases remain in the stationary state. Without additional convection the
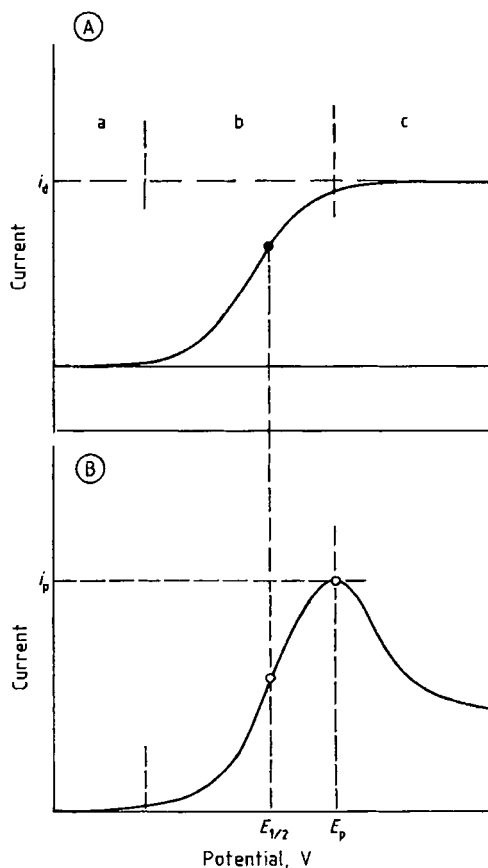
Figure 2. Current – voltage curve at electrodes
A) Stationary; B) Nonstationary diffusion layer thickness
Potential ranges for a) Electrode polarization; b) Electrode reaction; c) Limiting diffusion-controlled current i_d and peak current i_p
E_p = Peak potential: $E_{1/2}$ = Half-wave potential

thickness of the diffusion layer, and hence dc/dx, changes with time (Fick's second law). In the course of the electrode reaction, as a result of the electrochemical conversion, the diffusion layer thickness increases as $\sqrt{\pi D t}$. This means that the diffusion current, after reaching the limiting value i_p, decreases with time according to the Cottrell equation:

$$i(t) = DnFA\frac{c_a}{\sqrt{\pi D t}} \tag{5}$$

(Fig. 2 B).

In *polarography* with the DME, the relationships are even more complicated as both the con-

centration gradient and the diffusion layer width change with time. This is a consequence of the growth of the drop and hence the electrode surface area being in the opposite direction to the diffusion direction. With regard to the electrode reaction the two effects are superimposed, the time average being a constant limiting diffusion-controlled current corresponding to Figure 2 A. Further information on i_{lim} is provided by the Ilkovič equation (Section 25.2.1, Eq. 8).

The signals proportional to the analyte concentration, i_d and i_p, are analytically important.

Aside from the diffusion current, currents can occur in voltammetric/polarographic studies which are controlled by reaction kinetics, catalysis, and adsorption.

Kinetic Currents. In kinetic currents, the limiting current is determined by the rate of a chemical reaction in the vicinity of the electrode, provided this precedes the cell reaction. Electrochemically inactive compounds are converted into reducible or oxidizable forms (time-dependent protonation and deprotonation processes, formation and decomposition of complexes, etc.). Conversely, during a chemical reaction after the cell reaction, the product of the electrode reaction is converted to an electrochemically inactive form without influence on the current. However, owing to the changed equilibria between the concentrations of the oxidized and reduced forms at the electrode surface, the half-wave and peak potentials are shifted (Section 25.2.1). In evaluating kinetic effects, cyclic voltammetry can be helpful (Section 25.2.4).

Catalytic Currents. For analytical studies, catalytic currents, of which two types are known, are of great interest. In the first case the product of the cell reaction is returned, by a chemical reaction, to the initial state of the analyte. As a result, the analyte concentration at the electrode surface is always high, which results in a considerable increase of the limiting current.

The currents based on the catalytic evolution of hydrogen are the best known. The catalysts are electrochemically inactive heavy metals (primarily salts of the platinum group metals) or organic compounds. They lower the hydrogen overvoltage on the mercury working electrode and so shift hydrogen evolution to more positive potentials. The catalytic current associated with this depends, within limits, on the catalyst concentration. Because of the large number of substances that can

lower the hydrogen overvoltage, catalytic currents can only be used to determine concentration under controlled conditions.

Adsorption currents arise when the analyte, in its original form or after chemical or electrochemical conversion, is adsorbed on the working electrode. Adsorption and desorption lead to capacitance changes in the electrochemical double layer and so to analytically usable current signals (Section 25.2.3).

The theoretical principles of voltammetry and polarography are covered in several monographs [1] – [7].

25.2. Techniques

In recent years, voltammetric measurement techniques have continued to improve. Some of the improvements have had to await the advance of electronics before being used in practical analysis. Methods with modulated direct current (d.c.) ramps, e.g., differential pulse polarography, as well as alternating current (a.c.) and square-wave polarography, etc., are worth mentioning. New possibilities for controlling experiments and evaluating current signals have followed from digital electronics. The basic principles of analytical voltammetry and polarography are still best explained by way of classical d.c. polarography.

25.2.1. Direct Current Polarography

Direct current polarography (DCP), introduced by JAROSLAV HEYROVSKY in 1923, is characterized by the recording of current – voltage curves in an unstirred sample solution with the DME as the working electrode. With a slow linear change of voltage, stepwise polarograms are obtained, with marked oscillations caused by the growth and fall of the mercury drop. For better evaluation, the curves can be damped with electronic filters (Fig. 3).

The mercury flows with a mass flow rate m (mg/s) and a periodic drop time t_d (s) from a glass capillary. The surface area A of the mercury drop is greatest at the end of the drop time and, at time t during the growth of the drop is:

$$A = k\,m^{2/3}t^{2/3} \tag{6}$$
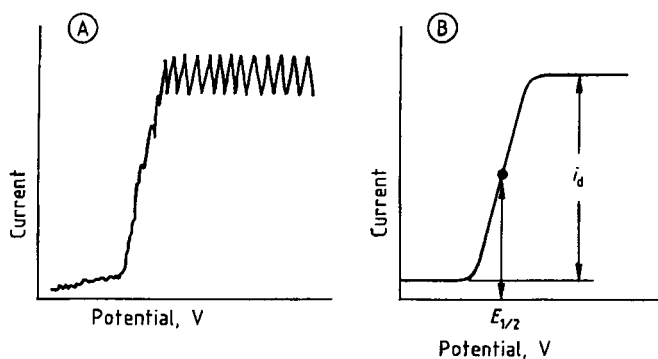
Figure 3. Direct current polarograms
A) Undamped; B) Damped

With this expression and the Cottrell equation (5), the following relation is obtained for the magnitude of the instantaneous current at time t_d for an individual drop:

$$i = k' m^{2/3} t_d^{1/6} n D^{1/2} c_a \qquad (7)$$

The constant k' includes all constant quantities, including the Faraday constant, and also takes account of the conditions for spherical diffusion. In 1935, it was calculated by Il.covič to have the numerical value 0.708.

The mean of the limiting diffusion current in the DC polarogram is obtained from Equation (7) by integration over the entire drop time:

$$i_d = \frac{1}{t_d} \int_0^{t_d} i \cdot dt = 0.607 \cdot n \cdot m^{2/3} \cdot D^{1/2} \cdot t_d^{1/6} \cdot c_a \qquad (8)$$

This is the Ilcovic equation, and applies in the voltage range in which all analyte species reaching the electrode surface by diffusion are either reduced or oxidized. The product $m^{2/3} t_d^{1/6}$ is dependent on the particular glass capillary used and is called the capillary constant.

The characteristic shape of the DC polarogram can be described by the Nernst equation:

$$E = E'_0 + \frac{RT}{nF} \ln \frac{c_{a_{ox}}}{c_{a_{red}}} \qquad (9)$$

on condition that, for the particular electrode potential E, the mean current magnitude at an individual mercury drop is considered.

For the case of reduction $Ox + e^- \rightarrow Red$:

$$i = k \left(c_{a_{ox}} - c_{a_{ox}}^0 \right) \qquad (10)$$

where $c_{a_{ox}}^0$ indicates the concentration of the analyte on the electrode surface (corresponding to c_s in Eq. 3). The same applies for the concentration $c_{a_{red}}^0$, which increases as $c_{a_{ox}}^0$ decreases; i.e., the sum $c_{a_{ox}}^0 + c_{a_{red}}^0$ corresponds to the total analyte concentration $c_{a_{ox}}$.

For the limiting diffusion current, at which $c_{a_{ox}}^0 = 0$ and $c_{a_{red}}^0 = c_{a_{ox}}$, it follows from Equation (10) that:

$$i_d = k c_{a_{ox}} \qquad (11)$$

Thus:

$$c_{a_{ox}}^0 = c_{a_{ox}} - \frac{i}{k} = \frac{i_d}{k} - \frac{i}{k} \qquad (12)$$

and:

$$c_{a_{red}}^0 = \frac{i}{k} \qquad (13)$$

When the expressions for $c_{a_{ox}}{}^0$ (Eq. 12) and $c_{a_{red}}{}^0$ (Eq. 13) are inserted in the Nernst equation (9):

$$E = E'_0 + \frac{RT}{nF} \ln \frac{i_d - 1}{i} \qquad (14a)$$

or:

$$E = E'_0 + \frac{0.059}{n} \log \frac{i_d - 1}{i} \qquad (14b)$$

For $i = i_d/2$, $E = E_{1/2}$ by definition. $E_{1/2}$, the half-wave potential in the DC polarogram is determined by the electrochemical properties of the redox system concerned.

According to Equation (14), the slope of the current curve depends on the electron transfer number n. Conversely, n can be calculated from
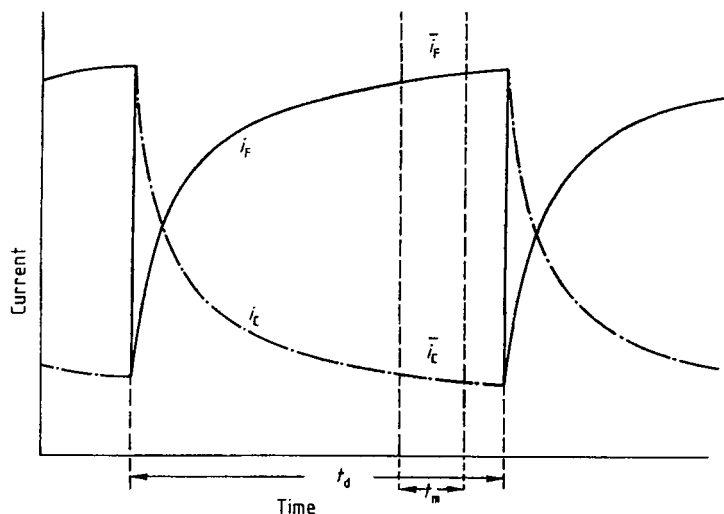
Figure 4. Change with time of i_F and i_C during the drop life t_d with the measurement or sampling interval t_m and the mean values $\bar{i}_F$ and $\bar{i}_C$ over the sampled range

the slope. However, the slope also depends on the reversibility of the electrode reaction.

If the activity coefficients f_{ox} and f_{red} for $c_{a_{ox}}$ and $c_{a_{red}}$ are inserted into the Nernst equation (9) as well as the diffusion coefficients $D_{ox}^{1/2}$ and $D_{red}^{1/2}$ (see Eq. 8), there follows for the half-wave potential:

$$E_{1/2} = E_0 - \frac{RT}{nF} \ln \frac{f_{red}D_{ox}^{1/2}}{f_{ox}D_{red}^{1/2}} \qquad (15)$$

As f and D are determined by the composition of the supporting electrolyte solution, the half-wave potential also depends on this composition.

Direct current polarography is a basic method, from which improved analytical techniques of polarography and voltammetry have been developed. With DCP, inorganic and organic analytes can be analyzed with a sensitivity of about 10^{-5} mol/L. For determining several analytes simultaneously, their half-wave potentials must be at least 100 mV apart.

On closer examination of DC polarograms, the current measured in the voltage range of the charge transfer reaction consists of two parts, the true measurement signal and an interference or noise signal. The measurement signal is the diffusion current, whose value depends on the electrode process and which, because of its origin, is called the Faradaic current i_F. The interference signal is the charging current or capacitive current i_C,

which is a consequence of charging the electrochemical double layer. In the case of the DME, the charging current is given by the equation

$$I_C = (dE/dt)C_DA + (dA/dt)C_DE \qquad (16)$$

where C_D is the double layer capacitance per unit area, E the potential applied to the cell and A the area of the electrode.

The first term of this equation allows for the effect of the rate of change of the electrode potential on the charging current, while the second term gives the increase in i_C due to the growing mercury drop.

With deceasing analyte concentration i_F becomes smaller and approaches i_C. The $i_F : i_C$ signal-to-noise ratio is the determining factor for the sensitivity of DCP. If more sensitive polarographic determinations are to be performed, this ratio must be increased by using techniques in which either the Faradaic current is enhanced or the charging current is minimized, or where both can be achieved.

During drop growth (drop life t_d), i_F and i_C are subject to changes in opposite directions. As shown in Figure 4, i_F increases with increasing surface area of the mercury drop, while i_C becomes smaller.

If the current is not measured over the whole drop life (as in classical DCP), but over a short period of time at the end of the life of a mercury
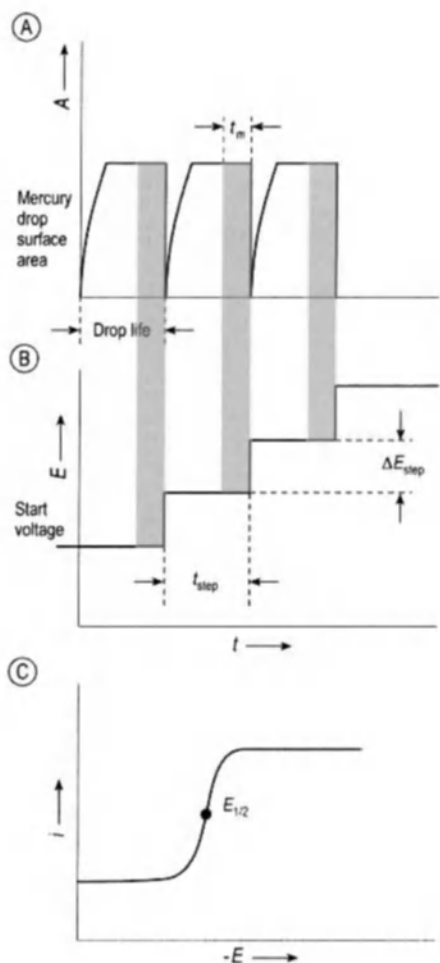
Figure 5. Measurement technique of current sampled DC polarography with the static mercury drop electrode as working electrode

(A) Variation of surface area with time for the SMDE;
(B) Potential – time profile for the staircase voltage applied to the electrode; (C) Current – potential response (polarogram).
t_m Measuring interval; t_{step} Time interval of the voltage step; ΔE_{step} Voltage step of the applied staircase waveform

drop where the growth of surface area is least the current averaged over the measurement (sampling) interval t_m has an improved ratio of $i_F : i_C$. This is the basis of the *current sampled DC-polarography* using a linear scan.

Two recent developments in the measurement technique have resulted in a major reduction in the charging current. Firstly, with the introduction of digitally controlled instruments, the linear voltage ramp applied to the electrodes has been replaced with a stepped (staircase) waveform. The voltage

steps are synchronized with the drop formation and the current is sampled at the end of the drop life. Since the voltage is constant when the current is measured, there is no contribution to the charging current from this source (see Eq. 16). The second innovation which led to a reduction in the charging current is the *static mercury drop electrode* (SMDE). In this electrode the mercury drop is extruded rapidly and its growth then stopped after which the area of the mercury electrode remains constant. The area – time profile for the SMDE, the potential – time waveform with the sampling period, and the resulting DC polarogram is shown in Figure 5. The current sampled DC polarography (sometimes called Tast Polarography) is not only more sensitive than classical DCP (ca. 5×10^{-7} mol/L), but also leads to polarograms without oscillations.

25.2.2. Pulse Techniques

With voltammetric pulse methods, BARKER and GARDNER [31] found very effective means of reducing the unwanted capacitive (noise) current, markedly improving the sensitivity of polarographic and voltammetric determinations.

A periodically changing square-wave potential with constant or increasing amplitude ΔE_p is applied to the working electrode. During the pulse duration t_p, the currents i_F and i_C that flow in response to the potential change ΔE_p decay in different ways. The Faradaic current decreases according to $t^{-1/2}$, (Eq. 5) while the capacitive current decreases according to

$$i_C = \frac{\Delta E_p}{R} \exp\left(-\frac{t}{R \cdot C_D} \right) \qquad (17)$$

(Eq. 17 corresponds to the capacitor formula with discharging resistance R and double-layer capacitance C_D).

Figure 6 illustrates the variations of i_F and i_C during the duration t_p of the square pulse ΔE_p. If the current is measured only toward the end of the pulse, essentially only i_F is recorded since i_C has effectively decayed to zero.

The methods based on the application of square voltage pulses differ in the frequency and height of the pulses applied and in the principle of measurement.

Pulse methods may be used with the DME, the HMDE, the SMDE as well as with solid electrodes and the different types of modified electrodes.
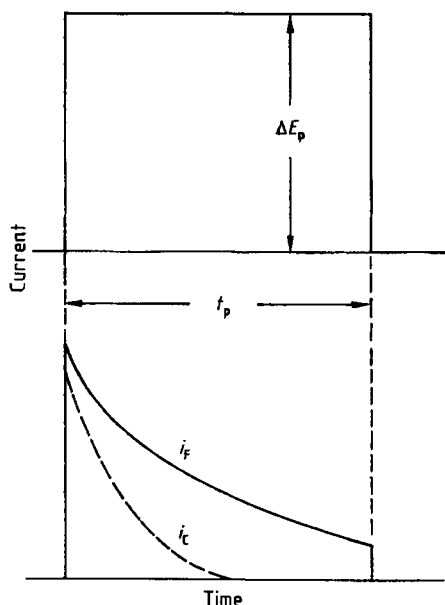
Figure 6. Principle of the pulse methods, according to BARKER

Figure 7. Square-wave alternating voltage and current signals in square-wave polarography
A) Waveform with pulse tilt: B) Capacitive current $i_{C\sim}$; C) Faradaic current $i_{F\sim}$
t_m = Measuring interval

Except for square-wave polarography, normally only one pulse is applied to each drop in analytical studies with the DME or SMDE. This pulse is synchronized with the drop period and is applied in each case towards the end of the drop life.

Square-Wave Polarography. In the oldest technique of square-wave polarography (SWP), a periodic rectangular alternating voltage with a frequency of 125 Hz and an amplitude in the range of $\Delta E_p = 5 - 30$ mV is superimposed on the linear rising d.c. ramp. The current is sampled over a measurement interval t_m near the end of each square-wave half cycle. This current is then amplified and recorded as a function of the applied d.c. voltage. Peak-shaped polarograms are obtained in which the peak current corresponds to the DC polarographic half-wave potential.

Sampling the current near the end of each half cycle ensures that the charging current arising from the sudden step in potential at the beginning of the half cycle has decayed away and does not contribute to the measured current. However, when using the DME, there is still a charging current flowing during t_m as a result of the slight increase in area of the mercury drop (see Eq. 16). This is minimized by using a tilted square-wave as shown in Figure 7.

In the modern version of square-wave polarography the linear d.c. potential ramp is replaced by

a d.c. staircase potential and a SMDE replaces the DME. The drop life is synchronized with each step in the staircase potential. The constant potential of each voltage step is modulated with a small amplitude (ΔE_p) alternating square-wave voltage of frequency f towards the end of the step as shown in Figure 8.

The current is measured over a number of cycles in the measurement period t_m, twice in each cycle, at point 1 at the positive end of the pulse and at point 2 at the negative end. The average of the differences between the current values $i_1 - i_2$ for each measurement cycle is plotted against the potential of the voltage ramp and gives a peak-shaped polarogram, as before.

For reversible electrode processes, the peak current is dependent on the superimposed rectangular voltage:

$$i_p = kn^2 D^{1/2} \Delta E_p c_a \qquad (18)$$

where k is constant at a given frequency.

The square-wave polarograms shown in Figure 9 highlight the influence of n and ΔE_p on the height and width of the peaks.

For reversible processes, sensitivities of about 10^{-8} mol/L can be achieved with SWP and the peak-to-peak resolution is $40 - 50$ mV. The sensitivity decreases rapidly with increasing irreversibility of the electrode process.

Figure 9. Dependence of the peak height on the square-wave voltage amplitude ΔE_p and on the number n of electrons transferred, n [4]
a)–c) $n = 1$ (thallium); d)–f) $n = 3$ (bismuth)
a), d) $\Delta E_p = 40$ mV; b), e) $\Delta E_p = 20$ mV; c), f) $\Delta E_p = 10$ mV

Figure 8. Digitally controlled SWP using the SMDE
A) Variation of the mercury drop surface area with the measurement period t_m; B) Potential – time waveform with the measuring points 1 and 2; C) Current potential response (polarogram). Typical values of parameters: $\Delta E_p \sim 25$ mV, $f \sim 125$ Hz (in 40 ms), $\Delta E_{step} \sim 10$ mV, $t_{step} \sim 10$ s
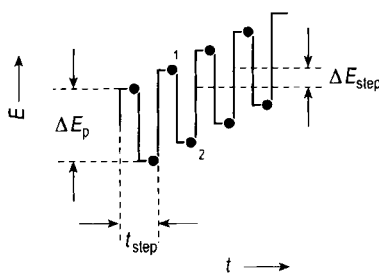
Figure 10. Measurement technique of square-wave voltammetry
ΔE_{step} = Voltage step of the staircase voltage ramp. (ca. 10 mV) ΔE_p = Amplitude of superimposed pulse, ca. 50 mV; t_{step} = wave form period, ca. 5 – 10 ms; 1 and 2 Measurement points

Normally the voltage scan rate of SWP measurements is about 20 mV/s, but much higher scan rates up to 1000 mV/s can readily be programmed making the method ideal for fast voltammetric measurements.

A high scan rate is the characteristic feature of square-wave voltammetry (SWV), first described by OSTERYOUNG and OSTERYOUNG [32]. In this technique the whole measurement process can be carried out with an extremely rapid change of voltage on a single mercury drop. SWV thus enables very short analysis times and is particularly important as a detection method in flow systems.

A relatively large square wave signal of amplitude $\Delta E_p = 50$ mV is superimposed on a stepped voltage ramp (staircase) with voltage steps of about 10 mV. The duration of the square wave

cycle is equal to that of the staircase voltage steps t_{step} and is usually within the range 5 – 10 ms.

As can be seen from Figure 10, two values of current are determined for each square wave cycle. Measurements are made at point 1 at the positive end of the cycle and point 2 at the negative end. The difference between the two current values $i_1 - i_2$ is plotted against the potential of the voltage ramp and gives a peak-shaped current –voltage curve, as in SWP.

The pulse times, which are in the millisecond range, enable the potential to be scanned at extremely high rates of up to 1200 mV/s. Only one mercury drop is required for each measurement process. However, this rapid procedure is achieved
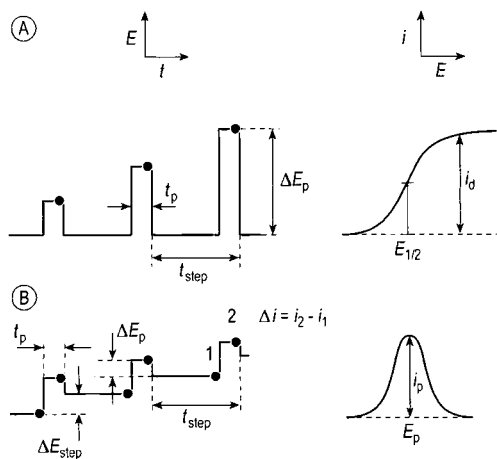
Figure 11. Potential – time waveforms and current – potential responses of A) normal pulse polarography and B) differential pulse polarography (typical values of parameter see text)

at the cost of sensitivity, as the ratio of Faradaic to capacitive currents is lowered by the short pulse times.

The state of development and capability of square-wave polarography/voltammetry has been reviewed [8], [33].

Normal Pulse Polarography. In normal pulse polarography (NPP), the voltage of the working electrode is changed, not by means of a direct voltage ramp (as in SWP), but by increasing rectangular pulses superimposed on an initial, constant voltage. The application of the pulses is synchronized with the mercury drop and occurs just before the end of the drop life. Only one voltage pulse with a duration of 30 – 60 ms is applied to each drop. The amplitude ΔE_p increases from one pulse to the next and reaches a maximum of 1000 mV. To eliminate i_C the current measurement is made ca. 10 – 15 ms before the end of the pulse (Fig. 11).

The measured current determined in pulse polarography is plotted against the pulse amplitude and, as in DCP, yields a curve in the form of a wave. The height of the wave is proportional to the analyte concentration, and the half-wave potential corresponds to the DCP value. Peak-shaped curves can be obtained when the current differences between successive pulses are plotted against the

voltage. Sensitivities are ca. 10^{-7} mol/L; the peak resolution is reported to be ca. 100 mV.

Differential Pulse Polarography. The most important pulse method is differential pulse polarography (DPP). In this technique a stepped voltage rise (modern staircase technique) is used and a rectangular voltage pulse with constant amplitude ΔE_p of 10 – 100 mV is applied to each mercury drop at the end of its drop time; the pulse duration t_p is 40 – 60 ms.

As shown in Fig. 11 B two current measurements are made in DPP: the first i_1 at measurement point 1, immediately before the application of voltage pulse and the second i_2 at point 2 near the pulse end. Both measurements are made on the same mercury drop and also over the same surface area, when the SMDE is used as the working electrode. The difference $\Delta i_1 = i_2 - i_1$ is plotted against the applied d.c. voltage (staircase) and gives a peak-shaped polarogram, because $\Delta i_1/\Delta E$ reaches a maximum in the region of $E_{1/2}$.

Determinations by DPP are on average ten times more sensitive than determinations with NPP. The sensitivity of DPP is about 10^{-8} mol/L and for irreversible processes it falls off only slightly to 5×10^{-8} mol/L. The loss of sensitivity by irreversibility is thus smaller than in the case of other pulse methods. For reversible processes the peak height in DP (differential pulse) polarograms is proportional to the concentration to be analyzed, c_a:

$$i_p = \frac{n^2 F^2}{4RT} A\, c_a\, \Delta E_p \sqrt{\frac{D}{\pi t_p}} \qquad (19)$$

and depends not only on the quantities already mentioned, but on the amplitude of the voltage pulse ΔE_p and the pulse duration t_p. The DP polarograms shown in Figure 12 highlight the increase of peak heights with increasing ΔE_p.

The theory, measurement techniques, and analytical applications of pulse methods have been reported [9].

25.2.3. Alternating Current Polarography

In this technique, a linear voltage ramp is modulated with a sinusoidal alternating voltage of small amplitude ($\Delta E_\sim = 10 - 100$ mV) and low frequency ($f = 5 - 100$ Hz). The superimposed alternating voltage causes an alternating current $i_\sim$, whose size depends on the instantaneous value of
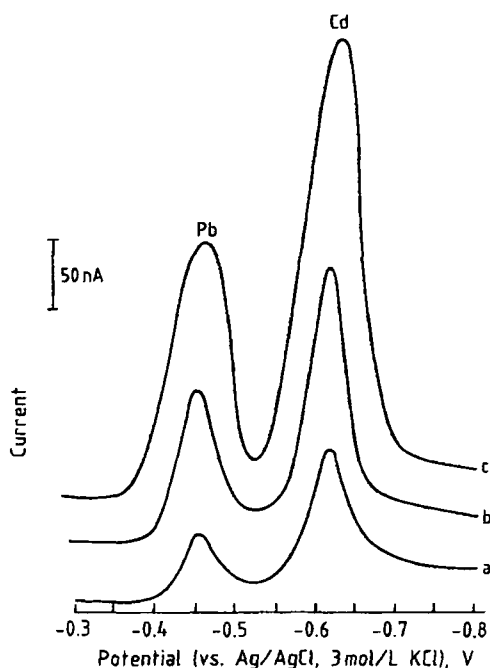
Figure 12. Influence of the pulse amplitude on DP polarograms: Pb and Cd (10 mg/L) in acetate buffer solution pH 2.8
a) $\Delta E_p = 10$ mV; b) $\Delta E_p = 25$ mV; c) $\Delta E_p = 75$ mV

the linear voltage ramp, and is greatest at the half-wave potential. On plotting the selectively measured alternating current against the applied voltage, the peak-shaped AC polarogram is obtained (Fig. 13).

In reversible processes, the peak potential of an AC polarogram is identical with the DC polarographic half-wave potential. The peak current is

$$i_p = \frac{n^2 \cdot F^2}{4RT} (2\pi \cdot D)^{1/2} \cdot A \cdot \Delta E_\sim \cdot f^{1/2} \cdot c_a \qquad (20)$$

and is thus dependent on the usual variables as well as being influenced by the frequency f, the amplitude $\Delta E_\sim$ of the superimposed alternating voltage and the area of the electrode A. The value of i_p also depends on the reversibility of the cell reaction and decreases with increasing irreversibility.

Owing to the high capacitive current component in AC polarography, caused by the periodic charging and discharging of the double layer, the sensitivity is limited to about 10^{-5} mol/L.

Compared with the alternating voltage applied to the electrode, the capacitive current shows a phase shift of $90°$ ($\pi/2$). Since the Faradaic alter-

nating current $i_{F_\sim}$ has a phase shift of only about $45°$, the ratio $i_{F_\sim} : i_{C_\sim}$ and therefore the sensitivity of AC polarographic determinations can be improved by phase-selective rectification. Instruments designed for AC polarography generally have provision for phase selective current measurements to be made (*AC1 polarography*).

By using phase selective measurement of a higher-harmonic alternating current (e.g., 2nd harmonic with a frequency $2f$) increased sensitivity (10^{-7} mol/L) is obtained because of the marked reduction in the capacitive current component of the higher harmonic (*AC2 or second harmonic wave polarography*).

A phase-selective second harmonic wave polarogram is shown in Figure 14. In contrast to the AC1 polarograms, it has two current peaks. For an invariant residual current (Fig. 14 A), the distance between the two current peaks $i_p^+ + i_p^- = i_p$ can be used for analytical purposes. For a rising residual current (Fig. 14 B), the current components i_p^+ and i_p^- are measured over the extrapolated course of the residual current.

Tensammetry is a variant of AC polarography. In this method, the capacitive current component, whose value is determined by the double layer capacitance, is measured instead of the Faradaic alternating current. Capacitance changes are caused by the adsorption and desorption of surface active compounds at the electrode surface. Commonly, two almost symmetrical current peaks that are particularly well defined and known as tensammetric peaks are obtained. In the potential region between the peaks the compound is adsorbed. As a result, the capacitance of the electrochemical double layer is lowered and the capacitive current is smaller than the current of the supporting electrolyte solution (Fig. 15).

Tensammetric studies are important for the characterization of interfacial problems and for the analysis of surface-active compounds (e.g., surfactants) that are neither reducible nor oxidizable.

25.2.4. Linear-Sweep and Cyclic Voltammetry

Both linear-sweep voltammetry (LSV) and cyclic voltammetry (CV) are based on recording the current during a linear change of voltage at a stationary working electrode. The rate of change of voltage $v = dE/dt$ is relatively high, in the range
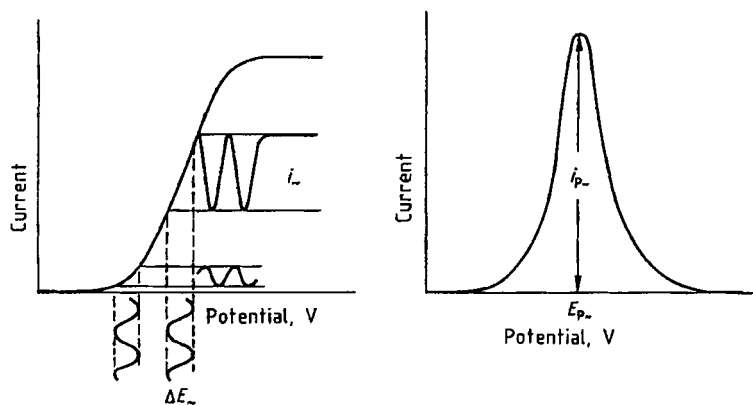
Figure 13. Waveform and response of AC polarography
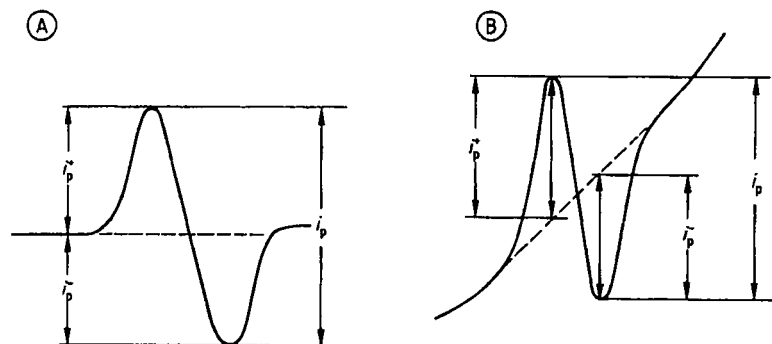
Figure 14. Current profile and evaluation of second harmonic wave polarograms
A) Horizontal residual current; B) Rising residual current

1 – 100 mV/s. The difference is that in the case of CV the voltage is returned to the starting potential; CV is therefore also known as triangle-sweep voltammetry.

Figure 16 compares the variation of voltage with time and the response for the two methods. The peak currents, for both LSV and CV, are proportional to the analyte concentration:

$$i_p = k\, c_a v^{1/2} \tag{21}$$

and are influenced by the rate of change of voltage v. The sensitivity of LSV is about 10^{-7} mol/L and the resolution is about 50 mV.

Cyclic voltammetry is mainly used for studying the reversibility of electrode processes and for kinetic observations, and only sometimes for analytical purposes. The voltage cycle illustrated in Figure 16 ensures that the reaction products

formed at the potential $E_{P_{red}}$ on the cathodic path are reoxidized at $E_{P_{ox}}$ in the anodic sweep.

For a reversible redox process:

$$\Delta E_p = E_{P_{red}} - E_{P_{ox}} = \frac{-57}{n}\,\text{mV} \tag{22}$$

the position of the current peak in this case being independent of the voltage scan rate. The two peaks have equal heights. With increasing irreversibility, ΔE_p becomes greater.

For quasi-reversible processes and for a slow change of voltage, the difference is about $(60/n)$ mV, but it becomes greater for a faster sweep. For totally irreversible processes, the reduction product is not reoxidized, so the anodic current peak is not seen.

If the reversible charge transfer is followed by a chemical reaction during which an electrochemically active product is formed, a cyclic voltam-
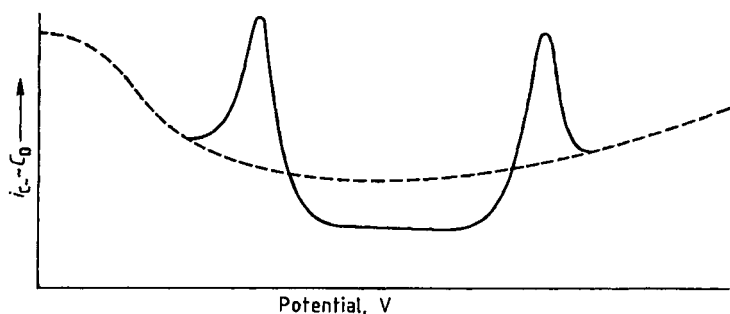
Figure 15. Tensammogram
Alternating current polarogram of the supporting electrolyte solution (---); Formation of current peaks after addition of a surfactant (——); $i_{c\sim}$ = Capacitive current; C_D = Double layer capacitance
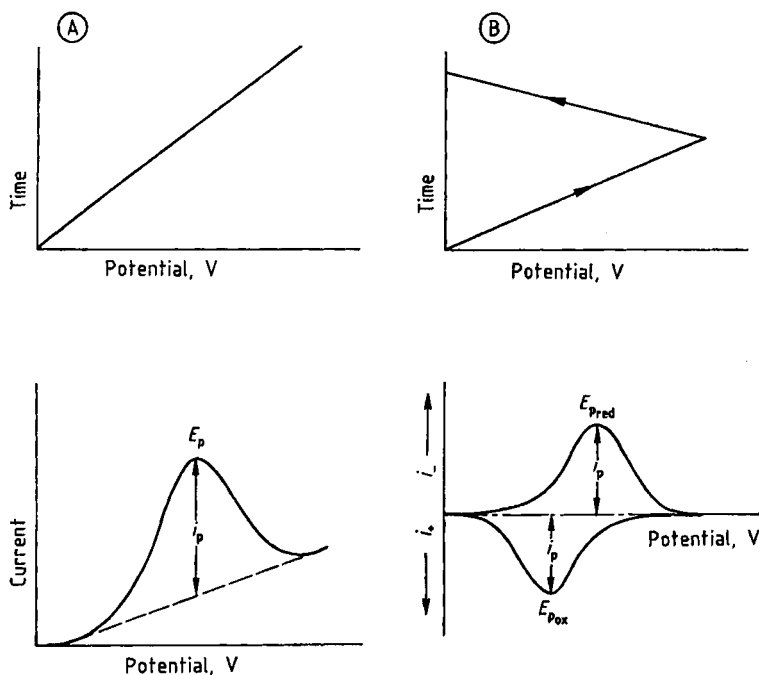
Figure 16. Principle of linear sweep voltammetry (A) and cyclic voltammetry (B)
Top: Variation of voltage with time; Bottom: Resulting current–potential curves

mogram of the type shown in Figure 17 is obtained.

In the first sweep the two peaks $E_{P_{red}}^1$ and $E_{P_{ox}}^1$ appear. The ratio of the peak heights is not, however, 1:1, as would be expected for a reversible process. Rather, $E_{P_{ox}}^1$ is smaller than $E_{P_{red}}^1$, since part of the electrolysis product is chemically converted and therefore no longer available for reoxidation. Since the formed product is electrochemically active, a second peak $E_{P_{ox}}^2$, which corresponds to the

oxidation of this product, appears in the complete anodic sweep. In the second sweep an additional cathodic peak $E_{P_{red}}^2$, which indicates the reduction of the previously oxidized compound, can also be recorded.

Cyclic voltammetry gives information on the redox behavior of electrochemically active species and on the kinetics of electrode reactions as well as offering the possibility of identifying reactive intermediates or subsequent products [10].
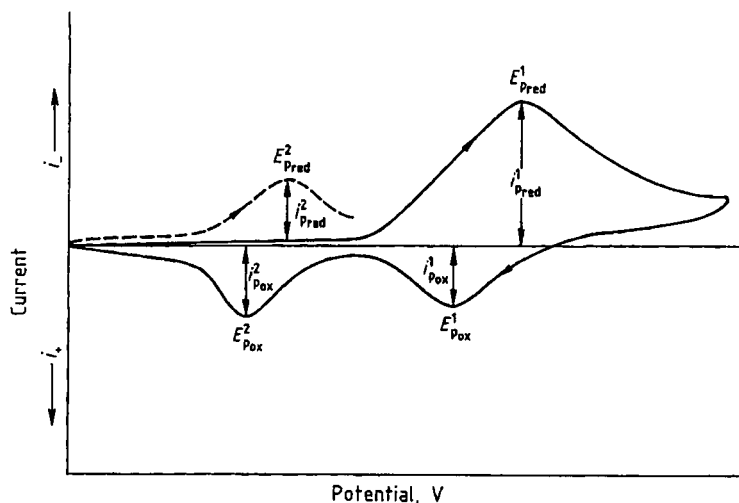
Figure 17. Cyclic voltammogram of a reaction with reversible charge transfer and subsequent chemical reaction
First sweep (———); Second sweep (– – –)
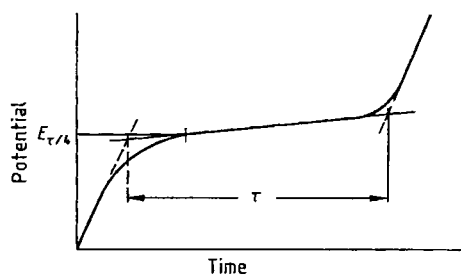
Figure 18. Chronopotentiometric potential – time curve with the transition time τ

25.2.5. Chronopotentiometry

In voltammetric studies, the current flowing through the working electrode is measured as a function of the applied voltage and the rate at which the voltage is scanned. In chronopotentiometry, the change of potential of the working electrode with time at constant current is followed. Chronopotentiometry has also been called "galvanostatic voltammetry" and can, therefore, be grouped with the methods described in this chapter.

Determinations are usually carried out under stationary conditions. The electrolytic processes occurring when current flows decide the ratio $c_{a_{ox}} : c_{a_{red}}$ of the analyte on the electrode surface, which leads to changes of potential according to the Nernst Equation (9). The potential – time plot

is illustrated in Figure 18, and depends on the diffusion behavior and analyte concentration.

The time between the steep sections of the curve is called the transition time τ, which, in diffusion-controlled reversible electrode processes, is described by the Sand equation:

$$\tau^{1/2} = \frac{\pi^{1/2} n F D^{1/2}}{2 j_0} c_a \tag{23}$$

where j_0 is the current density (A/cm^2), and $\tau^{1/2}$ corresponds to the limiting diffusion current in DC polarography. At constant j_0, the potential – time function is:

$$E = E_{\tau/4} + \frac{RT}{nF} \ln \frac{\tau^{1/2} - t^{1/2}}{t^{1/2}} \tag{24}$$

with $E_\tau/4 = E_{1/2}$ for a reversible reaction.

Chronopotentiograms were traditionally evaluated graphically, which with small transition times is difficult and often inaccurate. It is more convenient to measure the time values — which are proportional to concentration — electronically via the first or second derivative. In this way empirical time values are obtained which always differ to some degree from the theoretical transition times [34].

Normal chronopotentiometry has little importance for analysis. Determinations are possible in the range $10^{-4} - 10^{-5}$ mol/L at best. However, the technique is of use in understanding the methods of stripping chronopotentiometry and potentiomet-
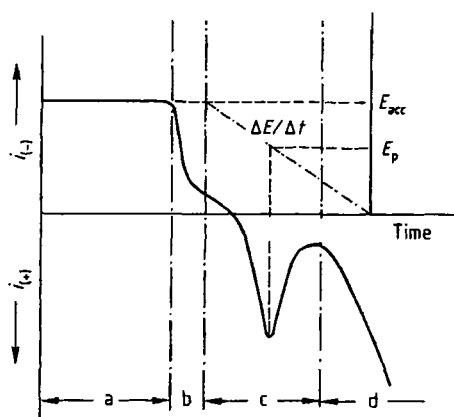
Figure 19. Stripping voltammetric determination
a) Accumulation time; b) Rest period; c) Determination or stripping step; d) Anodic dissolution of mercury

ric stripping analysis, which are important in trace analysis.

25.2.6. Stripping Techniques

The stripping techniques are so-called multi-stage combined procedures, with the characteristic that the voltammetric or chronopotentiometric determination is preceded by an electrochemical accumulation of analyte. The accumulation occurs on the surface of a stationary working electrode and leads to considerably improved performance. The methods enable analytical studies in the pico-trace range [11]–[13].

Stripping Voltammetry. The analyte is accumulated, as the metal or in the form of a known compound, by controlled potential electrolysis or adsorption and with constant convection of the sample solution. The subsequent stripping of the preconcentrated species from the electrode occurs voltammetrically as a linear voltage ramp is applied to the electrode, generally using one of the DC, AC, or DP measurement modes. The current peaks in the voltammograms result from either the reduction or oxidation of the accumulated products. Depending on this electrode reaction, a distinction is made between *anodic stripping voltammetry* (ASV) and *cathodic stripping voltammetry* (CSV).

According to the Randles–Sevcik equation:

$$i_p = k\, n^{3/2} A\, D^{1/2} v^{1/2} c \qquad (25)$$

the peak height i_p in the stripping voltammogram is (analogous to i_p in the linear sweep and cyclic voltammogram) proportional to the concentration c and is also influenced by the working conditions, particularly by the surface area A of the working electrode and the rate of change of voltage v during the stripping process. The proportionality to concentration relates to the amount of accumulated material or to that part of the electrolysis product which takes part in the stripping process. Consequently, i_p depends on two further influences: the accumulation time and the degree of stripping of the electrolysis product. By careful selection of all parameters, limits of determination can be well below 10^{-8} mol/L.

Various means of trace accumulation and determination are used in stripping voltammetry. In the simplest case the analyte is reduced to the metal and accumulated as an amalgam at a stationary mercury electrode (HMDE or TMFE). The determination step proceeds in the reverse direction to the accumulation and is based on the anodic stripping of the metal (reoxidation). This process of ASV is also known as *inverse voltammetry* and can be illustrated as follows:

$$M^{n+} + n e^- + (Hg) \xrightleftharpoons[\text{stripping}]{\text{accumulation}} M^0\,Hg$$
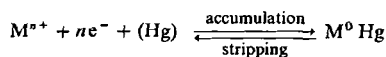
The current profile and the sequence of steps for inverse voltammetric determination of the analyte M^{n+} are shown in Figure 19.

The analyte is deposited from the stirred sample solution at a voltage E_{acc}, as metal on the mercury electrode. The accumulation time is followed by a rest period, during which the solution is unstirred and the cathodic current falls to a small residual value. During the determination step, the voltage is ramped to more positive values (linear sweep, potential scan rate $v = \Delta E/\Delta t$) and, at the appropriate potential, the accumulated metal is reoxidized. The resulting current–potential curve displays a peak, with peak potential E_p. The height of the peak is proportional to the analyte concentration (Eq. 25).

The results of stripping voltammetric determinations are reproducible if the operating conditions are maintained exactly. This includes reproducible renewal of the electrode surface, uniform stirring of the solution, reproducible electrolysis and rest times, and exact adjustment of the electrolysis voltage and its rate of change.

The electrolysis time, adjusted to the analyte concentration to be determined, is usually 1–5 min.

Useful guidance on the selection of a suitable electrolysis voltage E_{acc} is given by the DC or DP voltammogram of the analyte in the particular supporting electrolyte solution. For an electrode reaction with $n = 1$, the voltage should be about 0.15 V more negative than the half-wave or peak potential of the analyte. Consequently, E_{acc} should lie in the voltage range of the limiting diffusion current.

The potentiostatic principle of accumulation electrolysis enables analytes present simultaneously to be determined either individually or together by adjustment of the accumulation potential. Simultaneous determinations require a difference of peak potentials $\Delta E_p > 100$ mV.

Those metals which can be determined are soluble in mercury and form amalgams (Pb, Cu, Cd, Sb, Sn, Zn, Bi, In, Mn, and Tl). Nobler metals such as Hg, Ag, Au, and Pt are also determinable by ASV using inert solid electrodes (metal or carbon electrodes; Chap. 25.3).

Metal ions such as arsenic(III), selenium and tellurium (Me_a^{n+}) may be determined by stripping voltammetry after adding a second metal, such as copper (Me_b^{m+}), to the test solution and co-electrolysing the two metals onto the surface of the HMDE. The copper acts as a co-deposition agent and facilitates the deposition of the analyte, Me_a, on the electrode surface as an inter-metallic compound. The analyte may then be stripped from the electrode either by oxidation (ASV) or by further reduction (CSV) to an anionic species according to the following reaction scheme.

$$M_a^n + M_b^{m'} + (m+n)e^- + (Hg) \longrightarrow M_a^0 M_b^0 (Hg) \qquad \text{Accumulation}$$

$$M_a^0 M_b^0 (Hg) \begin{cases} \xrightarrow{\text{CSV}} M_a^{p-} + M_b^0(Hg) - pe^- \\ \xrightarrow[\text{ASV}]{} M_a^{n+} + M_b^{m+} + (Hg_2^{2+}) + (m+n+2)e^- \end{cases} \quad \text{Stripping}$$
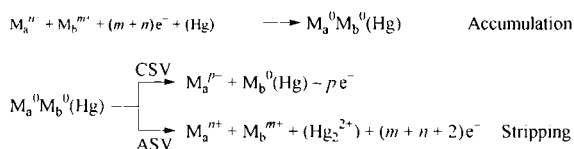
Arsenic, selenium and tellurium are three such elements which may be determined by cathodic stripping voltammetry after having been reductively co-deposited with copper [38]. A characteristic of the cathodic stripping voltammograms of these three elements is that only a single current peak, which arises from the further reduction of the deposited analyte to As^{3-}, Se^{2-} or Te^{2-}, respectively, is observed. In this case the determination

by CSV is more selective than by ASV, since in the anodic dissolution, additional current signals are obtained, which arise from the oxidation of the copper and possibly also of mercury [35]. Typical stripping voltammograms for the determination of arsenic by both ASV and CSV after deposition from a copper containing solution are shown in Figure 20. The determination limits for the CSV determination of As, Se or Te in the presence of copper were found to be 0.5 µg/L for As and 0.2 µg/L for Se and Te. Note that only As(III) is determined since under these experimental conditions As(V) is not electroactive. The small amount of arsenic present in water samples is mainly in the +5 oxidation state as a result of oxidation by oxygen. Recently it has been found that if the test solution contains D-mannitol, As(V) can be reduced electrolytically to As(0). This is the basis of a method for determining As(V) in presence of As(III) by cathodic stripping voltammetry [36].

The technique of CSV can be applied to the determination of anions forming as accumulation products sparingly soluble Hg(I) salts on the electrode surface. These include the halides, pseudo-halides, and oxometallates (vanadate, chromate, tungstate, molybdate).

During the subsequent cathode stripping step, the Hg_2^{2+} ion in the sparingly soluble compound is reduced with formation of a voltammetric current peak (Fig. 21). The peak height is proportional to the accumulated amount of the anion to be determined.

Organic substances can also be determined by this indirect procedure, provided they form insoluble Hg_2^{2+} or Hg^{2+} compounds. These are principally thiols, thioureas, thiobarbiturates, dithiocarbamates, and thioamides.

In adsorptive stripping voltammetry (AdSV) metal chelates and organic molecules are accumulated by adsorption at the surface of the working electrode. If these compounds are electrochemically active, i.e., if they are reducible or oxidizable, their subsequent voltammetric determination is possible. By this principle of so-called adsorptive stripping voltammetry, organic and organometallic compounds are determined in the ultra-trace range. This technique is particularly important for the trace analysis of metals that are not readily deposited as the element on mercury electrodes [13].

Two approaches have been used to effect the adsorption of metal ions onto the electrode surface
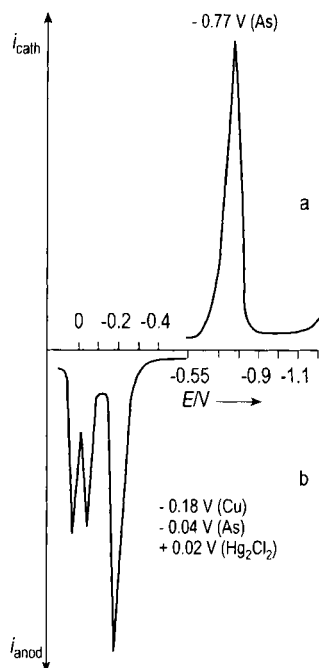
Figure 20. The differential pulse cathodic (a) and anodic (b) stripping voltammograms of arsenic. The cell solution was 0.1 mol/L HCl containing 2×10^3 mol/L Cu^{2+} and 10 µg/L As. Voltage scan rate = 100 mV/s, $E_{acc} = -0.55$ V; $t_{acc} = 1$ min

Figure 21. Determination of anions by CSV

as the metal chelate. The simplest is to add an excess of a suitable complexing agent to the test solution prior to the accumulation step. This is the most common approach and is used preferably with mercury or glassy carbon working electrodes. A selection of frequently used complexing agents is listed in Table 1.

An alternative approach is to modify the surface of a glassy carbon, carbon paste or graphite electrode with the complexing agent. The metal ion is then accumulated by reaction with this modified surface (see p. 807).

The potentiostatically accumulated species is determined voltammetrically by reduction or oxidation of the central atom or of the ligand of the metal chelate complex or via catalytic hydrogen evolution. Adsorbed organic molecules can be determined in an analogous way by the oxidation or reduction of their electroactive functional groups.

The quantity of the adsorbed species c_{ad} present after the accumulation time t_{acc} on the surface of the mercury drop area A and with the radius r determines the value of the voltammetric peak current i_p. Moreover, c_{ad} is proportional to

the concentration c_a of the analyte in solution for low surface coverage. For diffusion-controlled processes, which are assumed here in order to simplify the relationships, the following dependence results for the peak in the stripping voltammogram:

$$i_p = k A c_{ad} = k A c_a \left(\frac{D}{r} t_{acc} + 2 \sqrt{\frac{D}{\pi}} t_{acc}^{1/2} \right) \qquad (26)$$

The peak current increases linearly with $t_{acc}^{1/2}$ until the electrode surface is saturated, and reaches its maximum value at:

$$i_{p(max)} = k A c_{ad(max)} \qquad (27)$$

$c_{ad(max)}$ being proportional to the accumulation time $t_{acc(max)}^{1/2}$, and corresponds to the maximum surface coverage.

The linear proportionality

$$i_p \sim c_a t_{acc}^{1/2} \qquad (28)$$

applies in AdSV only for the lower and middle µg/L range and for short accumulation times $< t_{acc(max)}^{1/2}$. In routine AdSV practice, departures from linearity at fairly high analyte concentrations are controlled, either by dispensing with the usual stirring of the sample solution during accumulation or by dilution of the sample solution.

The voltage dependence of the adsorption of neutral molecules is obviously connected with the interaction studied by A. N. FRUMKIN between the adsorption energy of a surface-active molecule and the value of the electrode potential determined by the surface charge. The zero charge potential in

Table 1. Complexing agents used for the determination of metal ions by adsorptive stripping voltammetry

Complexing agent	Element
Reduction of the central metal ion	
1,2-Dihydroxybenzene (catechol)	U, Cu, Fe, V, Ge, Sb, Sn, As
1,3-Butanedionedioxime (dimethylglyoxime)	Co, Ni, Pd
8-Hydroxyquinoline (oxine)	Mo, Cu, Cd, Pb, U
2-Hydroxy-2,4,6,-cycloheptatriene (tropolone)	Mo, Sn
2,5-Dichloro-3,6-dihydroxy-1,4-benzoquinone (chloranilic acid)	U, Mo, Sn, V, Sb
N-Nitroso-N-phenylhydroxylamine (cupferron)	U, Mo, Tl
Reduction of the ligand	
o-Cresolphthalexone (OCP)	Ce, La, Pr
5-Sulfo-2-hydrobenzeneazo-2-naphthol (Solochrome Violet RS - SVRS)	Al, Fe, Ga, Ti, Y, Zr, V, Tl, Mg, alkali and alkaline earth metals
1,2-Dihydroxyanthraquinone-3-sulfonic acid (Alizarin Red S)	Al
Chromazurol B (MB9, Mordant Blue 9)	Th, U
Catalytic hydrogen evolution	
Formazone	Pt

the electrocapillary curve of the mercury is critical for the adsorption of neutral molecules. Negatively charged surface-active ions are chiefly adsorbed at potentials corresponding to the positive branch of this curve, and the cations at potentials that lie in the negative range (see textbooks on electrochemistry).

Stripping Chronopotentiometry (SCP). With stripping chronopotentiometry, as with stripping voltammetry, the preliminary electrolytic accumulation of the analyte significantly increases the sensitivity of the technique. The difference between the two stripping techniques is that, in stripping chronopotentiometry the change of potential of the working electrode with time during the stripping of the electrolysis product (e.g., of a deposited metal) is recorded. Stripping is performed by application of an anodic current. The output is an "inverse" potential–time curve from which the transition time τ is obtained, as this is a function of concentration, analogous to the determination of the peak current i_p in the stripping voltammogram. From a comparison of the two curves in Figure 22 it can be seen that $E_{\tau/4}$ in the stripping chronopotentiogram corresponds to the peak potential E_p in the stripping voltammogram.

For the accumulation step, the HMDE and the TMFE are mainly used. The transition times for determinations by stripping chronopotentiometry at these electrodes are approximately:

$$\tau_{HMDE} = \frac{n \cdot F \cdot c_{amalg} \cdot r_d}{j_0} \tag{29}$$

and

$$\tau_{TMFE} = \frac{n F c_{amalg} \cdot l}{j_0} \tag{30}$$

Here, c_{amalg} is the amalgam concentration after accumulation, r_d the radius of the mercury drop, l the thickness of the mercury film on the electrode surface, and j_0 the current density during the stripping process.

In addition to the time for which the electrolysis potential is applied during the accumulation step, the current density during stripping is also of crucial importance for the limits of detection: as j_0 decreases, the determinations become more sensitive. Limits of detection of ca. 10^{-8} mol/L can be achieved [14].

Figure 23 illustrates the form of the curve for the determination of Cd, Tl, and Pb by stripping chronopotentiometry. The normal potential–time curve (A) is compared with the second derivative (B), from which the transition times can be more simply and accurately obtained. A TMFE coated in situ is used for the accumulation.

For the highest possible sensitivities in stripping chronopotentiometry, and to avoid systematic errors, the sample solutions must not contain oxidizing impurities. Care must therefore be taken to remove the oxygen present in the solution.

This precondition does not apply if the anodic stripping of the accumulated elements is replaced by chemical oxidative stripping. With this variant the technique is known as *Potentiometric Stripping Analysis* (PSA). The working electrode
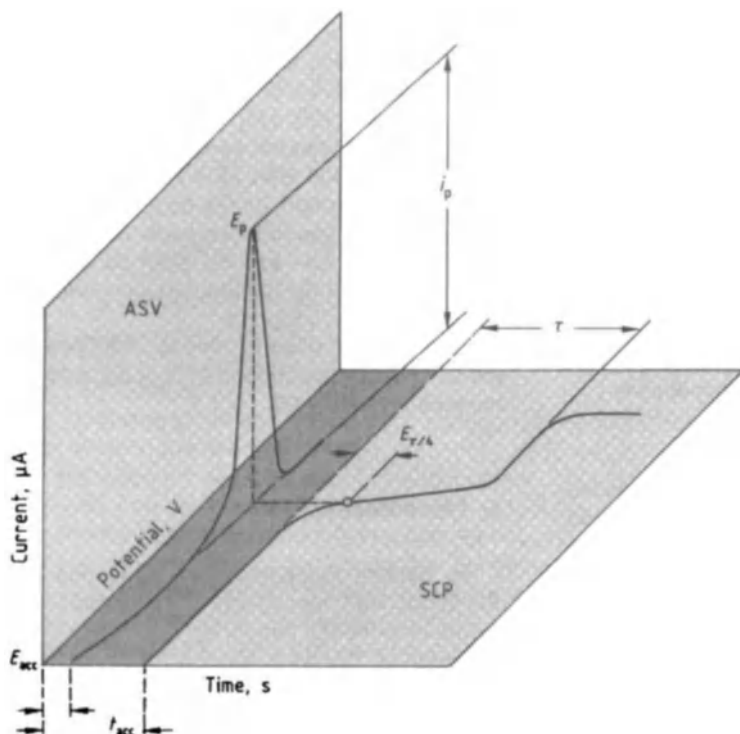
Figure 22. Comparison of anodic stripping voltammetry and stripping chronopotentiometry

chiefly used for this is the TMFE; the oxidants are Hg^{2+} salts or oxygen dissolved in the sample.

Potentiometric stripping analysis is carried out in several stages. After electrochemical generation of the mercury film on a graphite substrate, the elements to be determined are accumulated by electrolysis at constant potential. The next stage is the oxidation of the deposited elements by the oxidant present in the solution. For this, the current circuit is disconnected. The deposited analytes are stripped in the order of their electrochemical potentials. Anodically deposited precipitates can similarly be stripped by chemical reduction. In all cases, potential – time curves with transition times proportional to concentration result [39] – [41].

Recent developments have led to computer-assisted differential PSA $[E = f\,(dt/dE)]$ which has become important for the detection of trace elements [42]. Figure 24 compares the normal potential – time curve and the differential curve calculated from it for the simultaneous determination of the four elements.

The theory of stripping chronopotentiometry and differential potentiometric stripping analysis has been reviewed [43].

25.3. Instrumentation

In modern equipment, the measuring arrangement for polarographic or voltammetric studies consists of a measuring stand and a digital computer. The latter controls the more or less automated techniques and also analyzes the current signals.

The measuring stand includes the potentiostat locked into analog electronics, the current measurement amplifier, and the digital – analog and analog – digital converters. It also provides a holder for the electrochemical cell, which consists of measuring vessel, electrodes, stirrer, etc.

The *potentiostat* is the voltage source for the DC supply (ca. 4 V) as well as for the superimposition of voltage pulses (NPP and DPP modes) and
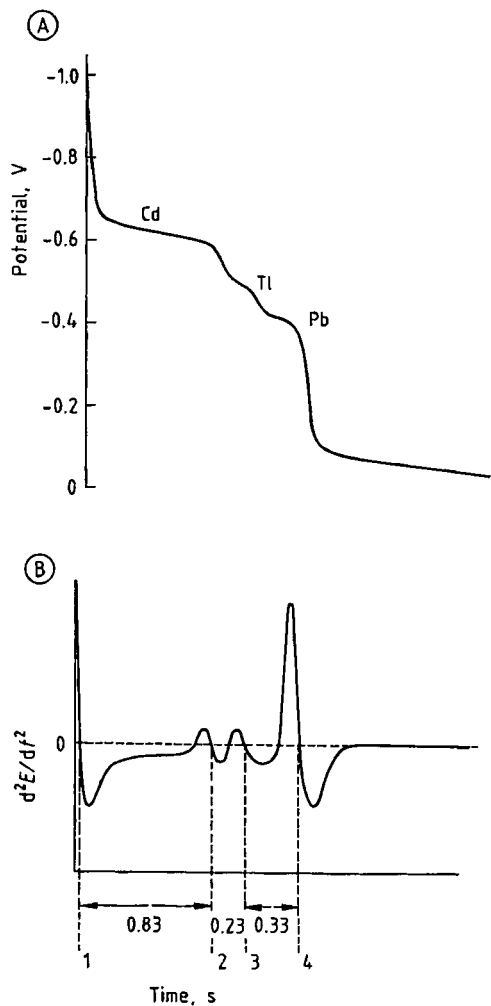
Figure 23. Simultaneous determination of Cd
$(5 \times 10^{-7}$ mol/L), Tl $(2 \times 10^{-7}$ mol/L), and Pb $(3 \times 10^{-7}$ mol/L)
in 0.05 mol/L HCl + 0.5 mol/L NaCl + Hg^{2+} 100 mg/L by
stripping chronopotentiometry at a TMFE (electrolysis time
10 min; stripping current strength 10 μA)
A) Potential – time curve; B) Second derivative [14]

Figure 24. Potential – time curve (A) and differential plot (B)
for the determination of zinc, cadmium, lead, and copper by
PSA [42]

a sine-wave or square wave alternating voltage
(ACP and SWP mode). The basic circuit diagram
is shown in Figure 25.

When recording the voltammograms or pola-
rograms using three electrodes, the current is flow-
ing between the working electrode and auxiliary
electrode. Compared with the working electrode,
the auxiliary electrode has a much greater surface
area and thus a smaller current density. The refer-
ence electrode is an electrode of the second kind,
and the potential of the working electrode is meas-
ured against this reference electrode. Potential

data for voltammetric and polarographic measure-
ments are reported relative to calomel or silver –
silver chloride reference electrodes.

The potentiostatic measuring arrangement with
three electrodes (Fig. 25) has superseded the clas-
sical measuring arrangement with two electrodes,
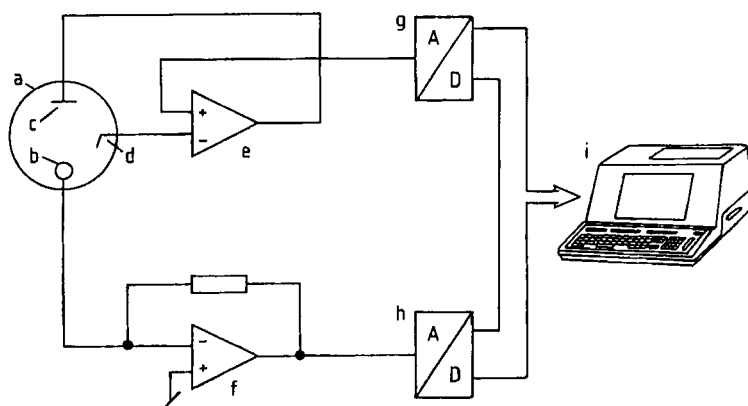
Figure 25. Circuit diagram of a polarograph for various polarographic/voltammetric measuring techniques (Courtesy of Metrohm)
a) Measuring cell; b) Working electrode; c) Auxiliary electrode; d) Reference electrode; e) Potentiostat; f) Current measurement amplifier; g) Digital – analog converter; h) Analog – digital converter; i) Computer

as with the latter the ohmic potential drop in the sample solution cannot be compensated adequately.

In addition to carbon and the noble metals, the principle material used for the working electrode is mercury, especially for capillary and film electrodes. Because of the high overvoltage of hydrogen on mercury, the latter electrodes ensure a relatively wide potential range for voltammetric/polarographic determinations in various supporting electrolyte solutions. If capillary electrodes are used, the electrode surface can be renewed easily and reproducibly as with the DME, the SMDE and the HMDE.

Dropping Mercury Electrode. The DME, originally proposed by HEYROVSKY and used for polarographic studies, consists of a glass capillary, 10–20 cm long, with internal diameter 0.04–0.1 mm, and is connected by a tube to the mercury storage vessel. The distance between the capillary opening and the mercury level determines the drop time and influences to different degrees the types of current useful for polarographic determinations [1]. The drop time also depends on the electrode potential and the surface tension of the mercury. The surface tension is mainly influenced by the medium, and particularly the solvent, into which the mercury flows. The drop time must be held constant during the measurement and should be 2–8 s. In modern instruments the mercury drop is knocked off mechani-

cally by tapping on the capillary at a predetermined constant time.

Static Mercury Drop Electrode. Recent developments have led to the static mercury drop electrode (SMDE), which today is the preferred electrode for polarographic analysis. Compared with the DME, the drops from this type of electrode can be renewed more frequently, about once or twice per second. A microvalve with opening times of 20–200 ms controls the mercury inflow to the capillary. During each open period a mercury drop forms, whose surface area remains constant after the valve closes.

In Figure 26 the increase of the electrode surface area with time during the drop life is compared for the DME and the SMDE. The maximum electrode surface area is reached much more rapidly with the SMDE than with the DME so that the measurement can be carried out at constant electrode surface area. This results in a reduction of the capacitive current and increased analytical sensitivity.

The greatest advantage of this new type of electrode is its ability to function as a stationary and a nonstationary mercury electrode, as well as the possibility of synchronizing the electrode (drop) formation with the programm-control arrangement. The reproducibility of the mercury metering of an SMDE is important for its dependability in operation. In a commercial model (Multi-Mode electrode. Metrohm AG, Switzerland), the mercury is pushed through the capillary by a ni-
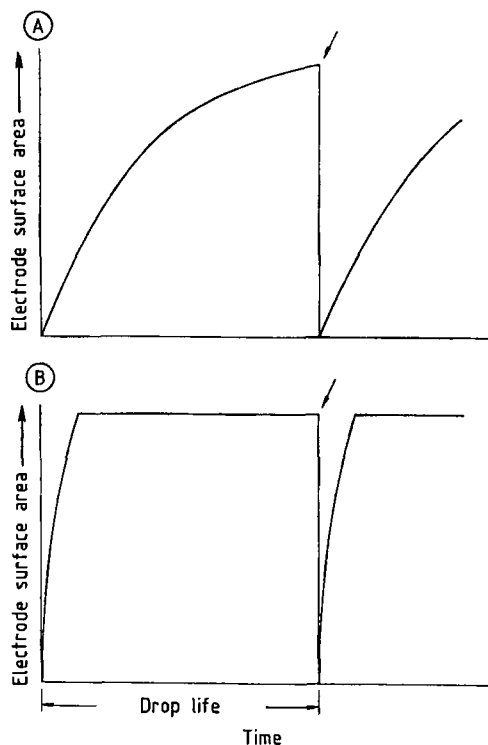
Figure 26. Drop growth as a function of time; comparison of the DME (A) with the SMDE (B)
Drop fall indicated by arrows

Figure 27. Construction of a static mercury electrode (Courtesy of Metrohm)
a) N_2 supply for membrane control; b) N_2 supply for overpressure; c) Blind hole; d) Stopper for topping up; e) Electrode connection; f) Hg reservoir; g) Adjusting screw; h) Steel spring; i) PTFE membrane; j) Steel needle; k) Rubber packing (damper); l) Seal; m) Capillary

trogen overpressure and a needle valve controls the mercury flow.

Figure 27 shows the construction of the pneumatically controlled Multi-Mode electrode. The electrode is suitable for all voltammetric/polarographic work for which a DME, SMDE, or HMDE is required.

Thin mercury film electrodes (TMFE) are used chiefly for ultra-trace analysis by stripping voltammetry. Glassy carbon has proved itself to be the best substrate onto which a mercury film can be formed. The mercury is deposited by electrolysis, either in a separate operation before the voltammetric determination or, in the case of stripping voltammetry, together with the amalgam-forming metal during the accumulation by electrolysis (in situ coated TMFE, Section 25.2.6; the mercury film generated should be $1-100$ µm thick). Voltammetric determinations with the TMFE can be more sensitive than with the HMDE, with improved resolution of neighboring peaks.

Carbon and graphite electrodes are employed principally in the anodic potential range. Glassy carbon electrodes (GCE) and carbon paste electrodes (CPE) are mostly used and are available commercially. They can be used either as stationary or rotating electrodes. *Glassy carbon electrodes* consist of pins or rods $2-8$ mm diameter, cemented or pressed into a glass or plastic holder. *Carbon paste electrodes* consist of mixtures of spectroscopically pure carbon or graphite powder and water-insoluble, highly viscous organic liquids, e.g., nujol, silicone oil, or naphthalene derivatives. The mixture is ground to a paste, transferred to a glass tube, and expressed from the tube with a piston. The surplus carbon paste must be rubbed off and the surface polished. In voltammetric studies the composition of CPE influences the peak potential and the peak current. Glassy carbon electrodes and carbon paste electrodes are also used as working electrodes in voltammetric detectors for liquid chromatography ($\rightarrow$ Liquid Chromatography).

Noble Metal Electrodes. Working electrodes of noble metals are often used for studies in the

anodic potential range, or for determining elements with which mercury electrodes interfere (e.g., voltammetric Hg determination at an Au electrode). The electrochemical properties of noble metal electrodes and the reproducibility of the analytical values obtained with them depends on their physical and chemical pretreatment. In comparison with mercury and carbon electrodes, noble metal electrodes are of secondary importance for voltammetric determinations.

Modified Electrodes. When the surface of a working electrode has been chemically changed or treated in order to change its electrochemical properties or to improve its sensitivity and selectivity in an analytical determination, the electrode is referred to as a *modified electrode*. The modification of a voltammetric working electrode, usually a carbon electrode, is done by fixing (immobilizing) reagents on the electrode surface. The reagent that is used to modify the surface is referred as the *modifier*. One of the reasons for the development of modified electrodes is to use them to replace mercury electrodes (for environmental reasons) and so facilitate the further development of voltammetry, and in particular stripping voltammetry, as a "leading edge" technique. A major role of the modifier is to improve the selectivity of the electrode and aid the binding and enrichment of the analyte on the electrode surface.

There are several methods used to modify electrode surfaces: adsorption of reagents on the surface, covalent bonding by reaction between the modifier and functional groups on the electrode surface, incorporation of the modifier within a gel or an electrically conducting polymer film, physically coating the electrode surface with the modifier, e.g., an enzyme, or mixing the modifier with carbon paste.

Carbon paste electrodes have the advantage that the surface can be quickly renewed by simply removing the used surface layer with a tissue or knife. Organic complexing agents, ion exchangers, solid phase extractants and biological materials may be used as modifiers. The modifiers must have a low solubility in the analyte solution and not be electroactive in the working potential range [44].

Modified screen printed electrodes may be prepared either by incorporating the modifier in the carbon containing "ink" or by fixing it to the screen printed electrode surface. A mercury film, electrolytically deposited on a screen electrode has

been used to determine heavy metals by ASV or AdSV (e.g., uranium [45]).

Thick-film graphite electrodes can be modified in a similar way to screen printed electrodes to give *thick film modified graphite electrodes* (TFMGEs). After modifying the surface in different ways the electrodes can be used for the stripping voltammetric determination of a number of heavy metals [46].

Microelectrodes. In general, the carbon and mercury working electrodes commonly used in voltammetry have surface areas in the mm^2 range. This means that the linear dimensions of these electrodes are an order of magnitude greater than the thickness of the diffusion layer and the transport of the analyte to the electrode takes place mainly by linear diffusion. However, if the lateral dimensions of the electrode are similar to or less than those of the diffusion layer (< 0.1 mm), the diffusion of the analyte to the electrode is then spherical rather than linear. Electrodes with such small dimensions in the lower μm range are referred to as *microelectrodes*. Microelectrodes are becoming more important as working electrodes in voltammetric trace analysis.

The miniaturization of the working electrode results in a higher current density at the electrode surface, but also in small currents flowing in the electrochemical processes. The small area means that the charging current i_c, which is proportional to the area of the electrode (Eq. 16), is also small. Thus as the electrode size decreases, the signal to noise ratio (or sensitivity) increases. In addition, note the very fast response to potential changes (response time) at a microelectrode.

A method of overcoming the problem of measuring the small currents is to join many microelectrodes in parallel and so obtain a much larger current output. An arrangement with many (up to 10 000) single microelectrodes is called a *microelectrode array*. By wiring n single electrodes in any array together in parallel, the output current will be n times that of a single microelectrode, i.e., it can be in the μA range instead of in the pA – nA range. Such an array, while having the advantages of a single microelectrode, produces a current which is large enough to be measured with conventional polarographic equipment.

With microelectrode arrays (mainly made from platinum or carbon by various techniques) it is possible to extend the application of analytical voltammetry to the direct determination of analytes in non-aqueous samples without adding a

supporting electrolyte. In stripping voltammetry, the accumulation of the analyte on the electrode can be achieved in a very short time without stirring the solution because of the higher mass transport [47].

Reference and Auxiliary Electrodes. The function of the reference electrode is to monitor the potential of the working electrode and the function of the auxiliary electrode is to complete the electrical circuit for the current generated at the working electrode. The voltage of the auxiliary electrode (Pt wire or graphite rod) with respect to the working electrode is electronically adjusted via the potentiostat to the desired voltage plus the iR drop in the solution so that the potential difference between the working and reference electrode is maintained at the required value. Usually the area of the auxiliary electrode is at least 50 times greater than the working electrode.

The potential of the reference electrode must be constant and independent of the composition of the solution. Therefore the reference electrodes used in polarographic and voltammetric practice are electrodes of the second kind, of which the silver – silver chloride electrodes (with potassium chloride of a defined concentration, for example Ag/AgCl, 3 M KCl) are the most important. Previously calomel electrode (for example the saturated calomel electrodes, SCE) were used.

Measuring Cells. The commonest cells have a maximum volume of 25 mL. The vessels are generally made of glass; for demanding trace analyses they consist of quartz or plastic (principally polytetrafluoroethylene, PTFE). Micro-cells with a capacity of 1 mL or less are used for special microanalytical purposes and are constructed individually. Flow cells for analyses in connection with liquid chromatography or for flow injection analysis (FIA) are becoming increasingly important.

25.4. Evaluation and Calculation

Voltammograms and polarograms are evaluated either graphically or by calculation. For accurate determination of the current value — which is proportional to concentration — from the current wave or from the current peak, the residual current must be subtracted from the signal. For routine analysis, it is often sufficient to interpolate the residual current in the potential range of the voltammetric/polarographic current signal.

Figure 28 compares various practical curves and the methods for their manual interpretation. Such current – voltage curves, and others, can be evaluated more rapidly and with improved accuracy with computer-controlled instruments. The best values are obtained when a polynomial describing the approximate shape of the baseline is used. The coefficients of the polynomial are determined from the experimental points which lie on both sides of the current peak [4].

This evaluation becomes straightforward with equipment in which the current – potential curves are recorded digitally at a high data rate. The first step in the data processing consists of the automatic search for stray experimental data points. These are then eliminated and replaced by the mean value of both neighboring data points. The line of best fit of the current – potential curve is then smoothed within a sliding window along the voltage axis. The smoothing factor that can be applied essentially depends on the number of data points available. The principle of the curve smoothing is illustrated in Figure 29.

After smoothing, the curve is automatically differentiated (first derivative) and in this form used for peak recognition. This is achieved by seeking out successive maxima and minima in the differentiated curve (Fig. 30). In the usual voltammetric presentation (increasing negative potentials to the right), a maximum followed by a minimum indicates a reduction peak, whereas the reverse order indicates an oxidation peak. The potentials of the maximum and minimum in the di/dE versus E plot are noted and the peak width is the difference between these values while the peak potential is the mid-point between them. The qualitative identification of the analyte is achieved by comparing the peak potential ($\pm$ a pre-set tolerance) with data stored in the computer memory. For a peak to be identified it must be greater than a pre-determined minimum peak height. After the peak is recognized a baseline is constructed and its value at the peak potential is subtracted from the maximum value of current to give the peak height. The result is presented as a peak or wave height, or as the peak area, all of which are proportional to the concentration of the analyte.

When the baseline is curved the polynomial which approximately describes this baseline is used to establish it mathematically over the potential region across the base of the peak. This potential region is determined from the potentials at
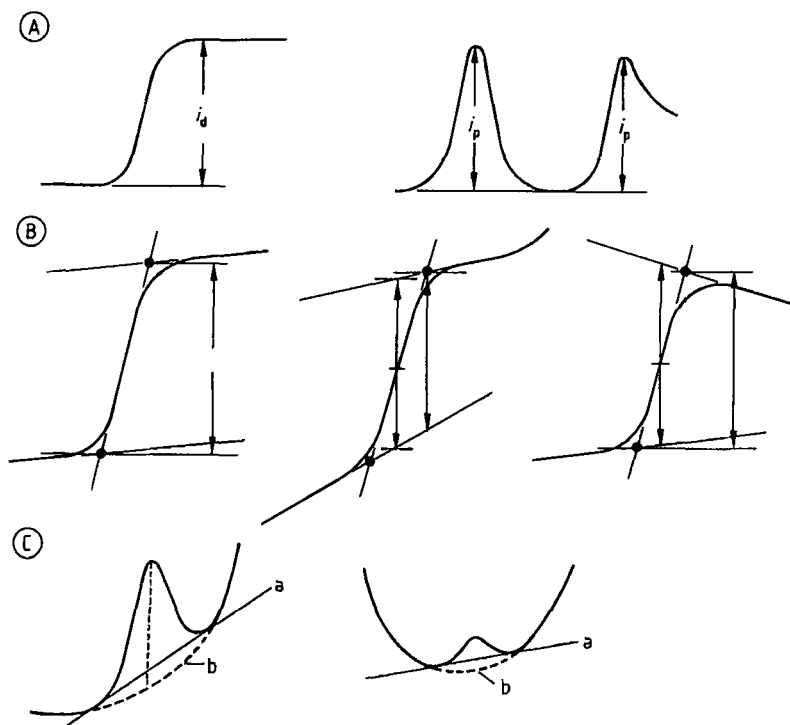
Figure 28. Interpretation of polarograms and voltammograms
A) Ideal conditions: linear trend of residual current; B) Evaluation of wave heights under nonideal conditions; C) Evaluation of peak currents with strongly curved residual current.
a) Tangent method; b) Evaluation with French curve [5]
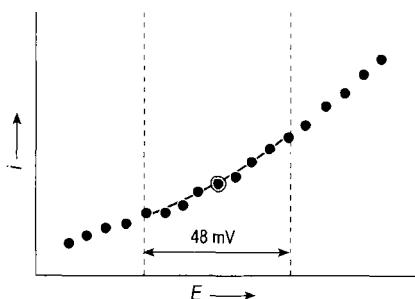
Figure 29. Curve smoothing using a voltage window with 9 data points

which the change in the gradient (becomes more positive) of the current – potential curve occur. The same procedure is used for the evaluation of closely adjacent peaks.

The peak or wave height (h) or peak area (A) calculated by the above procedure can be converted into the desired concentration units either by a *calibration curve* (c versus h or c versus A) or by the method of *standard additions*. When using a calibration curve its slope and linearity must be checked at least daily, also when experimental parameters are changed.

When using the method of standard additions, an aliquot, v, of a standard solution of known analyte concentration c is added (manually with a micropipette or automatically with a motorized burette) to a Volume V of the sample solution in which the analyte concentration is c_a. If h is the peak or wave height of the sample before the standard addition, and H the height after that, then the analyte concentration is given by

$$C_a = hcv / H(V + v) - Vh \tag{31}$$

The highest accuracy can be obtained by using several (at least three) standard additions, provided the total concentration always lies within the linear range of the standard curve, when plotting h versus c, and extrapolating the straight line
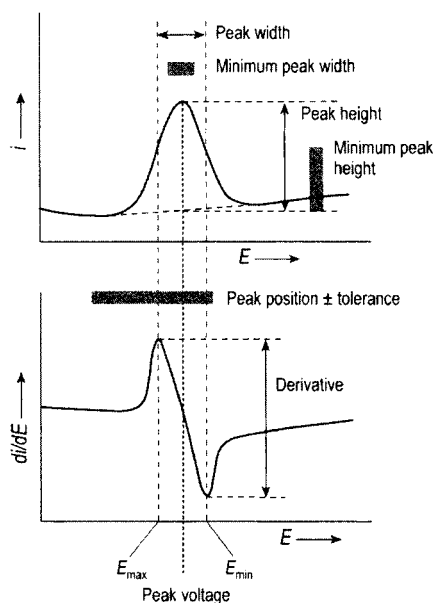
Figure 30. Peak recognition and characterization (Courtesy of Metrohm).
Peak potential $= (E_{max} + E_{min})/2$
Peak width $= E_{min} - E_{max}$

plot back to the c axis. The distance from the origin to the intercept corresponds to c_a.

The reproducibility of voltammetric/polarographic determinations depends not only on the evaluation, but also on the quality of the electrode and the measurement set-up. With modern equipment, reproducibility is $0.5 - 2.0\%$.

A characteristic source of error is the temperature dependence of voltammetric currents. All equations for the proportionality of polarographic currents to concentration contain the diffusion coefficient, which increases with temperature. The rate constants of preceding and succeeding chemical reactions are also affected by temperature and this must be taken into account when evaluating kinetic and catalytic currents. Electrode reactions that are associated with analyte adsorption processes likewise show a specific temperature dependence. The temperature influence differs and leads to different temperature coefficients for the individual voltammetric methods [48].

In simple diffusion-controlled electrode reactions, deviations between $+1.5$ and $+2.2\%/°C$ occur. For kinetic and catalytic currents, the temperature coefficients can be up to $+70\%/°C$. Currents that are based on adsorption processes can

have negative temperature coefficients, because of decreasing adsorption with increasing temperature.

25.5. Sample Preparation

Before performing the voltammetric/polarographic determination, in some cases the analyte must be separated from the matrix and from other interfering components, and converted into a measurable form. For total element determinations, sample digestion is normally required. To separate organic analytes from the matrix, extraction methods and chromatographic separation are mainly used.

Heavy metals in association with organic materials are usually in a bound form, from which they must be liberated prior to analysis. This is carried out by dry or wet ashing, whereby the organic matrix must be completely destroyed. To avoid systematic errors, care must be taken in all forms of sample treatment to ensure that the elements to be determined remain completely in the sample and that sample contamination from digestion reagents and vessels is avoided.

Wet digestions are carried out only with high-purity acids. If there is a risk that an element will volatilize, the treatment is carried out under reflux, or better in pressure digesters with quartz or plastic inserts; heating is carried out electrically or by microwave radiation. Sample digestion with nitric acid at high temperatures is favored owing to the increased oxidation power. Figure 31 shows a high-pressure ashing apparatus for the digestion of organic and inorganic samples with acid (HNO_3, HCl, $HClO_4$, H_2SO_4, HF, or mixed acids) in sealed quartz vessels under high pressure (up to 13 MPa) and high temperature (up to 320 °C). The digester is contained in a heated autoclave and filled with nitrogen (ca. 10 MPa) in order to compensate the internal pressure resulting from the digestion reaction.

For the digestion of biological samples, combustion in oxygen (*dry ashing*) can also be useful under suitable conditions. The use of oxygen of highest purity ensures low contamination digestion. Some forms of equipment operate with reflux condensation to avoid losses of elements by volatilization; cooling with liquid nitrogen is particularly suitable. Organic samples can be digested at relatively low temperature in high-frequency oxygen plasmas.
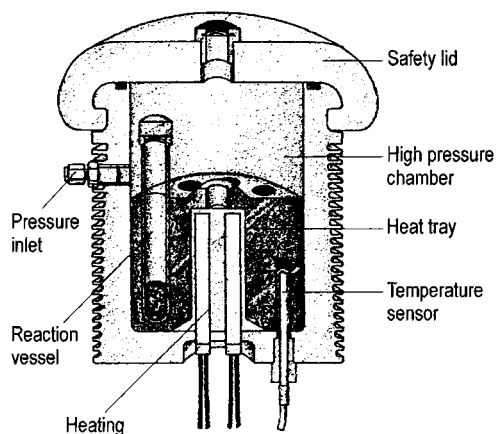
Figure 31. High-pressure ashing apparatus HPA-S (Courtesy of Perkin Elmer)

Figure 32. Ultraviolet digester (Courtesy of Metrohm)
a) Lamp; b) Digester vessels with stopcocks; c) Cooling chamber with water inlet and outlet; d) Fan; e) Cables

Voltammetry and polarography are particularly important for the analysis of water samples for heavy metals. In surface waters with organic contamination, they can only be analyzed after digestion. Sampling and storage is of prime importance for correct analytical results.

Water samples are stored in flasks of polyvinyl chloride (PVC) or PTFE, cleaned with 50 % nitric acid before use. The flask is subsequently prerinsed with the sample water. Contaminated samples are filtered through membrane filters directly into the sample flask and acidified, e.g., with high purity nitric acid to pH 1 – 2; the filter residue is analyzed separately. Water samples with little or no organic contamination are mixed with the chosen supporting electrolyte and immediately placed in the voltammetric/polarographic set-up.

Contaminated water samples are prepared for electrochemical heavy metal trace analysis either by UV photolysis or by oxidative microwave digestion.

Ultraviolet photolysis is based on the generation of OH radicals, which react with organic compounds and degrade them. Hydrogen peroxide is used as an initiator. The acidified water sample plus hydrogen peroxide is irradiated in quartz vessels. An arrangement for the digestion of one to six samples is shown in Figure 32.

When nitrate-containing samples are UV-irradiated, nitrite is formed by reduction. The voltammetrically active nitrite causes a broad current – voltage peak that can cause interference (e.g., with

the determination of zinc). This must be taken into account when examining water samples acidified with HNO_3. Another acid should be used or the nitrite formed should be reduced to nitrogen with sulfamic acid (NH_2SO_3H) immediately after the end of the photolysis.

The complete destruction of organic substances in water samples by UV photolysis depends on the nature and extent of the contamination, and on the intensity of the irradiation. The irradiation time required for waste water samples is usually 1 h (see Fig. 33).

Microwave Digestion. Oxidative microwave digestion is suitable for the rapid degradation of interfering organic substances in surface waters. For this purpose the samples are heated with the oxidizing agent Oxisolv (Merck, Germany) in sealed PTFE vessels in a microwave oven. During the thermal treatment, the pH value of the sample passes through the wide range of pH values (from 11 to 2). Wastewater samples are generally completely digested after 50 – 60 s, so that the voltammetric/polarographic metal trace analysis can be carried out completely free of interference.

If the metals are determined by the highly sensitive method of adsorptive stripping voltammetry, the digestion time must be increased by 1 – 2 min.
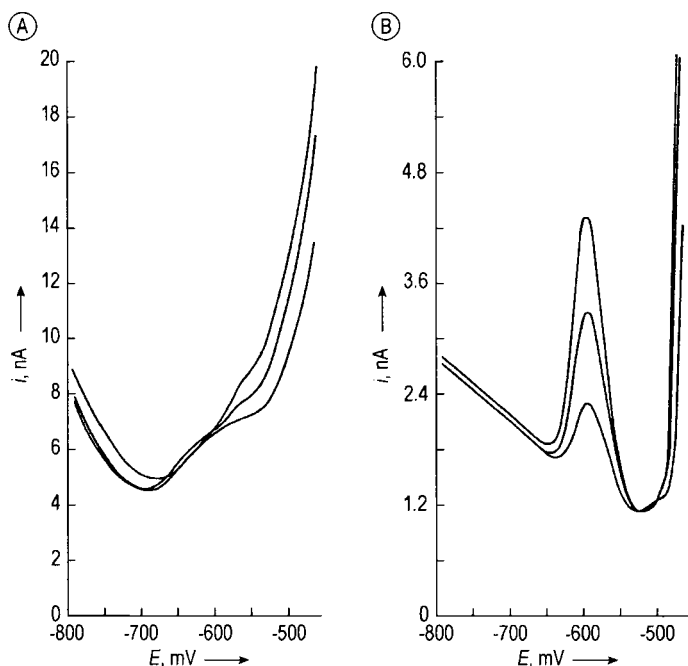
Figure 33. Determination of Cd in municipal wastewater by DPASV before (A) and after (B) 1 h UV irradiation [49]

Figure 34 shows the voltammograms for the determination of molybdenum in a river water sample; the digestion time was 2×50 s at 260 W microwave power [50].

25.6. Supporting Electrolyte Solution

The solution for the voltammetric/polarographic determination of elements and organic compounds always contains a supporting electrolyte in order to minimise the migration current and reduce the ohmic voltage drop. The supporting electrolytes used are mainly chlorides, nitrates, and sulfates of Li, Na, K, perchlorates of Li, Na, and salts of tetraalkylammonium bases (NR^+X^-; R = methyl, ethyl, butyl; X = Cl, ClO_4, Br, I). Moreover, acids (HCl, H_2SO_4), bases (LiOH, NaOH, $NR_4^+OH^-$), and frequently buffer solutions are used as they simultaneously provide for regulation of pH during electrode reactions. Supporting electrolyte concentrations are commonly between 0.1 and 1 mol/L.

If complexation occurs between the analyte and the supporting electrolyte or an added ligand, crucial changes in electrochemical behavior can

occur. As a result of complexation, the analyte can become electrochemically inactive, or its behavior may change in such a way that the value of the polarographic half-wave or peak potential becomes more negative. For the simple complexing reaction of a hydrated cation $M(H_2O)_x$ with ligand L and coordination number i:

$$M(H_2O)_x + i\,L \rightleftharpoons ML_i + xH_2O$$
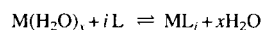
and with the stability constant K of the complex:

$$K = \frac{[ML_i]}{[M][L]^i}$$

the shift of the half-wave or peak potential is given by:

$$\Delta E_{1/2} = \Delta E_p = -\frac{0.059}{n}\log K - \frac{0.059}{n}i\log[L] \quad (32)$$

Stability constants and coordination numbers of complexes can be determined from the shift of the polarographic half-wave or peak potentials.

The complexing behavior of supporting electrolytes is analytically important for the resolution of overlapping current signals and for masking interfering sample components. Specific additions
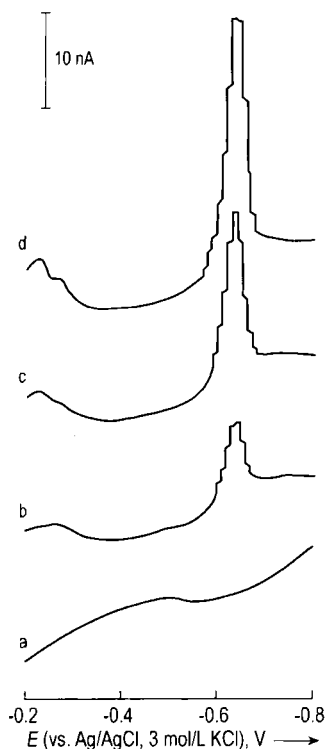
Figure 34. Adsorptive stripping voltammograms for Mo in river water after microwave digestion [50]
a) Untreated sample; b) After microwave digestion + Oxisolv; c) + 2 ng/mL Mo; d) + 4 ng/mL Mo Conditions: 10^{-3} mol/L chloranilic acid, pH 2.7, stirred 1 min, accumulation at –0.2 V
Calculated sample concentration of Mo 1.82 ng/mL

of suitable complexing agents are therefore used to improve the selectivity of polarographic determinations, as demonstrated in Figure 35 by the separation of the DPP peaks of Tl/Pb and Cd/In. Similar phenomena are known in multielement determinations by stripping voltammetry. By changing the composition of the supporting electrolyte solution after the simultaneous accumulation by electrolysis and use of suitable complexing agents, the metals are stripped at different potentials and can be determined from their separated peaks [5].

Simple ligands are used as complexing agents, e.g., CN^-, OH^-, NH_3, oxalate, halides, as well as stronger complexing agents, such as ethylenediaminetetraacetic acid (EDTA), citrate, tartrate, etc.

A high salt content, which can arise after sample digestion or as a result of preceding separation operations, changes the viscosity η of the supporting electrolyte solution and hence the diffusion behavior of the analyte. With increasing salt content the polarographic currents become smaller and displaced to more negative potentials. The peaks in the stripping voltammograms are subject to the same influence, except that the potentials become more positive. Small amounts of inert salts, as can occur in natural samples, generally have no influence on polarographic or voltammetric determinations. Sometimes, as in the case of heavy metal determination in seawater by stripping voltammetry, electrolyte addition becomes superfluous.

The useful voltage range for voltammetric/polarographic determinations depends significantly on the supporting electrolye used. The reduction of the cation of the supporting electrolyte restricts the cathodic working range, as hydrogen evolution from acidic supporting electrolyte solutions depends on the working electrode material. The high overvoltage of H^+ discharge on mercury is therefore particularly advantageous.

The anodic range is determined by oxidation of water or of the supporting electrolyte anion, as well as by oxidation of the electrode material. Anodic dissolution of mercury can even begin at ca. $- 0.4$ V if anions that form poorly soluble $Hg(I)$ salts are present.

Oxygen must be removed (by deaeration) from the supporting electrolyte solution before each voltammetric/polarographic determination. Oxygen is present in aqueous solutions at ca. 10^{-3} mol/L, depending on the solubility conditions, and is reduced at the working electrode in two stages. The course of the reduction is pH dependent and can be described by the following equations.

1) Reduction in acid solution:
$$O_2 + 2H^+ + 2e^- \longrightarrow H_2O_2$$
$$H_2O_2 + 2H^+ + 2e^- \longrightarrow 2H_2O$$
2) Reduction in alkaline solution:
$$O_2 + 2H_2O + 2e^- \longrightarrow H_2O_2 + 2OH^-$$
$$H_2O_2 + 2e^- \longrightarrow 2OH^-$$
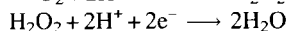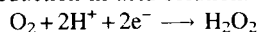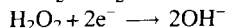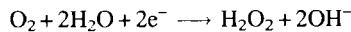
In a DC polarogram the first stage lies between $- 0.1$ and $- 0.3$ V, depending on the pH value, and the second stage occurs between $- 0.7$ and $- 1.3$ V.

The currents caused by the reduction of oxygen interfere with the current – voltage signals of many analytes. In addition, there is interference due to possible subsequent reactions of the H_2O_2 and of the OH^- ions formed.

The dissolved oxygen is removed by introducing pure nitrogen or argon for $5 - 10$ min. Oxygen
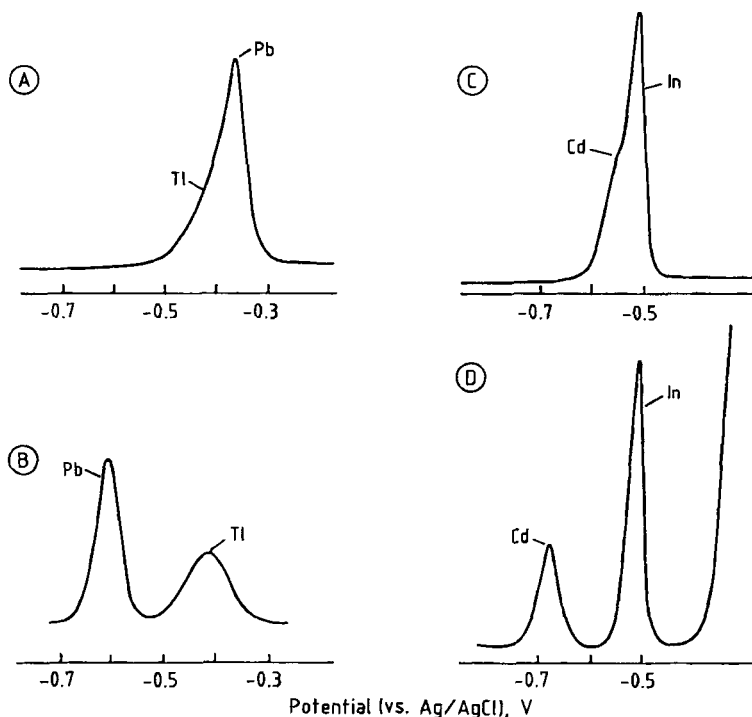
Figure 35. Influence of complexation on the peak separation in differential pulse polarographic curves [5]
10^{-4} mol/L Pb^{2+} and Tl^+ in 0.1 mol/L KCl (A) and 0.1 mol/L KCl + 0.1 mol/L NaOH (B) 10^{-4} mol/L Cd^{2+} and In^{3+} in 0.1 mol/L HCl (C) and 0.1 mol/L HCl + 2 mol/L KI (D)

can also be removed from alkaline solutions by addition of Na_2SO_3.

Inorganic ions are determined almost exclusively in aqueous supporting electrolyte solutions. For organic and organometallic compounds, either aqueous supporting electrolyte solutions mixed with organic solvents or pure organic solvents are used (e.g., alcohols, acetonitrile, propylene carbonate, dimethylformamide, and dimethylsulfoxide). Figure 36 illustrates the potential ranges that are available when using these solvents for anodic and cathodic studies at various working electrodes.

Organic solvents are also used for the extraction of metal chelates, which can then be analyzed in the extract, directly or after adding a polar solvent (in order to raise the solubility of the supporting electrolyte). The procedure known as *extractive voltammetry/polarography* combines the selectivity of partition methods with the possibility of carrying out the determination in organic solvents for many analytes. Using this procedure, the analyte can be simultaneously accumulated and separated from interfering components [15].

25.7. Application to Inorganic and Organic Trace Analysis

Voltammetry and polarography are used primarily for the trace analysis of inorganic and organic substances. Because of the introduction of the powerful pulse methods and stripping techniques, and of advances in instrumentation, these measurement techniques are also of interest for routine analysis. With modern polarographs the most important voltammetric or polarographic methods can be selected by simple switching. Investigations are therefore possible with the same measuring instrument in a concentration range extending over at least six decades. Ions and molecules can be determined and the resulting signals give qualitative and quantitative information as well as providing insight into the bonding state of the analyte.

Elements and Inorganic Ions. The determination of elements and inorganic ions is based chiefly on reduction and oxidation processes that
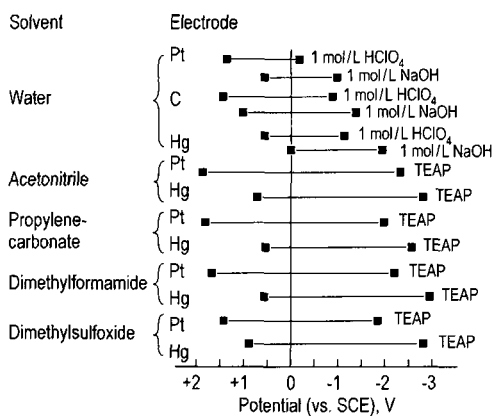
Figure 36. Polarographically usable voltage ranges with Hg, C, and Pt electrodes in various solvents and supporting electrolytes [16]
TEAP = Tetraethylammonium perchlorate

can take place in combination with kinetic, catalytic, or adsorptive processes. The reduction of simple or complexed ions of metals that are soluble in mercury is usually reversible and leads to readily interpretable current–potential curves. Examples of this are the polarographic determinations of cadmium, copper, thallium, zinc, and lead. Less important for routine analysis is the possibility of determining mercury, silver, and gold voltammetrically at noble metal or carbon electrodes.

Multivalent ions are often reduced in several steps, the mechanism and degree of reversibility of the electrode reaction in most cases being dependent on the composition of the supporting electrolyte solution. Two-step current–potential curves are obtained with chromium(VI), copper(II), iron(III), molybdenum(VI), and vanadium(V), among others, when suitable supporting electrolyte solutions are used. In a multi-step process, the shape and height of the polarographic wave or peak associated with each step is determined by the number of electrons involved and the reversibility of that step. For analytical purposes, the best sensitivity is obtained when a supporting electrolyte can be found in which the multi-electron exchange occurs in one step at a single potential, resulting in a large single wave or peak in the polarogram. This is illustrated in the DP-polarographic reduction of Cr(VI) shown in Figure 37. In 0.1 mol/L KNO_3 supporting electrolyte (curve A), Cr(VI) is reduced in two steps — firstly to Cr(III) at -0.3 V and then to Cr(II) at -1.2 V. In 0.1 mol/L Na_2EDTA solution (curve B), only

one peak is observed corresponding to the four electron reduction of Cr(VI) to Cr(II). Because the three-electron reduction to Cr(III) is irreversible the first peak in curve A is smaller than second peak for the reversible one-electron reduction of Cr(III) to Cr(II). The largest peak, obtained for the four-electron reduction of Cr(VI) to Cr(II) in EDTA solution (curve B), gives a determination limit of 10^{-6} mol/L for the determination of chromium.

For those systems which give well behaved catalytic currents, fairly sensitive polarographic and voltammetric analyses are possible. Examples are the determination of Cr(VI) or Mo(VI) in the presence of nitrate ions or of Fe(III) in the presence of bromate ions. For example, the determination limit for Cr(VI) using DPP can be lowered to ca. 10^{-8} mol/L when Na_2EDTA containing nitrate ions is used as the supporting electrolyte [62]. The Cr(II) formed by the reduction of Cr(VI) at the working electrode is immediately oxidized back to the trivalent state by the nitrate ions and is available to be reduced again in the electrode reaction. Titanium, tungsten, uranium and vanadium can also be determined via catalytic currents in the presence of nitrate or chlorate ions with limits of detection in the µg/L (ca. 10^{-8} mol/L) range [4], [5], [63].

Voltammetric/polarographic methods are suitable both for the simultaneous determination of several analytes and for the determination of single analytes in a sample solution. The possibilities are a consequence of the different potentials at which the peak currents occur and the selection of a suitable voltage range for the analysis. Superimposition or overlap of responses can be eliminated by using another supporting electrolyte. Complexing compounds that are expected to change drastically the redox behavior of the analyte or of the interfering component are suitable for the purpose.

The half-wave potentials determined for selected elements in various supporting electrolyte solutions are compared in Figure 38.

Voltammetric/polarographic methods are not very important for anion analysis. Only bromate, iodate, and periodate can be determined by means of polarographic reduction currents. Indirect determinations are possible for those anions that form compounds of low solubility or stable complexes with the Hg_2^{2+} ions formed in the anodic oxidation of the electrode mercury. The current caused by the dissolution of the mercury is proportional to the concentration of these anions in the sample
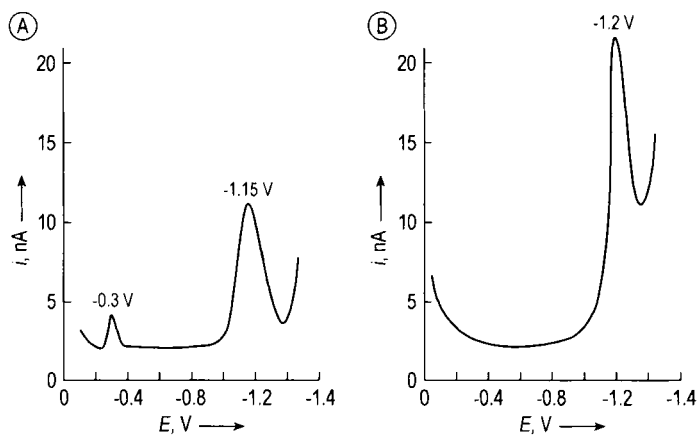
Figure 37. DP-polarogram of chromium(VI) in different supporting electrolytes
A) 0.5 mg/L Cr(VI) in 0.1 M KNO_3; B) 0.5 mg/L Cr(VI) in 0.1 M Na_2EDTA Electrode – SMDE; Potential scan from – 0.1 V to – 1.4 V (vs. Ag/AgCl, 3 M KCl)
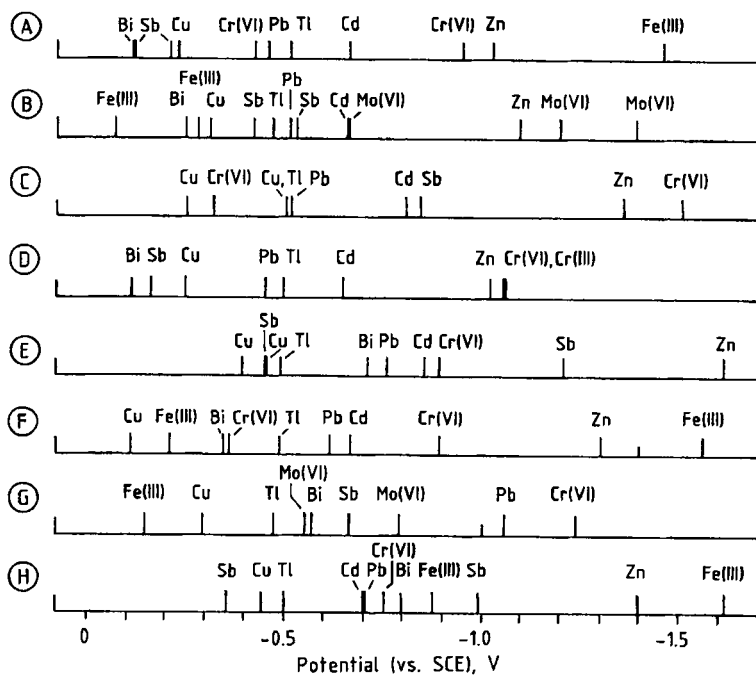
Figure 38. Polarographic half-wave potentials of selected elements measured in the following supporting electrolyte solutions
A) 1 mol/L KCl; B) 2 mol/L $CH_3COOH + 2$ mol/L CH_3COONH_4; C) 1 mol/L $NH_3 + 1$ mol/L NH_4Cl; D) 1 mol/L HCl; E) 1 mol/L NaOH; F) 0.5 mol/L Na tartrate (pH 9); G) 0.1 mol/L Na_2EDTA; H) 0.1 mol/L Na citrate + 0.1 mol/L NaOH

solution. Halides, sulfide, and cyanide can be determined in this way [2], [5].

The most important and most successful application of voltammetry/polarography is the determination of traces of heavy metals in aquatic environmental samples. In only lightly polluted

water samples, e.g., in rainwater, drinking water, and seawater, the determinations can often be carried out without further sample preparation, i.e., directly after sampling and adjustment to the required pH value. More heavily polluted samples must be digested, which is usually performed by

Sampling

Insoluble components ——— No

Filtration

Determination of insoluble components

Determination of soluble components

Filter and residue

Filtrate

Acidification

Wet digestion or low-temperature ashing in activated oxygen

No

Interference

Yes

Digestion by UV photolysis

Take aliquots

Determination of Tl only

Determination of Zn, Cd, Pb, Cu

Determination of Ni, Co

pH Adjustment (4.6) with KCl/sodium acetate solution

pH Adjustment (4.6) with KCl/sodium acetate solution

pH Adjustment (9.3) with NH_4Cl/NH_3 solution

Complexation of Pb with EDTA solution

Complexation with dimethylglyoxime and imidodiacetic acid

DPP: c > 0.1 mg/L
ASV: c < 0.1 mg/L

AdSV: c < 5 µg/L or after dilution

Figure 39. Flow diagram of the procedure for the polaro-graphic/voltammetric determination of heavy metals in drinking water, groundwater, and surface water according to DIN 38 406 Part 16

UV irradiation or in a microwave oven (Chap. 25.5). There is particular interest in determining those metals whose concentration limits in natural waters or wastewaters are regulated by law and accordingly must be controlled (regulations governing purity of drinking water and wastewater discharge). In DIN 38 406, Part 16, for example (German Standard Method), the Zn, Cd, Pb, Cu, Tl, Ni, and Co concentrations are monitored with the aid of polarographic and voltammetric methods of determination: the procedure is shown in Figure 39.

The method illustrated in Figure 39 is suitable for determinations between the lower µg/L and middle mg/L ranges. It is used for monitoring the heavy metal content in polluted surface waters and identifies the metal content of solutions as well as the constituents that occur in association with suspended matter or colloidal particles. The generally very small amounts of Co and Ni are determined as dimethylglyoxime complexes by AdSV, while for the other metals DPP or ASV is used, depending on the concentration. Sample preparation depends on the state of the sample.

The methods of voltammetry/polarography are important for determining the concentrations of metals in different oxidation states and for characterizing physicochemical forms of bonding (*speciation analysis*) [17], [18].

The metal content determined by ASV in untreated water samples at natural pH values or in buffered solutions is usually lower than the total content and is described as the biologically available fraction. This is the fraction that is present in the form of labile inorganic complexes or hydrated ions. The metal content combined with organic ligands can only be determined voltammetrically after digestion by UV photolysis. In this way labile and nonlabile or combined metal species can be distinguished.

For the ultra-trace analysis of heavy metals in water the AdSV method, in addition to ASV, is especially powerful [19]–[21]. The limits of determination are generally in the middle to upper ng/L range; Table 2 gives a survey of the best-known methods.

Most AdSV determinations are carried out, after sample digestion by UV irradiation, by recording cathodic stripping voltammograms. Oxidative microwave digestion with Oxisolv (Chapter 25.5) is also suitable for sample preparation. Only rarely can trace concentrations of metals in natural samples such as seawater be determined without sample digestion. An example is the determination of

Table 2. Ultra-trace analysis of metals in natural waters by AdSV

Analyte	Ligand	Electrolyte	Sample pretreatment	Detection limit, ng/L	Reference
Aluminum	DASA[c]	BES[d] pH 7.1	UV irradiation (not seawater)	30	[52]
Antimony	chloranilic acid	pH 1.0 adjusted with HCl	UV irradiation	550	[53]
Gallium	solochrome violet RS	acetate buffer pH 4.8	UV irradiation (not seawater)	80	[54]
Molybdenum	chloranilic acid	pH 2.7 adjusted with HCl	UV irradiation (not seawater)	20	[50]
Platinum	formazone	formaldehyde + hydrazine + 0.5 mol/L H_2SO_4	UV irradiation	0.01	[55]
Thorium	mordant blue S	acetate buffer pH 6.5	UV irradiation	100	[56]
Tin	pyrocatechol	acetate buffer pH 4.2	UV irradiation (not seawater)	5	[57]
Titanium	mandelic acid	$KClO_3 + NH_3$ pH 3.2	UV irradiation	0.05	[58]
Uranium	pyrocatechol	acetate buffer pH 4.7	separation by ion exchange	240	[59]
		HEPES[a] + 0.5 mol/L NaOH pH 6.8	UV irradiation (not seawater)	70	[60]
	oxine	0.01 mol/L PIPES[b] + 0.5 mol/L NaOH pH 6.7 – 6.9	UV irradiation (not seawater)	500	[61]
	chloranilic acid	pH 2.5 adjusted with HCl	unnecessary	25	[64]
Vanadium	pyrocatechol	PIPES[b] pH 6.9	UV irradiation	5	[65]

[a] HEPES = N-2-hydroxyethylpiperazine-N-2-ethanesulfonic acid.
[b] PIPES = piperazine-N,N'-bis(2-ethanesulfonic acid) monosodium salt.
[c] DASA = 1,2-dihydroxyanthraquinone-3-sulfonic acid.
[d] BES = N,N'-bis-(2-hydroxyethyl)-2-aminoethanesulfonic acid

uranium using chloranilic acid (2,5-dichloro-3,6-dihydroxy-1,4-benzoquinone) (Fig. 40) as complexing agent, which has recently been developed. This method enables interference-free determinations in river water samples to be carried out [64].

In contrast, the determination of *platinum* by AdSV with formazone (a condensation product of formaldehyde and hydrazine) suffers interference from surface-active substances which can be present even in the purest drinking water at trace level. The unusual feature of the extremely sensitive determination of platinum by AdSV is that the complex of Pt(II) with formazone, after adsorptive accumulation at the HMDE, reduces the hydrogen overvoltage on the mercury surface. At the same time, catalytic currents are developed that can be used for the trace analysis of platinum in the ng/L range [55]. The Pt determination is carried out in water samples after prior UV photolysis, and in biological materials after digestion with HNO_3 in a high-pressure asher.

A further field of application of AdSV is the determination of heavy metal traces in high-purity chemicals. The methods originally developed for water analysis (Table 2) can be used without modification for the purity control of salts of the alkali and alkaline earth metals without having to take account of interference by the salt matrix [66].

Arsenic, selenium, and tellurium can be determined at the trace level by cathodic stripping voltammetry (CSV) after having been reductively codeposited with copper. A characteristic of the stripping technique of these three elements is that only a single current peak which arises from the further reduction of the deposited analyte to As^{3-}, Se^{2-} or Te^{2-} is observed. This method is suitable for the determination and speciation of As(V) and As(III) in water samples [36], selenium in minerals [38] and biological samples [68], and for the simultaneous determination of selenium and tellurium (Fig. 41) [37].

Potentiometric stripping analysis (PSA) and chronopotentiometric stripping analysis (CPS) are an alternative techniques that may be used for the determination of metals such as Bi, Cd, Cu, Pb, Sn, Tl, and Zn in water samples [67]. With microprocessor controlled equipment, limits of detection are in the sub-µg/L range. The main advantage of PSA and CPS is that atmospheric oxygen dissolved in an aqueous sample does not have to be removed and that in many cases, in-
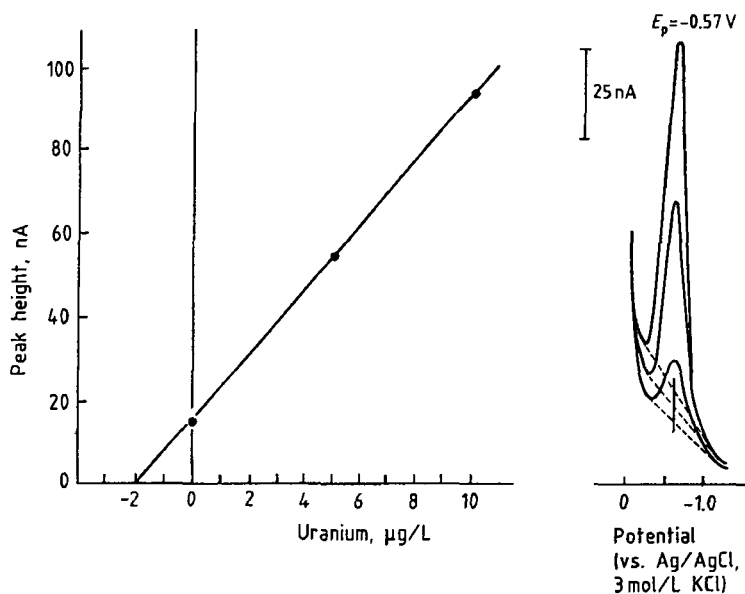
Figure 40. Direct determination of low concentrations of uranium in river water by AdSV [64]
Sample contains 2.4×10^{-4} mol/L chloranilic acid, adjusted with HCl to pH 2.5; $E_{acc} = +0.15$ V; $t_{acc} = 90$ s; standard additions method

cluding those of wine, urine, and blood, direct determination of cadmium and lead is possible [69] – [71].

The determination of trace concentrations of heavy metals by stripping voltammetry in body fluids and organ samples is used to assess their physiological function. Methods have been developed for determining toxic metals by ASV and PSA in blood, serum, and urine. Stripping methods are also important for monitoring the heavy metal content in food and drink. The samples are mineralized by dry ashing or by wet digestion. The choice of the appropriate decomposition procedure depends on the type of test and the form of the elements to be determined [4], [5].

Organic compounds. The polarographic and voltammetric determination of organic compounds usually depends on the reduction or oxidation of functional groups. In addition there are some applications where the determination is based on the reaction of organic molecules with the electrode material (mercury) or on the change in the double layer capacitance associated with the organic species being adsorbed on the electrode surface.

Organic functional groups which are electroactive and the oxidations or reductions of which are the bases of the analytical determinations are listed

in Table 3 together with the most important classes of compounds containing these groups. The ranges for the half-wave or peak potentials of selected groups of organic compounds are presented in Figure 42.

Compounds with reducible functional groups predominate. Polyaromatic hydrocarbons, aromatic hydroxy compounds and amines, as well as amides and various nitrogen heterocyclic compounds can be determined anodically. These processes are in many cases pH dependent. Determinations based on redox processes are therefore carried out in buffered solutions. The composition and concentration of the buffer systems generally has no influence on the position of the half-wave or peak potentials. If its concentration is sufficiently high, the buffer simultaneously performs the function of the supporting electrolyte.

The polarographic analysis of aldehydes, organic acids, monosaccharides, and substituted pyridine derivatives is based on kinetic currents, if a rate-determining chemical reaction precedes or follows the actual electrode reaction. Alkaloids, proteins, and many organic sulfur compounds reduce the overvoltage of the hydrogen reaction on the mercury drop electrode and are determined via catalytic hydrogen waves. In native protein molecules, the sulfhydryl and disulfide groups are

of the metal ions or by a newly formed complex. Both NTA and EDTA are determined according to DIN 38413 Part 5 after addition of bismuth(III) nitrate to the sample acidified to pH 2 with nitric acid. The polarographic curve is shown in Figure 43.

The determination of *surface-active substances* by tensammetry is particularly sensitive owing to adsorption on the electrode surface and the associated change of the double layer capacitance [23].

Voltammetric trace analysis of *alkaloids and other pharmaceuticals* is carried out after adsorptive accumulation at the electrode surface at constant potential [72]. The compounds that can be determined by AdSV in the submicromolar and nanomolar concentration range also include various pesticides [73]–[75]. The importance of AdSV for the trace analysis of adsorbable molecules has been summarized [24].

Organic substances that form slightly soluble or coordination compounds with mercury ions (Table 3) are determined polarographically via anodic waves or voltammetrically after anodic accumulation. Agrochemicals containing thiourea are also determined by this CSV technique; the limits of determination are reported to be $10^{-7}-10^{-8}$ mol/L [76].

Electrochemically inactive organic compounds can be converted into an electrochemically active state by nitration, nitrosation, oxidation, hydrolysis, or some other preliminary chemical reaction, and can then be determined polarographically or voltammetrically. Examples of this are polarographic determinations of alkylbenzenesulfonates in natural water samples after extraction and nitration of the aromatic ring [77] and determinations of beta-receptor blocking agents in tablets after conversion to the corresponding *N*-nitroso derivatives [78].

Aliphatic and aromatic *nitro compounds* are reduced at relatively positive potentials via the hydroxylamines to the corresponding amines. The carcinogenic 4-nitroquinoline-*N*-oxide is determined by DPP in the presence of 4-hydroxyaminoquinoline-*N*-oxide and 4-aminoquinoline-*N*-oxide via the reduction of the nitro group [79]. Nitrazepam, parathion, nitrofurantoin, and the nitroimidazoles in blood plasma or urine are also determined via the reduction of the respective nitro groups.

The *nitroso group* is generally more easily reducible than the nitro group. *N*-Nitrosamines, e.g., *N*-nitroso-*N*-methylaniline in blood, serum, or albumin, are determined in the μg/kg range

Figure 41. Cathodic stripping voltammograms of selenium and tellurium [37]
Supporting electrolyte solution: 0.5 mol/L $(NH_4)_2SO_4$ + 4×10^{-3} mol/L EDTA + H_2SO_4(pH 4.5) + 1 mg/L Cu^{2+}; t_{acc} = 1 min, E_{acc} = –0.4 V, sensitivity 200 nA fc
a) 1 μg/L Se,10 μg/L Te; b) 1 μg/L Te,10 μg/L Se; c) 1 μg/L Se standard addition to 1 mg/L Te; d) 1 μg/L Te standard addition to 1 mg/L Se

masked and therefore do not catalyze this electrode process. Not until denaturation by the action of acids, bases, UV radiation, etc., do the catalytic waves appear as a result of exposure of these groups. This phenomenon can be used to determine denatured DNA in natural DNA samples [22].

Indirect polarographic methods are used to determine nitrilotriacetic acid (NTA) and other *polyamino acids*, e.g., EDTA (auxiliary agents for detergents and cleansers). The methods are based on the formation of stable heavy metal complexes with polarographic properties that are different from those of the acids. Bismuth and cadmium salts are used, added in small excess to the sample. For determination of the acids, use is made of the current signals that are caused either by the excess

Table 3. Functional groups for the polarographic and voltammetric determination of organic compounds

Reducible groups	Compounds
$\diagdown$C=C$\diagup$	unsaturated aliphatic and (poly)aromatic hydrocarbons with conjugated double or triple bonds and with multiple double bonds
$-C\equiv C-$	
$-\overset{\mid}{\underset{\mid}{C}}-X$	halogen-substituted aliphatic and aromatic hydrocarbons with the exception of fluoro compounds
$\diagdown$C=O	aliphatic and aromatic aldehydes, ketones, and quinones
$-O-O-$	aliphatic and aromatic peroxides and hydroperoxides
$-NO_2$	aliphatic and aromatic nitro and nitroso compounds
$-NO$	
$-N=N-$	azo compounds
$\diagdown$C=N$-$	benzodiazepines, pyridines, quinolines, acridines, pyrimidines, triazines, oximes, hydrazones, semicarbazones
$-S-S-$	disulfides
$\diagdown$C=S	thiobenzophenones
$-SO-$	diaryl and alkylaryl sulfoxides
$-SO_2-$	sulfones
$-SO_2NH-$	sulfonamides
$-\overset{\mid}{\underset{\mid}{C}}-Me$	organometallic compounds

Oxidizable groups	Compounds
$A-\overset{\mid}{\underset{\mid}{C}}-H$	poly(aromatic) hydrocarbons (A = aryl)
$-OH$	phenols
$-NH_2$	aromatic amines
$-CO-N\diagdown$	amides

Groups reacting with mercury ions	Compounds
$-SH$	thiols
$-N-\overset{\overset{\displaystyle S}{\parallel}}{C}\diagdown_{S^-}$	dithiocarbamates
$\overset{-NH\diagdown}{\underset{-NH\diagup}{}}C=S$	thioureas, thiobarbiturates
$-C\overset{\overset{\displaystyle S}{\parallel}}{\diagdown_{NH_2}}$	thioamides
$\overset{-NH\diagdown}{\underset{-NH\diagup}{}}C=O$	derivatives of barbituric acid and of uracil
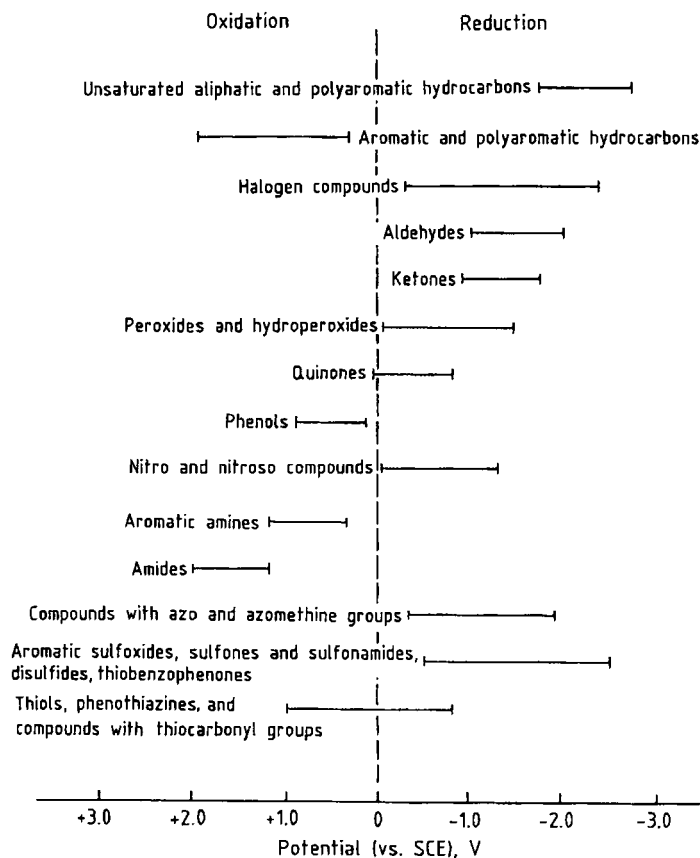
Figure 42. Potential ranges for the polarographic and voltametric half-wave and peak potentials of organic compounds

by DPP [80]. The polarographic and voltammetric behavior of the nitrosamines, quinones, steroid hormones, and imidazoles is of interest, particularly in connection with cancer research [25].

Knowledge about the bio-transformation of *drugs* and their excretion is important pharmacologically. Numerous methods for the polarographic and voltammetric determination of pharmaceuticals and their metabolites have been developed for the assessment of such relationships. The determination is generally preceded by separation from the matrix by liquid – liquid extraction with diethyl ether or ethyl acetate.

1,4-Diazepines are determined polarographically via their azomethine groups. In addition to flurazepam, further representatives of this class of pharmaceutically active agents can be determined polarographically. These include diazepam, chlordiazepoxide, bromazepam, lorazepam, and chlo-

razepam; the limits of determination with DPP are in the μg/L range [4], [5].

Organic compounds with oxidizable functional groups (see Table 3) are mostly determined, after separation by liquid (column) chromatography or by high-performance liquid chromatography, with the aid of a voltammetric (amperometric) detector (→ Liquid Chromatography). These compounds, which can be determined in liquid and solid biological samples, in foods, or in aquatic environmental samples, include catecholamines, sulfonamides, phenothiazines, dopamines, estrogens and other hormones, tocopherols, and various groups of agricultural pesticides. Liquid – liquid extraction or solid phase extraction are also used for sample preparation; the determinations are carried out in the ng/L range (→ Liquid Chromatography).

Polarography and voltammetry are of importance in the *analysis of organometallic* compounds
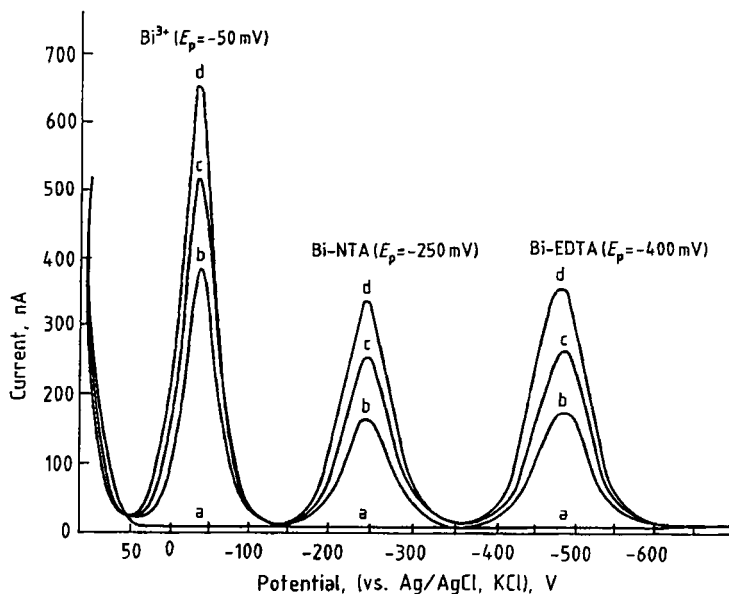
Figure 43. Polarographic determination of NTA and EDTA according to DIN 38 413 Part 5, standard additions method
a) Residual current; b) After Bi^{3+} addition; c) First addition of Bi-NTA/EDTA solution; d) Second addition of Bi-NTA/EDTA solution

in environmental samples. In contrast to spectroscopic methods which only give information on the total metal contents in a sample, the voltammetric techniques are capable of characterizing and determining the various forms of those compounds as they exist in the sample (speciation).

Representative of the organometallic compounds are the R–Me–compounds where Me = As, Hg, P, Pb, Sb, Se, Sn, Te, and Ti; the peak potentials are listed in [30].

Organomercury compounds of the type RHgX generally are reduced in two steps according to the mechanism.

$$RHg^+ + e^- \rightleftharpoons RHg^{\bullet} + e^-$$

with branches:

+H⁺, pH < 7 → RH + Hg

pH > 7 → RHg^- →(+H⁺) RH + Hg

$$\downarrow$$
$$(RHg)_2$$
$$\downarrow$$
$$R_2Hg + Hg$$
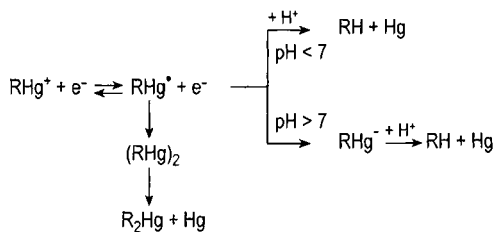
Likewise organo-arsenic, antimony, and tin compounds either singly or in admixture can be determined in the µg range by polarography and particularly mono-, di-, and tributyltin in the ng

range, based on solid phase extraction and adsorptive cathodic stripping voltammetry [81].

The polarographic or voltammetric determination of organic compounds of biological, pharmaceutical, or environmental significance have been summarized [27] – [29].

25.8. References

General References

[1] J. Heyrovsky, J. Kuta: *Grundlagen der Polarographie*, Akademie Verlag, Berlin 1965.
[2] A. M. Bond: *Modern Polarographic Methods in Analytical Chemistry*. Marcel Dekker, New York 1980.
[3] F. G. Thomas, G. Henze: *An Introduction to Analytical Voltammetry — Theory and Practice*, CSIRO, Melbourne (2001).
[4] G. Henze: *Polarographie und Voltammetrie; Grundlagen und analytische Praxis*. Springer Verlag, (2001).
[5] G. Henze, R. Neeb: *Elektrochemische Analytik*, Springer Verlag, Berlin 1986.
[6] J. Koryta, J. Dvorak: *Principles of Electrochemistry*, J. Wiley & Sons, New York 1987.
[7] M. R. Smyth, J. G. Vos: *Analytical Voltammetry*, Elsevier, New York 1992.
[8] G. C. Baker, A. W. Gardner: "Forty Years of Square-Wave Polarography," *Analyst (London)* **117** (1982) 1811.

[9] J. G. Osteryoung, M. M. Schreiner: "Recent Advances in Pulse Voltammetry," *CRC Crit. Rev. Anal. Chem.* **19** (Suppl. 1) (1988) 1.

[10] J. Heinze: "Cyclovoltammetrie – die Spektroskopie des Elektrochemikers," *Angew. Chem.* **96** (1984) 823.

[11] R. Neeb: *Inverse Polarographie und Voltammetrie*, Verlag Chemie, Weinheim 1969.

[12] J. Wang: *Stripping Analysis* VCH Verlagsgesellschaft, Weinheim 1985.

[13] R. Kalvoda, M. Kopanica: "Adsorptive Stripping Voltammetry in Trace Analysis," *Pure Appl. Chem.* **61** (1987) 97.

[14] G. Henze: "Stripping-Chronopotentiometrie," *Fresenius Z. Anal. Chem.* **315** (1983) 438.

[15] G. K. Budnikow, N. A. Ulakhovich: "Extraction Polarography and its Analytical Application," *Russ. Chem. Rev. (Engl. Transl.)* **49** (1980) 74.

[16] P. Rach, H. Seiler: *Polarographie und Voltammetrie in der Spurenanalyse*, Hüthig Verlag, Heidelberg 1985.

[17] G. Henze: "Application of Polarographic and Voltammetric Techniques in Environmental Analysis," *NATO ASI Ser. G* **23** (1990) 391.

[18] T. M. Florence: "Analytical Application to Trace Element Speciation in Water," *Analyst* **111** (1986) 489.

[19] C. M. G. van den Berg: "Adsorptive Cathodic Stripping Voltammetry of Trace Elements in Water," *Analyst* **114** (1989) 1527.

[20] C. M. G. van den Berg: "Electroanalytical Chemistry of Sea-Water," *Chem. Oceanogr.* **9** (1988) 195.

[21] C. M. G. van den Berg: "Potentials and Potentialities of Cathodic Stripping Voltammetry of Trace Elements in Natural Water," *Anal. Chim. Acta* **250** (1991) 265.

[22] E. Palecek: "Polarographic Techniques in Nucleic Acid Research," in W. F. Smyth (ed.): "Electroanalysis in Hygiene, Environmental, Clinical and Pharmaceutical Chemistry," *Anal. Chem. Symp.* **2** (1980) 79.

[23] P. M. Bersier, J. Bersier: "Polarographic Adsorption Analysis and Tensammetry: Toys or Tools," *Analyst* **113** (1988) 3.

[24] R. Kalvoda: "Polarographic Determination of Adsorbable Molecules," *Pure Appl. Chem.* **59** (1987) 715.

[25] "Electrochemistry in Cancer Research," *Anal. Proc. (London)* **17** (1980) 278 (Conference Proceedings)

[26] J. P. Hart: "Electroanalysis of Biologically Important Compounds," in *Analytical Chemistry*, Ellis Harwood Series.

[27] J. Wang: *Electroanalytical Techniques in Clinical Chemistry and Laboratory Chemistry*, VCH Verlagsgesellschaft, Weinheim 1988.

[28] W. F. Smyth: "Voltammetric Determination of Molecules of Biological, Environmental, and Pharmaceutical Importance," *CRC Crit. Rev. Anal. Chem.* **18** (1987) 155.

[29] W. F. Smyth, M. Smyth: "Electrochemical Analysis of Organic Pollutants," *Pure Appl. Chem.* **59** (1987) 245.

Specific References

[30] W. F. Smyth: *Voltammetric Determination of Molecules of Biological Significance*, John Wiley & Sons, New York 1992.

[31] G. C. Barker, A. W. Gardner, *Fresenius' Z. Anal. Chem.* **113** (1960) 73.

[32] J. G. Osteryoung, R. A. Osteryoung, *Anal. Chem.* **57** (1985) 101 A.

[33] E. J. Zachowski, M. Wojciechowski, J. Osteryoung, *Anal. Chim. Acta* **183** (1986) 47.

[34] G. Henze, R. Neeb, *Fresenius' J. Anal. Chem.* **310** (1982) 111.

[35] G. Henze, A. P. Joshi, R. Neeb, *Fresenius J. Anal. Chem.* **300** (1980) 267.

[36] G. Henze, W. Wagner, S. Sander, *Fresenius J. Anal. Chem.* **358** (1997) 741.

[37] G. Henze et al., *Fresenius' Z. Anal. Chem.* **295** (1979) 1.

[38] G. Henze, *Mikrochim. Acta* 1981, no. II, 343.

[39] D. Jagner, K. Aren, *Anal. Chim. Acta* **100** (1978) 375.

[40] D. Jagner, *Anal. Chem.* **50** (1978) 1924.

[41] L. Kryger, *Anal. Chim. Acta* **120** (1980) 19.

[42] A. Graneli, D. Jagner, M. Josefson, *Anal. Chem.* **52** (1980) 2220.

[43] T. Garai et al., *Electroanalysis* **4** (1992) 899.

[44] K. Kalcher, "Chemically Modified Carbon Paste Electrodes in Voltammetric Analysis," *Electroanalysis* **2** (1990) 419.

[45] J. Wang, T. Boamin, R. Setiadji, *Electroanalysis* **6** (1994) 317.

[46] Kh. Brainina, G. Henze, N. Stojko, N. Malakhova, Chr. Faller, "Thick-film Graphite Electrodes in Stripping Voltammetry," *Fresenius J. Anal. Chem.* **364** (1999) 285.

[47] J. Schiewe, A. M. Bond, G. Henze: "Voltammetrische Spurenanalyse mit Mikroelektroden" in G. Henze, M. Köhler, J. P. Lay (eds.): *Umweltdiagnostik mit Mikrosystemen*, Wiley-VCH, Weinheim 1999, p. 121.

[48] H. Y. Chen, R. Neeb, *Fresenius' Z. Anal. Chem.* **319** (1984) 240.

[49] M. Kolb, P. Rach, J. Schäfer, A. Wild, *Fresenius' J. Anal. Chem.* **342** (1992) 341.

[50] M. Karakaplan, S. Gücer, G. Henze, *Fresenius' J. Anal. Chem.* **342** (1992) 186.

[51] D. J. Myers, J. Osteryoung, *Anal. Chem.* **45** (1973) 267.

[52] C. M. G. van den Berg, K. Murphy, J. P. Riley, *Anal. Chim. Acta* **188** (1986) 177.

[53] W. Wagner, S. Sander, G. Henze, *Fresenius J. Anal. Chem.* (in press).

[54] J. Wang, J. M. Zadeii, *Anal. Chim. Acta* **185** (1986) 229.

[55] C. M. G. van den Berg, G. S. Jacinto, *Anal. Chim. Acta* **211** (1988) 129.

[56] J. Wang, J. M. Zadeii, *Anal. Chim. Acta* **188** (1986) 187.

[57] S. B. Adeloju, *Anal. Sci.* **7** (1991) 1099.

[58] K. Yokoi, C. M. G. van den Berg, *Anal. Chim. Acta* **245** (1991) 167.

[59] N. K. Lam, R. Kalvoda, M. Kopanica, *Anal. Chim. Acta* **154** (1983) 79.

[60] C. M. G. van den Berg, Z. Qiang Huang, *Anal. Chim. Acta* **164** (1984) 209.

[61] C. M. G. van den Berg, M. Nimmo, *Anal. Chem.* **59** (1987) 924.

[62] J. Zarebski, *Chemia Analityczna* **22** (1977) 1037.

[63] L. Meites: *Polarographic Techniques*, 2nd ed., Interscience, New York 1961.

[64] S. Sander, G. Henze, *Fresenius J. Anal. Chem.* **349** (1994) 654.

[65] C. M. G. van den Berg, Z. Q. Huang, *Anal. Chem.* **56** (1984) 2383.

[66] R. Naumann, W. Schmidt, G. Höhl, *Fresenius J. Anal. Chem.* **343** (1992) 746.

[67] J. M. Estela, C. Tomas, A. Cladera, V. Cerda, *Crit. Rev. Anal. Chem.* **25** (1995) 91.

[68] S. B. Adeloju, A. M. Bond, *Anal. Chem.* **56** (1984) 2397.

[69] D. Jagner, M. Josefson, S. Westerlund. K. Aren, *Anal. Chem.* **53** (1981) 1406.

[70] H. Huiliang. D. Jagner, L. Renman. *Talanta* **34** (1987) 539.

[71] C. Hua, D. Jagner, L. Renman, *Anal. Chim. Acta* **197** (1987) 25.

[72] R. Kalvoda, *Anal. Chim. Acta* **138** (1982) 11.

[73] M. Pedrero. F. J. Manuel de Villena, J. M. Pingarron. L. M. Polo, *Electroanalysis* **3** (1991) 419.

[74] Ch. Li, B. D. James, R. J. Magee, *Electroanalysis* **2** (1990) 63.

[75] M. A. Goicolea, J. F. Arrenz, R. J. Barrio, Z. G. de Balugera, *Fresenius' J. Anal. Chem.* **339** (1991) 166.

[76] M. R. Smyth, J. G. Osteryoung, *Anal. Chem.* **49** (1977) 2310.

[77] J. P. Hart, W. F. Smyth, B. J. Birch. *Analyst* **104** (1979) 853.

[78] M. A. Korany, H. Riedel, *Fresenius' Z. Anal. Chem.* **314** (1983) 678.

[79] J. S. Burmich. W. F. Smyth. *Anal. Proc. (London)* **17** (1980) 284.

[80] H. M. Pylypiw, G. W. Harrington, *Anal. Chem.* **53** (1981) 2365.

[81] J. Schwarz, G. Henze, F. G. Thomas. *Fresenius J. Anal. Chem.* **352** (1995) 474, 479.

26. Thermal Analysis and Calorimetry

STEPHEN B. WARRINGTON, Anasys, IPTME, Loughborough University, Loughborough, United Kingdom (Chap. 26)

GÜNTHER W. H. HÖHNE, Dutch Polymer Institute, Eindhoven University of Technology, Eindhoven, The Netherlands (Chap. 26.2)

26.1. Thermal Analysis

26.1.1. General Introduction

26.1.1.1. Definitions

Thermal analysis (TA) has been defined as "a group of techniques in which a physical property of a substance and/or its reaction products is measured as a function of temperature while the substance is subjected to a controlled temperature programme" [1]. The formal definition is usually extended to include isothermal studies, in which the property of interest is measured as a function of time.

The definition is a broad one, and covers many methods that are not considered to fall within the field of thermal analysis as it is usually understood. The present chapter will be restricted to the major techniques as currently practiced. Since all materials respond to heat in some way, TA has been applied to almost every field of science, with a strong emphasis on solving problems in materials science and engineering, as well as fundamental chemical investigations. TA is applicable whenever the primary interest is in determining the effect of heat upon a material, but the techniques can also be used as a means of probing a system to obtain other types of information, such as composition.

The aim of this chapter is to give an overview of the main TA methods and their applications,

Table 1. Processes that can be studied by thermogravimetry

Process	Weight gain	Weight loss
Ad- or absorption	*	
Desorption		*
Dehydration/desolvation		*
Sublimation		*
Vaporization		*
Decomposition		*
Solid – solid reactions		*
Solid – gas reactions	*	*

with sufficient references to provide reader access to more detailed information. The following list summarizes properties subject to investigation and the corresponding thermal analysis techniques:

Mass	Thermogravimetry (TG)
Temperature	Differential thermal analysis (DTA)
Enthalpy	Differential scanning calorimetry (DSC)
Evolved gas	Evolved gas analysis (EGA)
Physical dimensions	Thermodilatometry (TD)
Mechanical properties	Thermomechanical analysis (TMA) Dynamic mechanical analysis (DMA)
Optical properties	Thermoptometry
Electrical properties	Thermoelectrometry
Magnetic properties	Thermomagnetometry

26.1.1.2. Sources of Information

Two journals (the *Journal of Thermal Analysis and Calorimetry* and *Thermochimica Acta*) devote their contents entirely to TA; the *Proceedings* of the (now) four-yearly Conferences of the International Confederation for Thermal Analysis and Calorimetry (ICTAC) constitute an excellent additional source of research papers. Specific information regarding *Proceedings* volumes for the nine ICTAC Conferences between 1965 and 1991 is available in [1]. The most complete listing of worldwide TA literature is also found in the ICTAC handbook [1], which in addition gives addresses for national TA societies and important equipment suppliers. The most useful textbooks include [2] – [5].

There is a wealth of information available on the Internet; see, for example, http://www.users.globalnet.co.uk/~dmprice/ta/links/index.htm

26.1.2. Thermogravimetry

26.1.2.1. Introduction

Thermogravimetry (TG) is used to measure variations in mass as a function of temperature (or time). Processes amenable to study in this way are listed in Table 1. TG is one of the most powerful TA techniques from the standpoint of quantitative data, and for this reason it is often employed in combination with other measurements.

26.1.2.2. Instrumentation

The instrument used is a *thermobalance*. The schematic diagram in Figure 1 presents the main components of a typical modern unit. Component details vary according to the design, and choice of a particular instrument is usually dictated by requirements of the problem under investigation (temperature range, sensitivity, etc.). The balance mechanism itself is usually of the null-deflection type to ensure that the sample's position in the furnace will not change. The balance transmits a continuous measure of the mass of the sample to an appropriate recording system, which is very often a computer. The resulting plot of mass vs. temperature or time is called a *TG curve*. Balance sensitivity is usually of the order of one microgram, with a total capacity of as much as a few hundred milligrams. Furnaces are available that operate from subambient (e.g., –125 °C) or room temperature up to as high as 2400 °C. A furnace programmer normally supports a wide range of heating and cooling rates, often in combination, as well as precise isothermal control. The programming functions themselves are increasingly being assumed by computers. Most applications involve heating rates of $5 - 20\,°C\,min^{-1}$, but the ability to heat a sample as rapidly as $1000\,°C\,min^{-1}$ can be useful in the simulation of certain industrial processes, for example, or in flammability studies. Special control methods, grouped under the term *controlled-rate thermal analysis* (CRTA), are receiving increasing attention [6], and offer advantages in resolving overlapping processes and in kinetic studies.

The quality of the furnace atmosphere deserves careful attention. Most commercial thermobalances operate at atmospheric pressure. Vacuum and high-pressure studies normally require specialized equipment, either commercial or home-
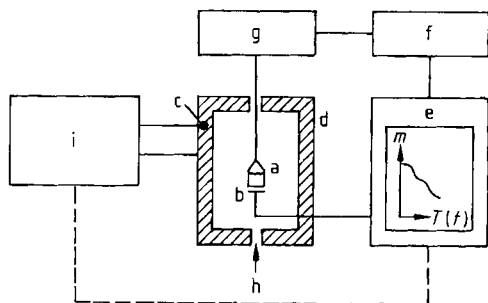
Figure 1. Schematic diagram of a thermobalance
a) Sample; b) Sample temperature sensor; c) Furnace temperature sensor; d) Furnace; e) Recorder or computer, logging sample mass, temperature, and time; f) Balance controller; g) Recording microbalance; h) Gas; i) Furnace temperature programmer
The computer may also control the furnace programmer

Figure 2. TG (A) and DTG (B) curves for calcium oxalate monohydrate
Sample mass 30 mg, heating rate $20\,°C\,min^{-1}$, argon atmosphere

made, as do experiments with corrosive gases [7]. The ability to establish an inert (oxygen-free) atmosphere is useful, as is the potential for rapidly changing the nature of the atmosphere.

The mode of assembly of the components varies; for example, the furnace might be above, below, or in line with the balance. Sample containers also vary widely in design; cylindrical pans are common, typically 5 – 8 mm in diameter and 2 – 10 mm high, though flat plates and semisealed containers may be used to investigate the effects of atmospheric access to a sample. Compatibility between the construction materials and the system under investigation must be carefully considered. Materials commonly available include aluminum, platinum, alumina, and silica. Temperature indication is normally provided by a thermocouple located near the sample container. Because of inevitable thermal gradients within the apparatus, an indicated temperature can never be taken as an accurate reflection of the temperature of the sample. Reproducible location of the thermocouple is vital; recommended calibration procedures have been described in [8]. Only in simultaneous TG – DTA instruments is direct measurement of the temperature possible.

26.1.2.3. Factors Affecting a TG Curve

WENDLANDT [2] has listed 13 factors that directly affect a TG curve, both sample- and instrument-related, some of which are interactive. The primary factors are heating rate and sample size,

an increase in either of which tends to increase the temperature at which sample decomposition occurs, and to decrease the resolution between successive mass losses. The particle size of the sample, the way in which it is packed, the crucible shape, and the gas flow rate also affect the progress of a thermal reaction. Careful attention to consistency in experimental details normally results in good repeatability. On the other hand, studying the effect of deliberate alterations in such factors as heating rate can provide valuable insights into the nature of observed reactions. All these considerations are equally applicable to other techniques as well, including DTA (Section 26.1.3).

26.1.2.4. Applications

TG has been applied extensively to the study of analytical precipitates for gravimetric analysis [9]. One example is calcium oxalate, as illustrated in Figure 2. Information such as extent of hydration, appropriate drying conditions, stability ranges for intermediate products, and reaction mechanisms can all be deduced from appropriate TG curves. Figure 2 also includes the first derivative of the TG curve, termed the *DTG curve*, which is capable of revealing fine details more clearly.

A TG method has also been devised for the proximate analysis of coal, permitting four samples per hour to be analyzed with results of the same precision and accuracy as BS and ASTM methods [10]. A typical curve is shown in Figure 3. The approach taken here can be generalized to the compositional analysis of many materials [11].
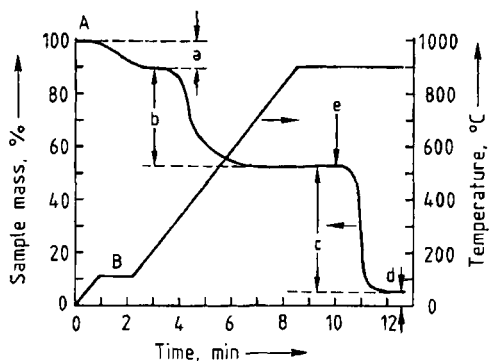
Figure 3. TG curve (A) illustrating a programmed-temperature method for the proximate analysis of coal, where the line B defines the temperature program
a) Moisture; b) Volatiles; c) Fixed carbon; d) Ash; e) Point at which the atmosphere was changed to air

Figure 5. DTG curves for four ASTM standard coal samples heated in air [12]
a) Lignite "A"; b) Sub-bituminous "A"; c) High-volatility bituminous "A"; d) Buckwheat anthracite

Figure 4. TG curves for four polymers in air, showing their relative stabilities
a) Silicone rubber; b) PVC; c) PTFE; d) Perspex

rapid assessment of new fuels (Fig. 5). The simulation of industrial processes, particularly heterogeneous catalysis, is another important area of application.

26.1.3. Differential Thermal Analysis and Differential Scanning Calorimetry

26.1.3.1. Introduction

Both differential thermal analysis (DTA) and differential scanning calorimetry (DSC) are concerned with the measurement of energy changes, and as such are applicable in principle to a wider range of processes than TG. From a practical standpoint DSC may be regarded as the method from which quantitative data are most easily obtained. The use of DSC to determine absolute thermodynamic quantities is discussed in Sections 26.2.3.2 and 26.2.4.1. Types of processes amenable to study by these methods are summarized in Table 2.

26.1.3.2. Instrumentation

Figure 6 shows the components of a typical DTA apparatus. The sample and an inert reference material (commonly alumina) are subjected to a common temperature program while the temperature difference between the two is monitored. In an ideally designed instrument, the temperature difference ΔT remains approximately constant if no reaction is taking place in the sample. When an endothermic process such as melting occurs, the

TG also facilitates comparisons of the relative stabilities of polymers (Fig. 4). An analysis of curves prepared at different heating rates makes it possible to extract kinetic data for estimating the service lifetimes of such materials. In another coal-related application, an empirical correlation has been established between characteristic temperatures obtained from DTG curves of coals examined under air and the performance of the same coals in steam-raising plants [12], permitting the

Table 2. Processes that can be studied by DTA/DSC

Process	Exothermic	Endother-mic
Solid – solid transitions	*	*
Solid – liquid transitions	*	*
Vaporization		*
Sublimation		*
Adsorption	*	
Desorption		*
Desolvation		*
Decomposition	*	*
Solid – solid reactions	*	*
Solid – liquid reactions	*	*
Solid – gas reactions	*	*
Curing	*	
Polymerizations	*	
Catalytic reactions	*	*

Figure 7. Typical DTA curve for the melting of a pure material

Figure 6. Schematic diagram of a classical DTA apparatus a) Reference thermocouple; b) Sample thermocouple; c) Heating block; d) ΔT amplifier; e) Recorder or computer, logging T_s, ΔT, and time; f) Furnace; g) Temperature programmer, which may be linked to the computer; h) Gas inlet T_s, sample temperature; T_r, reference temperature; $\Delta T = T_s - T_r$

sample temperature T_s lags behind that of the reference T_r; $\Delta T = T_s - T_r$ is thus negative, so a (negative) peak is produced on the corresponding *DTA curve*, which is a record of ΔT against either temperature or time (Fig. 7). Exothermic processes lead to peaks in the opposite direction. The peak *area* is related to both the energy change and the sample size according to the relationship

$$\Delta H \cdot m = K \int \Delta T \cdot dt$$

where m is the sample mass, and ΔH is the enthalpy of reaction. The proportionality constant K is temperature-dependent, and with classical DTA it is also influenced by thermal properties of the sample and container. Certain apparatus designs

minimize these influences, so that once the characteristics of $K = f(T)$ have been determined the ΔT signal can be conditioned (electronically or digitally) to give an output signal calibrated directly in power units; i.e., the sensitivity is nominally independent of temperature. This is the principle underlying heat-flux DSC. For a more complete description of this technique, as well as power-compensated DSC, see Section 26.2.3.2. A wide variety of commercial devices covers the temperature range from ca. −150 – 2500 °C, but high-temperature instruments are less sensitive than low-temperature designs due to constraints imposed by thermocouple technology. DSC instruments are usually limited to a maximum temperature of 700 °C, though heat-flux DSC has been extended to 1400 °C. DTA head designs vary enormously, but most accommodate samples in the range 5 – 200 mg in metal or ceramic containers. The thermocouples usually take the form of beads that fit into recesses in the containers, or plates upon which the containers rest. Nowadays the thermocouple is never actually placed in the sample, in contrast to the "classical" arrangement shown in Figure 6.

26.1.3.3. Applications

DSC is used extensively in polymer science [13]. A generalized DSC curve for a polymer is shown in Figure 8. Most polymers display a *glass transition*, in the course of which the material passes from a glassy to a rubbery state with a simultaneous increase in specific heat capacity.
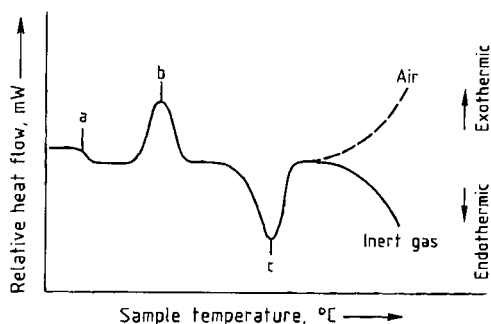
Figure 8. Generalized DSC curve for a polymer
a) Glass transition; b) Crystallization; c) Melting
The subsequent reactions may be endothermic or exothermic

Figure 10. Generalized DSC curve for the melting of an impure material
Fraction melted at temperature $Tx = ($area $ABE)/($area $ADC)$

Figure 9. DTA curve for a partially converted high-alumina cement concrete from a test drilling
Degree of conversion $=100\ y/(x+y)$

Glass-transition temperature measurements are used for material characterization and comparison, often providing information about such factors as thermal history, crystallinity, extent of cure, and plasticizer content of a polymer. Amorphous polymers usually exhibit a crystallization exotherm, and thermosetting polymers a curing exotherm, both of which can be analyzed kinetically to establish the thermal treatment required to obtain desired properties in a particular product. The melting endotherm is a measure of the extent of crystallinity, an important parameter related to mechanical properties. The temperature at which exothermic degradation begins under an oxidizing atmosphere is used to compare oxidative stabilities of different polymers or different stabilizers. Discrimination can be improved by isothermal measurements, in which the material under investigation is heated initially in an inert gas, after which the atmosphere is changed to air or oxygen and the time lapse prior to exothermic deflection is measured. The same approach is applicable to studies

of oils, greases, fats, etc., though high-pressure cells may be necessary to prevent volatilization.

The particular temperatures at which peaks are observed can be used for the identification of components in a mixture, and the size of a particular peak can be used for quantitative evaluation. Examples include the determination of quartz in clays and analysis of the constituents of polymer blends, both of which would be difficult to achieve by other means. DTA has been used to determine the presence and degree of conversion of high-alumina cement concrete (HAC), a material subject to severe loss of strength under certain atmospheric conditions due to solid-state reactions in the cement matrix. Figure 9 shows a DTA curve for partially converted HAC together with the way the data are used for a conversion measurement.

An analysis of the shape of the fusion peak from a relatively pure material (e.g., a pharmaceutical agent) can, with certain restrictions, lead to an estimate of purity [14]. Figure 10 shows an idealized melting curve for an impure material. By calculating the fraction (F) of material melted at a series of temperatures, and applying several simplifying assumptions, the mole fraction of an impurity can be calculated from the expression:

$$T_s = T_0 - \frac{RT_0^2 X_2}{\Delta H_f} \cdot \frac{1}{F}$$

where T_s = sample temperature, T_0 = melting point of the pure material, R = gas constant, X_2 = mole fraction of impurity, ΔH_f = enthalpy of fusion, and F = the fraction melted at temperature T_s. The cal-
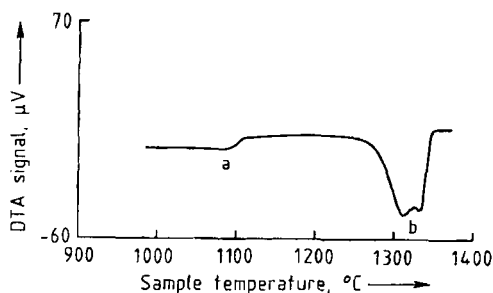
Figure 11. DTA curve for a nickel-based superalloy
a) Solution of the γ' phase; b) Melting 250 mg sample heated
at $10\,^\circ C\,min^{-1}$ in argon

culations are readily performed by commercial software.

DTA curves can be used to construct phase diagrams for mixtures. Measurement of a single point on the phase diagram may be sufficient for quality-control purposes. Figure 11 shows the DTA curve for a nickel-based superalloy, where the solvus temperature (the temperature corresponding to complete solution of one solid phase in another) of the γ' phase characterizes the heat treatment to which the alloy has been subjected.

Solid – solid reactions have also been successfully studied by these techniques, as have decompositions of high explosives. The applicability of DTA/DSC to very small samples is of obvious value here. Similarly, preliminary screening of potentially hazardous reaction mixtures is conveniently carried out in this way as a guide to the likelihood of exothermic reaction, the corresponding temperature range and magnitude, and possible effects of pressure and atmospheric conditions (see also Section 26.2.4.5).

26.1.3.4. Modulated-Temperature DSC (MT-DSC)

MT-DSC is an important recent development (see also Section 26.2.2.3) which offers significant advantages in the quality of results and gives additional information, particularly for polymers [15]. A sinusoidal modulation is superimposed on the normally linear temperature ramp, and the resulting signals are analyzed by a Fourier Transform technique. This allows the separation of "reversing" and "nonreversing" thermal events and simplifies the analysis of overlapping transitions such as those often found in the glass transition region. Other benefits include the ability to meas-

ure heat capacity changes pseudo-isothermally, as in, e.g., curing systems, and improved signal-to-noise ratio and baseline linearity.

26.1.4. Simultaneous Techniques

26.1.4.1. Introduction

TG or DTA/DSC alone rarely gives sufficient information to permit a complete interpretation of the reactions in a particular system; results must usually be supplemented by other thermal methods and/or general analytical data. Alternative thermal methods are best applied simultaneously, leading, for example, to TG and DTA information from the same sample under identical experimental conditions. This avoids ambiguity caused by material inhomogeneity, but also problems attributable to different experimental conditions with different instruments, which sometimes markedly affects the correspondence between TG and DTA curves. Other advantages include: (1) indication of the thermal stability of materials examined by DTA, which in turn makes it possible to correct measured heats of reaction for partially decomposed samples; (2) accurate temperature measurement in TG work, which is vital in kinetic studies; and (3) the detection of unsuspected transitions in a condensed phase, which may help explain puzzling features of the corresponding TG curve.

The range of simultaneous techniques has been reviewed by PAULIK [16]. TG–DTA and TG–DSC are the commonest simultaneous methods, followed by evolved gas analysis (Section 26.1.5).

26.1.4.2. Applications

TG – DSC curves showing the curing and decomposition of a polyimide resin are presented in Figure 12. Following a glass transition (a) and subsequent fusion of the material at ca. 120 °C (b), which are detected by DSC, the TG curve reveals the initial stages of an exothermic curing reaction (c) that occurs with a 7 % mass loss. Endothermic decomposition of this resin does not take place until ca. 430 °C, as shown again by TG (d), a finding that permitted assignment of the appropriate baseline for measuring the heat of curing.

26.1.5. Evolved Gas Analysis

Evolved gas analysis (EGA) measures the nature and/or quantity of volatile products released
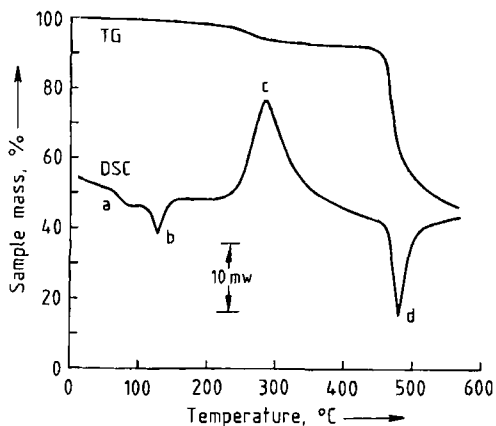
Figure 12. Simultaneous TG–DSC curves for an uncured polyimide resin
a) Glass transition; b) Melting; c) Curing; d) Degradation
10 mg sample heated at 10 °C min⁻¹ in argon

Figure 13. Simultaneous TG–MS curves for a brick clay
A) TG curve; B) MS data
Evolution of a) Water (18 amu); b) Carbon dioxide (44 amu); c) Sulfur dioxide (64 amu)
40 mg sample heated at 15 °C min⁻¹ in air

on heating. In practice it is often performed simultaneously with TG or TG–DTA, and is particularly useful in furnishing direct chemical information to augment the physical measurements of TG or DTA. Additional advantages that have emerged include specificity — a single decomposition can be followed against a background of concurrent processes — and sensitivity, which can be far greater than with TG alone [17], [18].

Instrumentation. Almost every imaginable type of gas detector or analyzer has been utilized in EGA, including hygrometers, nondispersive infrared analyzers, and gas chromatographs. Absorption of the products into solution permits analysis by coulometry, colorimetry, ion selective electrode measurements, or titrimetry. The most important analyzers are Fourier transform infrared (FTIR) spectrometers and, preeminently, mass spectrometers [19]. The two latter methods can be used to record spectra repetitively, thereby producing a time-dependent record of the composition of the gas phase, from which EGA curves can be constructed for selected species.

Applications. Simultaneous TG–MS curves for a brick clay are shown in Figure 13, which simulates a set of firing conditions [20]. After an initial loss of moisture, combustion of organic matter at ca. 300 °C gives characteristic peaks for CO_2, H_2O, and SO_2. More SO_2 results from the oxidation of iron sulfides. Clay dehydroxyla-

tion occurs in the range 350–600 °C, and calcite dissociates to give the CO_2 peak observed at higher temperatures. EGA is of great value in the interpretation of complex TG and DTA curves.

In another example (Fig. 14), the evolution of benzene from two samples of PVC, one containing a smoke retardant, was compared quantitatively by injecting known amounts of benzene into the purge gas stream before and after the experiments, thereby providing a standard for estimating the amount of benzene produced by the polymer [20]. The smoke retardant was found in this case to reduce benzene evolution by a factor of ca. 20.

26.1.6. Mechanical Methods

This section covers a family of techniques leading to dimensional or mechanical data for a material as a function of temperature or time. For a succinct review of the application areas see [22].
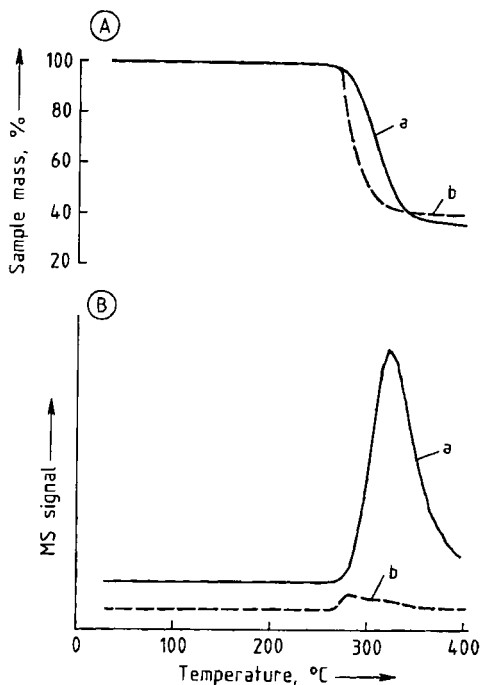
Figure 14. Simultaneous TG – MS curves showing the evolution of benzene (78 amu)
A) TG curve; B) MS curve
a) PVC without a smoke retardant; b) PVC with a smoke retardant (4 % MoO_3)
15 mg samples heated at 10 °C min^{-1} in argon

Figure 15. Modes of deformation subject to examination by TMA
A) Penetration; B) Extension; C) Flexure
a) Sample; b) Probe; c) Clamps; d) Load

Thermodilatometry (TD) measures dimensional changes as a function of temperature in materials subject to negligible loads. A probe, which is held in light contact with the heated sample, is connected to a sensitive position sensor, usually a linear variable differential transformer (LVDT). In addition to providing expansion coefficients the technique can also indicate phase changes, sintering, and chemical reactions. Major application areas include metallurgy and ceramics.

Thermomechanical Analysis. The equipment used in thermomechanical analysis (TMA) is similar in principle to that for TD, but provision is made for applying various types of load to the specimen, so that penetration, extension, and flexure can be measured. This approach to analyzing such modes of deformation is illustrated schematically in Figure 15. The technique finds most use in polymer studies, as in the determination of glass-transition and softening temperatures for thin films and shrinkage characteristics of fibers.

Dynamic Mechanical Analysis. Polymers exhibit both elastic and viscous behavior under stress. In dynamic mechanical analysis (DMA) a sinusoidally modulated stress, often flexural (though measurement is also possible under tensile, compressive, shear, or torsional conditions) is applied to a specimen of material maintained under a specified temperature regime. Displacement transducers measure strain induced in-phase with the stress, as well as strain that lags behind. The former gives a measure of the sample's modulus, or stiffness, and the latter reflects damping characteristics. Like most other thermoanalytical techniques, DMA can be used to generate quantitative physical data, but it is also invoked in a comparative sense: to monitor the effects of additives, for example, or in quality control. It is particularly well suited to the examination of engineering composites, since gelation and curing behavior is sometimes more easily followed by DMA than by alternative techniques such as DSC.

26.1.7. Less Common Techniques

Thermoptometry. This group of techniques includes:

1) *Thermomicroscopy*, in which a material is observed under reflected, transmitted, or polarized transmitted light. The corresponding apparatus is referred to as a *hot stage*, which may also incorporate a DSC sensor for additional information.

2) *Thermophotometry*, in which the intensity of reflected or transmitted light is measured.

3) *Thermoluminescence*, which provides a measure of the intensity of light emitted by the material itself.

Electrical and Magnetic Techniques. Thermomagnetometry is used to study the magnetic properties of materials by revealing apparent weight changes that accompany phase transitions in the presence of an applied magnetic field, or variations in the amount of a magnetic substance.

Thermoelectrometry comprises methods for the measurement of resistance or capacitance during heating. Some equipment in this category provides remote probes for in situ monitoring of the curing of thermosets.

Micro-thermal Analysis (μ-TA) A family of techniques has recently been introduced that enables a range of thermal methods to be applied to regions of a material only a few micrometers wide. They are based on the use of an atomic force microscope (AFM) with a heatable tip. Many materials are heterogeneous, and the microstructure can influence the material properties on the macroscale. After acquiring the conventional AFM topographic image, selected features can be subjected to DTA and TMA with the probe tip. The tip can also be used to pyrolyze the selected area, and the vapors are then analyzed by GC-MS [23].

26.2. Calorimetry

26.2.1. Introduction [24]–[30], [31], [32]

Calorimetry in the broadest sense means the quantitative measurement of energy exchanged in the form of heat during a reaction of any type. By contrast, thermal analysis (Chap. 26.1) is concerned only with the measurement and recording of temperature-induced changes or temperature differences. Since all chemical reactions and many physical changes (e.g., deformation, phase transformations) are associated with the uptake or release of heat, the quantitative investigation of heat exchange is a relatively simple and universal method for characterizing particular processes both in an overall sense and with respect to time.

Nevertheless, only in recent decades has calorimetry emerged from the laboratories of a few

thermodynamicists and specialists to become a widespread, convenient analytical method. The development of commercial calorimeters since the 1950s has led to rapid dissemination and application of the method even beyond the bounds of universities.

The explanation of this phenomenon, as with many other apparatus-based developments, lies in the refinement in measurement techniques made possible by modern electronics. As a rule, heat cannot be measured directly, but must be determined instead on the basis of a temperature change in the system under investigation. Very accurate classical thermometers tend to be very slow measuring devices. Resistance thermometers and thermocouples respond much more rapidly, but they require electronic amplification, and amplifiers with the required precision have become available only in recent decades.

A further explanation for the increasing importance of calorimetry is the development of personal computers, which relieve the user of the often laborious task of evaluating the experimental results, thereby opening the way for the method to become a routine laboratory technique.

In the sections that follow, calorimetry and the present state of instrument development will be described in such a way as to reveal both the possibilities and the limits of the method. The scope of the chapter rules out exhaustive treatment of the subject, so the goal is rather to delineate the method and guide the interested reader to other relevant technical literature.

26.2.2. Methods of Calorimetry

As noted above, heat that is released or consumed by a process cannot be measured directly. Unlike material quantities such as amount of substance, which can be determined with a balance, or the Volume of a liquid, which can be established with a liter measure, the quantity heat must be measured indirectly through its effect on a substance whose temperature it either raises or lowers. The fundamental equation for all of calorimetry describes the relationship between heat exchanged with a calorimeter substance and a corresponding temperature change:

$$\Delta Q = C(T) \cdot \Delta T \tag{1}$$

where ΔQ = exchanged heat (J), ΔT = observed temperature change (K), and $C(T)$ = heat capacity (J/K).

For heats (and therefore temperature changes) that are not too large, the temperature change is directly proportional to the exchanged heat. The proportionality constant is the heat capacity of the calorimeter substance (previously referred to as the "water value"). However, if the temperature change exceeds a few Kelvin, the temperature dependence of heat capacity stands in the way of a linear relationship, and a knowledge of the temperature function of the particular heat capacity in question is required in order to determine heat on the basis of a measured temperature difference.

The relationship above leads directly to the standard calorimetric methods for determining the heat of a process: either temperature is held constant by appropriate compensation for the heat effect, and the required compensation power is measured, or a temperature change is determined and used to calculate a corresponding value for exchanged heat. A precondition for the latter approach is accurate knowledge of the heat capacity as well as the heat transport properties of the measurement system.

26.2.2.1. Compensation for Thermal Effects

In the first method of heat measurement, temperature changes in the calorimeter substance are avoided by supplying or dissipating heat exactly equal in magnitude (but opposite in sign) to that associated with the process under investigation. As a rule, electrical energy is used to provide this compensation, either by introduction in the form of Joulean heat or dissipation through the Peltier effect. A combination approach is also possible, with an appropriate constant level of cooling and simultaneous controlled electrical heating. The requisite amount of compensation power can today be determined with a high degree of precision. It was once customary to invoke for compensation purposes "latent" heat derived from a phase change; in other words, heat released in a process was measured by using it to melt a substance such as ice, with subsequent weighing of the resulting water and conversion of the data into energy units based on the known heat of fusion for ice.

The method of compensation is advantageous, since it permits measurements to be carried out under quasi-isothermal conditions, thereby avoiding heat loss from the calorimeter to the surroundings by heat transport processes. Furthermore,

there is no need for a calibrated thermometer, only a sensor sufficiently sensitive to temperature changes to provide adequate control over the compensation power.

26.2.2.2. Measurement of a Temperature Difference

In this alternative method of heat measurement, which is also indirect, a measured temperature difference is used to calculate the amount of heat exchanged. A distinction can be made between temporal and spatial temperature-difference measurements. In *temporal* temperature-difference measurement the temperature of a calorimeter substance is measured before and after a process, and a corresponding heat is calculated on the basis of Equation (1) (which presupposes an accurate knowledge of the heat capacity). In the *spatial* method a temperature difference between two points within the calorimeter (or between the calorimeter substance and the surroundings) is the quantity of interest. The basis for interpretation in this case is from Fourier's law the equation for stationary heat conduction (Newton's law of cooling):

$$\Phi = \lambda(T) \cdot A \cdot \Delta T \cdot l^{-1} \tag{2}$$

where Φ = heat flow rate (W), $\lambda(T)$ = thermal conductivity (W m^{-1} K^{-1}), A = cross-sectional area (m^2), ΔT = temperature difference (K), and l = length (m).

From this expression it follows that the heat flow rate through a heat-conducting material is proportional to the corresponding temperature difference. If this temperature difference is recorded as a function of time, then for a given thermal conductivity one acquires a measure of the corresponding heat flow rate, which can be integrated to give a total heat. The technique itself is very simple, but the result will be correct only if the measured temperature difference accurately reflects the total heat flow rate and no heat is lost through undetected "heat leaks". Careful calibration and critical tests of the calorimeter are absolutely necessary if one hopes to obtain reliable results.

A fundamental disadvantage of the temperature-difference method is the fact that nonisothermal conditions prevail, so the precondition of a stationary state is fulfilled only approximately. The potential therefore exists for nonlinear phenomena in the heat-transport process, which *might*

result in conditions outside the range of validity for Equation (2). Furthermore, with nonisothermal studies it is impossible to rule out completely the presence of undetected leaks. It must also be pointed out that in the case of nonisothermal studies the thermodynamic quantities of interest cannot be established according to the methods of reversible thermodynamics, at least not without further consideration. Strictly speaking, *irreversible thermodynamics* should be applied to nonisothermal calorimetry. Although the distinction is generally unimportant in normal calorimetric practice, it becomes essential in cases where time-dependent processes are involved with timescales which are comparable to those of the experiment, e.g., in the *temperature-modulated* mode of operation.

26.2.2.3. Temperature Modulation

The temperature-modulated mode of operation has been well known for many decades in calorimetry [33], but became well established only during the 1990s, when commercial DSC was modified this way [34]. The idea is to examine the behavior of the sample for periodic rather than for isothermal or constant-heating-rate temperature changes. In this way it is possible to obtain information on time-dependent processes within the sample that result in a time-dependent generalized (excess) heat capacity function or, equivalently, in a complex frequency-dependent quantity. Similar complex quantities (electric susceptibility, Young's modulus) are known from other dynamic (dielectric or mechanical) measurement methods. They are widely used to investigate, say, relaxation processes of the material.

In calorimetry there are three established temperature modulation methods:

1) The sample is heated with a periodically changing heat flow from a electrical heater [35] or chopped light beam [36], and the temperature change of the sample (magnitude and phase shift) is measured as the response signal (*AC calorimetry*).

2) A thermal wave is sent into the sample from a metallic film that is evaporated onto its surface and which heats it periodically. The temperature change is measured by using the same metal film as a resistance thermometer. The third harmonic of the resulting temperature oscillations is proportional to the power input and depends on the product of the heat capacity and

thermal conductivity of the sample in a characteristic way (*3ω method* [37]).

3) The sample (or furnace) temperature is controlled to follow a set course with superimposed periodical changes, and the heat flow rate is measured via the differential temperature between sample and reference (*temperature-modulated differential scanning calorimetry, TMDSC* [38]).

In all three cases, both the magnitude of the periodic part of the response signal and its phase shift with respect to the stimulating signal are measured, and this results in a complex quantity (temperature, heat flow rate, and heat capacity, respectively). As heat transport always requires time, the frequency of temperature modulation is normally limited to a maximum of 0.2 Hz to allow the heat to flow through the sample properly. For the 3ω method, however, frequencies of several kilohertz are possible because the heat source and the temperature probe are identical in this case. At the lower end, the frequency range is only limited by the noise threshold (sensitivity) of the sensors in question. As a rule only two decades are available. Consequently, only time-dependent processes with timescales within this window can be followed in the calorimeter.

For quasistatic conditions it follows from Equation (1) that:

$$\frac{dQ}{dt} = \Phi = C_p \cdot \frac{dT}{dt} \tag{3}$$

where the real quantities heat flow rate Φ and temperature change (heating rate) are strictly proportional, with the real heat capacity C_p as proportionality factor. This implies an infinite propagation velocity of thermal waves and an infinitely fast excitation time of the vibrational states that characterize the (static) heat capacity of a sample. Both conditions are fulfilled for normal solids within the above-mentioned frequency range. In cases where processes are involved which need time, the heat flow rate and the temperature change are no longer proportional, and the process in question causes the heat flow to lag behind the temperature. This results formally in a time-dependent apparent heat capacity, and Equation (3) no longer holds; instead:

$$\Phi(t) = C_p(t) * \frac{dT(t)}{dt} \tag{4}$$

This appears very similar, but the operator * stands for the convolution product, which is an integral operator rather than a product [39].

Equation (4) can be Fourier transformed and then reads:

$$\Phi(\omega) = C_p(\omega) \cdot \text{Fourier}\left(\frac{dT(t)}{dt}\right) \tag{5}$$

The (complex) heat flow rate is given as a normal product of the complex heat capacity and the (complex) Fourier-transformed heating rate in the frequency domain. For a given $T(t)$ course (normally periodical with underlying constant heating rate), the complex heat capacity can be determined from the measured complex (magnitude and phase) heat flow rate function. The resulting C_p function is given either as real and imaginary part or as magnitude (absolute value) and phase. The two forms are mathematically equivalent. From the frequency dependence of these quantities the $C_p(t)$ behavior can be derived.

26.2.3. Calorimeters

There are many diverse types of calorimeters, differing with respect to measuring principle, mode of operation, and general construction. Measuring principles have been described in the preceding sections. In terms of mode of operation, two approaches can be distinguished: static calorimeters, in which the temperature remains constant or changes only in consequence of the heat of reaction, and dynamic calorimeters, in which the temperature of the calorimeter substance is varied deliberately according to a prescribed program (usually linear with time).

With regard to construction, a distinction can be made between single and differential calorimeters. Twin design is characteristic of the latter: that is, all components are present in duplicate, and the two sets are arranged as identical as possible. The reaction under investigation takes place in one side of the device, while the other side contains an inert reference substance. The output signal represents the difference between signals originating in the sample and reference side. With such a difference signal all symmetrical effects cancel (e.g., heat leaks to the surroundings).

26.2.3.1. Static Calorimeters

Considering first the static calorimeters, three types can be distinguished depending upon whether they are operated in an *isothermal, isoperibolic* (isoperibolic = uniform surroundings), or adiabatic manner.

26.2.3.1.1. Isothermal Calorimeters

In the *isothermal* mode of operation it is imperative that all thermal effects be somehow compensated. This is achieved either electrically or with the aid of a phase transition for some substance. Only phase-transition calorimeters can be regarded as strictly isothermal. In this case thermodynamics ensures that the temperature will remain precisely constant since it is controlled by a two-phase equilibrium of a pure substance. The most familiar example is the *ice calorimeter*, already in use by the end of the 18th century and developed further into a precision instrument about 100 years later by BUNSEN (Fig. 16). The liquid – gas phase transition has also been used for thermal compensation purposes; in this case a heat of reaction can be determined accurately by measuring the volume of a vaporized gas.

Phase-transformation calorimeters are easy to construct, and for this reason they are not available commercially. Such a device permits very precise determination of the heat released during a process. An important drawback is the fact that the isothermal method is limited to the few fixed temperatures corresponding to phase transitons of suitable pure substances.

This disadvantage does not apply to an isothermal device based on electrical compensation. Nevertheless, calorimeters of the latter type operate only in a quasi-isothermal mode, since electronic control systems depend for their response upon small deviations from an established set point, and a certain amount of time is required for changing the prevailing temperature. The use of modern circuits and components ensures that errors from this source will be negligible, however. Electrical compensation makes it possible to follow both endothermic and exothermic processes. In both cases the compensation power is readily measured and recorded or processed further with a computer. Isothermal calorimeters are used quite generally for determining heats of mixing and solution. Commercial devices are available that also support the precise work required for multiphase thermo-
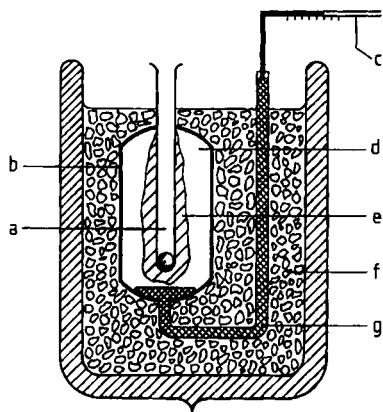
Figure 16. Ice calorimeter (based on [40])
a) Sample holder; b) Calorimeter vessel; c) Capillary with mercury thread; d) Water; e) Ice; f) Ice – water mixture; g) Mercury

dynamics. As a rule such calorimeters are of the single rather than the differential type.

26.2.3.1.2. Isoperibolic Calorimeters

An *isoperibolic calorimeter* is distinguished by the fact that it is situated within a thermostated environment. As a result, the temperature of the surroundings is kept constant, as is the effect of the surroundings on the calorimeter vessel. A well-defined thermal conduction path is maintained between the calorimeter vessel and its surroundings so that the temperature difference between the calorimeter vessel and the environment will be proportional to the heat flow rate. Temperature equilibration usually adheres to an exponential time function, and an integral taken over the entire course of this function is proportional to the heat of reaction. Such instruments often have exceptionally large time constants; that is, temperature equilibration occurs very slowly. Indeed, the greater the equilibration time, the more sensitive is the calorimeter. It is of course difficult with an exponential function to identify precisely the end point, so the integral over the measurement function and therefore the heat of interest are subject to corresponding degrees of uncertainty. The advantages lie in simplicity of construction and ease of producing a measurement signal (a straightforward temperature difference), which might for example be derived from a differential thermocouple and a simple set of amplifiers.

The best-known calorimeter of this type was developed by TIAN and CALVET (Fig. 17). Here the defined heat-conduction path to the thermostated surroundings (a large aluminum block) consists of a large number of differential thermocouples coupled in series (thermopile). This arrangement permits optimum determination of the heat flow rate to the surroundings, and such an instrument can be very sensitive (*microcalorimeter*).

If heat transport to the surroundings is restricted by suitable insulation, temperature equilibration becomes so slow that — with small corrections — temperature changes in the calorimeter substance alone can serve as the basis for determining the heat of a reaction. Calorimeters in this category include the classical *mixing calorimeter* (with heat exchange) in which the sample substance is brought into contact with a (usually) colder calorimeter substance, whereupon an intermediate "mixing temperature" is established (Fig. 18). Other examples of mixing calorimeters include the simpler types of *combustion calorimeters*, in which the calorimeter substance is water, and the *drop calorimeter* (see Fig. 24). The latter consists of a metal block whose temperature change subsequent to introduction of a sample of accurately known temperature provides information about the sample's heat capacity. The category of isoperibolic calorimeters also includes *flow calorimeters*, in which two fluid media (usually liquids containing the reactants) are first brought to the temperature of the thermostated surroundings and then combined. The ensuing reaction causes a measurable temperature change, the extent of which is proportional to the heat of the reaction (Fig. 19).

26.2.3.1.3. Adiabatic Calorimeters

An *adiabatic calorimeter* is equipped with an electronic control system designed to ensure to the greatest extent possible that the surroundings of the calorimeter cell remain at all times at precisely the same temperature as the substance under examination. The object is to prevent virtually all heat transport to the surroundings, thereby making it possible to establish the total heat effect from the change in temperature of the calorimeter substance. Good adiabatic calorimeters are difficult to construct, and they are not available commercially. Nevertheless, precision instruments of this type permit quantities such as specific heat capacities, as well as associated anomalies, to be meas-
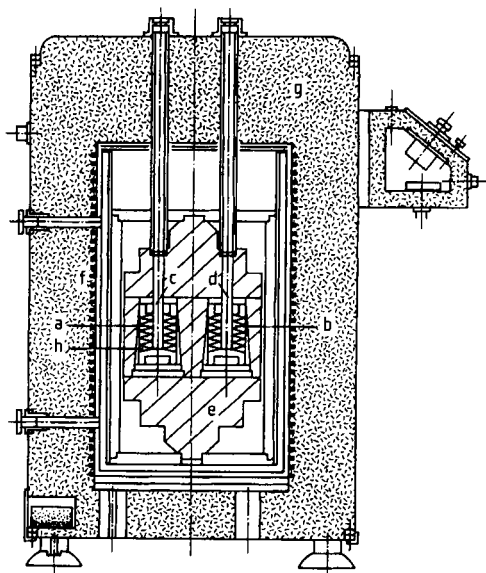
Figure 17. Tian – Calvet calorimeter
a, b) Heat conduction paths between measuring cells and block; c, d) Sample and reference containers; e) Isoperibolic block; f) Thermostatic jacket; g) Thermal insulation; h) Thermopiles

Figure 18. A classical mixing calorimeter (from Meyers Conversationslexicon, 1893)
a) Calorimeter vessel; b) Convection and radiation shield; c) Cover; d) Jacketing vessel (thermostated); e) Lifting mixer

ured very accurately (Fig. 20). The method entails an extraordinary level of effort, and it is practiced only by specialists working in a few laboratories in the world.

Simpler *adiabatic calorimeters* are distributed commercially in the form of adiabatic mixing and combustion calorimeters. These constitute very satisfactory routine instruments, especially since the adiabatic mode of operation (which entails measurement of only a single temperature difference) facilitates automated data acquisition and interpretation.

26.2.3.2. Scanning Calorimeters

Scanning calorimeters are devices in which the sample temperature is deliberately changed during the course of an experiment according to a prescribed program. As a rule, heating or cooling of the sample is caused to occur linearly with time, and the heat flow rate required to accomplish the desired change is determined and output as a function of time or temperature. Any chemical or physical change in the sample at a particular temperature will manifest itself as a change in the heat flow rate, leading to a "peak" in the measurement

curve whose area is proportional to the heat of the process (see also Section 26.1.3.2).

Such instruments have been on the market since the 1950s. Their great advantage is that they permit heats of reaction and transformations of every type to be measured both rapidly and with reasonable precision. Heating rates usually fall in the range 5 – 20 K/min, so a measurement extending over 300 K can be completed within an hour, providing a reliable indication of all changes in the sample over this temperature range, including heats of reaction. The rapid development of calorimetry in recent years is due largely to the perfection of this type of calorimeter.

Scanning calorimeters are designed almost without exception as differential devices. With regard to the principle of measurement, two varieties can be distinguished: *differential-temperature* and *power-compensated calorimeters*. Although the two types rely upon very different approaches to determining heat flow rates with respect to the sample, and are therefore subject to different sources of error, they are marketed under the single heading *differential scanning calorimeter* (DSC). They can even be operated in *temperature-modulated* mode.
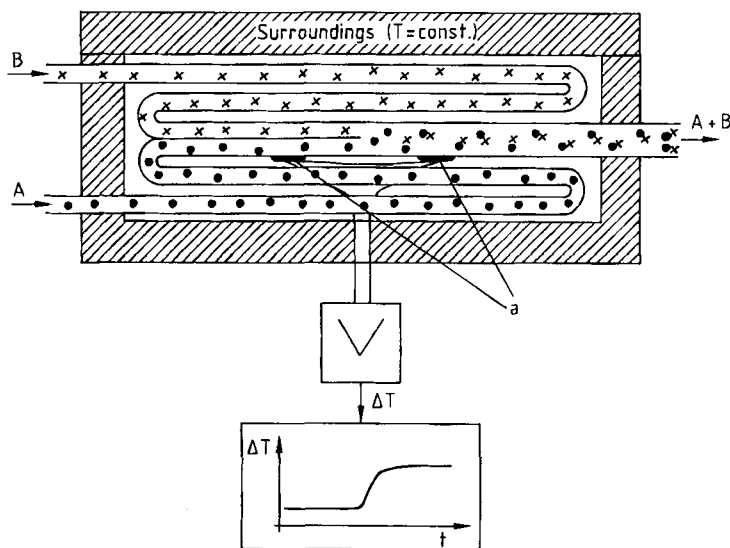
Figure 19. Flow calorimeter, with a chemical reaction occurring in the liquid
A, B reactants, A + B reaction product; a) Differential thermocouple

26.2.3.2.1. Differential-Temperature Scanning Calorimeters

In a differential-temperature scanning calorimeter the heat necessary for raising the sample temperature reaches the sample position via one well-defined heat conduction path (cf. Fig. 21). The measured signal is derived from a temperature difference generated along this path, which is proportional to the heat flow rate [see Eq. (2)]. This measuring and operating principle leads to the alternative designation *heat-flux calorimeter* for such a device. In order to minimize errors from heat leaks and other influences of the surroundings, calorimeters of this type are always based on a differential design, with full duplex construction. Many commercial models are available.

Any temperature difference between the sample and the reference is directly proportional to the corresponding differential heat flow rate, and so this is the quantity to be measured. The appropriate functional relationship follows directly from Equation (2):

$$\Phi = K(T) \cdot \Delta T \tag{6}$$

The heat flow rate of interest, Φ (in W), is proportional to the measured temperature difference ΔT (in K). Unfortunately, the proportionality con-

stant K (the calibration factor for the calorimeter in $K\,W^{-1}$) is itself a function of the thermal conductivity of the sample [see Eq. (2)], and therefore temperature dependent. Moreover, temperature differences between the sample and the reference result in variable amounts of heat loss to the surroundings via inevitable heat leaks. Heat exchange with the environment also occurs via radiation and convection. Since the two transport processes are linked to temperature differences in a nonlinear way, the calibration factor in the presumably linear calorimeter equation necessarily depends on ΔT, which is the quantity to be measured. Consequently, the calibration factor depends upon such sample parameters as mass, thermal conductivity, and specific enthalpy difference [31], so heats measured with these instruments must realistically be assigned an uncertainty of about five percent unless considerable effort has been expended in calibration [42]. Repeatability of the results is generally many times better, but the user should not make the mistake of assuming an equally high level of certainty. Another source of systematic error may develop during the measurement if some reaction causes a change in the heat capacity and/or the extent of heat transfer to the sample holder. Both phenomena contribute to changes in the baseline, with the result that the course of the latter cannot be plotted accurately
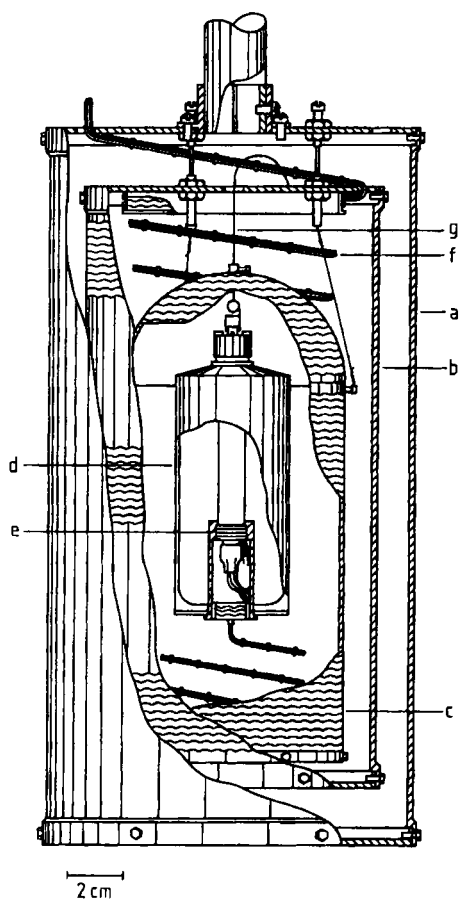
Figure 20. Construction of an adiabatic calorimeter (based on [41])
a) External radiation shield; b) Internal radiation shield; c) Adiabatic jacket; d) Calorimeter vessel; e) Sample suspension platform with resistance thermometer and heater; f) Electrical leads; g) Calorimeter suspension system

Figure 21. Operating principle of a temperature-difference scanning calorimeter (heat-flux calorimeter)
T_F furnace temperature, T_S sample temperature, T_R reference temperature, T_0 initial temperature, ΔT differential temperature $T_S - T_R$, β heating rate, t time
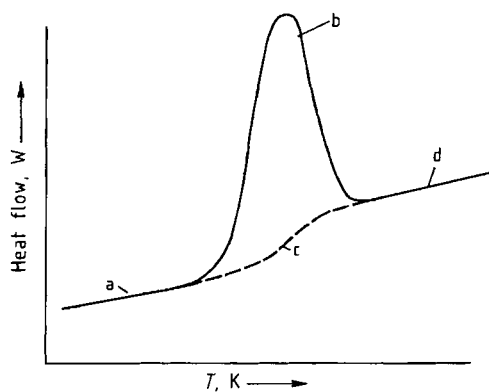
Figure 22. Output curve from a scanning calorimeter showing a shift in the baseline caused by variability in either c_p or the coefficient of heat transfer
a) Original baseline; b) Peak; c) Unknown baseline path; d) Final baseline

(Fig. 22). This in turn means that the area corresponding to the heat of reaction also cannot be determined exactly, leading to uncertainty in the associated heat [31]. Despite these restrictive comments, differential-temperature scanning calorimeters offer the advantages of a simple, transparent mode of operation and relatively low cost, and for this reason they account for a large share of the market. With appropriately critical supervision they are capable of fulfilling their assignment admirably.

26.2.3.2.2. Power-Compensated Scanning Calorimeters

Like differential-temperature devices, *power-compensated scanning calorimeters* always feature duplex construction, with sample and reference holders heated in such a way that at every
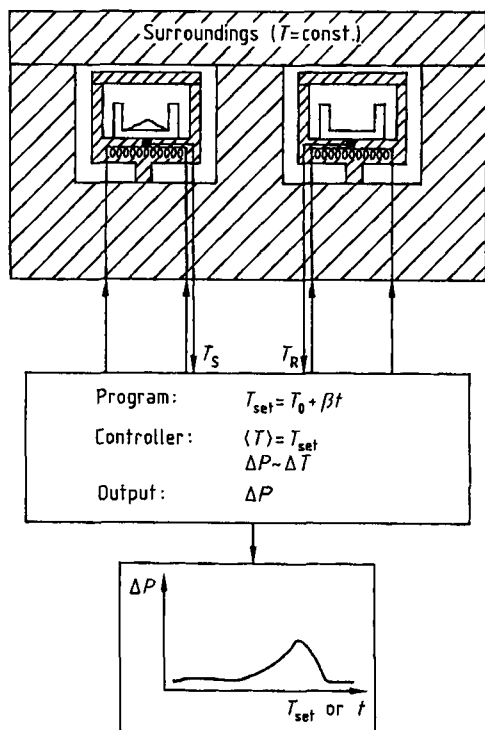
Figure 23. Operating principle of a power-compensated differential scanning calorimeter (Perkin–Elmer)
T_0 initial temperature; T_S sample temperature; T_R reference temperature; $\langle T \rangle$ mean temperature between sample and reference; ΔP differential power $P_S - P_R$; β heating rate; t time

instant the temperature of each corresponds almost exactly to a programmed set temperature. Any temperature differences that arise between sample and reference by reason of thermal events in the sample are compensated immediately by appropriate changes in the electrical heating (Fig. 23). The output signal is proportional to the instantaneous differential heat flow rate. Since temperature differences between sample and reference are essentially compensated at once, measurement errors due to undetected heat leaks can be largely avoided. Instruments of this type, the most widely distributed of which are manufactured by Perkin–Elmer, require a very complicated electronic control system, and are therefore rather expensive. The principal disadvantage lies in the fact that the electronic system is a black box for the average layman, thereby obscuring possible sources of error. These instruments also require that there be some deviation, albeit minimal, from the set temperature to guide the control circuit, so just as in the preceding case some systematic error must be anticipated (although certainly to a lesser degree), associated once again with temperature differences between sample and reference and nonlinear heat exchange. Accordingly, the level of systematic uncertainty for heats measured by power-compensated DSC can be estimated at $1-3\%$.

26.2.3.2.3. Temperature-Modulated Scanning Calorimeters

Differential-temperature and *power-compensated scanning calorimeters* are now commercially available to run in the temperature modulated mode (TMDSC). In both cases the temperature is controlled to follow the normal (linear in time) course but with an additional periodical temperature change of particular amplitude and frequency. The measured heat flow rate function is the sum of the "underlying" part and the periodical part. The former can be extracted by calculating the "gliding average" integral over a period along the measured curve. It is identical to the curve that would be obtained in this DSC when the temperature modulation is switched off. The periodical part, on the other hand, is the difference of the measured and the underlying curves. From the periodic part the complex C_p can be calculated by using Fourier analysis or other mathematical techniques. This enables time dependent-processes to be investigated in cases where their timescale falls within the frequency window of the TMDSC method.

For precise measurements it is necessary to correct (calibrate) the measured heat flow rate magnitude and phase shift for influences from the apparatus, as well as from the heat transport, which also is a time-dependent process and therefore influences the measured heat flow rate function [43].

26.2.4. Applications of Calorimetry

Modern calorimeters permit relatively rapid and truly precise measurement of heat exchanges in a wide variety of reactions. Since heat evolution is proportional to the conversion (extent of reaction) in a chemical, physical, or biological reaction, calorimetric measurement constitutes one method for quantitative evaluation of the reaction itself. Measurement is possible not only of the total heat (and therefore the total conversion) of a reaction but also of the course of the reaction

26.2.4.1. Determination of Thermodynamic Functions

Various *thermodynamic potential functions* (enthalpy, entropy, free energy, etc.) can be determined by integration on the basis of *specific heat capacities*. The accurate establishment of specific heat capacities at constant pressure, c_p, as a function of temperature is thus of fundamental importance. Different types of calorimeters are used for determinations of c_p depending upon accuracy requirements and the temperature range in question.

The most accurate results are obtained with adiabatic calorimeters, which are useful over a temperature range from fractions of a Kelvin to about 1000 K. These instruments permit specific heat capacities to be determined with very small uncertainties ($< 1\%$). Nevertheless, such measurements are so complex that they are rarely applicable to normal laboratory practice in industry, especially given the lack of commercial instruments.

The standard approach to c_p measurement therefore relies on the less precise but more easily constructed and operated mixing calorimeters, in particular drop calorimeters (Fig. 24). A (mean) specific heat capacity is calculated in this case from the heat released when a sample of known temperature is introduced (dropped) into an isoperibolic calorimeter. Charging temperatures may exceed 2500 K. Determining the temperature dependence of c_p requires many individual measurements at various charging temperatures. The uncertainty in such measurements is usually 3 – 5 %.

A truly elegant method for determining c_p at < 1000 K takes advantage of a scanning calorimeter [31], preferably a power-compensated DSC. The procedure involves raising the temperature of a sample in the calorimeter at a defined rate (typically 10 K/min). The same measurement is subsequently repeated, this time with empty crucibles. The difference between the two measurement curves provides a record of the heat flow rate into the sample, from which the specific heat capacity can be calculated. This method makes it possible conveniently to measure the $c_p(T)$ of a material over a temperature range of several hundred Kelvins within the course of a single day. However, it does demand a very high level of repeatability (baseline stability) for the DSC in question. In the case of differential-temperature (heat-flux) DSC — even with careful calibration — uncertainties of 5 – 10 % in the re-
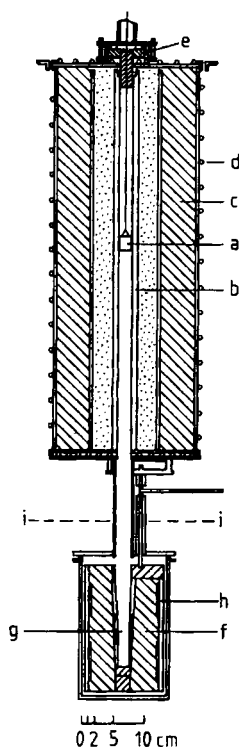
Figure 24. High-temperature drop calorimeter (based on [46])
a) Sample; b) Heated corundum tube; c) Magnesia insulation; d) Cooling coils; e) Sample mounting device; f) Copper block; g) Hole for receiving sample; h) Resistance thermometer; i) Level of the thermostated oil bath

with respect to time. This opens the way to studies of kinetics, problems related to stability and safety, and other areas in which temperature and time-dependencies in a reaction play a part.

The sections that follow describe several possible applications of calorimetry, although these should be regarded only as illustrative, since the scope of the present article rules out any treatment that is even approximately exhaustive. No such limits apply to the reader's imagination, however, and any process that generates (or consumes) heat can in principle be followed with a calorimeter. Most recently, enzymatic and bacterial manufacturing processes have assumed increasing importance in industry, and interest in biocalorimetry is growing accordingly [44], [45].

sults must be anticipated. With power-compensated DSC the systematic uncertainties tend to be in the range of 3 – 5 %. Uncertainties in the calculated thermodynamic potential functions of interest are comparable.

The TMDSC enables another elegant possibility of c_p (magnitude) determination from the amplitude of the modulated part of the measured heat flow rate function both in the isothermal and scanning modes of operation. This method is especially advantageous in cases of noisy signals with low sample masses or low heating rates. Precise calibration of the heat flow rate amplitude is a prerequisite for obtaining reliable results [43].

However, a calorimeter operating dynamically always indicates a temperature that deviates to a greater or lesser extent from the true sample temperature, resulting in further uncertainty. It must in any case be borne in mind that potential functions determined on the basis of thermodynamic equilibrium, and thus defined statically, may differ from functions determined dynamically, since equilibrium conditions are not maintained in DSC, and in the case of a reaction even stationary conditions are lacking.

If time-dependent processes are involved, then the process is clearly outside the scope of classical thermodynamics. The time-dependent (apparent) heat capacity, measured with, say, TMDSC would lead to time-dependent potential functions which must be interpreted in terms of *irreversible thermodynamics*. In such cases, a nonzero imaginary part of the complex heat capacity exists which is linked to the entropy production of the process in question (for details see [32]). Thus, temperature-modulated calorimetry makes it possible to determine time-dependent (irreversible) thermodynamic quantities.

26.2.4.2. Determination of Heats of Mixing

For liquid systems consisting of several components, multiphase thermodynamics provides the theoretical background for computing phase diagrams from heats of mixing for the relevant components. Phase diagrams in turn are interesting not only to thermodynamicists; they are also of great importance in chemical process engineering practice.

The determination of heats of mixing is a classical field of investigation for calorimetry, one that at the required level of precision is today incomparably less laborious than previously. Enthalpies of mixing were formerly determined in some cases indirectly via vapor pressure measurements, but today they are measured directly in (quasi)isothermal or isoperibolic calorimeters.

The art of conducting such an experiment successfully lies in rigorously excluding vaporization during the measurement, since specific heats of vaporization are orders of magnitude larger than the effects of mixing that are to be measured. As excess quantities (which characterize deviations from ideal behavior), heats of mixing are invariably small, and they must be determined with high precision in order to permit the construction of sufficiently precise phase diagrams. Isothermal calorimeters are therefore definitely to be preferred, since here the problem of incomplete determination of a heat of mixing — owing perhaps to undetected heat leaks — is less severe than with an isoperibolic instrument. Moreover, the isothermal heat of mixing is defined more precisely, and it lends itself more readily to mathematical treatment than in the case of parallel temperature changes.

Phase transitions in solid mixtures can be followed directly in a scanning calorimeter, since vapor pressures associated with the components are here so low that vaporization plays virtually no role and therefore does not interfere. Corresponding phase diagrams (i.e., liquidus–solidus curves) can be inferred from heat flow rate curves obtained by controlled heating and cooling of appropriate mixtures (cf. Section 26.1.3.3). Thermodynamically relevant quantities can in turn be determined indirectly by comparing measured phase diagrams with calculated diagrams.

26.2.4.3. Combustion Calorimetry

The determination of heats of combustion for fuels of all types is prescribed in detail by government regulation, since the "calorific value" or "heat content" of a particular coal, oil, or natural or manufactured gas determines its economic value. Combustion calorimeters or "bomb calorimeters" (usually operated adiabatically) are used for such measurements, whereby the sample is combusted in a "calorimetric bomb" (likewise standardized; cf. Fig. 25) in pure oxygen at $3 \cdot 10^6$ Pa to a defined set of end products. Methods of measurement and interpretation are also standardized, and they have been simplified to such an extent that the entire procedure can be automated.

The same instruments are useful for determining heats of combustion for a wide variety of
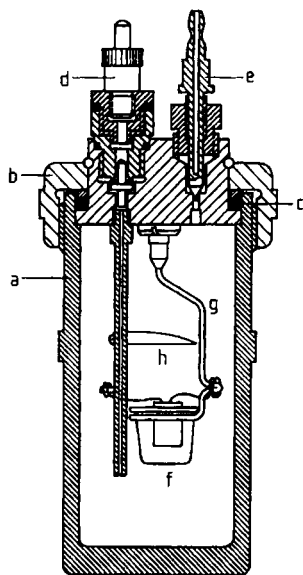
Figure 25. Example of a calorimetric bomb for use in a combustion calorimeter
a) Pressure vessel; b) Screw cap; c) Sealing ring; d) Inlet valve; e) Outlet valve; f) Sample crucible with sample; g) Crucible holder; h) Heat shield

(usually organic) substances, from which standard enthalpies of formation can be determined. For these calorimetric methods errors in measurement are very small ($< 1\%_0$), with a level of precision that is also appropriate for analytical purposes (e.g., purity or composition checks on batches of chemicals). Combustion calorimetry is useful in biology as well. For example, in determinations of the nutrient content of leaves and fruits as a function of season and place based on appropriate heats of combustion.

Many combustion calorimeters are available commercially. Instruments utilizing water as the calorimeter substance generally operate adiabatically, thereby avoiding the somewhat complicated Regnault – Pfaundler procedure (cf. [29]) for determining and interpreting prior and subsequent temperature patterns. "Dry" combustion calorimeters that avoid the use of water or other liquids have also recently come onto the market.

26.2.4.4. Reaction Calorimetry

Every chemical process is associated with some type of heat effect. Accurate knowledge of heat released or consumed is of great importance in process engineering. Moreover, heat evolution

is strictly coupled with the particular course of a reaction. If one is able to determine unambiguously the time dependence of the heat flow rate associated with a particular sample, one also gains access to the corresponding reaction rate law, which in turn opens the way to the reaction kinetics [31]. Knowledge of the appropriate kinetic parameters permits one to predict the course of the reaction under other sets of conditions, which is of great practical value. Heat of reaction determinations as well as kinetic studies are best carried out under isothermal (and isobaric) conditions, and for this reason it is advisable to select a (quasi)isothermal calorimeter with compensation for effects of the measurement process. One practical difficulty lies in the fact that reaction begins immediately after introduction of the sample at the appropriate temperature, even though the calorimeter itself is not yet at thermal equilibrium (and therefore not ready for use). One way out of this dilemma is to devise a system for bringing the components of the reaction together only after isothermicity has been achieved. With liquids the challenge is not particularly great, since pumps and stirrers can ensure rapid (and thorough) mixing of the reactants at any desired time. The piercing of a membrane (or an ampoule) is also a common expedient. Another possibility is the use of a specially designed reaction calorimeter. With one device of this type the first reactant is present from the outset in a suitable glass or metallic vessel, to which a second reactant is added with mixing after stationary conditions have been established (Fig. 26). With the aid of electronic temperature control the reaction can then be carried out in any way desired: isothermally, adiabatically, or under the influence of some temperature program. Measurement of the relevant quantities, process control, and thermochemical data interpretation are all accomplished with the aid of a personal computer. A second type of special reaction calorimeter, suitable only for liquid components, is so constructed that each reactant is pumped in through its own system of tubing, with the streams uniting after temperature equilibration (Fig. 27). The ensuing reaction causes a temperature change, from which the corresponding heat of reaction can be calculated.

Transport and mixing of liquid or gaseous components is accomplished rather easily, but the situation is more complex for solid substances or mixtures. Here there is little hope of satisfactorily mixing the components inside a normal calorimeter, so one is forced to resort to a scanning
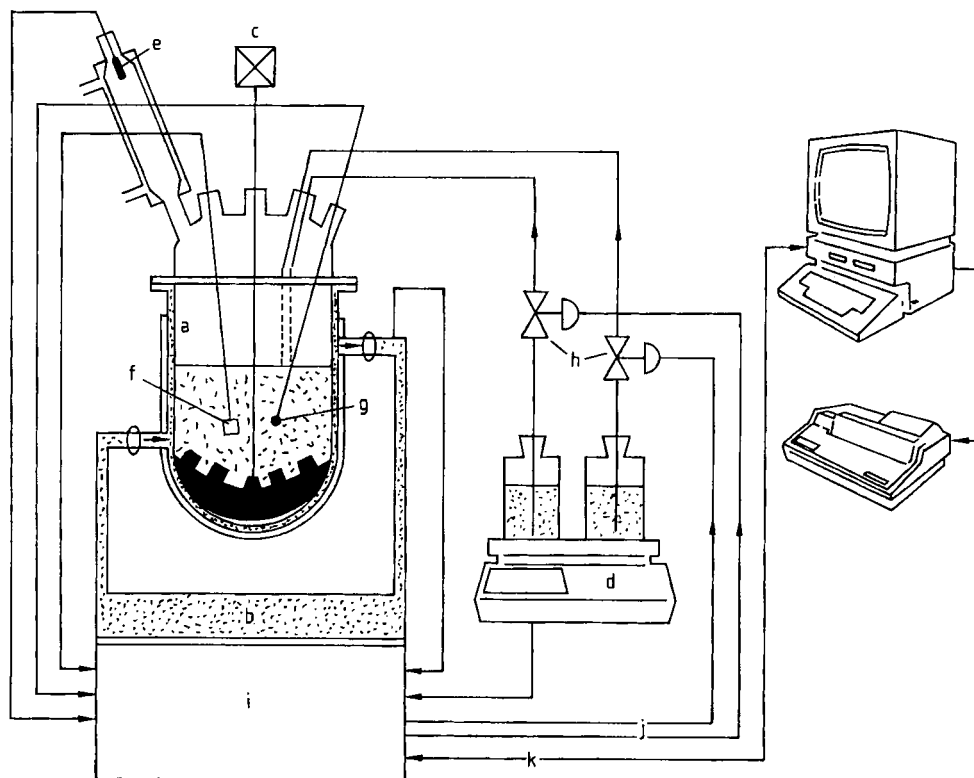
Figure 26. Operating principle of a reaction calorimeter (Contraves)
a) Reaction vessel; b) Heating and cooling bath; c) Stirrer; d) Balance; e) Pressure sensor for internal pressure; f) pH probe;
g) Temperature sensor; h) Dosing valve; i) Control unit; j) Control and monitoring signals; k) Measured signal
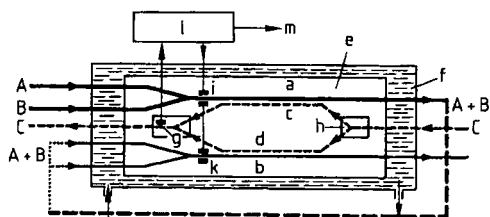
Figure 27. Schematic diagram of a flow calorimeter after
P. PICKER (Setaram)
A, B) Reactants; A + B) Reaction mixture; C) Heat-exchange
liquid
a) Reaction tube (sample tube); b) Reference tube;
c, d) Heat-exchange tubes; e) Vacuum jacket; f) Thermo-
stated liquid; g) Thermistor; h) Flow divider; i, k) Heating
elements; l) Measurement and control electronics; m) Output
signal

calorimeter. The calorimeter is charged at a tem-
perature below that at which the substances react,
after which the temperature is increased linearly
with time until reaction is complete. From a heat
flow rate – time function acquired at a well-de-
fined heating rate it is possible to establish both
the heat of reaction and, if need be, kinetic param-
eters for the process. One prerequisite, however, is
that the recorded function reproduces in an unbi-
ased way the true heat production of the sample.
This presents serious problems, since heat must
inevitably flow from the point at which it is gen-
erated to the site of the temperature sensors before
any measurement signal can appear. In the case of
a fairly rapid chemical process, where the rate
constant is comparable to the time constant of
the calorimeter, the measurement signal from the
recording device must be treated mathematically
("desmeared") [31] before it can be utilized for a
kinetic analysis. Owing to the smaller time con-

stant and more precise linearity of the corresponding temperature – time program, a scanning calorimeter with power compensation is preferred over a temperature-difference DSC.

26.2.4.5. Safety Studies

Calorimetric methods applicable to safety studies are also concerned with clarifying the kinetic behavior of substances, but in this case from a safety engineering viewpoint. The important role assigned recently to such studies justifies their being treated as a separate topic.

The vast majority of chemical substances are endothermic materials; that is, they release heat when they decompose, and their decomposition is accelerated by heat. Chemical reaction kinetics teaches that any reaction is capable of proceeding at temperatures considerably lower than the temperature at which the maximum reaction rate is achieved. Usually, however, the reaction rate under these conditions is so low that the process is not apparent. The more unstable a substance, the lower is its decomposition temperature, and therefore the greater its decomposition rate at room temperature. In the event that heat evolved during storage — attributable to traces of reaction occurring under storage conditions — exceeds the level at which heat is emitted to the surroundings, then the decomposition rate may increase very rapidly, possibly to the point of an explosion.

Thorough knowledge of decomposition kinetics is therefore essential for ensuring safe storage and handling. The requisite information can, in principle, be gained through the methods described in Section 26.2.4.4, but it is often impossible to avoid an uncontrolled outcome and simultaneous destruction of the measuring system. This obviously rules out the use of expensive apparatus, which has led to the development of special safety calorimeters that are either inexpensive and disposable or designed to withstand the effects of violent reactions.

Simple arrangements of this type clearly preclude precision measurements, but they are entirely adequate from the specific viewpoint of assessing storage safety at particular temperatures or with particular types of packaging. The simplest example is an exceedingly primitive isoperibolic calorimeter consisting of a glass ampoule, either open or fitted with a pressure-tight closure, mounted together with a thermocouple inside a small furnace — a device that anyone can easily assemble and calibrate. More ambitious designs

differ from standard calorimeters in that the sample holder (or measuring head) is incorporated into an autoclave of sufficient stability to permit studying the thermal behavior even of explosives. Included in this category are the so-called accelerating rate calorimeters (ARC), in which a sample is heated very slowly under quasiadiabatic conditions until the reaction intensifies by itself as a result of exothermic decomposition. The resulting temperature – time function can then be used to establish the corresponding heat effect and stability behavior.

26.3. References

[1] J. O. Hill (ed.): "For Better Thermal Analysis and Calorimetry," 3rd ed., International Confederation for Thermal Analysis and Calorimetry, 1991.
[2] W. Wendlandt: *Thermal Analysis,* 3rd ed., Wiley-Interscience, New York 1986.
[3] M. E. Brown: *Introduction to Thermal Analysis,* Chapman and Hall, London 1988.
[4] R. C. Mackenzie (ed.): *Differential Thermal Analysis,* Academic Press, London, vol. 1, 1970, vol. 2, 1972.
[5] M. E. Brown (ed.): *Handbook of Thermal Analysis and Calorimetry,* **vol. 1,** Elsevier, Amsterdam 1998.
[6] M. Reading in [5], pp. 423 – 443.
[7] S. P. Wolsky, A. W. Czanderna (eds.): *Microweighing in Controlled Environments,* Elsevier, Amsterdam 1980.
[8] ASTM Standard E 914-83.
[9] C. L. Duval: *Inorganic Thermogravimetric Analysis.* Elsevier, Amsterdam 1953.
[10] M. R. Ottaway, *Fuel* **61** (1982) 713.
[11] C. M. Earnest (ed.): *Compositional Analysis by Thermogravimetry,* ASTM STP 997, Philadelphia 1988.
[12] J. W. Cumming, J. McLaughlin, *Thermochim. Acta* **57** (1982) 253.
[13] E. Turi (ed.): *Thermal Characterization of Polymeric Materials,* Academic Press, San Diego 1997.
[14] R. L. Blaine, C. K. Schoff (eds.): *Purity Determination by Thermal Methods,* ASTM STP 838, Philadelphia 1984.
[15] M. Reading, *Trends Polym. Sci.* **1** (1993) 248.
[16] F. Paulik, J. Paulik, *Analyst (London)* **103** (1978) 417.
[17] W. Lodding (ed.): *Gas Effluent Analysis.* Marcel Dekker, New York 1967.
[18] S. B. Warrington in E. L. Charsley, S. B. Warrington (eds.): *Thermal Analysis – Techniques and Applications,* Royal Society of Chemistry, London 1992.
[19] M. R. Holdiness, *Thermochim. Acta* **75** (1984) 361.
[20] E. L. Charsley, C. Walker, S. B. Warrington, *J. Therm. Anal.* **40** (1993) 983 (Proc. 10th ICTAC Conf., Hatfield, 1992).
[21] V. Balek, J. Tolgyessy in C. L. Wilson, D. W. Wilson (eds.): *Comprehensive Analytical Chemistry,* vol. 12, part C, Elsevier, Amsterdam 1984.
[22] M. Reading in E. L. Charsley, S. B. Warrington (eds.): *Thermal Analysis – Techniques and Applications,* Royal Society of Chemistry, London 1992.
[23] T. Lever et al., *American Laboratory* **30** (1998) 15.

General References

[24] W. Hemminger, G. Höhne: *Calorimetry, Fundamentals and Practice*, Verlag Chemie, Weinheim 1984.
[25] E. Koch: *Non-Isothermal Reaction Analysis*, Academic Press, London 1977.
[26] J. P. McCullough, D. W. Scott (eds.): *Experimental Thermodynamics*, vol. 1: "Calorimetry of Non-reacting Systems," Butterworths, London 1968.
[27] B. Le Neindre, B. Vodar (eds.): *Experimental Thermodynamics*, vol. 2, Butterworths, London 1975.
[28] H. Spink, I. Wadsö: "Calorimetry as an Analytical Tool in Biochemistry and Biology," in D. Glinck (ed.): *Methods of Biochemical Analysis*, vol. 23, J. Wiley & Sons, New York 1976.
[29] S. Sunner, M. Mansson (eds.): "Combustion Calorimetry," *Experimental Chemical Thermodynamics*, vol. 1, Pergamon Press, Oxford 1979.
[30] H. Weber: *Isothermal Calorimetry*, Herbert Lang, Bern, Peter Lang, Frankfurt 1973.
[31] G. W. H. Höhne, W. Hemminger, H. J. Flammersheim, *Differential Scanning Calorimetry; An Introduction for Practitioners*, Springer-Verlag, Berlin, 1996.
[32] C. Schick, G. W. H. Höhne (eds.): "Temperature-Modulated Calorimetry", *Thermochim. Acta* **304/305** (1997) special issue.

Specific References

[33] E. Gmelin: "Classical Temperature-Modulated Calorimetry: A Review", *Thermochim. Acta* **304/305** (1997) 1 – 26.

[34] P. S. Gill, S. R. Sauerbrunn, M. Reading, *J.Thermal Anal.* **40** (1993) 931 – 939.
[35] P. F. Sullivan, G. Seidel *Phys. Rev.* **173** (1968) 679.
[36] P. Handler, D. E. Mapother, M. Rayl, *Phys. Rev. Lett.* **19** (1967) 356.
[37] O. M. Corbino, *Phys. Z.* **11** (1910) 413; **12** (1911) 292.
[38] H. Gobrecht, K. Hamann, G. Willers, *J. Physics E: Sci. Instr.* **4** (1971) 21 – 23.
[39] J. E. K. Schawe, G. W. H. Höhne, *Thermochim. Acta* **287** (1996) 213 – 223.
[40] R. W. Bunsen, *Ann. Phys.* **141** (1870) 1.
[41] E. D. West, E. F. Westrum Jr. in J. P. McCullough, D. W. Scott (eds.): *Experimental Thermodynamics*, **vol. 1**, Butterworths, London 1968, p. 354.
[42] S. M. Sarge et al.: "The Caloric Calibration of Scanning Calorimeters," *Thermochim. Acta* **247** (1994) 129 – 168.
[43] G. W. H. Höhne, *Thermochim. Acta* **330** (1999) 45 – 54.
[44] J. Lamprecht, W. Hemminger, G. W. H. Höhne (eds.): "Calorimetry in the Biological Sciences," *Thermochim. Acta* **193 (Special Issue)** (1991) 1 – 452.
[45] R. B. Kemp, B. Schaarschmidt (eds.): "Proceedings of Biological Calorimetry", *Thermochim. Acta* **251/252** (1995) special issues.
[46] T. B. Douglas, E. G. King in J. P. McCullough, D. W. Scott (eds.): *Experimental Thermodynamics*, **vol. 1**, Butterworths, London 1968, p. 297.

27. Surface Analysis

JOHN C. RIVIÈRE, Harwell Laboratory, AEA Technology, Didcot, United Kingdom

27.1. Introduction

Wherever the properties of a solid surface are important, it is also important to have the means to measure those properties. The surfaces of solids play an overriding part in a remarkably large number of processes, phenomena, and materials of technological importance. These include catalysis; corrosion, passivation, and rust; adhesion; tribology, friction, and wear; brittle fracture of metals and ceramics; microelectronics; composites; surface treatments of polymers and plastics; protective coatings; superconductors; and solid surface reactions of all types with gases, liquids, or other solids. The surfaces in question are not always external; processes occurring at inner surfaces such as interfaces and grain boundaries are often just as critical to the behavior of the material. In all the above cases, the nature of a process or of material behavior can be understood completely only if information about both surface composition (i.e., types of atoms present and their concentrations) and surface chemistry (i.e., chemical states of the atoms) is available. Occasionally, knowledge of the arrangement of surface atoms (i.e., the surface structure) is also necessary.

First of all, what is meant by a *solid surface*? Ideally the surface should be defined as the point at which the solid terminates, that is, the last atom layer before the adjacent phase (vacuum, vapor, liquid, or another solid) begins. Unfortunately such a definition is impractical because the effect of termination extends into the solid beyond the outermost atom layer. Indeed, the current definition is based on that knowledge, and the surface is thus regarded as consisting of that number of atom layers over which the effect of termination of the solid decays until bulk properties are reached. In practice, this decay distance is of the order of 5 – 20 nm.

By a fortunate coincidence, the depth into the solid from which information is provided by the techniques described here matches the above definition of a surface almost exactly. These techniques are therefore *surface specific*; in other words, their information comes only from that very shallow depth of a few atom layers. There are other techniques that can be *surface sensitive*, in that they would normally be regarded as techniques for bulk analysis, but have sufficient sensitivity for certain elements that those elements can be analyzed with these techniques even if they are present only on the surface.

Why should surfaces be so important? The answer is twofold. Firstly, because the properties of surface atoms are usually different from those of the same atoms in the bulk, and secondly, because in any interaction of a solid with another phase the surface atoms are the first to be encountered. Even at the surface of a *perfect single crystal* the surface atoms behave differently from those in the bulk simply because they do not have the same number of nearest neighbors; their electronic distributions are altered and hence their reactivity. Their structural arrangement is often also different. When the surface of a *polycrystalline or glassy multielemental solid* is considered, such as that of an alloy or a chemical compound, the situation can be very complex. The processes of preparation or fabrication may produce a material whose surface composition is quite different from that of the bulk, in terms of both constituent and impurity elements. Subsequent treatments (e.g., thermal and chemical) will almost certainly change the surface composition to something different again. The surface is highly unlikely to be smooth, and roughness at both micro- and macro-levels may be present, leading to the likelihood that many surface atoms will be situated at corners and edges and on protuberances (i.e., in positions of increased reactivity). Surfaces exposed to the atmosphere, which include many of those of tech-

Table 1. Surface-specific analytical techniques using primary particle excitation*

Emission	Excitation**			
	Photons $h\nu$	Electrons e^-	Ions n^+	Neutrals n^0
$h\nu$	RAIRS SERS ellipsometry SHG	SXAPS DAPS CLS IPES BIS	IBSCA GDOS	
e^-	XPS XAES UPS SRPS	AES SAM ELS HREELS CEELS AEAPS LEED RHEED	IAES PAES INS MQS	
n^+, n^-	PSD	ESD ESDIAD	SSIMS ISS RBS GDMS	FABMS
n^0	PSD	ESD	SNMS	

* Some techniques in Table 1 have angle-resolved variants, with the prefix AR, thus giving ARUPS, ARAES, ARELS, etc.
** For meanings of acronyms, see Listing 1.

Table 2. Surface-specific analytical techniques using non-particle excitation

Emission	Excitation*		
	Heat kT	High electrical field F	Mechanical force
n^+	TDS	APFIM POSAP	
n^-	TDS		
e^-		IETS STM, STS	
(Displacement)			AFM

* For meanings of acronyms, see Listing 1.

nological interest, will acquire a contaminant layer 1 – 2 atom layers thick, containing principally carbon and oxygen but also other impurities present in the local environment. Atmospheric exposure may also cause oxidation. Because of all these possibilities the surface region must be considered as a separate entity, effectively a separate quasi-two-dimensional phase overlaying the normal bulk phase. Analysis of the properties of such a quasi phase necessitates the use of techniques in which the information provided originates only or largely within the phase (i.e., the surface-specific techniques described in this article).

Nearly all these techniques involve interrogation of the surface with a particle probe. The function of the probe is to excite surface atoms into states giving rise to emission of one or more of a variety of secondary particles such as electrons, photons, positive and secondary ions, and neutrals. Since the primary particles used in the probing beam may also be either electrons or photons, or ions or neutrals, many separate techniques are

possible, each based on a different primary – secondary particle combination. Most of these possibilities have now been established, but in fact not all the resultant techniques are of general application, some because of the restricted or specialized nature of the information obtained and others because of difficult experimental requirements. In this article, therefore, most space is devoted to those surface analytical techniques that are widely applied and readily available commercially, whereas much briefer descriptions are given of the many others whose use is less common but that — in appropriate circumstances, particularly in basic research — can provide vital information.

Because the various types of particle can appear in both primary excitation and secondary emission, most authors and reviewers have found it convenient to group the techniques in a matrix, in which the columns refer to the nature of the exciting particle and the rows to the nature of the emitted particle [1] – [9]. Such a matrix of techniques is given in Table 1, in terms of the acronyms now accepted. The meanings of the acronyms, together with some of the alternatives that have appeared in the literature, are in Listing 1.

Listing 1. Meanings of surface analysis acronyms, and their alternatives, that appear in Tables 1 and 2.

1) Photon Excitation

RAIRS Reflection – absorption infrared spectroscopy (or IRAS IR reflection – absorption spectroscopy)
SERS Surface-enhanced Raman scattering
SHG   Second harmonic generation
XPS   X-ray photoelectron spectroscopy (or ESCA electron spectroscopy for chemical analysis)
XAES X-ray (excited) Auger electron spectroscopy

2) Electron Excitation

3) Ion Excitation

4) Neutral Excitation

5) Thermal Excitation

6) High Field Excitation

7) Mechanical Force

A few techniques, including one or two important ones, cannot be classified according to the nature of the exciting particle, since they do not employ primary particles but depend instead on the application either of heat or a high electric field. The latter techniques are listed in Table 2.

27.2. X-Ray Photoelectron Spectroscopy (XPS)

X-ray photoelectron spectroscopy is currently the most widely used surface analytical technique, and is therefore described here in more detail than any of the other techniques. At its inception by SIEGBAHN and coworkers [10] it was called ESCA (electron spectroscopy for chemical analysis), but the name ESCA is now considered too general, since many surface electron spectroscopies exist, and the name given to each one must be precise. Nevertheless, the name ESCA is still used in many places, particularly in industrial laboratories and their publications. Briefly, the reasons for the popularity of XPS are the exceptional combination of compositional and chemical information that it provides, its ease of operation, and the ready availability of commercial equipment.

27.2.1. Principles

The surface to be analyzed is irradiated with soft X-ray photons. When a photon of energy hv interacts with an electron in a level with binding energy E_B, the entire photon energy is transferred to the electron, with the result that a photoelectron is ejected with kinetic energy

$$E_{Kin} = hv - E_B - e\varphi \tag{1}$$

where $e\varphi$ is a small, almost constant, work function term.

Obviously hv must be greater than E_B. The ejected electron may come from a core level or from the occupied portion of the valence band, but in XPS most attention is focused on electrons in core levels. Since no two elements share the same set of electronic binding energies, measurement of the photoelectron kinetic energies provides an *elemental analysis*. In addition, Equation (1) indicates that any changes in E_B are reflected in E_{Kin}, which means that changes in the chemical environment of an atom can be followed by monitoring
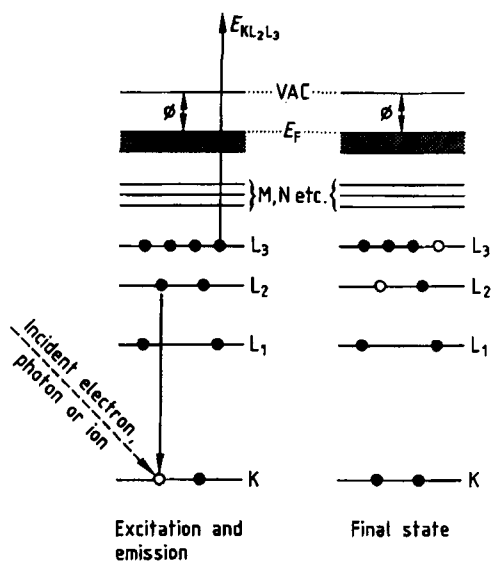
Figure 1. Auger process in an atom in a solid

changes in the photoelectron energies, leading to the provision of *chemical information*. XPS can analyze all elements in the periodic table except hydrogen and helium.

Although XPS is concerned principally with photoelectrons and their kinetic energies, ejection of electrons by other processes also occurs. An ejected photoelectron leaves behind a core hole in the atom. The sequence of events following the creation of the core hole is shown schematically in Figure 1. In the example, the hole has been created in the K shell, giving rise to a photoelectron whose kinetic energy would be $(hv - E_K)$, and is filled by an electronic transition from the L_2 shell. The energy $E_K - E_{L_2}$ associated with the transition can then either be dissipated as a characteristic X-ray photon or given up to an electron in a third shell, shown in this example as the L_3. The second of these possibilities is called the Auger process after its discoverer [11], and the resultant ejected electron is called an Auger electron and has an energy given by

$$E_{KL_2L_3} = E_K - E_{L_2} - E^*_{L_3} \tag{2}$$

$E^*_{L_3}$ is starred because it is the binding energy of an electron in the L_3 shell in the presence of a hole in the L_2 shell, that is, not quite the same as E_{L_3}.

X-ray photon emission (i.e., X-ray fluorescence) and Auger electron emission are obviously

competing processes, but for the shallow core levels involved in XPS and AES the Auger process is far more likely.

Thus in all X-ray photoelectron spectra, features appear due to both photoemission and Auger emission. In XPS, the Auger features can be useful but are not central to the technique, whereas in AES (see Chap. 27.3), Equation (2) forms the basis of the technique.

At this point the nomenclature used in XPS and AES should be explained. In XPS the *spectroscopic notation* is used, and in AES the *X-ray notation*. The two are equivalent, the different usages having arisen for historical reasons, but the differentiation is a convenient one. They are both based on the so-called $j-j$ coupling scheme describing the orbital motion of an electron around an atomic nucleus, in which the total angular momentum of an electron is found by summing vectorially the individual electron spin and angular momenta. Thus if l is the electronic angular momentum quantum number and s the electronic spin momentum quantum number, the total angular momentum for each electron is given by $j = l + s$. Since l can take the values 0, 1, 2, 3, 4,... and $s = \pm \frac{1}{2}$, clearly $j = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$, etc. The principal quantum number n can take values 1, 2, 3, 4,.... In *spectroscopic notation*, states with $l = 0, 1, 2, 3,...$ are designated $s, p, d, f, ...$, respectively, and the letter is preceded by the number n; the j values are then appended as suffixes. Therefore one obtains $1s, 2s, 2p_{1/2}, 2p_{3/2}, 3s, 3p_{1/2}, 3p_{3/2}$, etc.

In *X-ray notation*, states with $n = 1, 2, 3, 4,...$ are designated K, L, M, N,..., respectively, and states with various combinations of $l = 0, 1, 2, 3,...$ and $j = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$, are appended as the suffixes 1, 2, 3, 4,.... In this way one arrives at K, $L_1, L_2, L_3, M_1, M_2, M_3$, etc. The equivalence of the two notations is set out in Table 3.

In X-ray notation the Auger transition shown in Figure 1 would therefore be labeled KL_2L_3. In this coupling scheme, six Auger transitions would be possible in the KLL series. Obviously, many other series are possible (e.g., KLM, LMM, MNN). These are discussed more fully in Chapter 27.3, dealing with AES.

The reasons why techniques such as XPS and AES, which involve measurement of the energies of ejected electrons, are so surface specific, should be examined. An electron with kinetic energy E moving through a solid matrix M has a probability of traveling a certain distance before losing all or part of its energy as a result of an inelastic collision. Based on that probability, the average dis-

Table 3. Spectroscopic and X-ray notation

Quantum numbers			Spectroscopic state	X-ray suffix	X-ray state
n	l	j			
1	0	1/2	$1s$	1	K
2	0	1/2	$2s$	1	L_1
2	1	1/2	$2p_{1/2}$	2	L_2
2	1	3/2	$2p_{3/2}$	3	L_3
3	0	1/2	$3s$	1	M_1
3	1	1/2	$3p_{1/2}$	2	M_2
3	1	3/2	$3p_{3/2}$	3	M_3
3	2	3/2	$3d_{3/2}$	4	M_4
3	2	5/2	$3d_{5/2}$	5	M_5
	etc.		etc.	etc.	etc.

Figure 2. Compilation by SEAH and DENCH [12] of measurements of inelastic mean free path as a function of electron kinetic energy
The solid line is a least-squares fit

tance traveled before such a collision is known as the inelastic mean free path (imfp) $\lambda_M(E)$. The imfp is a function only of M and of E. Figure 2 shows a compilation of measurements of λ made by SEAH and DENCH [12], in terms of atomic monolayers as a function of kinetic energy. Note that both λ and energy scales are logarithmic. The important consequence of the dependence of λ on kinetic energy is that in the ranges of secondary electron kinetic energies used in XPS and AES, the values of λ are very small. In XPS, for example, typical ranges are 250 – 1500 eV, corresponding to a range of λ from about four to eight monolayers, while in AES, the energy range is typically 20 to 1000 eV, in which case λ would range from about two to six monolayers. What this means in practice is that if the photoelectron or the Auger electron is to escape into a vacuum and be detected, it must originate at or very near the surface of the solid. This is the reason why the electron spectroscopic techniques are surface specific.

27.2.2. Instrumentation

27.2.2.1. Vacuum Requirements

Electron spectroscopic techniques require vacua of the order of 10^{-8} Pa for their operation. This requirement arises from the extreme surface specificity of these techniques, mentioned above. With sampling depths of only a few atom layers, and elemental sensitivities down to 10^{-5} atom layers (i.e., one atom of a particular element in 10^5 other atoms in an atomic layer), the techniques are clearly very sensitive to surface contamination, most of which comes from the residual gases in the vacuum system. According to gas kinetic theory, to have enough time to make a surface analyt-

ical measurement on a surface that has just been prepared or exposed, before contamination from the gas phase interferes, the base pressure should be 10^{-8} Pa or lower, that is, in the region of ultrahigh vacuum (UHV).

The requirement for the achievement of UHV conditions imposes restrictions on the types of material that can be used for the construction of surface analytical systems, or inside the systems, because UHV can be achieved only by accelerating the rate of removal of gas molecules from internal surfaces by raising the temperature of the entire system (i.e., by baking). Typical baking conditions are 150 – 200 °C for several hours. Construction materials should not be distorted, lose their strength, produce excessive gas, or oxidize readily under such treatment, which should be repeated after each exposure of the inside to atmospheric pressure. Thus, the principal construction material is stainless steel, with mu-metal (76 % Ni, 5 % Cu, 2 % Cr) used occasionally where magnetic screening is needed (e.g., around electron energy analyzers). For the same reasons, metal seals, not elastomers, are used for the demountable joints between individual components—the sealing material usually being pure copper although gold is sometimes used. Other materials that may be used between ambient atmosphere and UHV are borosilicate glass or quartz for windows, and alumina for electrical insulation for current or voltage connections. In-

Table 4. Energies and linewidths of some characteristic low-energy X-ray lines

Line		Energy, eV	Width, eV
Y	M_ζ	132.3	0.47
Zr	M_ζ	151.4	0.77
Nb	M_ζ	171.4	1.21
Mg	K_α	1253.6	0.70
Al	K_α	1486.6	0.85
Si	K_α	1739.5	1.00
Y	L_α	1922.6	1.50
Zr	L_α	2042.4	1.70

side the system, any material is permissible that does not produce volatile components either during normal operation or during baking. Thus, for example, brass that contains the volatile metal zinc could not be used.

The production of vacua in the UHV region has been routine for many years now, but success still depends on attention to details such as the maintenance of strict cleanliness.

27.2.2.2. X-Ray Sources

The most important consideration in choosing an X-ray source for XPS is energy resolution. Equation (1) gives the relationship between the kinetic energy of the photoelectron, the energy of the X-ray photon, and the binding energy of the core electron. Since the energy spread, or linewidth, of an electron in a core level is very small, the linewidth of the photoelectron energy depends on the linewidth of the source, if no undue broadening is introduced instrumentally. Now in XPS the analyst devotes much effort to extracting chemical information by means of detailed study of individual elemental photoelectron spectra. Such a study needs an energy resolution better than 1.0 eV if subtle chemical effects are to be identified. Thus the linewidth of the X-ray source should be significantly smaller than 1.0 eV if the resolution required is not to be limited by the source itself.

Other considerations are that the source material, which forms a target for high-energy electron bombardment leading to the production of X rays, should be a good conductor—to allow rapid removal of heat—and should also be compatible with UHV.

Table 4 lists the energies and linewidths of the characteristic X-ray lines from a few possible candidate materials.

Not many sources with sufficiently narrow lines are available. Y M_ζ and Zr M_ζ would be suitable, but their line energies are too low to be useful for general application, although they have been used for special purposes. Silicon is a poor thermal conductor and is not easy to apply as a coating on an anode, so only the Mg K_α and Al K_α lines remain, which are in fact the two used universally in XPS.

For efficient production of X rays by electron bombardment, exciting electron energies that are at least an order of magnitude higher than the line energies must be used, so that in Mg and Al sources accelerating potentials of 15 kV are employed. Modern sources are designed with dual anodes, one anode face being coated with magnesium and the other with aluminum, and with two filaments, one for each face. Thus, a switch from one type of X irradiation to the other can be made very quickly.

To protect the sample from stray electrons from the anode, from heating effects, and from possible contamination by the source enclosure, a thin (≈ 2-μm) window of aluminum foil is interposed between the anode and the sample. For optimum X-ray photon flux on the surface (i.e., optimum sensitivity), the anode must be brought as close to the sample as possible, which means in practice a distance of ≈ 2 cm. The entire X-ray source is therefore retractable via a bellows and a screw mechanism.

The X radiation from magnesium and aluminum sources is quite complex. The principal K_α lines are in fact unresolved doublets and should correctly be labeled $K_{\alpha_{1,2}}$. Many satellite lines also exist of which the more important in practice are the $K_{\alpha_{3,4}}$, separated from the principal line by ≈ 10 eV and roughly 8 % of its intensity, and the K_β, lying about 70 eV higher in energy. In addition, the spectrum is superimposed on a continuous background called bremsstrahlung radiation, arising from inelastic processes.

Removal of satellites, elimination of the bremsstrahlung background, and separation of the $K_{\alpha_{1,2}}$ doublet can be achieved by monochromatization, shown schematically in Figure 3. The X-ray source is positioned at one point on a spherical surface, called a Rowland sphere after its originator, and a very accurately ground and polished quartz crystal is placed at another point. X rays from the source are diffracted from the quartz, and by placing the sample at the correct point on the Rowland sphere, the K_{α_1} component can be selected to be focused on it, according to

Characteristic X ray (X)   Al K$_\alpha$ characteristic   High resolution XPS
+ satellites (S) + background (B)   only
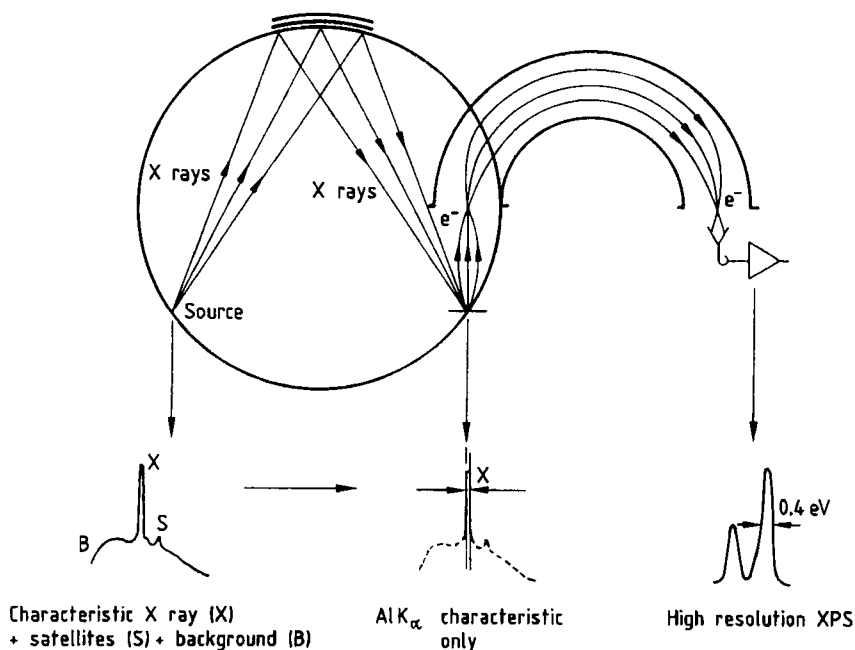
Figure 3. Schematic of X-ray monochromatization to remove satellites, eliminate bremsstrahlung background, and separate the Al K$_{\alpha_{1,2}}$ doublet
Courtesy of Kratos Analytical.

the well-known Bragg relation for diffraction. Quartz is a very convenient diffracting medium for Al K$_\alpha$ since the spacing between the $10\bar{1}0$ planes is exactly half the wavelength of the X radiation, and for first-order diffraction the Bragg angle is 78.5°. Since the width of the Al K$_{\alpha_1}$ line is < 0.4 eV, the energy dispersion needed around the surface of the sphere implies that the Rowland sphere should have a diameter of at least 0.5 m. Accurate positioning of source, crystal, and sample is essential. Although an XPS spectrum will be much "cleaner" when a monochromator is used, because satellites and background have been removed, the photon flux at the sample is much lower than that from an unmonochromatized source operating at the same power. Against this must be set the greatly improved signal-to-background level in a monochromatized spectrum.

27.2.2.3. Synchrotron Radiation

The discrete line sources described above for XPS are perfectly adequate for most applications, but some types of analysis require that the source be tunable (i.e., that the exciting energy be vari-

able). The reason is to allow the photoionization cross section of the core levels of a particular element or group of elements to be varied, which is particularly useful when dealing with multielement semiconductors. Tunable radiation can be obtained from a synchrotron.

In a synchrotron, electrons are accelerated to near-relativistic velocities and constrained magnetically into circular paths. When a charged body is accelerated, it emits radiation, and when the near-relativistic electrons are forced into curved paths, they emit photons over a continuous spectrum. The general shape of the spectrum is shown in Figure 4. For a synchrotron with an energy of several gigaelectronvolts and a radius of some tens of meters, the energy of the emitted photons near the maximum is of the order of 1 keV (i.e., ideal for XPS). As can be seen from the universal curve, plenty of usable intensity exists down into the UV region. With suitable monochromators on the output to select a particular wavelength, the photon energy can be tuned continuously from about 20 to 500 eV. The available intensities are comparable to those from conventional line sources.
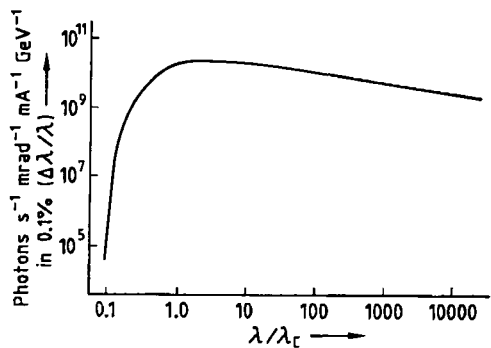
Figure 4. Normalized spectrum of photon energies emitted from a synchrotron
λ_c = wavelength characteristic of the individual synchrotron

Figure 5. Schematic of a concentric hemispherical analyzer (CHA) [13]

27.2.2.4. Electron Energy Analyzers

In electron spectroscopic techniques — among which XPS is the most important — analysis of the energies of electrons ejected from a surface is central. Because of the low kinetic energies involved in the techniques, analyzers using magnetic fields are undesirable. Therefore the energy analyzers used are exclusively of the electrostatic deflection type. The two that are now universally employed are the concentric hemispherical analyzer (CHA) and the cylindrical mirror analyzer (CMA). Since both have been used in XPS, both are described here, although in practice the CHA is more suitable for XPS, and the CMA for AES.

Energy analyzers cannot be discussed without a discussion of energy resolution, which is defined in two ways. *Absolute resolution* is defined as ΔE, the full width at half-maximum (FWHM) of a chosen peak. *Relative resolution* is defined as the ratio R of ΔE to the kinetic energy E of the peak energy position (usually its centroid), that is, $R = \Delta E/E$. Thus absolute resolution is independent of peak position, but relative resolution can be specified only by reference to a particular kinetic energy.

In XPS, closely spaced peaks at any point in the energy range must be resolved, which requires the same absolute resolution at all energies. In AES, on the other hand, the inherent widths of peaks in the spectrum are greater than in XPS and the source area on the surface is far smaller, both of which lead to operation at constant relative resolution.

Concentric Hemispherical Analyzer (CHA). The CHA is shown in schematic cross section in

Figure 5 [13]. Two hemispheres of radii R_1 (inner) and R_2 (outer) are positioned concentrically. Potentials $-V_1$ and $-V_2$ are applied to the inner and outer hemispheres, respectively, with V_2 greater than V_1. The source S and the focus F are in the same plane as the center of curvature, and R_0 is the radius of the equipotential surface between the hemispheres. If electrons of energy $E = eV_0$ are injected at S along the equipotential surface, they will be focused at F if

$$V_2 - V_1 = V_0 \left(\frac{R_2}{R_1} - \frac{R_1}{R_2} \right) \qquad (3)$$

If electrons are injected not exactly along the equipotential surface, but with an angular spread $\delta\alpha$ about the correct direction, then the energy resolution is given by

$$\Delta E/E = (W_1 + W_2)/4R_0 + (\delta\alpha)^2 \qquad (4)$$

where W_1 and W_2 are the respective widths of the entrance and exit slits. In most instruments, for convenience in construction purposes, $W_1 = W_2 = W$, whereupon the resolution becomes

$$\Delta E/E = W/2R_0 + (\delta\alpha)^2 \qquad (5)$$

In XPS the photoelectrons are retarded to a constant energy, called the pass energy, as they approach the entrance slit. If this were not done, Equation (5) shows that to achieve an absolute resolution of 1 eV at the maximum kinetic energy of about 1500 eV (using Al K_α radiation), and with a slit width of 2 mm, would require an analyzer with an average radius of about 300 cm, which is impracticable. Pass energies are selected in the
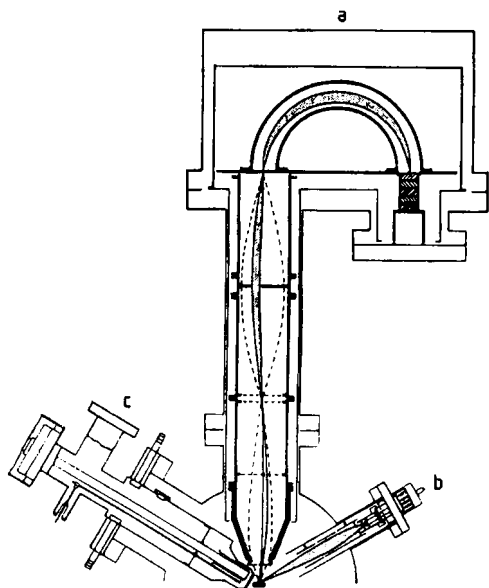
Figure 6. Typical configuration in a modern XPS spectrometer
Courtesy of Leybold
a) XPS – Auger analyzer; b) Auger electron gun; c) X-ray source

Figure 7. Schematic of a cylindrical mirror analyzer (CMA)

$$\frac{E}{e\,V} = \frac{1.31}{\ln\,(r_2/r_1)} \tag{6}$$

As in the CHA, therefore, scanning the deflection potential $-V$ and recording the signal as a function of electron energy provide the distribution in energy of electrons leaving the sample surface.

If the angular spread of the acceptance angle is $\delta\alpha$, then the relative energy resolution of a CMA is

$$\Delta E/E = 0.18 W/r_1 + 1.39\,(\delta\alpha)^3 \tag{7}$$

where W is the effective slit width. For $\delta\alpha \approx 6°$, a typical value, W can be replaced by the source size (i.e., the area on the sample from which electrons are accepted).

In *AES* the inner cylinder is at earth potential, and primary electrons originate in an integral electron gun built into the middle of the analyzer. The source size in AES is very small, because of the focused primary beam, and the important property of the CMA for AES is therefore its very high transmission. In *XPS* the entire sample surface is normally flooded with X rays, in which case the source size is the entire acceptance area of photoelectrons, and Equation (7) shows that the relative energy resolution would be too poor. The important parameter in XPS is not the transmission alone, but the luminosity, which is the product of transmission, acceptance area, and acceptance solid angle.

The single-stage CMA depicted in Figure 7 is therefore not suitable for XPS, but a variation of the CMA that is suitable is the double-pass CMA shown in Figure 8 [14]. Two CMAs are placed in series, so that the exit aperture for the first is the entrance aperture for the second. Electrons from

range $20-100$ eV for XPS, which allows the analyzer to be built with a radius of $10-15$ cm.

Modern XPS spectrometers employ a lens system on the input to the CHA, which has the effect of transferring an image of the analyzed area on the sample surface to the entrance slit of the analyzer. Such a spectrometer is illustrated schematically in Figure 6.

Cylindrical Mirror Analyzer (CMA). In the CMA shown schematically in Figure 7, two cylinders of radii r_1 (inner) and r_2 (outer) are accurately positioned coaxially. Annular entrance and exit apertures are cut in the inner cylinder, and a deflection potential $-V$ is applied between the cylinders. Electrons leaving the sample surface at the source S with a particular energy E on passing into the CMA via the entrance aperture are then deflected back through the exit aperture to the focus F. For the special case in which the acceptance angle $\alpha = 42.3°$, the first-order aberrations vanish, and the CMA becomes a second-order focusing device. The relationship between the electron energy E and the deflection potential is then

Figure 8. Schematic of double-pass CMA designed by PALMBERG [14] for either XPS or AES
a) Inner and outer magnetic shields; b) Outer cylinder; c) Inner cylinder; d) Rotary motion for aperture change; e) Electron multiplier; f) Second stage; g) First stage; h) Spherical retarding grids; i) Sample; j) X-ray source; k) Electron gun
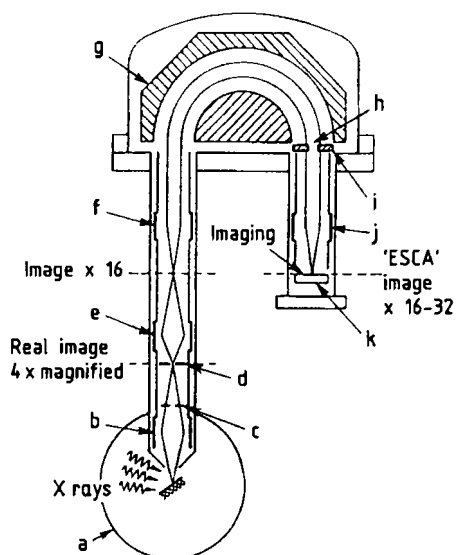
Figure 9. Schematic of the electron-optical arrangement designed by COXON et al. [15] for the production of energy-resolved two-dimensional images in XPS
a) Spherical mu-metal chamber; b) Lens 1; c) Objective aperture; d) Field aperture; e) Lens 2; f) Lens 3; g) 180° Hemispherical analyzer; h) Hole; i) Spectrum detector 1; j) Lens 5; k) Image detector 2

the sample surface are retarded to a constant pass energy, as in the CHA, by spherical retardation grids in front of the entrance aperture, centered on the source area on the sample. The first mesh is grounded, and the second mesh is at the potential of the inner cylinder, which in the XPS mode

of operation is not grounded, as in AES, but floating at the scanning potential. Adequate energy resolution is achieved in the second stage because the effective source area, being defined not by the irradiated sample surface but by the size of the first stage exit aperture, is sufficiently small. Entrance and exit apertures can be changed in the second stage by an externally operated mechanism.

27.2.2.5. Spatial Resolution

The principal disadvantage of conventional XPS is the lack of spatial resolution; the spectral information comes from an analyzed area of several square millimeters and is, therefore, an average of the compositional and chemical analyses from that area. On the other hand, the majority of samples of technological origin are inhomogeneous on a scale much smaller than that of normal XPS analysis, and obtaining chemical information on the same scale as the inhomogeneities would be very desirable.

Major steps have been taken recently in the direction of improved XPS spatial resolution by making use of the focusing properties of the CHA described above. One of the more successful configurations that have evolved for this purpose is shown in Figure 9, from the work of COXON et al. [15]. The instrument incorporating it is called the ESCASCOPE. Compared to a conventional XPS system, it has additional lenses (3 and 5) on the input and output sides of the CHA. Lenses 1 and 2 collect electrons from the analyzed area and
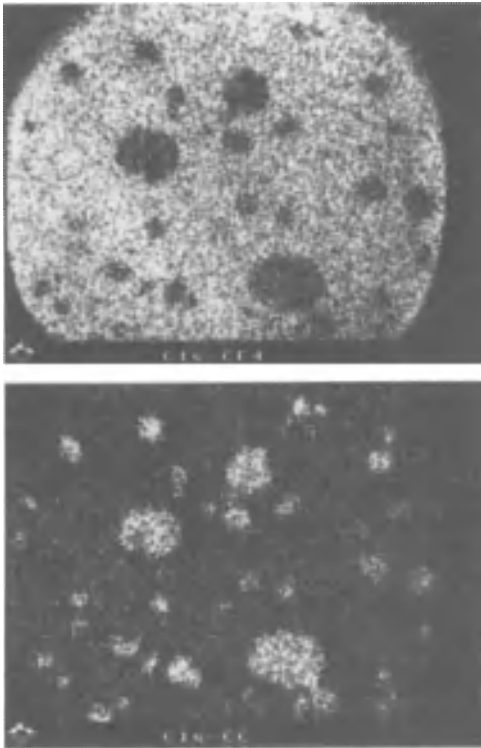
Figure 10. Chemical state images obtained with the ES-CASCOPE in Figure 9, from a contaminated fluoropolymer [15]
A) Image in contribution to C 1 s from C – F bonding;
B) Image in contribution to C 1 s from C – C bonding

Figure 11. "Wide," or "survey," spectrum from almost clean copper, using Al K_α radiation

present a real image of the sample, magnified about 16 times, at the object plane of lens 3.

Lens 3 is so arranged that the distance to its object plane is equal to its focal length, so that electrons reach the object plane (also the entrance slit of the analyzer) traveling in straight lines. This means that electrons leaving a point on the sample surface within a narrow angular range, pass into the analyzer in a parallel beam inclined at a particular angle to the object plane, with no intensity variation in the plane; another way of putting this is to say that a Fourier transformation of the surface image exists at the entrance slit.

The CHA (lens 4) then forms an image of the Fourier-transformed surface image in its own image plane (the exit slit) and introduces energy dispersion at the same time, as discussed above. The angular information is conserved. Photoelectrons pass through a hole in the center of the image plane and into lens 5, which is arranged similarly to lens 3. The Fourier-transformed image is then

inverted so that a real image is formed in the image plane of lens 5, and this image will be an energy-filtered two-dimensional image of the sample surface. The image is detected via a position-sensitive detector consisting of two 40-mm-diameter channel plates, a phosphor screen, and a TV camera. Since the photoelectron kinetic energies can be selected at will in the CHA, images in electron energies characteristic of a particular element, or of a particular chemical state of an element, can be formed.

An example of chemical state imaging in the ESCASCOPE is given in Figure 10, also from [15]. The material is a contaminated fluoropolymer, the contamination being in the form of dark spots of 10 – 80-µm size. Both images in Figure 10 are in the C 1 s photoelectron peak, but of different chemical states. The upper is in that contribution arising from carbon bound to fluorine, and the lower from carbon bound to carbon (i.e., graphitic in nature). The complementarity of the images indicates that the contamination is graphitic in nature.

27.2.3. Spectral Information and Chemical Shifts

Figure 11 shows a *wide* or *survey XPS spectrum*, that is, one recorded over a wide range of energies, in this case 1000 eV. The radiation used was unmonochromatized Al K_α, at 1486.6 eV, and the surface is that of almost clean copper. Such a spectrum reveals the major features to be found,
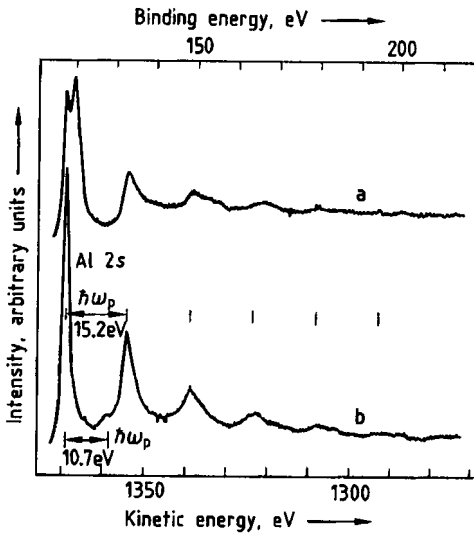
Figure 12. "Narrow," or "detail," spectra from aluminum [16]
a) Slightly oxidized aluminum; b) Clean aluminum
Successive bulk plasmon losses, and a surface plasmon loss are shown associated with the Al 2 s peak

but to investigate minor or more detailed features, spectra are acquired over much more restricted energy ranges, at better energy resolution; the latter spectra are called *narrow* or *detail scans.*

The most prominent features in Figure 11, and in most photoelectron spectra, are peaks arising from excitation of core-level electrons according to Equation (1). Thus, at the left-hand end, two intense peaks are found at about 553 and 533 eV, corresponding to photoelectrons from the $2p_{3/2}$ and $2p_{1/2}$ levels of copper, respectively, the separation of 20 eV being the spin–orbit splitting. At the right-hand, high kinetic energy end, are three other copper photoelectron peaks at about 1363, 1409, and 1485 eV, which arise from the $3s$, $3p$, and $3d$ levels, respectively. The other prominent features associated with copper are the three Auger peaks $L_3M_{4,5}M_{4,5}$, $L_3M_{2,3}M_{4,5}$, and $L_3M_{2,3}M_{2,3}$ appearing at 919, 838, and 768 eV, respectively. As pointed out in Section 27.2.1, the creation of a core hole by any means of excitation can lead to ejection of an Auger electron, so that in XPS Auger features form a significant contribution to the spectrum. If Auger and photoelectron peaks happen to overlap in any spectrum, they can always be separated by changing the excitation (e.g., from Al K_α to Mg K_α, or vice versa) since the Auger peaks are invariant in energy whereas the photoelectron peaks must shift

with the energy of the exciting photons according to Equation (1).

In addition to features due to copper in Figure 11 are small photoelectron peaks at 955- and 1204-eV kinetic energies, arising from the oxygen and carbon $1s$ levels, respectively, due to the presence of some contamination on the surface. Other minor features can be seen associated with the major peaks. Those situated about 10 eV toward higher kinetic energy occur because unmonochromatized X radiation was used; they are photoelectron peaks excited by the $K_{\alpha_{3,4}}$ satellite already mentioned (see Section 27.2.2.2.). They would disappear if a monochromator were used. Toward the lower kinetic energy of the major peaks (e.g., visible some 10–30 eV below the Cu $2p$ peaks) are the so-called *plasmon loss peaks.* Peaks of this type are shown more clearly in Figure 12, from BARRIE [16], in narrow scans from an aluminum surface. They arise when outgoing electrons of sufficient energy, originating in any way, lose energy by exciting collective oscillations in the conduction electrons in a solid. The oscillation frequency, and hence the energy involved, is characteristic of the nature of the solid and is called the plasmon frequency (or energy). The plasmon loss determined by the three-dimensional nature of the solid is called a "bulk" or "volume" loss, but when it is associated with the two-dimensional nature of the surface, it is called a "surface" loss. In Figure 12 (b), which is from clean aluminum, examples of both types of plasmon loss from the Al $2s$ peak can be seen—a bulk loss of 15.2 eV and a surface loss of 10.7 eV. Theoretically, the surface plasmon loss energy should be a factor of $2^{1/2}$ less than the bulk plasmon loss energy. Where the bulk loss is very pronounced, as in a free-electron-like metal such as aluminum, a succession of peaks can be found at multiples of the loss energy, as indicated in Figure 12, with progressively decreasing intensity. Spectrum (a) in Figure 12, from slightly oxidized aluminum, reveals the rapid quenching of the surface plasmon on oxidation and also shows a chemical shift in the Al $2s$ photoelectron peak.

Chemical shift is the name given to the observed shift in energy of a photoelectron peak from a particular element when the chemical state of that element changes. When an atom enters into combination with another atom or group of atoms, an alteration occurs in the valence electron density, which might be positive or negative according to whether charge is accepted or donated, causing a consequent alteration in the electrostatic potential affecting the core electrons. Therefore the
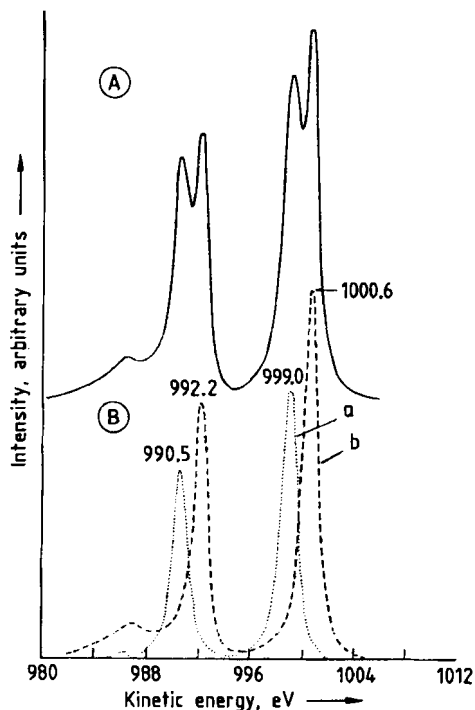
Figure 13. Example of a chemical shift in the Sn $3d$ peak, exhibited by a very thin layer of Sn oxide on Sn metal
A) Spectrum after linear background subtraction; B) Spectrum resolved into its respective components
a) Sn^{n+}; b) Sn^0

Figure 14. Example of Cu$2p$ "shake-up" peaks [17]
A) Present in a Cu^{2+} compound (CuO); B) Absent in a Cu^+ compound (Cu_2O)

binding energies of the core electrons change, giving rise according to Equation (1) to shifts in the corresponding photoelectron peaks. Tabulation of the chemical shifts experienced by any one element in a series of pure compounds of that element thus enables its chemical state to be identified during analysis of unknown samples. Many such tabulations have appeared, and a major collation of them is found in [2]. The identification of chemical state in this way is the principal advantage of XPS over other surface analytical techniques.

An example of a spectrum exhibiting a chemical shift is that of the tin $3d$ peaks in Figure 13. A thin layer of oxide on the metallic tin surface allows photoelectrons from both the underlying metal and the oxide to appear together. Resolution of the doublet $3d_{5/2}$, $3d_{3/2}$ into the components from the metal (Sn^0) and from the oxide Sn^{n+} is shown in Figure 13 B. The shift in this case is 1.6–1.7 eV. Curve resolution is an operation that can be performed routinely in data processing systems associated with photoelectron spectrometers.

Another type of satellite often seen in XPS spectra is the so-called *shake-up satellite*. Such satellites occur when formation of a compound leads to the presence of unpaired electrons in $3d$ or $4f$ levels; they can also occur in organic compounds according to the degree of conjugation in C=C double bonds, particularly in aromatic compounds. Figure 14, from SCROCCO [17], compares the copper $2p$ photoelectron spectra from CuO and Cu_2O. Since Cu^{2+} as in CuO has an open shell configuration, the $2p$ spectrum shows strong shake-up satellites, whereas that from Cu_2O does not because Cu^+ has a closed shell. In many cases, such satellite structures can be used to diagnose the oxidation state(s).

Other spectral features occasionally seen in XPS spectra arise from *multiplet splitting*. The latter occurs when an element has unpaired electrons either in the valence band or in localized core levels very close to the valence band (as in, for example, some of the rare-earth elements). In the $4f$ electron levels of the rare earths, correlation between spin and angular electron momenta can occur in many different ways, leading to multiplets at different energies, most of which are resolvable [18] in XPS as a multipeak spectrum.
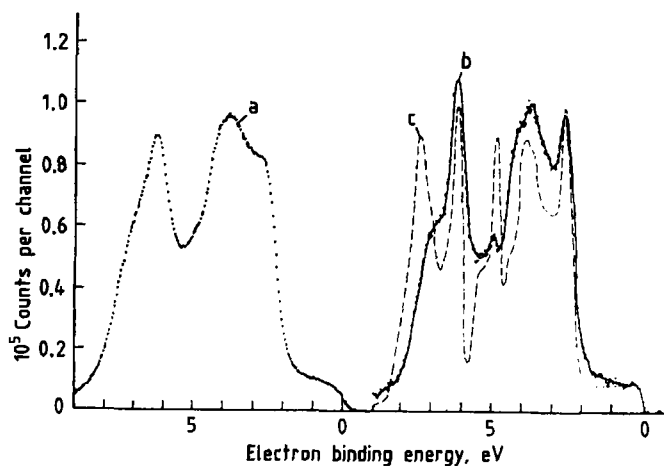
Figure 15. Valence band spectra of gold [19]
a) Using monochromatized Al K_α radiation, and high-energy resolution; b) Result of deconvoluting instrumental broadening from the raw data of a); c) Result of a theoretical calculation of the valence band structure

Photoelectrons of highest kinetic energy in a spectrum (i.e., those of lowest binding energy) originate in the shallowest levels of the solid, that is, in the valence band. Thus the photoelectron peak labeled $3\,d$ in the copper spectrum in Figure 11 corresponds to the valence band spectrum, although no detailed structure can be seen in such a "survey" spectrum because the energy resolution used there is not adequate. If much better resolution is used together with a monochromatized source, the valence bands of many metals reveal interesting detail, as shown [19] in Figure 15 for gold. However, for a systematic study of valence band structure, XPS is not ideal, because of both photoionization cross-section effects and the high kinetic energies involved, and the techniques of ultraviolet photoelectron spectroscopy (UPS) and synchrotron radiation photoelectron spectroscopy (SRPS) are preferred.

27.2.4. Quantification and Depth Profiling

27.2.4.1. Quantification

If an X-ray photon of energy $h\nu$ ionizes a core level X in an atom of element A in a solid, the photoelectron current I_A from X in A is

$$I_A(X) = K \cdot \sigma(h\nu{:}E_B) \cdot \beta(h\nu{:}X) \cdot \bar{N}_A \cdot \lambda_M(E_x)$$
$$\cdot \cos\theta \qquad (8)$$

where $\sigma(h\nu{:}E_B)$ is the photoelectric cross section for ionization of X (binding energy E_B) by photons of energy $h\nu$; $\beta(h\nu{:}X)$ the asymmetry parameter for emission from X by excitation with photons of energy $h\nu$; $\bar{N}_A$ the atomic density of A averaged over depth of analysis; $\lambda_M(E_x)$ the inelastic mean free path in matrix M containing A, at kinetic energy E_x, where $E_x = h\nu - E_B$; and θ the angle of emission to surface normal. K is a constant of proportionality that contains fixed operational parameters such as incident X-ray flux, transmission of analyzer at kinetic energy E_x, and efficiency of detector at kinetic energy E_x, kept fixed during any one analysis.

Values of the total cross section σ for Al K_α radiation, relative to the carbon $1\,s$ level, have been calculated by SCOFIELD [20] (Fig. 16), and of the asymmetry parameter β by REILMAN et al. [21] (Fig. 17, plotted by SEAH [22]). SEAH and DENCH [12] have compiled many measurements of the inelastic mean free path, and for elements the best-fit relationship that they found was

$$\lambda_M(E) = 538E^{-2} + 0.41\,(aE)^{1/2} \text{ monolayers} \qquad (9)$$

where a is the monolayer thickness in nanometers.

In principle, therefore, the surface concentration of an element can be calculated from the intensity of a particular photoelectron emission, according to Equation (8). In practice, the *method of relative sensitivity factors* is in common use.
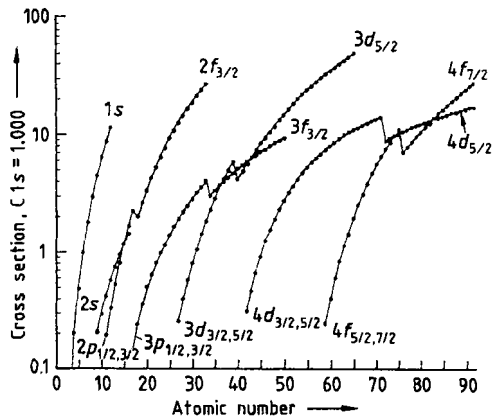
Figure 16. Photoionization cross-section values, for Al K_α radiation, relative to that of the C 1 s taken as unity [20]

Figure 17. Values of the asymmetry parameter β (Eq. 8) for Al K_α radiation [21]

The intensity of a particular elemental photoelectron peak, normally the F 1 s, is taken as standard and put equal to unity, and all intensity measurements are referred to it. Thus, from Equation (8), the ratio $I_A(X)/I_F(1\,s)$, with $I_F(1\,s)$ equal to unity, is

$$I_A(X) = \left[\frac{\sigma(h\nu{:}E_x) \cdot \beta(h\nu{:}X) \cdot \lambda_A(E_x)}{\sigma(h\nu{:}E_{1s}) \cdot \beta(h\nu{:}\,1s) \cdot \bar{N}_F \cdot \lambda_F(E_{1s})} \right] \bar{N}_A \quad (10)$$

When $I_A(X)$ is measured from pure A and a well-characterized fluorine standard is used, the expression in square brackets is the relative sensitivity factor of A. Tables of such empirically derived sensitivity factors have been published by several authors, the most comprehensive being that in [23].

27.2.4.2. Depth Profiling

Very often, in addition to analytical information about the original surface, information is required about the distribution of chemical composition to depths considerably greater than the inelastic mean free path. Nondestructive methods such as variation of emission angle or variation of exciting photon energy are possible, but they are limited in practice to evaluation of chemical profile within depths of only ca. 5 nm. To obtain information from greater depths requires destructive methods, of which the one used universally is removal of the surface by ion bombardment, also called sputtering. The combination of removal of the surface by sputtering and of analysis by either

XPS, AES, or SSIMS is termed depth profiling. In AES and SSIMS the sputtering and analytical operations can occur simultaneously, so that a continuous composition profile is obtained if a single element is being monitored, or a quasi-continuous one if several elements are monitored. In XPS, which is inherently slower and where chemical information is required along with the depth, profiling is recorded in a stepwise fashion, with alternate cycles of sputtering and analysis. Examples of profiles through oxide films on pure iron and on Fe–12Cr–1Mo alloy are shown in Figure 18, in which the respective contributions from the metallic and oxide components of the iron and chromium spectra have been quantified [24]. In these examples the oxide films were only ≈ 5 nm thick on iron and ≈ 3 nm thick on the alloy.

Current practice in depth profiling is to use positively charged argon ions at energies between 0.5 and 10 keV, focused into a beam of 2–5-μm diameter, which is then rastered over the area to be profiled. In XPS the bombarded area should be significantly greater than the analyzed area so that only the flat bottom of the sputtered crater is analyzed. Ion current densities are variable between 5 and 50 μA/cm². Uniformity of sputtering, and therefore the depth resolution of the profile, can be improved by rotating the sample during bombardment.

During the sputtering process, many artifacts are introduced that can affect quantification; some of these derive from the nature of the sample itself (e.g., roughness, crystalline structure, phase distribution, electrical conductivity), and others are ra-
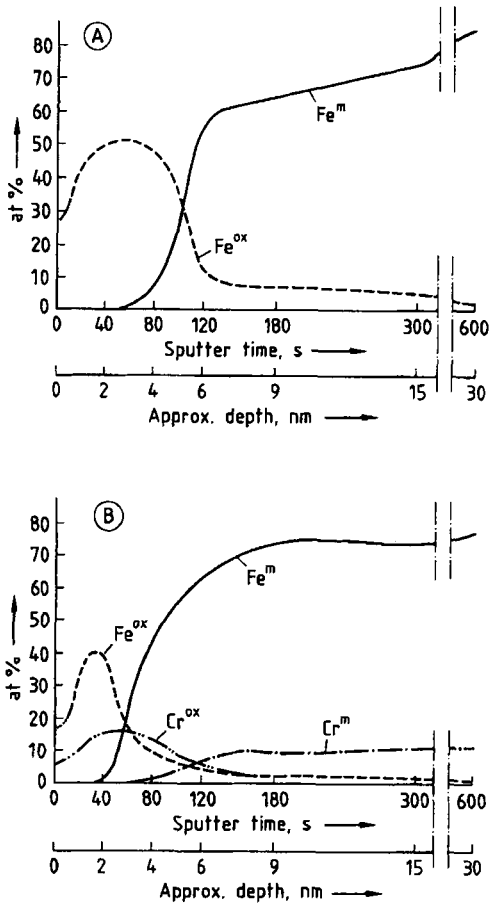
Figure 18. Depth profiles of thin (3–5-nm) oxide films [24] on
A) Pure iron; B) Fe–12 Cr–1 Mo alloy

diation induced (e.g., atomic mixing, preferential sputtering, compound reduction, diffusion and segregation, sputter-induced topography). This is an extensive subject in its own right; a discussion of all aspects is given in [25].

27.2.5. The Auger Parameter

Section 27.2.3 shows that in an XPS spectrum, X-ray-excited Auger peaks are often as prominent as the photoelectron peaks themselves. For many elements, the chemical shifts in Auger peaks are actually greater than the shifts in photoelectron peaks. The two shifts can be combined in a very useful quantity called the Auger Parameter α^*, first used by WAGNER [26] and defined in its modified form [27] as

$$\alpha^* = E_K(jkl) + E_B(i) \tag{11}$$

where $E_K(jkl)$ is the kinetic energy of the Auger transition jkl, and $E_B(i)$ is the binding energy of an electron in level i. The Auger and photoelectron peak energies must be measured in the same spectrum.

The Parameter α^* is independent of photon energy and has the great advantage of being independent of any surface charging. It is therefore measurable for all types of material from metals to insulators. Extensive tabulations of α^* have appeared, based on a large number of standard materials, and the most reliable have recently been collected by WAGNER [28].

As defined in Equation (11), α^* is purely empirical; any prominent and conveniently situated Auger and photoelectron peaks can be used. If α^* is defined slightly more rigorously, by replacing $E_B(i)$ by $E_B(j)$ in Equation (11), it is related closely to a quantity called the extraatomic relaxation energy R^{ea}, since $\Delta\alpha^*$ can be shown to equal $2\Delta R^{ea}(j)$. In other words, the change in Auger Parameter between two chemical states is equal to twice the change in the extraatomic relaxation energies in the two states, associated with the hole in the core level j. Thus α^* has an important physical basis as well as being useful analytically.

27.2.6. Applications

XPS has been applied in virtually every area in which the properties of surfaces are important. Given below are a few selected examples from some typical areas; for more comprehensive discussions of the applications of XPS, see [1], [3]–[9], [23], [29]–[33].

27.2.6.1. Catalysis

An understanding of the basic mechanisms of catalysis has been one of the major aims of XPS from its inception. To this end the technique has been used in two ways: (1) in the study of "real" catalysts closely resembling those used industrially; (2) in the study of "model" catalysts, where the number of parameters has been reduced by the use of well-characterized crystal surfaces.

For industrial catalysts the chemical information obtainable from XPS is all important, since the course and efficiency of the catalyzed reaction will be governed by the states of oxidation of the elements at the surface. Thus XPS is used to an-
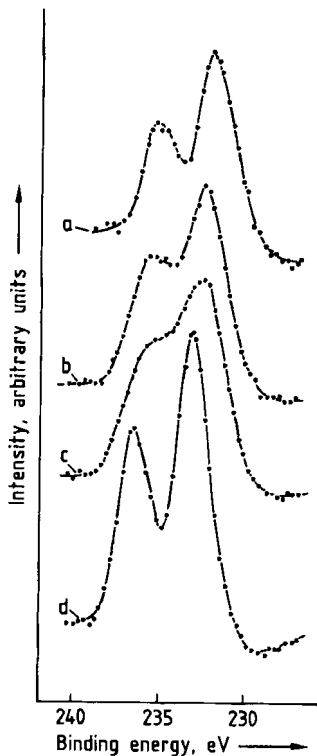
Figure 19. The Mo 3 d XPS spectra of model catalysts in the Mo – Sm – O system [34]
Mo : Sm ratios are a) 0.25; b) 1.33; c) 4.00; d) 8.00

alyze both elemental and chemical composition as a function of bulk composition, of surface loading where the active material is deposited on a support, of oxidation and reduction treatments, and of time in reactor under operating conditions. Figure 19 demonstrates the effect on the chemical state of molybdenum in the surface of Mo – Sm – O catalysts of various compositions [34]. For the highest concentration of molybdenum (spectrum d) the Mo 3 d doublet is very similar to that of MoO₃, but at the samarium-rich composition (spectrum a), it is more like that of a molybdate or a mixed oxide. The effect of successive reduction treatments on a cobalt – molybdenum – alumina catalyst is shown in the changes in the Mo 3 d spectra of Figure 20 [35]. The air-fired catalyst in spectrum a contains entirely Mo(VI) at the surface, but with repeated hydrogen reduction the spectrum changes radically and can be curve resolved (Fig. 21) into Mo(VI), (V), and (IV) contributions. The chemical state of the surface in spectrum f contains approximately equal proportions of the three oxidation states.
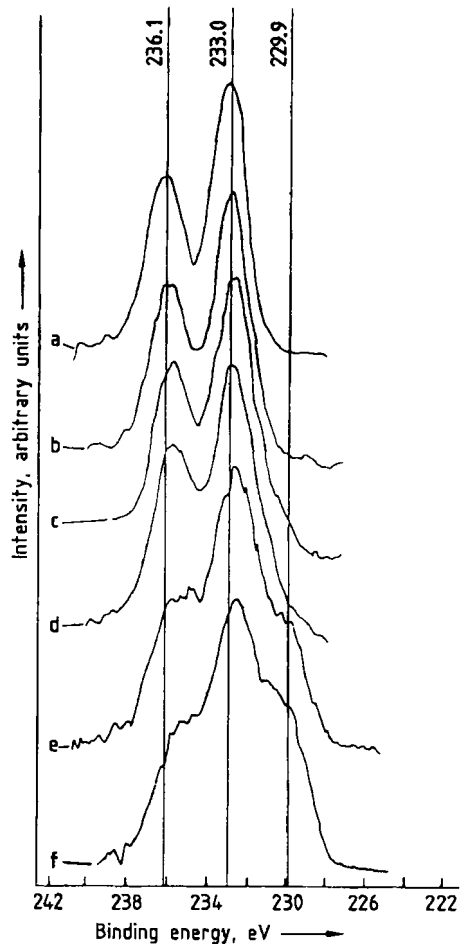
Figure 20. Changes in the Mo 3 d XPS spectrum from a cobalt – molybdenum – alumina catalyst during successive reduction treatments in hydrogen at 500 °C [35]
a) Air-fired catalyst; b) Reduction time 15 min; c) 50 min; d) 60 min; e) 120 min; f) 200 min

Other recent applications of XPS to catalytic problems include catalyst – support interactions [36] – [38], variation of surface catalyst loading [39] – [42], effect of reduction [43] – [45], and characterization of catalyst active surface species [46] – [51].

27.2.6.2. Corrosion and Passivation

Corrosion products that are formed as thin layers on metal surfaces in either aqueous or gaseous environments, along with the nature and stability of passive and protective films on metals and alloys, have also been major areas of application of
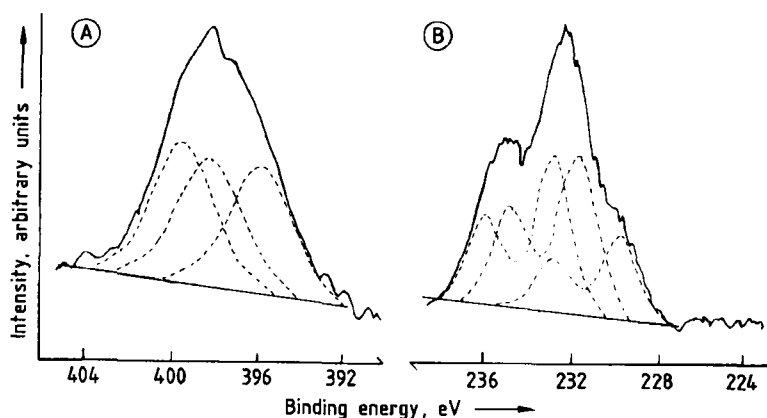
Figure 21. Examples of curve resolution of the Mo 3*p* (A) and 3*d* (B) spectral envelopes carried out by PATTERSON et al. [35] during the catalyst reduction treatment described in Figure 20

XPS. Just as in catalysis, XPS has been used in two ways, one in which materials corroded or passivated in the natural environment are analyzed, and another in which well-characterized, usually pure metal surfaces are studied after exposure to controlled conditions. Unlike catalysis, however, in which the two approaches tend to run parallel and not to overlap, in corrosion significant overlap and feedback from one to the other occur.

More effort has probably been devoted to study of the corrosion and passivation properties of Fe – Cr – Ni alloys such as stainless steels and other transition-metal alloys than to any other metallic system. The type of spectral information obtainable from an Fe – Cr alloy of technological origin, carrying an oxide and contaminant film after corrosion, is shown schematically in Figure 22 [52]. Actual XPS spectra recorded from a sample of the alloy Stellite-6 (a Co – Cr alloy containing some tungsten) treated in lithiated and borated water at 300 °C are shown as a function of treatment time in Figure 23. Stellite-6 is a hard alloy used in nuclear reactors. The high-temperature aqueous solution simulates light-water reactor conditions. The spectra reveal the removal of tungsten from the surface with increasing treatment time, a decrease in cobalt, and the appearance of nickel and iron deposited from stainless steel components elsewhere in the water circuit.

Spectra from an Fe – Cr alloy first polished, then passivated, and finally ion etched to remove the passive film are shown in Figure 24. Each spectrum has been resolved into the inelastic background and the various oxidation states of the element. The authors, MARCUS and OLEFJORD [53]

were even able to separate the oxide and hydroxide contributions to the chromium spectrum from the passive film.

An example [54] of typical spectra recorded during comparison of the gaseous oxidation of pure chromium and of an Ni – Cr alloy is shown in Figure 25. There the Cr 2*p* spectra from the two materials have been recorded as a function of exposure to oxygen. Note that pure chromium does not oxidize as rapidly at room temperature as chromium in the alloy.

XPS has also been applied recently in this field to copper films for the study of microbial effects [55], the behavior of UO_2 in mineral water [56], corrosion-resistant amorphous alloys [57] – [59], and the hydrolysis of fluorozirconate glasses [60].

27.2.6.3. Adhesion

The adhesion of metal and ink to polymer, as well as the adhesion of paint and coating to metal, are of vital importance in several technologies. In the aircraft industry, aluminum-to-aluminum adhesion is being employed increasingly. The strength and durability of an adhesive bond depend absolutely on the way in which the adhesive compound interacts with the surfaces to which it is supposed to adhere, which in turn often involves pretreatment of the surfaces to render them more reactive. The nature and degree of the reactivity are functions of the chemical states of the adhering surfaces, states that can be monitored by XPS.

In electronic packaging technology, metal – polymer adhesion has been important for some years. Many studies have been devoted to the in-
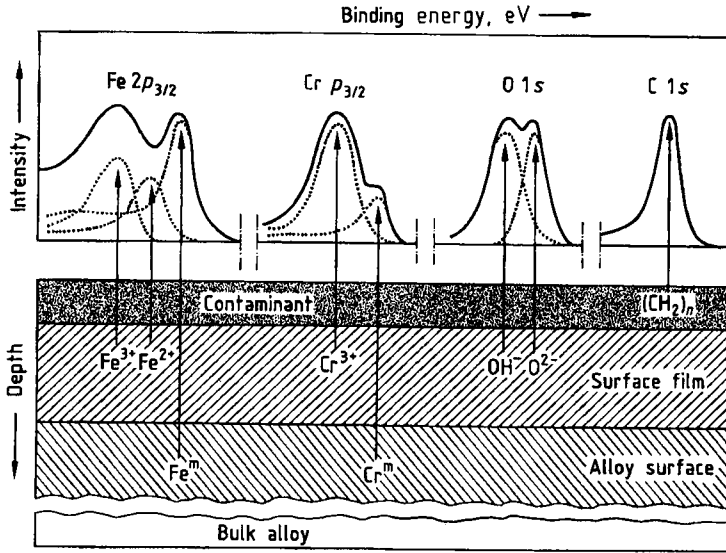
Figure 22. Schematic of type of information obtainable from XPS spectra from an Fe – Cr alloy with an oxide film underneath a contaminant film [52]
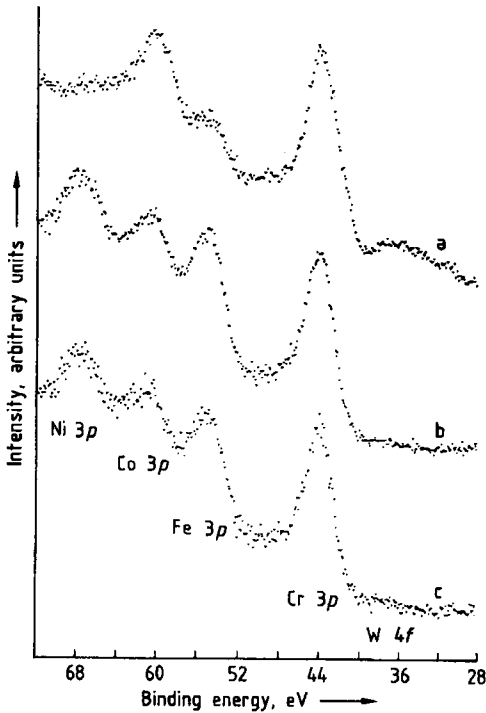
Figure 23. Transition-metal XPS spectra recorded from the surface of the alloy Stellite-6 after treatment for increasing lengths of time in lithiated and borated water at 300 °C [52] a) 4 h; b) 168 h; c) 360 h

teractions between metals and polymers, and that of CHIN-AN CHANG et al. [61] is a good example. They deposited 200-nm-thick films of copper, chromium, titanium, gold, and aluminum on various fluorocarbon polymer films, and found that adhesion in terms of peel strength was highest for titanium. Figure 26 shows the reason why on the perfluoroalkoxy polymer tested, the C 1 s spectra after deposition reveal that for titanium, and to a lesser extent chromium, a strong carbide bond is formed as evidenced by the peak at 282 – 283 eV. With the other metals, very low adhesion was found, and no carbide-like peaks occurred.

The path of failure of an adhesive joint can give information about the mechanism of failure if analysis of the elemental and chemical composition can be carried out along the path. Several authors have performed such analyses by loading the adhesive joint until it fractures and then using XPS to analyze each side of the fracture. Long-term weathering of the joint can be simulated by immersion in water at elevated temperature. The work of BOERIO and ONDRUS [62] demonstrates the nature of the results obtainable, an example of which is given in Figure 27. Silicon, carbon, and oxygen spectra were recorded from each side of the fracture face of an iron –epoxy resin joint before and after weathering in water for 7 d at 60 °C. Before weathering the failure path was as much
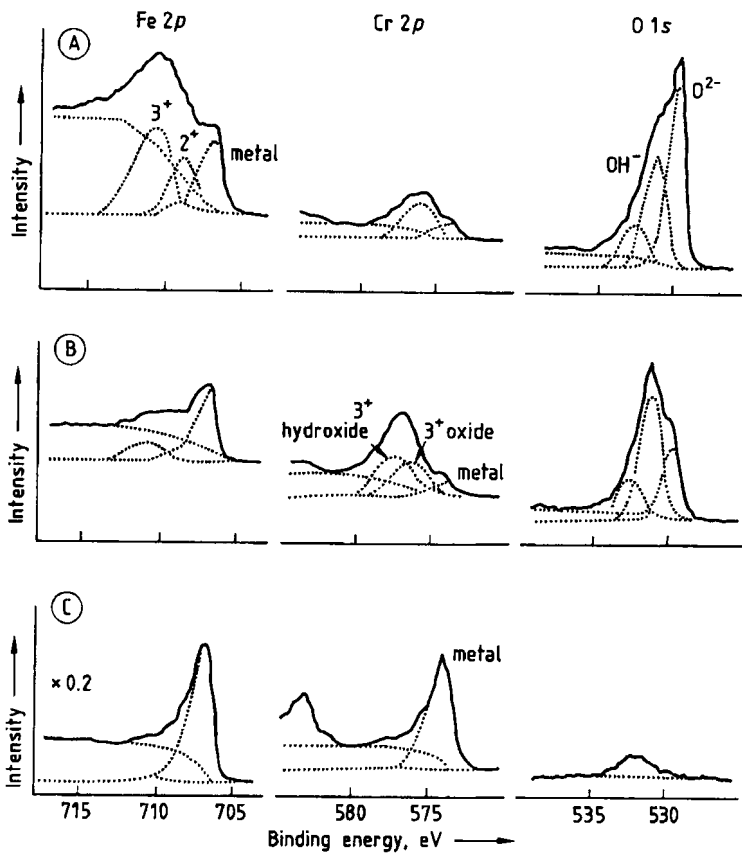
Figure 24. XPS spectra from an Fe–Cr alloy [53]
A) Mechanically polished, exposed in H_2O and air; B) Passivated in 0.5 M H_2SO_4 at 740-mV standard hydrogen electrode for 1 h;
C) Ion etched $\approx$ 10 nm

through the adhesive itself as through the iron oxide – epoxy resin interface, but after weathering, failure occurred almost entirely near the interface.

Also studied by XPS recently have been the measurement of energies of adhesion of aluminum and silicon on molybdenum [63], the ion-enhanced adhesion of nickel to glassy carbon [64], the effects of pretreatments on paint – metal adhesion [65], and the adhesion of titanium thin films to various oxide surfaces [66].

27.2.6.4. Superconductors

In the years since the discovery of superconducting oxides in 1986, more papers have probably been published describing the application of XPS to superconductors than in any other field. One reason for this frenzied activity has been the search for the precise mechanism of superconduc-

tivity in new materials, and an essential piece of information concerns the chemical states of the constituent elements, particularly copper, which is common to all of them. Some theories predict that small but significant amounts of the Cu^{3+} state should occur in superconductors, but other models disagree. Comparison by STEINER et al. [67] of the Cu $2p_{3/2}$ spectra from copper compounds containing copper as Cu^0 (metal), Cu^+, Cu^{2+}, and Cu^{3+}, with the spectra from the superconducting oxide $YBa_2Cu_3O_7$ and from the base material La_2CuO_4, showed conclusively that the Cu^{3+} contribution to the superconducting mechanism must be negligible. Some of their comparative spectra are shown in Figure 28. The positions of the Cu $2p_{3/2}$ peak in $YBa_2Cu_3O_7$ and La_2CuO_4, along with the satellite structure, are very similar to that in CuO (i.e., Cu^{2+}, rather than in $NaCuO_2$, which contains Cu^{3+}).
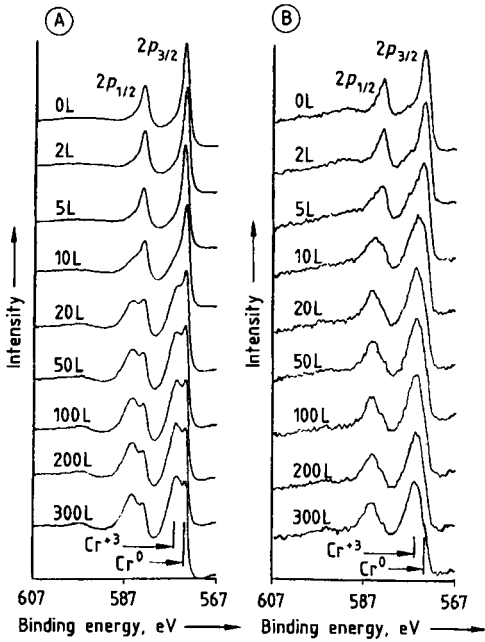
Figure 25. The Cr 2p XPS spectra during exposure to oxygen at room temperature [54] of
A) Pure Cr; B) Ni – Cr alloy
L = Langmuir (10^{-6} Torr · s)
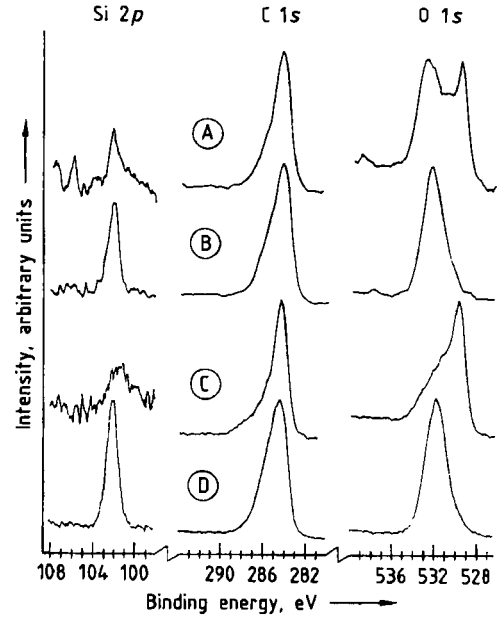
Figure 27. The Si 2p, C 1s, and O 1s XPS spectra from the adherend and adhesive sides of an iron – epoxy resin joint before and after weathering treatment [62]
A) and B) Adherend and adhesive before weathering; C) and D) Adherend and adhesive after immersion in water at 60 °C for 7 d
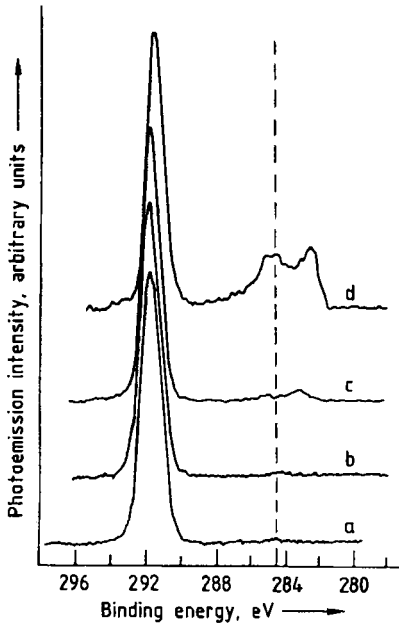
Figure 26. The C 1s XPS spectra recorded by CHIN-AN CHANG et al. [61] from perfluoroalkoxy polymer (PFA)
a) Before deposition; b) Deposition of copper; c) Deposition
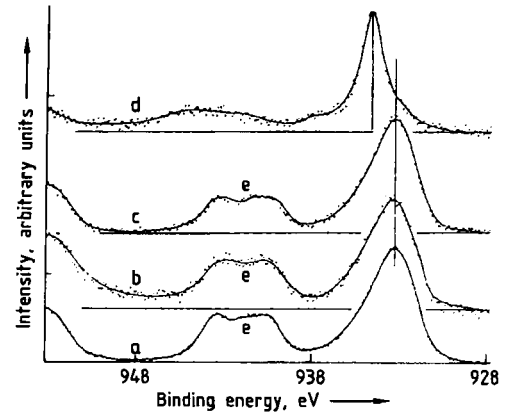
Figure 28. Comparison of the Cu 2$p_{3/2}$ and satellite XPS spectra from several copper compounds with the spectrum from the superconducting oxide YBa$_2$Cu$_3$O$_7$ [67]
a) CuO, Γ = 3.25 eV; b) La$_2$CuO$_4$, Γ = 3.30 eV;
c) YBa$_2$Cu$_3$O$_7$, Γ = 3.20 eV; d) NaCuO$_2$, Γ = 1.60 eV;
e) Cu^{2+} satellites
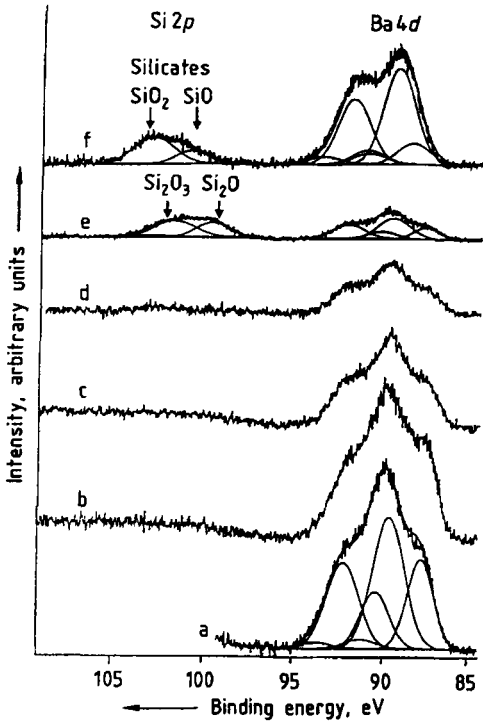
Figure 29. The Si $2p$ and Ba $4d$ XPS spectra [68] from the
a) Clean surface of the superconducting oxide $YBa_2Cu_3O_{7-x}$;
b) – e) Same surface after increasing deposition of b) 0.8 nm
Si; c) 2.8 nm Si; d) 4.4 nm Si; e) 6.0 nm Si; f) Deposited
surface after heating to 250 °C in oxygen (8×10^{-3} mbar)

Recent effort in superconductor research has
been devoted mostly to attempts to fabricate thin
films with the same superconducting properties as
the bulk material, and to prepare suitable electrical
contact to the surfaces of superconductors. XPS
has been of great use in both areas through its
ability to monitor both composition and chemical
state during the processes either of thin supercon-
ducting film formation, or of the deposition of thin
conducting and semiconducting films onto a su-
perconducting surface. Typical of the latter studies
is that of ZIEGLER et al. [68], who deposited in-
creasing amounts of silicon onto epitaxial films of
$YBa_2Cu_3O_{7-x}$. Figure 29 from their paper shows
the Si $2p$ and Ba $4d$ spectra during the deposition
of silicon and after heating the films in oxygen at
250 °C once 6 nm of silicon had accumulated. The
Ba $4d$ spectrum at the bottom can be resolved into
three $4d_{5/2,3/2}$ doublets, corresponding — with in-
creasing energy — to barium in the bulk supercon-
ductor, in the surface of the superconductor, and as
carbonate. With increasing silicon thickness the

barium spectrum attenuates but does not change
in character, but silicon does not appear in the
spectrum until nearly 5 nm has been deposited,
suggesting migration into the superconductor at
room temperature. At a thickness of 6 nm, the Si
$2p$ spectrum can be separated into suboxide con-
tributions, indicating reaction of silicon with the
$YBa_2Cu_3O_{7-x}$ surface. After heating in oxygen the
Si $2p$ spectrum changes to that characteristic of
SiO_2 and silicates, while the Ba $4d$ spectrum also
changes by increasing in intensity and conforming
mostly to that expected of a barium silicate. As a
result of the latter changes the superconducting
properties of the film were destroyed. The Y $3d$
and Cu $2p$ spectra establish that yttrium and
copper oxides are formed as well.

Other surface reactions studied have been
those of gold [69] and silver [70] with Bi_2Sr_2Ca-
Cu_2O_{8+y}, of niobium with YBa_2Cu_3-O_{7-x} [71],
and of lead with $YBa_2Cu_3O_{7-x}$ [72]. Character-
istics of superconductors in the form of thin films
prepared in several ways have been recorded by
XPS for $YBa_2Cu_3O_{7-x}$ [73]–[77], $GdBa_2Cu_3O_{7-x}$
[78], $Nd_{2-x}Ce_xCuO_{4-y}$ [79], $Rb_xBa_{1-x}BiO_3$ [80],
and $Tl_2Ca_2Ba_2Cu_3O_{10}$ [81]. Detailed electronic
studies have been made of the bulk superconduc-
tors $Bi_2Sr_2Ca_{1-x}Nd_xCu_2O_y$ [82]–[83], $Nd_{2-x}Ce_x$-
CuO_{4-y} [84]–[87], $ErBa_2Cu_4O_8$ [88], and
$Tl_2Ba_2CaCu_2O_8$ [89].

27.2.6.5. Interfaces

The chemical and electronic properties of ele-
ments at the interfaces between very thin films and
bulk substrates are important in several technolog-
ical areas, particularly microelectronics, sensors,
catalysis, metal protection, and solar cells. To
study conditions at an interface, depth profiling
by ion bombardment is inadvisable since both
composition and chemical state can be altered by
interaction with energetic positive ions. Normal
procedure is therefore to start with a clean or other
well-characterized substrate and deposit the thin
film onto it slowly at a chosen temperature while
XPS monitors the composition and chemical state
by recording selected characteristic spectra. The
procedure continues until no further spectral
changes occur, as a function either of film thick-
ness, of time elapsed since deposition, or of
changes in substrate temperature.

A good example is the study of the reaction at
the interface of the rare-earth element thulium
with the silicon (111) surface, carried out by GOK-
HALE et al. [90]. Figure 30 shows the Si $2p$ spec-
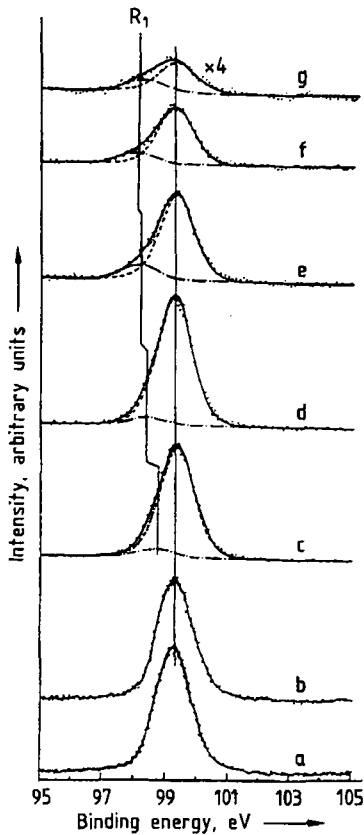
Figure 30. The Si $2p$ XPS spectra with Mg K$_\alpha$ excitation during deposition of thulium to a thickness of 12.5 nm at room temperature [90]
a) Clean silicon; b) 0.2 nm Tm; c) 0.5 nm Tm; d) 2 nm Tm; e) 4 nm Tm; f) 6.5 nm Tm; g) 12.5 nm Tm

trum from the substrate as thulium is deposited on it at room temperature. With the Mg K$_\alpha$ radiation used for the analysis, the kinetic energy of the Si $2p$ electrons is about 1155 eV, corresponding to an inelastic mean free path of about 2 nm. The Si $2p$ peak is still plainly detectable after 12.5 nm of thulium have been deposited, and in addition a new feature appears on the low binding energy side of the peak, increasing in magnitude and shifting to lower energy with thickness. Raising the temperature in stages of the thickest film and use of X-ray diffraction, established that the energetic final position labeled R$_1$ corresponded to Tm$_5$Si$_3$ and the initial energetic position found at a thickness of 0.5–1.0 nm, to TmSi$_2$. However, the magnitude of the residual Si $2p$ peak was too great at all stages to be accounted for by compound formation alone, indicating the formation

of thulium clusters followed by their coalescence into islands. Measurement of the Tm $4d_{5/2}$ binding energy during the interaction also showed that the valence of thulium was 3+ at all temperatures and coverages.

Recent studies of interfaces by XPS have also included those of SiO$_2$ [91], [92], samarium [93], and amorphous hydrogenated carbon [94] on silicon; of silver, indium, and aluminum on GaP [95]; of tin [96] and RbF–Ge interlayers [97] on GaAs; of platinum on TiO$_2$ [98]; of titanium, niobium, and nickel on Al$_2$O$_3$[99]; of chromium and iron on W(110) [100]; and of ZrO$_2$ on nickel [101].

Other important areas of application of XPS have included polymers [102]–[106], lubricating films [107], [108], glasses [109]–[113], ion implantation [114]–[118], and cleaning [119]–[121] and passivation [122]–[123] of semiconductor surfaces.

27.3. Auger Electron Spectroscopy (AES)

After XPS, AES is the next most widely used surface analytical technique. As an accepted surface technique, AES actually predates XPS by two to three years, since the potential of XPS as a surface-specific technique was not recognized immediately by the surface science community. Pioneering work was carried out by HARRIS [124] and by WEBER and PERIA [125], but the technique as it is known today is basically the same as that established by PALMBERG et al. [126].

27.3.1. Principles

The surface to be analyzed is irradiated with a beam of electrons of sufficient energy, typically in the range of 5–30 keV, to ionize one or more core levels in surface atoms. After ionization the atom can then relax by either of the two processes described in Section 27.2.1 for XPS, that is, ejection of a characteristic X-ray photon (fluorescence) or ejection of an Auger electron. Although these are competing processes, for shallow core levels ($E_B < 2$ keV) the probability of the Auger process is far higher. The Auger process is described schematically in Figure 1, which points out that the final state of the atom is doubly ionized and that the kinetic energy of the electron arising from the Auger transition ABC is given by

$$E_{ABC} = E_A - E_B - E_C^* \qquad (12)$$

where E_A, E_B, and E_C are binding energies of the A, B, and C levels in the atom, E^*_C being starred because it is the binding energy of C modified by an existing hole in B.

Equation (12) shows that the Auger energy is a function only of atomic energy levels. Since no two elements have the same set of atomic binding energies, analysis of Auger energies provides elemental identification. Even if levels B and C are in the valence band of the solid, analysis is still possible because the dominant term in Equation (12) is always the binding energy of A, the initially ionized level.

Equation (12) also shows that the heavier the element (i.e., the greater the number of atomic energy levels), the more numerous are the possible Auger transitions. Fortunately, large differences exist in the probabilities between different Auger transitions, so that even for the heaviest elements, only a few intense transitions occur, and analysis is still possible.

In principle, chemical as well as elemental information should be available in AES, since the binding energies appearing in Equation (12) are subject to the same chemical shifts as measured in XPS. In practice, since the binding energies of three levels are involved, extracting chemical information from Auger spectra is, in most cases, still too difficult. However, significant progress is being made in this area, and improved theory and modern data processing methods are likely to enable valuable chemical information to be derived soon.

The reasons AES is one of the surface-specific techniques have been given in Section 27.2.1, with reference to Figure 2. The normal range of kinetic energies recorded in an AES spectrum would typically be from 20 to 1000 eV, corresponding to inelastic mean free path values of 2 to 6 monolayers.

The nomenclature used in AES has also been mentioned in Section 27.2.1. The Auger transition in which initial ionization occurs in level A, followed by the filling of A by an electron from B and ejection of an electron from C, would therefore be labeled ABC. In this rather restricted scheme, one would thus find in the KLL series the six possible transitions KL_1L_1, KL_1L_2, KL_1L_3, KL_2L_2, KL_2L_3, and KL_3L_3. Other combinations could be written for other series such as the LMM, MNN, etc.

27.3.2. Instrumentation

27.3.2.1. Vacuum Requirements

The same considerations discussed in Section 27.2.2.1 for XPS apply to AES.

27.3.2.2. Electron Sources

The energy of the exciting electrons does not enter into Equation (12) for the Auger energy, unlike that of the exciting X-ray photon in XPS; thus the energy spread in the electron beam is irrelevant. What is relevant is the actual energy of the primary electrons because of the dependence on that energy of the cross section for ionization by electron impact. For an electron in a core level of binding energy E_B, this dependence increases steeply with primary energy above E_B, passes through a maximum about $3-5$ times E_B, and then decreases to a fairly constant plateau. It is illustrated in Figure 31, in which various theoretical and experimental cross sections are shown for ionization of the Ni K shell, from CASNATI et al. [127]. The implication of this dependence is that for efficient ionization of a particular core level, the primary energy should be about five times the binding energy of an electron in that level. Since in most samples for analysis, several elements are found in the surface region with core-level binding energies extending over a large range, choice of a primary energy that is sufficiently high to ionize all the core levels efficiently is advisable. Hence for *conventional AES*, primary energies are typically in the range $5-10$ keV. In *scanning Auger microscopy (SAM)*, much higher primary energies are used, in the range of $25-50$ keV, because in that version of AES as small an electron spot size must be achieved on the sample as possible, and the required focusing can be effected only at such energies.

For the primary energy range $5-10$ keV used in *conventional AES*, the electron emitter is thermionic, usually a hot tungsten filament, and focusing of the electron beam is carried out electrostatically. Typically, such an electron source would be able to provide a spot size on the specimen of about 0.5 μm at 10 keV and a beam current of about 10^{-8} A. The beam can normally be rastered over the specimen surface, but such a source would not be regarded as adequate for SAM. Sources for *SAM* may be of either the thermionic or the field emission type, the latter being partic-
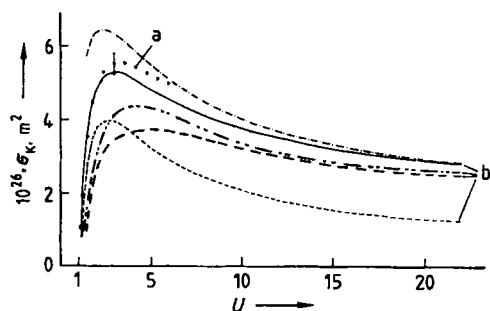
Figure 31. Electron impact ionization cross section for the Ni K shell, as a function of reduced electron energy U [127] $U = E_p/E_K$, E_p being the primary electron energy and E_K the binding energy of the K shell
a) Experimental points; b) Semiempirical or theoretical curves

ularly advantageous in achieving minimum spot sizes since it is a high-brightness source. Focusing is performed electromagnetically, and at optimum performance a modern electron gun for SAM would be capable of a spot size of 15 nm at 30 keV and a beam current of about 10^{-10} A. The beam is rastered over the surfaces at scan rates variable up to TV rates, with the dimensions of the scanned area variable as well.

27.3.2.3. Electron Energy Analyzers

The electron energy analyzer found to be most suitable for AES, the cylindrical mirror analyzer (CMA), has already been described in Section 27.2.2.4. The important property of this analyzer for AES is its very high transmission, arising from the large solid angle of acceptance; for a typical angular spread $\delta\alpha$ of 6°, the transmission is about 14 %. Since the source area (i.e., spot size) in AES is so small, as many Auger electrons as possible must be accepted into the analyzer, compatible with adequate resolution.

An example of the arrangement of a single-stage CMA in a modern system for AES analysis is shown in Figure 32, taken from [128]. The double-pass CMA described in Section 27.2.2.4 (see Fig. 8) can also be used for AES by grounding the spherical retardation meshes and the inner cylinders and by setting the variable internal apertures to minimum. Although the double-pass CMA is not ideal for either AES or XPS, use of the same analyzer, in its different modes of operation, for both techniques is very convenient. Similarly, the concentric hemispherical analyzer

(CHA), also described in Section 27.2.2.4, is often used for both techniques for the sake of convenience, even though it is much more suitable for XPS than AES.

Until recently, all Auger spectra were presented in energy distributions differentiated with respect to energy, rather than in the direct energy distributions used in XPS [i.e., $dN(E)/dE$ rather than $N(E)$]. The effect of differentiating is to enhance the visibility of Auger features, as demonstrated in Figure 33, in which the differential distribution of secondary electron energies from boron is shown above the corresponding $N(E)$ distribution. The differentiation was performed by a synchronous detection technique, in which a small voltage modulation at high frequency was superimposed on the deflecting potential as the latter was scanned, and then the second harmonic of the detected signal was amplified preferentially. An indication of this technique is given in Figure 32. Nowadays the $N(E)$ distribution is usually recorded digitally, and the differentiation is carried out by computer. However, with the rapid improvement in the signal-to-noise characteristics of detection equipment, the trend in presenting Auger spectra is increasingly toward the undifferentiated $N(E)$ distribution.

Note from Figure 33 that although the true position of the boron KLL Auger peak in the $N(E)$ distribution is at 167 eV, the position in the $dN(E)/dE$ distribution is taken for purely conventional reasons to be that of the negative minimum (i.e., at 175 eV). The other spectral features in Figure 33 are described in Section 27.3.3.

27.3.3. Spectral Information

As stated above, the range of primary energies used typically in conventional AES is 5 – 10 keV, and the ionization cross section passes through a maximum at 3 – 5 times the binding energy of an electron in the ionized core level. Therefore the core levels that can be ionized efficiently are limited. Thus K-shell ionization efficiency is adequate up to $Z \approx 14$ (silicon), L_3 up to $Z \approx 38$ (strontium), M_5 up to $Z \approx 76$ (osmium), and so on (for the much higher primary energies used in SAM, the ranges are correspondingly extended). This means that in various regions of the periodic table, characteristic Auger transitions occur that are most prominent under typical operating conditions.
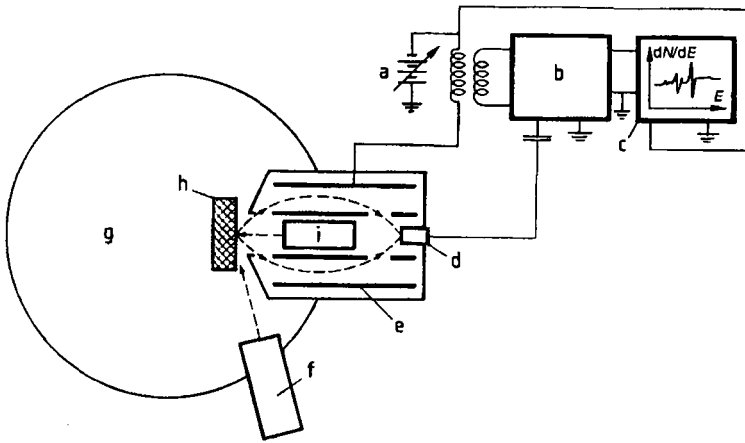
Figure 32. Typical experimental arrangement for AES using a single-pass CMA [128]
a) Voltage sweep; b) Lock-in amplifier; c) Display; d) Electron multiplier; e) Cylindrical mirror analyzer; f) Ion gun; g) UHV chamber; h) Sample; i) Electron gun
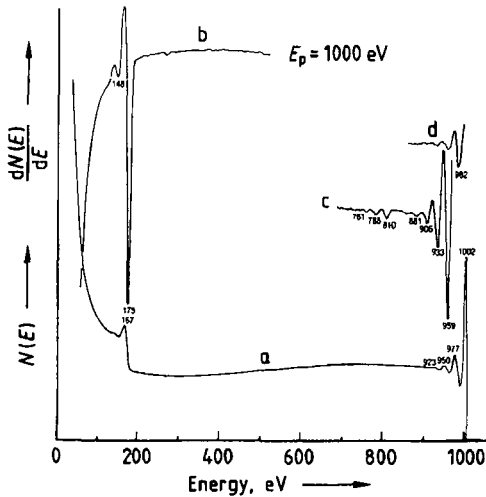
Figure 33. Secondary electron distribution from boron
a) In the $N(E)$ mode; b)–d) In the $dN(E)/dE$ mode, where the differentiation reveals the plasmon satellites

Figures 343536 show some of these prominent Auger features. They have all been taken from [129], and recorded using a CMA, in the differential distribution. Examples of the KLL Auger series are shown in Figure 34, of the LMM series in Figure 35, and of the MNN series in Figure 36. In the KLL series the most prominent feature is a single peak arising from the $KL_{2,3}L_{2,3}$ transition, with minor features from other Auger transitions to lower energy. The LMM series is characterized

by a triplet arising from $L_3M_{2,3}M_{2,3}$, $L_3M_{2,3}M_{4,5}$, and $L_3M_{4,5}M_{4,5}$ transitions in ascending order of kinetic energy, with another intense peak, the $M_{2,3}M_{4,5}M_{4,5}$, appearing at very low energy. The characteristic doublet seen in the MNN series arises from the $M_{4,5}N_{4,5}N_{4,5}$ transitions, in which the doublet separation is that of the core levels M_4 and M_5.

Note that AES transitions involving electrons in valence band levels are often written with a V rather than the full symbol of the level. Thus $L_3M_{2,3}M_{4,5}$ and $L_3M_{4,5}M_{4,5}$ would often appear as $L_3M_{2,3}V$ and L_3VV, respectively, and similarly $M_{2,3}M_{4,5}M_{4,5}$ as $M_{2,3}VV$. In Figure 35the increase in intensity of the L_3VV peak relative to the other two, upon going from chromium to iron, is due to the progressive increase in the electron density in the valence band.

Chemical effects in Auger spectra are quite commonly observed, but difficult to interpret compared to those in XPS since additional core levels are involved in the Auger process. Some examples of the changes to be seen in the KLL spectrum of carbon in different chemical environments are given in Figure 37 [130].

Associated with prominent features in an Auger spectrum are the same types of energy loss feature, the plasmon losses, that are found associated with photoelectron peaks in an XPS spectrum. As described in Section 27.2.3, plasmon energy losses arise from excitation of modes of collective oscillation of the conduction electrons by outgoing secondary electrons of sufficient energy. Successive plasmon losses suffered by the backscattered
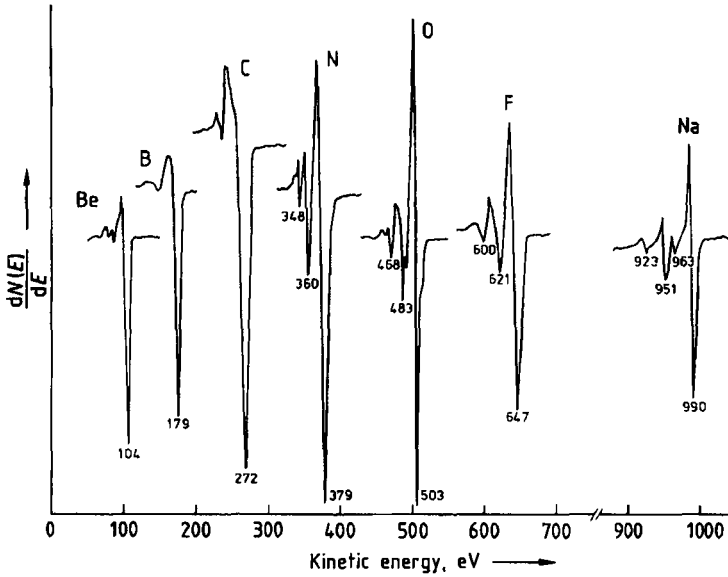
Figure 34. KLL Auger series characteristic of the light elements [129]
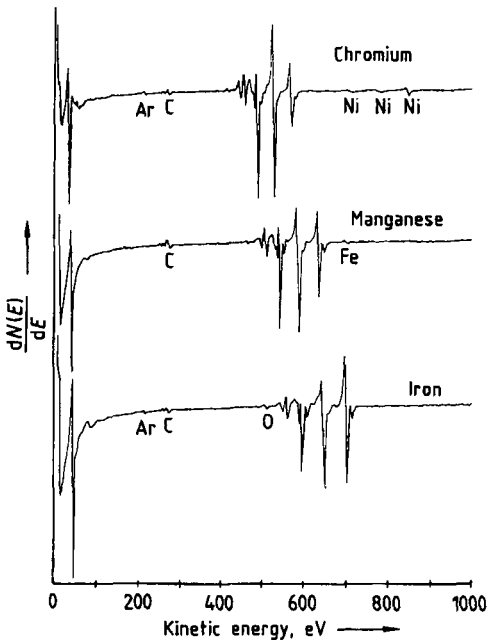
Figure 35. LMM and MMM Auger series in the middle of the first series of transition elements [129]

primary electrons can be seen in both the $N(E)$ and the $dN(E)/dE$ spectra of Figure 33, from boron. The same magnitude of plasmon loss, ≈ 27 eV, is also associated with the KLL Auger peak of boron. Plasmon losses are also present in Figures 34 35 36 37, but they are difficult to disentangle from minor Auger features.

Another type of loss feature not often recorded in Auger spectra can also be seen in Figure 33, in the $dN(E)/dE$ spectrum at 810 eV. This type arises from core level ionization and forms the basis for the technique of core electron energy loss spectroscopy (CEELS). A primary electron interacting with an electron in a core level can cause excitation either to an unoccupied continuum state, leading to complete ionization, or to localized or partly localized final states, if available. Since the primary electron need not give up all its energy in the interaction, the loss feature appears as a step in the secondary electron spectrum, with a tail decreasing progressively to lower energies. Differentiation, as in Figure 33, accentuates the visibility. In that figure, the loss at 810 eV corresponds to the K-shell ionization edge of boron at about 190 eV.
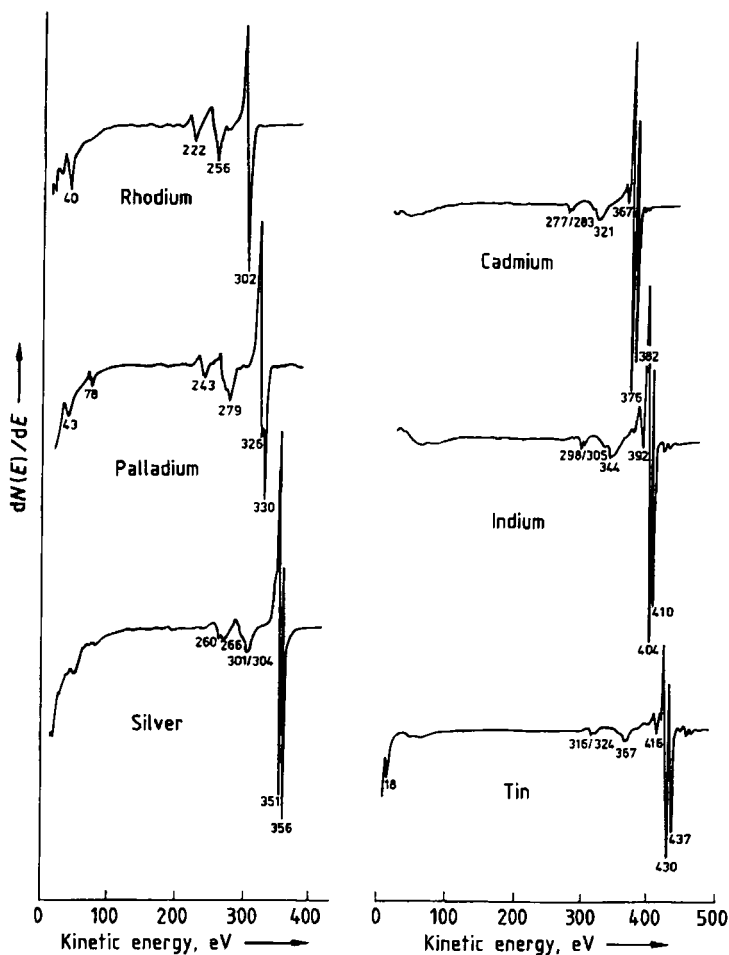
Figure 36. MNN Auger series in the second series of transition elements [129]

27.3.4. Quantification and Depth Profiling

27.3.4.1. Quantification

If ionization of a core level X in an atom A in a solid matrix M by a primary electron of energy E_p gives rise to the current I_{XYZ} of electrons produced by the Auger transition XYZ, then the Auger current from A is

$$I_{XYZ} = K \cdot \sigma(E_p{:}E_B)$$
$$\cdot \{1 + r_M(E_B{:}\alpha)\}\bar{N}_A \cdot \lambda_M(E_{XYZ}) \cdot \cos\theta \qquad (13)$$

where $\sigma(E_p : E_B)$ is the cross section for ionization of level X with binding energy E_B by electrons of energy E_p; r_M the backscattering factor that takes account of the additional ionization of X with binding energy E_B by inelastic electrons with energies between E_p and E_B; α the angle of incident electron beam to surface normal; $\bar{N}_A$ the atomic density of A averaged over depth of analysis; $\lambda_M(E_{XYZ})$ the inelastic mean free path in matrix M containing A, at the Auger kinetic energy E_{XYZ}; and θ the angle of Auger electron emission to surface normal.

Similarly to Equation (8), K is a constant of proportionality containing fixed operational parameters such as incident electron current density, transmission of the analyzer at the kinetic energy E_{XYZ}, efficiency of the detector at the kinetic energy E_{XYZ}, and the probability of the Auger transition XYZ.
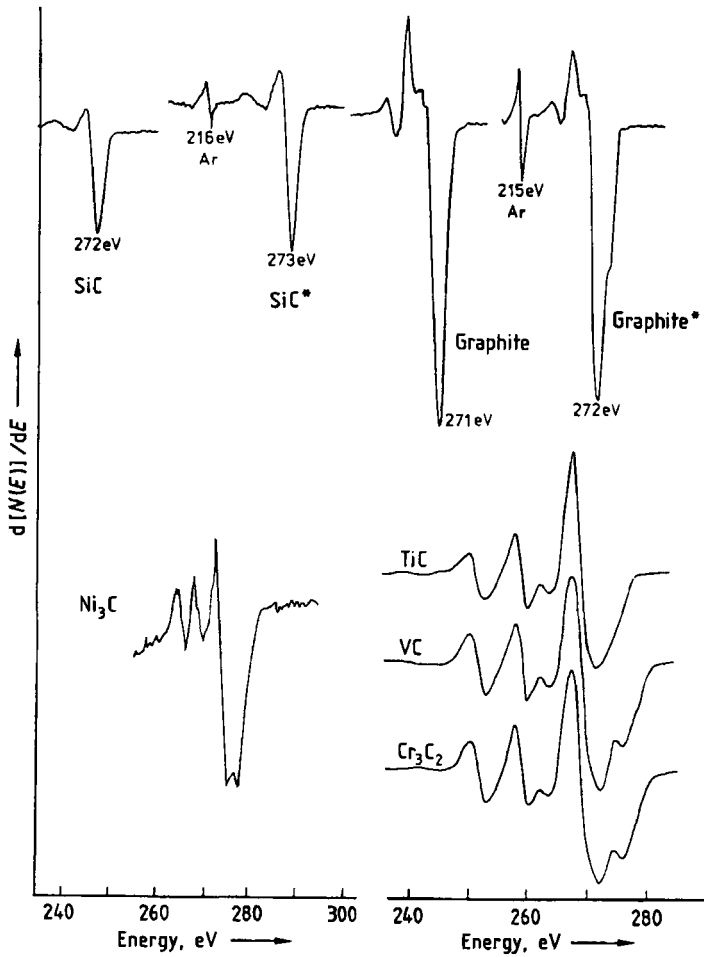
Figure 37. Examples of the effect of different chemical states on the KLL Auger spectrum of carbon [130] (SiC* and graphite* denote Ar$^+$-bombarded surfaces of SiC and graphite, respectively)

The cross section for electron impact ionization has already been mentioned in Section 27.3.2.2 in connection with electron sources, and various experimental and theoretical cross sections have been shown in Figure 31 for the particular case of the K shell of nickel. The expression for the cross section derived by CASNATI et al. [127] gives reasonably good agreement with experiment, with the earlier expression of GRYZINSKI [131] also being useful.

For the inelastic mean free path, Equation (9) given for XPS also applies.

The new factor in Equation (8) compared to that for XPS is the backscattering factor r_M. ICHIMURA and SHIMIZU [132] have carried out extensive Monte Carlo calculations of the backscatter-

ing factor as functions of primary beam energy and angle of incidence, of atomic number and of binding energy, and a selection of their results for $E_p = 10$ keV at normal incidence is shown in Figure 38. The best fit of their results to experiment gives the following relationship:

$$r = \left(2.34 - 2.1Z^{0.14}\right)U^{-0.35}$$
$$+ \left(2.58Z^{0.14} - 2.98\right) \quad \text{for} \quad \alpha = 0^\circ$$

$$r = \left(0.462 - 0.777Z^{0.20}\right)U^{-0.32}$$
$$+ \left(1.5Z^{0.20} - 1.05\right) \quad \text{for} \quad \alpha = 30^\circ$$

$$r = \left(1.21 - 1.39Z^{0.13}\right)U^{-0.33}$$
$$+ \left(1.94Z^{0.13} - 1.88\right) \quad \text{for} \quad \alpha = 45^\circ \quad (14)$$
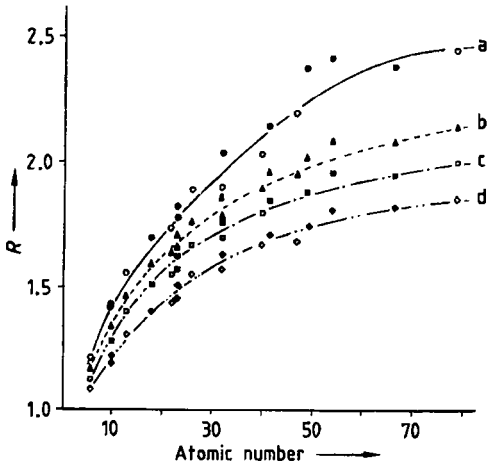
Figure 38. Values of the backscattering factor R (i.e., $1 + r$) calculated as a function of atomic number for an electron beam of 10 keV at normal incidence by ICHIMURA et al. [132] a) $E_B = 0.1$ keV; b) $E_B = 0.5$ keV; c) $E_B = 1.0$ keV; d) $E_B = 2.0$ keV

where Z is the atomic number and $U = E_p/E_B$.

Thus, as for XPS, in principle the average surface concentration $\bar{N}_A$ can be calculated from the measurement of the Auger current, according to Equation (13). Again, as in XPS, relative sensitivity factors are generally used. The Auger current $\bar{N}_A^s$ for the same transition XYZ in a standard of pure A is measured under the same experimental conditions as in the analysis of A in M, whereupon the ratio of Auger currents is

$$\bar{N}_A/\bar{N}_A^s = \frac{I_{XYZ}(1 + r_S)\lambda_s}{I_{XYZ}^s(1 + r_M)\lambda_M} \tag{15}$$

since the ionization cross section σ is the same for both standard and sample. Then, as before, Equation (15) can be reduced to

$$\bar{N}_A = S_A I_{XYZ} \tag{16}$$

where

$$S_A = \frac{\bar{N}_A^s(1 + r_s)\lambda_s}{I_{XYZ}^s(1 + r_M)\lambda_M} \tag{17}$$

S_A is then the relative sensitivity factor. Normally, values of S_A are derived empirically or semiempirically. Tables of such sensitivity factors have been published by PAYLING [133] and by MROCZKOWSKI and LICHTMAN [134] for differential Auger spectra, and for the direct, undifferen-

tiated, spectra by SATO et al. [135]. A comprehensive discussion of the assumptions and simplifications involved and the corrections that should be applied is given in [136], [22]. Because many workers still measure Auger current as the peak-to-peak height in the differential distribution, which takes no account of peak width or area, quantification in AES is still not as accurate as that in XPS.

27.3.4.2. Depth Profiling

As in XPS, elemental distributions near the surface, but to depths greater than the analytical depth resolutions, are frequently required when using AES. Again, the universally employed technique is that of progressive removal of the surface by ion sputtering, with simultaneous or quasi-simultaneous analysis of the exposed surface. Since the peak-to-peak height of an Auger peak in the differential distribution can be recorded very quickly by a CMA, it is normal in AES depth profiling, when only 10 – 12 individual peaks need to be recorded, to perform simultaneous sputtering and analysis. Such an operation is called multiplexing; the CMA is programmed to switch rapidly from one energetic peak position to the next, spectral recording being within a chosen energy window at each position. Should many more peaks be needed, or an entire energy spectrum, the ion erosion must be stopped during acquisition, and the profiling is quasi simultaneous as in XPS.

In modern ion guns using differential pumping, the pressure of argon (the gas normally used) in the analysis chamber can be maintained below 5×10^{-5} Pa during depth profiling, thus reducing the possibility of damage to the electron detector and the electron gun filament in the CMA. Such a gun is shown schematically in Figure 39. The source region around the filament (g) and anode (f) is differentially pumped, and positive ions are extracted through a cone-shaped extractor (b), then accelerated through a small aperture (e) to an energy selected between 0.5 and 5 keV. Focusing occurs in two tubular lenses (c) down to a spot size of 10 μm at 5 keV, and the beam can be rastered by two pairs of deflector plates (d) at the end of the gun. The area of raster on the specimen can be chosen between 1×1 mm² and 10×10 mm². According to the emission from the coated nonburnout iridium filament, the beam current can be selected from 1 to 10 μA.

Since the ion intensity is not uniform across the beam from an ion gun, but approximately Gauss-
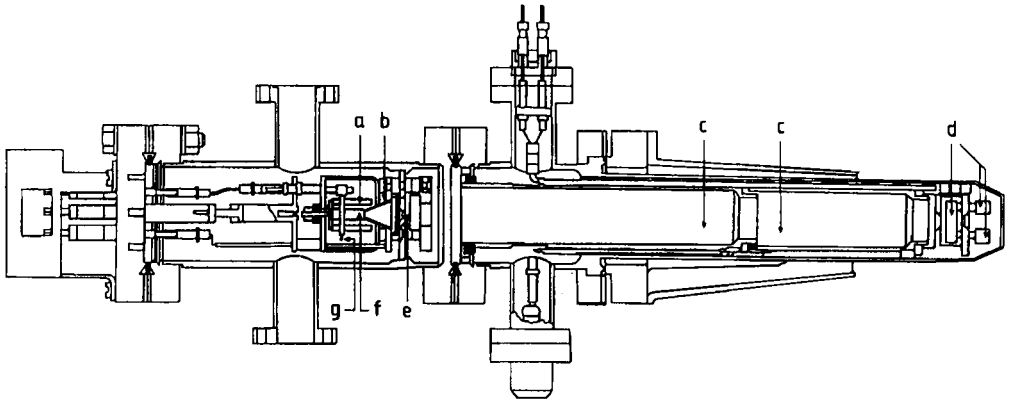
Figure 39. Ion source using a thermionic emitter in a collision chamber for ionization
Courtesy of Leybold
a) Electron repeller; b) Extractor; c) Tubular lenses; d) Deflector plates; e) Aperture; f) Anode; g) Filament

ian in distribution, in AES depth profiling either the area of the focused electron beam on the sample must be coincident with the region of maximum ion intensity in the focused ion beam or, preferably, the ion beam must be rastered. The analyzed area should then be centered within the rastered area to ensure that the flat bottom of the sputtered crater is analyzed and not the sloping sides.

The depth resolution achievable during profiling depends on many variables, and the reader is referred, as before, to the comprehensive discussion in [25].

A variation on depth profiling that can be performed by modern scanning Auger instruments (see Section 27.3.5) is to program the incident electron beam to jump from one preselected position on a surface to each of many others in turn, with multiplexing at each position. This is called *multiple point analysis*, which provides an elemental map after each sputtering step or each period of continuous sputtering. The set of maps can then be related to each other in a computer frame-store system to derive a three-dimensional analysis of a selected micro volume.

27.3.5. Scanning Auger Microscopy (SAM)

If an incident electron beam of sufficient energy for AES is rastered over a surface in a manner similar to that in a scanning electron microscope (SEM), and if the analyzer, usually a

CMA, is set to accept electrons of Auger energies characteristic of a particular element, then a visual display whose intensity is modulated by the Auger peak intensity will correspond to the distribution of that element over the surface. The result is called an *Auger map or image*. Care must be taken in interpreting the intensity distribution, since the Auger intensity depends not only on the local concentration of the element but on the topography as well, because surface roughness can affect the inelastic background underneath the Auger peak. Therefore Auger maps are customarily presented as variations of the ratio of peak intensity divided by the magnitude of the background on the high-energy side, which can be carried out easily by computer.

In SAM the thrust of development is toward ever-better spatial resolution, which means the smallest possible focused spot size on the sample compatible with adequate Auger signal-to-noise ratio in the image, for acquisition within a reasonable length of time. Electrostatic focusing of the incident electron beam is inadequate on its own, and the electron guns used for SAM all employ electromagnetic focusing. Similarly, focusing electrons of low energy into a sufficiently small spot is difficult, so the energies that must be used in SAM are in the range of 25 – 50 keV, with beam currents of the order of 1 nA or less. An electron gun operating at such energies and currents with electromagnetic focusing would be capable of producing a minimum spot size of about 15 nm. An illustration of the application of SAM at moderate resolution is shown in Figure 40 [137]. The surface

was that of a fractured compact of SiC to which boron and carbon had been added to aid the sintering process. The aim of the analysis was to establish the uniformity of distribution of the additives and the presence or absence of impurities. The *Auger maps* show not only very nonuniform distribution of boron (Fig. 40 A) but also a strong correlation of boron with sodium (Fig. 40 C), and a weaker correlation of boron with potassium (Fig. 40 B). Point analyses shown to the right at points A and B marked on the images reveal the presence of sulfur and calcium in some areas as well. In this case the sintering process had not been optimized.

27.3.6. Applications

Like XPS, the application of AES has been very widespread, particularly in the earlier years of its existence; more recently, the technique has been applied increasingly to those problem areas that need the high spatial resolution that AES can provide and XPS as yet cannot. Since data acquisition in AES is faster than in XPS, it is also employed widely in routine quality control by surface analysis of random samples from production lines of, for example, integrated circuits.

27.3.6.1. Grain Boundary Segregation

One of the original applications of AES, and still one of the most important, is the analysis of grain boundaries in metals and ceramics. Very small amounts of impurity or dopant elements in the bulk material can migrate under appropriate temperature conditions to the boundaries of the grain structure and accumulate there. In that way the concentration of minor elements at the grain boundaries can become much higher than in the bulk, and the cohesive energy of the boundaries can be so altered that the material becomes brittle. Knowledge of the nature of the segregating elements and their grain boundary concentrations as a function of temperature can be used to modify fabrication conditions and thereby improve the strength performance of the material in service.

GORETZKI [138] has discussed the importance and the analysis of internal surfaces such as grain boundaries and phase boundaries, and given several examples. He emphasizes that to avoid ambiguity in interpretation of the analysis, the internal surfaces must be exposed by fracturing the material under UHV conditions inside the electron spectrometer. If this is not done, atmospheric contamination would immediately change the surface condition irrevocably. One of the examples from GORETZKI's paper is given in Figure 41, in which Auger spectra in the differential mode from fracture surfaces in the embrittled (Fig. 41 A) and unembrittled (Fig. 41 B) states of a 12 % Cr steel containing phosphorus as an impurity, are compared. Embrittlement occurs when the steel is held at 400 – 600 °C for a long time. Note in the spectrum from the embrittled material not only the greatly enhanced phosphorus concentration at the grain boundary, but also increased amounts of chromium and nickel with respect to iron. By setting the analyzer energy first to that of the phosphorus peak, and then to that of the principal chromium peak, Auger maps could be recorded showing the distribution of phosphorus and chromium over the fracture surface. The two maps showed high correlation of the two elements, indicating association of phosphorus and chromium at the grain boundary, possibly as a compound.

An example of the application of SAM to fracture in ceramic materials has already been given in Section 27.3.5 (Fig. 40). Other recent applications in the study of grain boundary segregation in metals have included the effects of boron, zirconium, and aluminum on Ni_3Al intermetallics [139], the segregation of phosphorus and molybdenum to grain boundaries in the superalloy Nimonic PE 16 as a function of aging treatment [140], and the effect of addition of aluminum on the concentration of oxygen at grain boundaries in doped and sintered molybdenum rods [141].

27.3.6.2. Semiconductor Technology

With ever-increasing miniaturization of integrated circuits goes a need for the ability to analyze ever smaller areas so that the integrity of fabrication at any point on a circuit can be checked. This checking process includes not only microdetails such as continuity of components in the form of thin films and establishment of the correct elemental proportions in contacts, diffusion barriers, and Schottky barriers, but also the effectiveness of more macrotreatments such as surface cleaning. In all these types of analyses, depth profiling (see Section 27.3.4.2) is used extensively because, in general, elemental compositional information is required, not information about chemical state. As mentioned previously, chemical state information tends to become blurred or destroyed by ion bombardment.
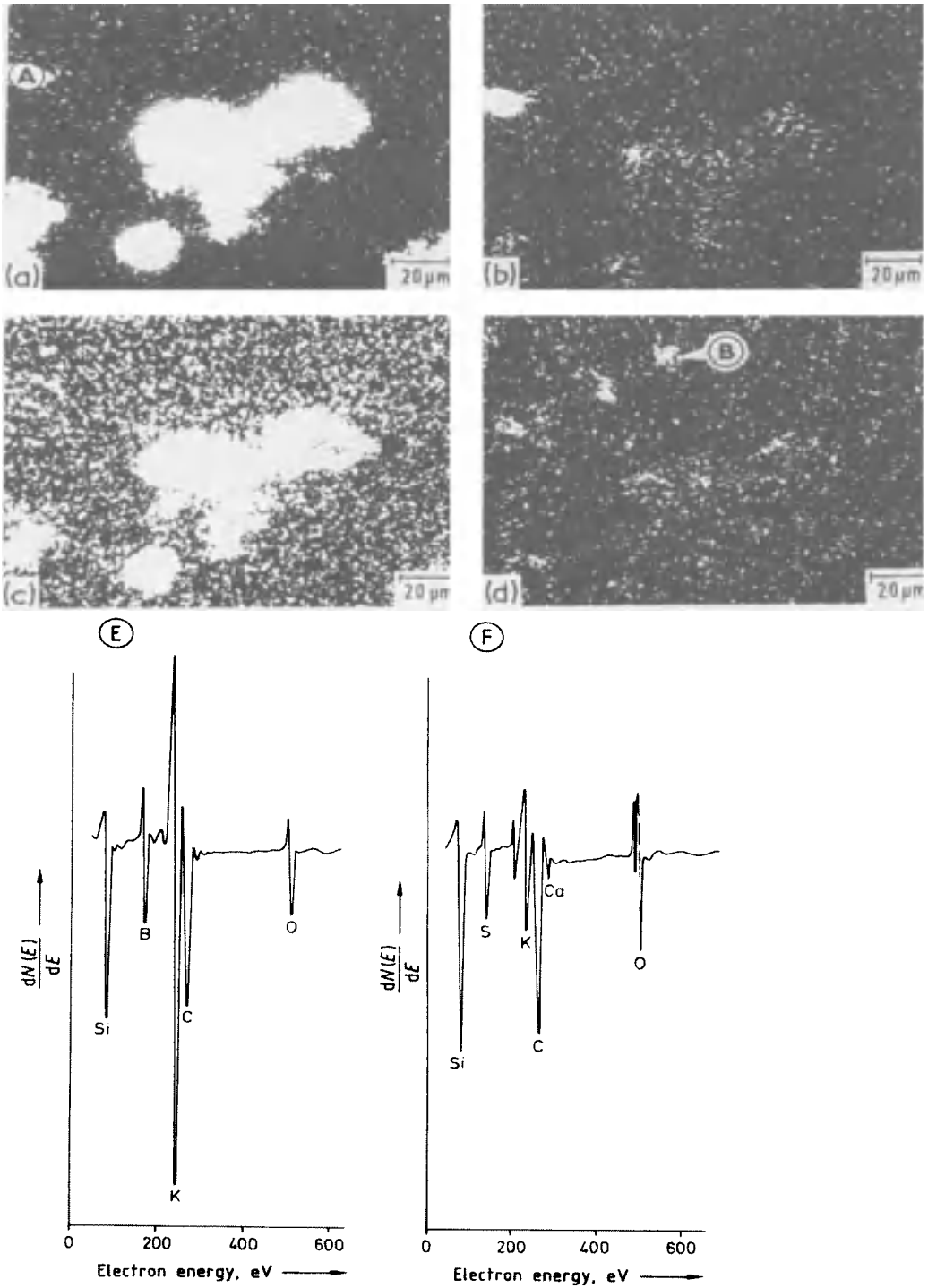
Figure 40. SAM map of fractured SiC after sintering with B addition [137]
a)–d) Elemental maps in boron, potassium, sodium, and oxygen, respectively; E), F) Point analyses at points ○A and ○B, respectively
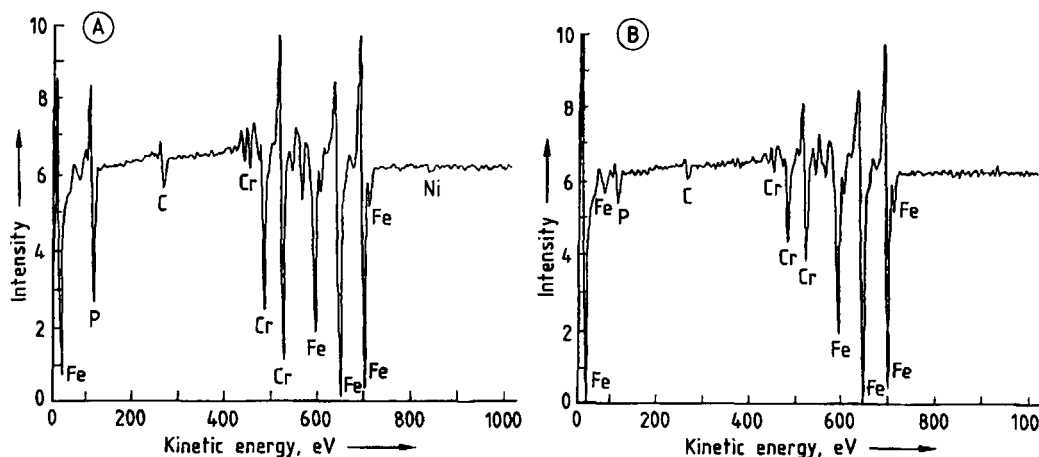
Figure 41. Auger spectra from fracture surfaces of a 12 % Cr steel [138]
A) Embrittled state; B) Unembrittled state

Various materials have been prescribed as *diffusion barriers* to prevent one circuit material from diffusing into another (usually Si) at the temperatures needed for preparation of certain components such as oxide films in integrated circuits. Thus a Ta – N film has been found to act as an efficient barrier between the metallizing alloy Al_3Ta and a silicon substrate, although SASAKI et al. [142] have shown that the barrier properties depend on the stoichiometry of the Ta – N film. These authors prepared structures of 30 nm of Ta – N on silicon, followed by 200 nm of Al_3Ta on top of the Ta – N, with the Ta – N intermediate having various stoichiometries. The composite structures were then heated to 500 and 600 °C, and analyzed by AES with depth profiling. The results for a Ta – N intermediate barrier formed by sputtering in an Ar + 5 % N_2 mixture are shown in Figure 42. The profiles before heat treatment (Fig. 42 A) and after 550 °C treatment (Fig. 42 B) indicate sharp interfaces between the various layers, with little interdiffusion. After 600 °C heat treatment (Fig. 42 C), however, very substantial intermixing has taken place, and the interfaces have virtually disappeared. Clearly the Ta – N film has not acted as an effective diffusion barrier above 550 °C. On increasing the N_2 content in the sputtering gas to Ar + 10 % N_2, and performing the same profiles, no intermixing was evident even at 600 °C, demonstrating that Ta – N can certainly be used as a diffusion barrier, but that care must be taken to achieve a minimum N : Ta stoichiometry. Ti – N is also used as a diffusion barrier between aluminum and silicon, and the analysis of

Ti – N stoichiometry in contact holes in advanced integrated circuits has been studied by PAMLER and KOHLHASE [143]. Since the holes are themselves of 2 – 3-μm diameter, analysis of their interior surface requires special techniques.

The nature of *metallic contacts to compound semiconductors* has also been an interesting problem. Ohmic contacts must be made to semiconducting surfaces, otherwise circuits could not be fabricated. However, during fabrication, excessive interdiffusion must not occur, since otherwise the electrical characteristics of the semiconductor cannot be maintained. Among the many studies of the reactions of metallizing alloys with compound semiconductors are those of PROCOP et al. [144] on Au – Ge contacts to GaAs, of REIF et al. [145] on Ti – Pt – Au contacts to GaAs, and of STEIN et al. [146] on Au – Pt – Ti contacts to GaAs and InP.

AES has also been applied to a study of the oxidation of copper in printed circuit boards [147], of sulfur passivation of GaAs [148], of cleaning of germanium and GaAs with hydrogen dissociated by a noble-gas discharge [149], and of segregation and corrosion of silver layers electroplated onto light emitting diode (LED) lead frames as part of the final packaging [150].

27.3.6.3. Thin Films and Interfaces

The nature of the interface formed between very thin metallic films and substrates of various types has been studied extensively by AES, just as it has by XPS (see Section 27.2.6.5). The degree and extent of interaction and interdiffusion can be
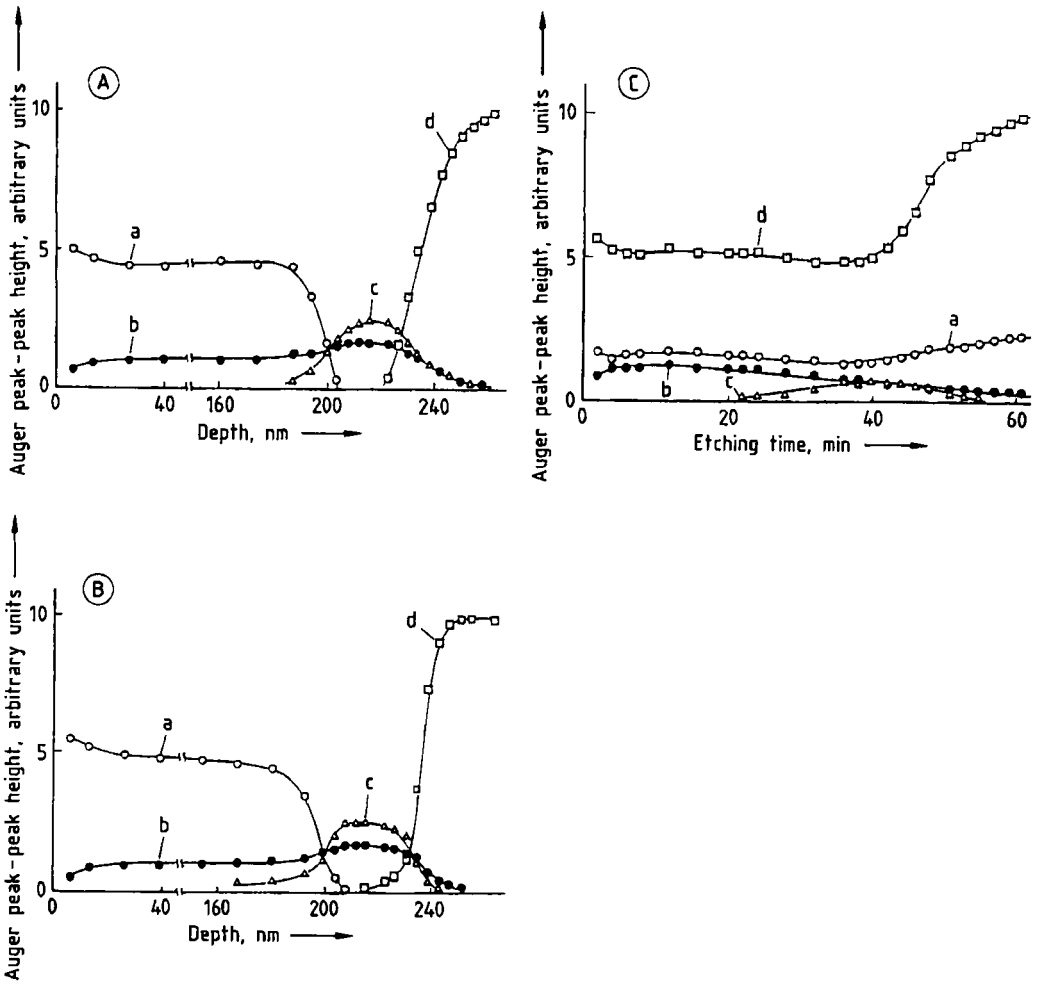
Figure 42. AES depth profiles through a structure consisting of the metallizing alloy Al₃Ta on top of a Ta – N diffusion barrier on a silicon substrate [142]
A) Before heat treatment; B) After heating to 550 °C; C) After heating to 600 °C
a) Aluminum; b) Tantalum; c) Nitrogen; d) Silicon

established by AES with depth profiling, although chemical information is normally absent, but the development of the interface is usually studied by continuous recording of Auger spectra during film deposition. Changes in the $M_{4,5}VV$ Auger spectrum from a palladium substrate as europium was deposited on it are shown in Figure 43 [151]. Up to a coverage of about two monolayers, no change occurred in the position or shape of the palladium spectrum, but for greater thicknesses the spectrum became narrower, indicating a decrease in the width of the palladium valence band arising from filling of the band. Such filling is due to an inter-

diffusion process between europium and palladium occurring even at room temperature.

In some cases, interaction of film with substrate does not occur at room temperature but occurs only after annealing at high temperature. A good example is that of a 10-nm-thick titanium film on an SiO_2 substrate. In Figure 44 [152], the Ti LMM Auger spectra are shown in the $N(E)$ (i.e., undifferentiated) mode, rather than the more usual differentiated $dN(E)/d(E)$ mode seen in Figures 41 and 43. Before annealing, the titanium film shows the typically sharp structure in the LMM region, but after annealing to 900 °C the
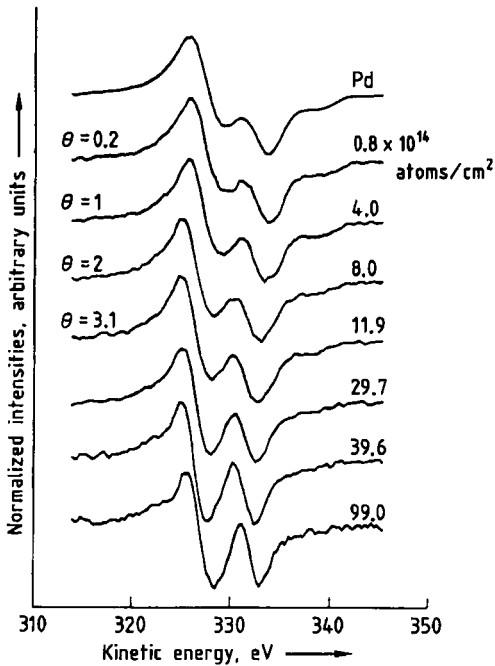
Figure 43. Changes in the Pd $M_{4,5}VV$ Auger spectrum during the deposition of increasing amounts of europium [151]

θ = coverage (monolayers)

Figure 44. Changes in the Ti LMM Auger spectra, in the $N(E)$ mode, from a 10-nm titanium film on SiO_2 after annealing to increasingly higher temperatures [152]

spectra change character completely. The small L_3VV peak near 450 eV has disappeared; the L_3MV peak has reduced in intensity, with the appearance of an additional peak at lower kinetic energy; and the LMM peak has lost its fine structure and broadened. All these features, like the measured O : Ti ratio, are consistent with the appearance of TiO_2, indicating complete oxidation of titanium as a result of reaction with SiO_2.

Other interesting thin-film studies using AES have included the growth and interaction of silicon on nickel [153], and of palladium on silicon [154], and the preparation and characterization of radio-frequency (rf) sputtered TiN_x [155], WC [156], and Al_3Ta [157] films.

27.3.6.4. Superconductors

Although AES has not been applied quite as extensively as XPS to the new high-T_c superconducting oxides, the number of applications since their discovery has also been large. The majority of applications have consisted of routine depth profiling measurements aimed at checking the extent of reaction of various materials with the superconductor. Recently such measurements, rou-

tine though they may be, have become even more important in analyzing uniformity and concentration during attempts to prepare superconductors in the form of thin films. This is because the thin films must of course be deposited on a substrate and must also be annealed to achieve their correct superconducting properties. At the annealing temperature, reaction with the substrate may occur, which must be avoided if possible. A favored substrate, because of its alleged inertness, is $SrTiO_3$. Typical of the depth profiles recorded to check possible interaction are those shown in Figure 45 [158]. The $YBa_2Cu_3O_{7-x}$ film was deposited by ion beam codeposition from sources of Cu_2O, Y_2O_3, and $BaCO_3$, together with oxygen ion bombardment, onto an $SrTiO_3$ substrate held at 350 °C, to a thickness of 100 nm. Figure 45 A shows the peak-to-peak Auger intensities as a function of sputter time, and Figure 45 B the same after conversion of yttrium, barium, copper, and oxygen intensities to atomic concentrations. Notice in the upper profile the sharpness of the interface, with relatively little diffusion of strontium and titanium into the superconducting film and none of yttrium, barium, and copper into the $SrTiO_3$. In the lower profile, note the almost correct stoi-
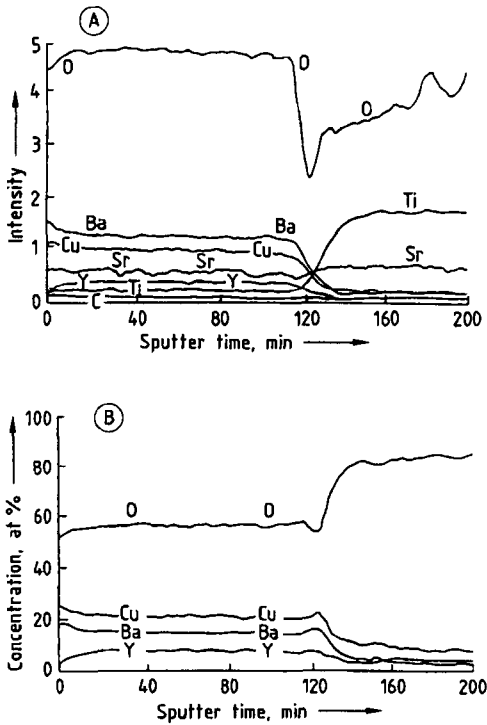
Figure 45. AES depth profiles through a superconducting $YBa_2Cu_3O_{7-x}$ film 100 nm thick on an $SrTiO_3$ substrate held at 350 °C [158]
A) Raw peak-to-peak (p-p) intensity data; B) Conversion of raw data to atomic concentrations

chiometry of the film and the uniformity through the film.

The effectiveness of so-called *buffer layers* has also been examined by AES; these are inert layers interposed between the superconducting film and a reactive substrate such as silicon. ZrO_2 is a much-used buffer material, and its role in preventing interdiffusion between $YBa_2Cu_3O_{7-x}$ films and silicon has been studied by AARNINK et al. [159] using not only AES but RBS as well. Another possibility is RuO_2, whose effectiveness was established by JIA and ANDERSON [160]. Other types of study of superconductors using AES have included the surfaces produced by fracture of sintered $YBa_2Cu_3O_{7-x}$ [161], a comparison of the surfaces of pure $YBa_2Cu_3O_{7-x}$ and its silver-added composite [162], and a quantitative comparison of superconducting bulk crystals and thin films [163].

27.3.6.5. Surface Segregation

The surface composition of alloys, compounds, and intermetallics may change profoundly during heat treatment as a result of segregation of either or both constituent and impurity elements. Surface enhancement may also occur through chemical forces during oxidation or other reaction. The mechanism of segregation is similar to that of grain boundary segregation already discussed, but of course at the free surface material is much more likely to be lost through reaction, evaporation, etc., so that excessive continued segregation can change the bulk properties as well as those of the surface. Nearly all such studies have been carried out on metals and alloys, and AES has been applied to the problem almost from its inception.

Sometimes the segregation of metallic and nonmetallic elements together to surfaces can result in the formation of surface compounds, as demonstrated by UEBING [164]. He heated a series of Fe – 15 % Cr alloys containing either 30 ppm nitrogen, 20 ppm carbon, or 20 ppm sulfur to 630° – 800 °C and obtained the typical Auger spectra shown in Figure 46. In each case an excess of both chromium and the nonmetal occurs in the spectrum, over that expected if the bulk concentration were maintained. Furthermore, some chemical information available from AES can be used here, because the fine structure associated with the nitrogen and carbon KLL spectra, if not that of sulfur, is characteristic of those elements as the nitride and carbide, respectively. Thus the cosegregation of chromium with nitrogen and with carbon has led to the formation of surface nitrides and carbides, with the same conclusion likely but not so certain, for sulfur. The surface nitride was identified as CrN and the carbide as CrC, the latter being hypothetical but probably stable as a surface compound.

Other recent surface segregation studies involving AES, often with depth profiling, have been carried out on the alloys Ag – Pd [165]; $Au_{0.7}Cu_{0.3}$ [166]; Fe – 40Cr – 3Ru, Fe – 40Cr – 3Pt, and Fe – 40Cr, with a platinum overlayer [167]; Cr – 5Pt and Cr – 15Pt [168]; and phosphorus segregation in pure iron using XPS and LEED as well [169].

AES has also been used to study the implantation of nickel in silicon [170], the biological corrosion of condenser tubes [171], the surfaces
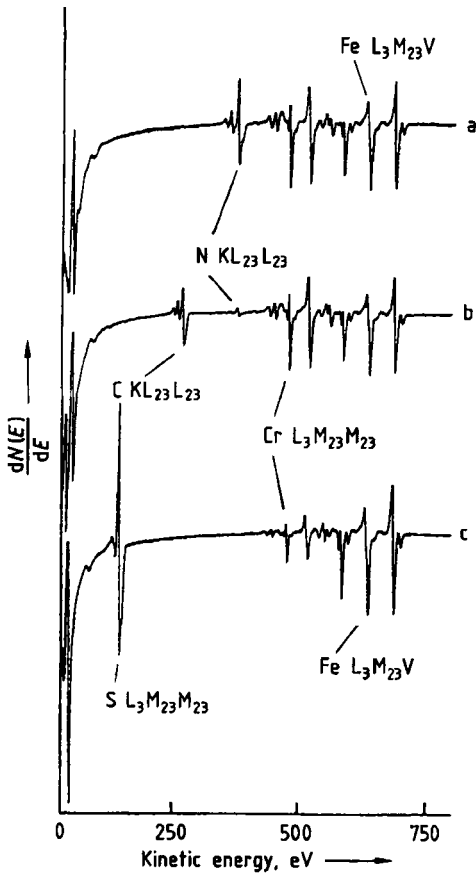
Figure 46. Formation of surface compounds on Fe – 15 % Cr alloys by cosegregation of chromium and a nonmetallic element [164]
a) Nitride; b) Carbide; c) Sulfide

of amorphous metals [172], [173], and superhard a-C : H coatings [174].

27.4. Static Secondary Ion Mass Spectrometry (SSIMS)

SSIMS originated with BENNINGHOVEN [175] in 1969 (i.e., at roughly the same time as AES). The prefix "static" was added by him to distinguish the technique from "dynamic" SIMS — the difference between the two lying in the incident current densities used, of the order of 1 – 10 nA/ cm^2 for SSIMS but much higher for dynamic SIMS. In addition, dynamic SIMS usually operates

in an imaging or scanning mode, whereas scanning SSIMS has only recently appeared.

SSIMS ranks with XPS and AES as one of the principal surface analytical techniques; its use, although expanding, is probably still rather less than that of the other two. Problems in interpretation continue, but its sensitivity for some elements is much greater than that of the other two techniques; in addition, a large amount of chemical information is available from it in principle.

27.4.1. Principles

A beam of positive ions irradiates a surface, leading to interactions that cause the emission of a variety of types of secondary particle, including secondary electrons, Auger electrons, photons, neutrals, excited neutrals, positive secondary ions, and negative secondary ions. SSIMS is concerned with the last two of these, *positive* and *negative secondary ions*. The emitted ions are analyzed in a mass spectrometer, giving rise to a positive or negative mass spectrum consisting of parent and fragment peaks characteristic of the surface. The peaks may be identifiable as arising from the substrate material itself, from contamination and impurities on the surface, or from deliberately introduced species adsorbed on the surface.

When a heavy energetic particle such as an argon ion at several kiloelectronvolts encounters a surface, it cannot be stopped immediately by the first layer of atoms but continues into the surface until it comes to a halt as a result of losing energy by atomic and electronic scattering. Along the way to its stopping place, the ion displaces some atoms from their normal positions in the solid structure, and as they recoil, these atoms in turn displace others, which also displace additional atoms, and so on, resulting in a complex sequence of collisions. Depending on the energy absorbed in an individual collision, some atoms are displaced permanently from their normal positions, whereas others return elastically after temporary displacement. The whole sequence is called a collision cascade and is illustrated schematically, along with some other processes, in Figure 47 [176]. As the cascade spreads out from the path of the primary ion, its effect may eventually reach atoms in the surface layer, and if by that stage enough kinetic energy is left in the collisional interaction, bonds may be broken and material leave the surface as atoms or clusters. The material is then said
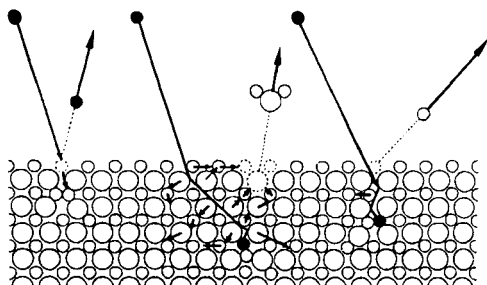
Figure 47. Schematic of various possible heavy-particle emission processes at a solid surface on ion bombardment [176]

Figure 48. Schematic of the design and operation of a liquid-metal ion source (LMIS) [178]
a) Metal ions; b) Extractor; c) Liquid-metal film; d) Capillary tube; e) Liquid metal; f) Needle

to be sputtered; the above model of the sputtering process is from SIGMUND [177].

Although SIGMUND's collision cascade model is able to explain much about the parameters involved in the removal of material from a surface by sputtering, it cannot predict the degree of positive or negative ionization of the material. In truth, a unified theory of secondary ion formation in SSIMS does not yet exist, although many models have been proposed for the process. Since the secondary-ion yields (i.e., the probabilities of ion formation) can vary by several orders of magnitude for the same element in different matrices, or for different elements across the periodic table, lack of a means of predicting ion yields in given situations is a hindrance to both interpretation and quantification. Some theoretical and semitheoretical models have had success in predictions over very restricted ranges of experimental data, but they fail when extended further. This is not an appropriate place to describe the theoretical models; for further information see, e.g., [1].

27.4.2. Instrumentation

27.4.2.1. Ion Sources

Two types of source are in use according to the type of instrument employed for SSIMS. In the first, *positive ions from a noble gas* (usually argon) are produced either by electron impact or in a plasma created by a discharge. The ions are then extracted from the source region, accelerated to the chosen energy, and focused in an electrostatic ion optical column. An ion gun of this type using electron impact ionization is shown in Figure 39.

In the second type, *positive ions from a liquid metal* (almost always gallium) are produced in the

manner shown schematically in Figure 48 [178]. A fine needle (f) (tip radius $\approx 5\,\mu m$) of refractory metal passes through a capillary tube (d) into a reservoir of liquid metal (e). Provided the liquid metal wets the needle, which can be ensured by fabrication procedures, the liquid is drawn through the tube and over the tip by capillary action. Application of $5-10$ kV between the needle and an extractor electrode (b) a short distance away draws the liquid up into a cusp because of the combined forces of surface tension and electrostatic stress. In the region just above the cusp, ions are formed and are accelerated through an aperture in the extractor, from where they may be focused in an ion optical column onto the sample surface. Gallium is invariably used as the liquid metal because its melting point is 35 °C and it remains liquid at room temperature; other metals have been tried but require heating of the reservoir to melt them.

Argon-ion guns of the electron impact type operate in the energy range $0.5-5$ keV at $1-10\,\mu A$, and provide optimum spot sizes on a surface of about 10 μm. The discharge-type argon-ion guns, often using the duoplasmatron arrangement, operate up to 10 keV, with currents variable up to 20 μA and, according to the particular focusing lens system used, can provide spot sizes down to about 2 μm. The *gallium liquid metal ion source (LMIS)* is what is known as a *high-brightness source*, which means that ion production occurs in a very small volume. As a result the ion beam can be focused to a fine spot, and at $8-10$ keV a spot size of 0.2 μm can be achieved,
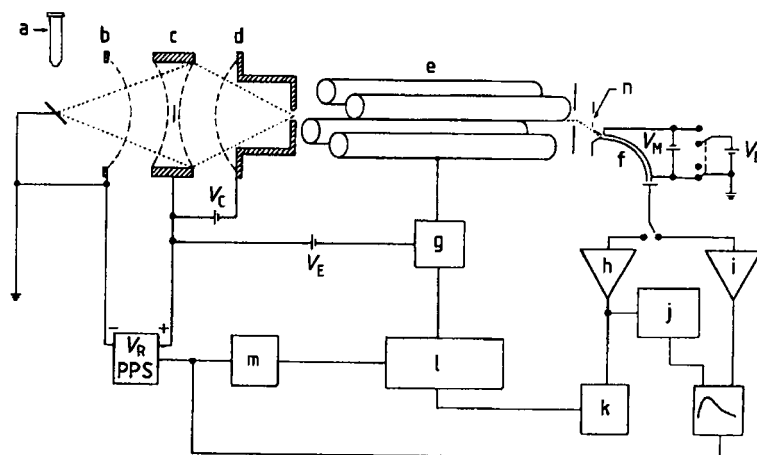
Figure 49. Experimental arrangement for SSIMS used by KRAUSS and GRUEN [179] employing a quadrupole mass spectrometer for mass analysis and a retarding-field analyzer for prior energy selection
a) Ion gun; b) – d) Lenses 1 – 3; e) Quadrupole mass spectrometer; f) Charge detecting electron multiplier; g) Quadrupole power supply; h) Pulse amplifier; i) Iso amplifier; j) Rate meter; k) Multichannel analyzer; l) Mass programmer; m) Sawtooth generator; n) Exit hole for neutrals

while at 30 keV it reduces to about 50 nm. All the ion gun optical columns provide deflection plates for rastering the ion beam over areas adjustable from many square millimeters to a few tens of square micrometers.

27.4.2.2. Mass Analyzers

As with ion sources, two types of ion mass analyzer are in use, and indeed each type of source goes with each type of analyzer. The argon-ion sources are used with quadrupole mass spectrometers and the gallium LMIS with the time-of-flight (TOF) mass spectrometer.

Quadrupole Mass Spectrometer. The quadrupole mass spectrometer (see also, →Mass Spectrometry) has been in use for many years as a residual gas analyzer (with an ionizing hot filament) and in desorption studies and SSIMS. It consists of four circular rods, or poles, arranged so that they are equally spaced in a rectangular array and exactly coaxial. Figure 49 indicates the arrangement, as depicted by KRAUSS and GRUEN [179]. Two voltages are applied to the rods, a d.c. voltage and an rf voltage. When an ion enters the space between the rods, it is accelerated by the electrostatic field, and for a particular combination of d.c. and rf voltages, the ion has a stable trajectory and passes to a detector. For other combinations of voltages, the trajectory diverges rapidly

and the ion is lost either by hitting one of the poles or by passing between them to another part of the system. The mass resolution is governed by the dimensions of the mass spectrometer, the accuracy of construction, and the stability and reproducibility of the ramped voltage. The quadrupole mass spectrometer is compact, does not require magnets, and is entirely ultrahigh vacuum compatible — hence its popularity. It does have disadvantages, however, in that its transmission is low and decreases with increasing mass number.

Time-of-Flight Mass Spectrometer. Because the quadrupole mass spectrometer is unsuitable for the analysis of large molecules on surfaces or of heavy metals and alloys, the TOF mass spectrometer has been developed for use in SSIMS by BENNINGHOVEN and coworkers [180]. The TOF mass spectrometer works on the principle that the time taken for a charged particle such as an ion to travel in a constant electrostatic and magnetic field from its point of origin to a detector is a function of the mass-to-charge ratio of the ion. For the same charge, light ions travel the fixed distance more rapidly than heavy ions. For good mass resolution, the flight path must be sufficiently long (1 – 1.5 m), and very sophisticated high-frequency pulsing and counting systems must be employed to time the flight of the ion to within a few nanoseconds.
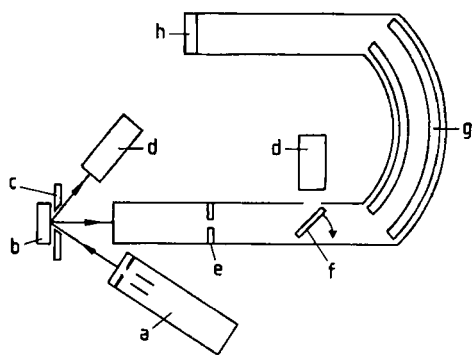
Figure 50. Schematic of the imaging time-of-flight SSIMS system designed by Eccles and Vickerman [181]
a) Pulsed ion gun; b) Target; c) Extraction electrode; d) SED; e) Variable iris; f) Movable plate; g) Energy-compensated flight tube; h) Detector

The original design [180] used pulsed beams of argon ions, but with the commercial development of the LMIS (see Section 27.4.2.1) the capability of the TOF system has been extended significantly. The principle of LMIS operation allows the beam of $^{69}Ga^+$ and $^{71}Ga^+$ ions to be focused to a probe of 50-nm minimum diameter, while being pulsed at frequencies up to 20 kHz and rastered at the same time. Mass filtering selects either of the isotopes before the beam strikes the sample. When an LMIS is operated at beam currents of 10 – 100 pA, the possibility is opered up of performing scanning SSIMS and of producing secondary-ion images in selected masses. A system designed for such *TOF imaging SSIMS* is shown schematically in Figure 50, from Eccles and Vickerman [181]. Pulsing of the ion source is achieved by rapid deflection of the ion beam across a small aperture, the pulse length being variable between 4 and 50 ns. The flight path length is 1.57 m, part of which includes an energy compensator to ensure that all ions of the same mass, but of different energies, arrive at the same time. Secondary electron detectors (SED) allow topographical images produced by ion-induced secondary electrons to be generated.

With such a TOF imaging SSIMS instrument, the useful mass range is extended beyond 10 000 amu; parallel detection of all masses occurs simultaneously; up to four secondary-ion images can be acquired simultaneously; and within each image, up to 20 mass windows (very narrow mass ranges) can be selected. The amount of data generated in a short time is enormous, and very so-

phisticated and capacious systems are required to handle and process the data.

27.4.3. Spectral Information

A SSIMS spectrum, like any other mass spectrum, consists of a series of peaks of varying intensity (i.e., ion current) appearing at certain mass numbers. The masses can be allocated on the basis either of atomic mass and charge, in the simple case of ions of the type $M''^{\pm}_i$, where both n and i can each equal 1, 2, 3,…, or of molecular mass and charge, in the more complex case of cluster ions of type $M_iX''^{\pm}_j$, where again n, i, and j can each equal 1, 2, 3,… In SSIMS, many of the more prominent secondary ions happen to have $n = i = 1$, which makes allocation of mass numbers to fragments slightly easier. The masses can be identified as arising either from the substrate material itself, from deliberately introduced molecular or other species on the surface, or from contamination and impurities on the surface. Complications in allocation often arise from isotopic effects. Although some elements have only one principal isotope, in many others the natural isotopic abundance can cause difficulties in identification.

An example showing both the allocation of peaks and the potential isotopic problems appears in Figure 51, in which the positive and negative SSIMS spectra from a nickel specimen have been recorded by Benninghoven et al. [182]. The specimen surface had been cleaned by standard procedures, but SSIMS reveals many impurities on the surface, particularly alkali and alkaline-earth metals, for which it is especially sensitive. For reasons discussed in Section 27.4.4, the peak heights cannot be taken to be directly proportional to the concentrations on the surface; thus, for example, the similar heights of the Na^+ and $^{58}Ni^+$ peaks should not lead to the conclusion that there is as much sodium as nickel on the surface. Note in Figure 51 the appearance of isotopes of nickel, potassium, and chlorine in their correct natural abundances and also, in the positive ion spectrum, peaks due to M_2^+ ions, also with isotopic effects. Since the intensity scale is logarithmic, the minor peaks are exaggerated, but identification problems do exist.

The relationship between what is recorded in a SSIMS spectrum and the chemical state of the surface is not as straightforward as in XPS and AES. Because of the large number of cluster ions that appear in any SSIMS spectrum from a mul-
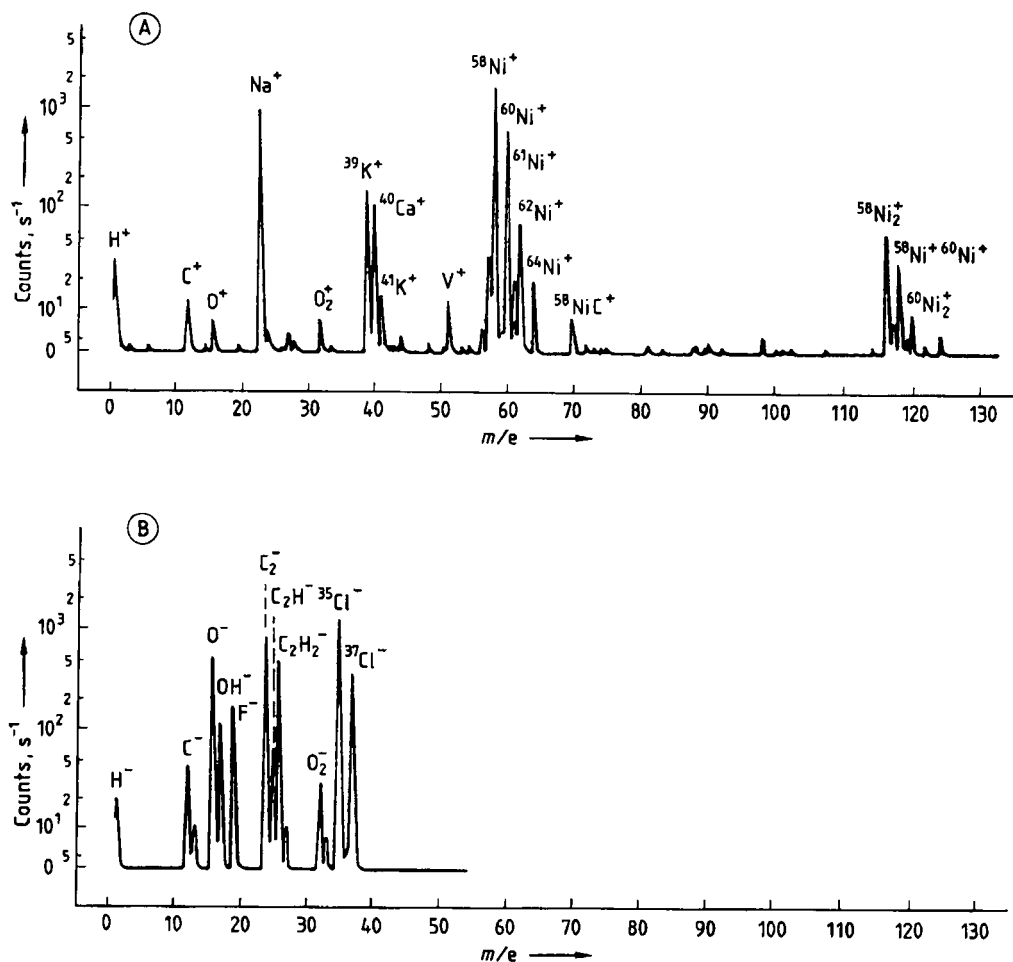
Figure 51. SSIMS spectra from a "cleaned" nickel target, under conditions of 3 keV primary argon ion energy and 5×10^{-9} A primary current [182]
A) Positive SSIMS spectrum; B) Negative SSIMS spectrum

ticomponent surface (e.g., during the study of a surface reaction), a great deal of chemical information is obviously available in SSIMS, potentially more than in XPS. The problem in using the information from a cluster ion lies in the uncertainty of knowing whether or not the cluster is representative of a group of the same atoms situated next to each other in the surface. For some materials, such as polymers, the clusters observed are definitely characteristic of the material, as seen in Figure 52 from BLETSOS et al. [183], where the SSIMS spectrum from polystyrene measured in a TOF mass spectrometer reveals peaks spaced at regular 104 mass unit intervals, corresponding to the polymeric repeat unit. On the other hand, for

example during oxidation studies, many metal – oxygen clusters are normally found in both positive and negative secondary-ion spectra, whose relationship to the actual chemical composition is problematic. Figure 53 from McBREEN et al. [184] shows the SSIMS spectra recorded after interaction of oxygen with half a monolayer of potassium adsorbed on a silver substrate. In both positive and negative spectra K_iO_j clusters are found that the authors were unable to relate to the surface condition, although quite clearly much chemical information is available in the spectra. The use of pure and well-characterized standards can sometimes help in the interpretation of complex spectra.
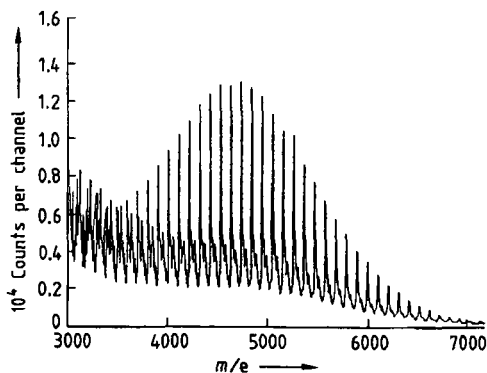
Figure 52. Positive SSIMS spectrum of polystyrene (M_n = 5100, $M_{n, calc.}$ = 4500, $M_{n, corr.}$ = 5600) from time-of-flight mass spectrometer [183]

27.4.4. Quantification

The secondary-ion current at a particular mass number from element A in a matrix B can be written as

$$I^{\pm}_{A(B)} = K \cdot S \cdot C_A \cdot \alpha^{\pm}_{A(B)} \tag{18}$$

where S is the total sputtering yield of B; C_A the concentration of A in B; and $\alpha^{\pm}_{A(B)}$ the probability of positive or negative ionization of element A in matrix B.

As for XPS and AES (Eq. 8 and Eq. 13, respectively) K is a constant containing the various instrumental and other factors that are maintained approximately fixed during an analysis.

The *sputtering yield S* can be measured from elemental and multielemental standards under various operational conditions and can therefore be known to reasonable accuracy for the material being analyzed by judicious interpolation between standards. However, the *ionization probability* α cannot be measured easily, and attempting to measure it for a set of standards and then trying to interpolate would be of no use, because it is very matrix and concentration dependent and can vary very rapidly. In fact, it represents the principal obstacle in achieving proper quantification in SSIMS.

Because of problems in the measurement of α, quantification is normally carried out either by relative sensitivity factor methods or by semi-empirical approximations to α based on one of the ionization theories.

Relative Sensitivity Factors. In SSIMS, unlike XPS and AES, elemental sensitivities are taken as
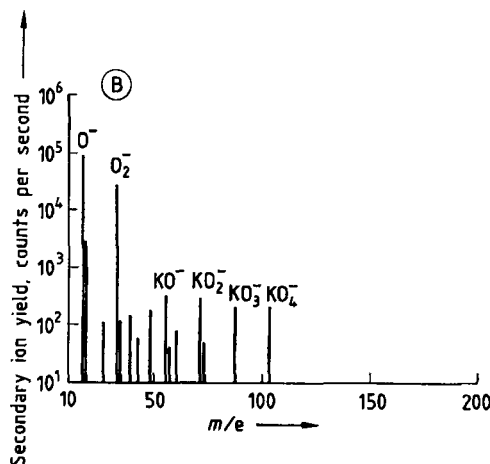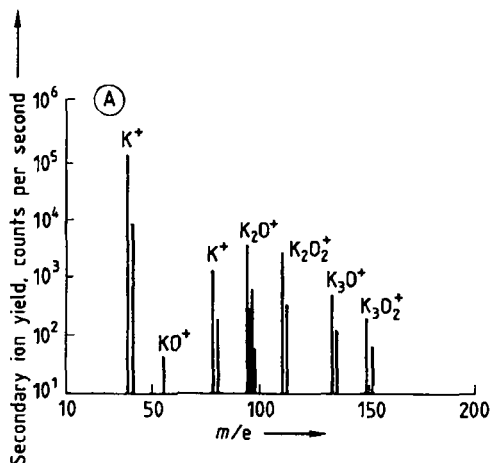
Figure 53. SSIMS spectra following interaction of oxygen with half a monolayer of potassium adsorbed on silver, at 300 K [184]
A) Positive SSIMS spectrum; B) Negative SSIMS spectrum

relative to the intensity of the signal from an internal reference element, rather than to that of a single chosen element such as fluorine or carbon. The internal reference would normally be the major constituent of the material. If the signal from the reference is $I^{\pm}_{R(B)}$, it can be written

$$I^{\pm}_{A(B)} / I^{\pm}_{R(B)} = \frac{K \cdot S \cdot C_A \cdot \alpha^{\pm}_{A(B)}}{K \cdot S \cdot C_R \cdot \alpha^{\pm}_{R(B)}} \tag{19}$$

or

$$I^{\pm}_{A(B)} / I^{\pm}_{R(B)} = \varrho_{AR} C_A / C_R \tag{20}$$

where ϱ_{AR} is the relative sensitivity factor of A with respect to R. Of course ϱ_{AR} contains matrix-dependent quantities, but their variations are damped to some extent by virtue of taking ratios, and in practice ϱ_{AR} is assumed constant for low concentrations of A (e.g., < 1 at %). It can be evaluated from measurements on a well-characterized set of standards containing A in known dilute concentrations. The accuracy of the method is not as good as XPS and AES when relative sensitivity factors are used.

Semiempirical Approximation. The method that adopts a semiempirical approximation to α uses the *Andersen – Hinthorne local thermodynamic equilibrium (LTE)* model [185] for estimating the degree of ionization. The model does not take account of any details of individual ionization processes but assumes that the region at and near the surface involved in sputtering can be approximated by a dense plasma in local thermodynamic equilibrium. The plasma has an associated temperature T. The ratio of the concentrations of two elements X and Y in a matrix B, if sufficiently dilute, can then be written as

$$C_X/C_Y \sim \left(I^{\pm}_{X(B)}/I^{\pm}_{Y(B)}\right) \exp\left\{(I_{P(X)} - I_{P(Y)})/kT\right\} \qquad (21)$$

where $I^{\pm}_{X(B)}$, $I^{\pm}_{Y(B)}$ are the positive secondary-ion currents of X and Y, respectively, in B; and $I_{P(X)}$, $I_{P(Y)}$ the first ionization potentials of neutral atoms X and Y.

T is the only adjustable parameter in Equation (21). Measurement of the secondaryion currents of all components then provides a set of coupled equations for $C_X/C_Y, \ldots, C_N/C_Y$, which, together with the identity $C_X + C_Y + \ldots C_N = 100 \%$, enables the concentrations to be found. The model does not have a sound physical basis, but the method has been found to be surprisingly accurate for some materials (e.g., steels). However, the values of T, the effective plasma temperature, that are necessary for agreement are sometimes ridiculously high.

27.4.5. Applications

In general, SSIMS has not been used for quantification of surface composition because of all the uncertainties described above. Its application has been more qualitative in nature, with emphasis on its advantages of high surface specificity, very high sensitivity for certain elements, and multiplicity of chemical information. Increasingly, it

is being used in the high spatial resolution TOF imaging mode and also in parallel with a complementary technique, such as XPS, on the same system.

27.4.5.1. Oxide Films

The higher surface specificity of SSIMS compared to XPS and AES (i.e., 1 – 2 monolayers versus 2 – 8 monolayers) can be useful in establishing more precisely the chemistry of the outer surface than would be possible with the latter techniques. Although from details of the O 1s spectrum XPS could give the information that OH was present on a surface as well as oxide, and from the C 1s spectrum that hydrocarbons as well as carbides were present, only SSIMS could say which hydroxide and which hydrocarbons. In the growth of oxide films for various purposes (e.g., passivation or anodization), such information is valuable since it provides a guide to the quality of the film and the nature of the growth process.

One important area in which the properties of very thin films play a significant role is adhesion of aluminum to aluminum in the aircraft industry. For bonds of the required strength to be formed, the aluminum surfaces must be pretreated by anodizing them, which should create not only the right chemical conditions for adhesion but the right structure in the resultant oxide film as well. Films formed by anodizing aluminum in phosphoric acid under a.c. conditions have been studied by TREVERTON et al. [186] using SSIMS and XPS together. Knowledge of both the chemistry of the outer monolayers in terms of aluminum oxides and hydroxides, and of the nature and degree of any contamination introduced by the anodizing process, was important. Figure 54from their paper compares the positive and negative SSIMS spectra after anodizing for 3 and 5 s at 20 V and 50 Hz. Both sets of spectra show that the longer anodizing treatment caused significant reduction in hydrocarbon contamination (peaks at 39, 41, 43, 45 amu in the positive spectrum, and 25, 41, 45 amu in the negative), and marked increases in oxide- and hydroxide-associated cluster ions at 44, 61, 70, and 87 amu in the positive spectrum and 43 and 59 amu in the negative. At higher masses, peaks allocated to $Al_2O_4H^-$, $Al_2O_3^-$, and $Al_3O_5^-$ were also identified. Residual PO_2^- and PO_3^- from the anodizing bath were reduced at the longer anodizing time. The low level of contaminant is important since the adhesive coating can then wet the surface
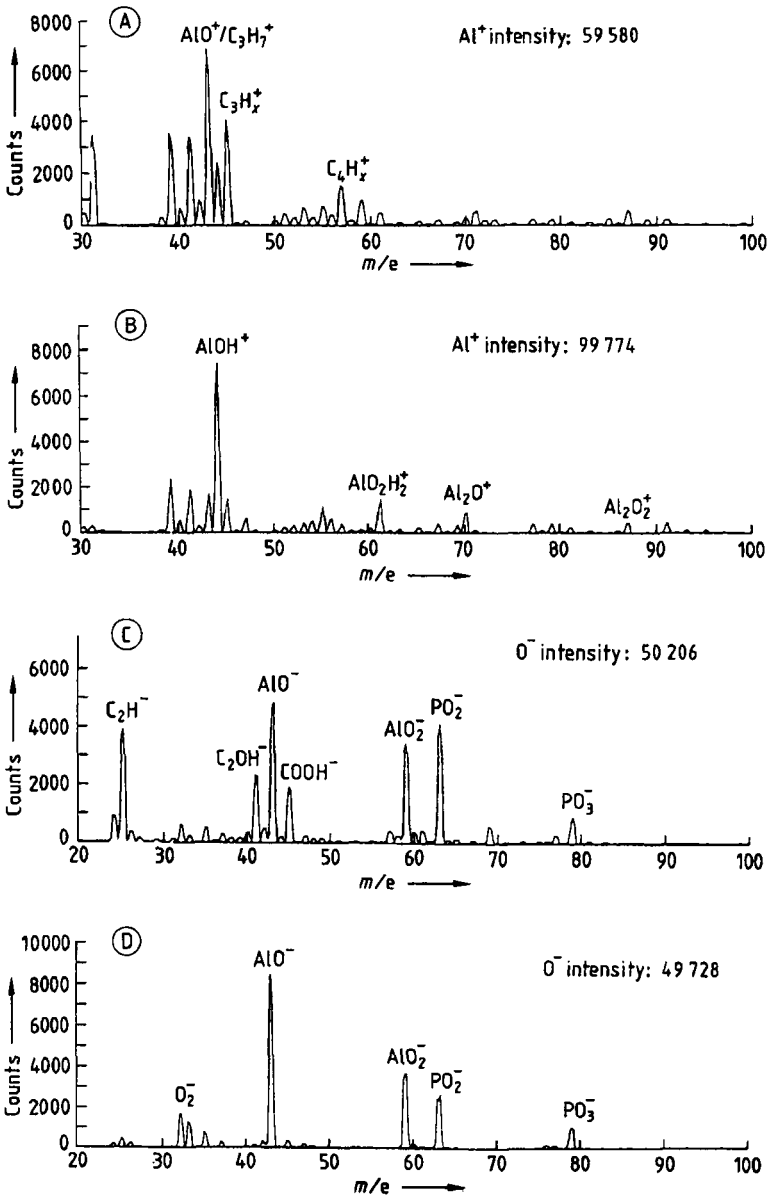
Figure 54. SSIMS spectra from oxide films formed on aluminum after anodization treatment in phosphoric acid for 3 or 5 s [186] A) Positive SSIMS spectrum, 3 s; B) Positive SSIMS spectrum, 5 s; C) Negative SSIMS spectrum, 3 s; D) Negative SSIMS spectrum, 5 s

more effectively, leading to improved adhesion. In addition, the change in aluminum chemistry from AlO^+ (mass 43) to $AlOH^+$ (mass 44) on going from 3- to 5-s anodizing was significant, and not an effect that could have been established by XPS.

27.4.5.2. Interfaces

The protection of steel surfaces by paint depends significantly on the chemistry of the paint – metal interface. The system has many variables since the metal surface is normally pretreated in a variety of ways, including galvanizing and phos-
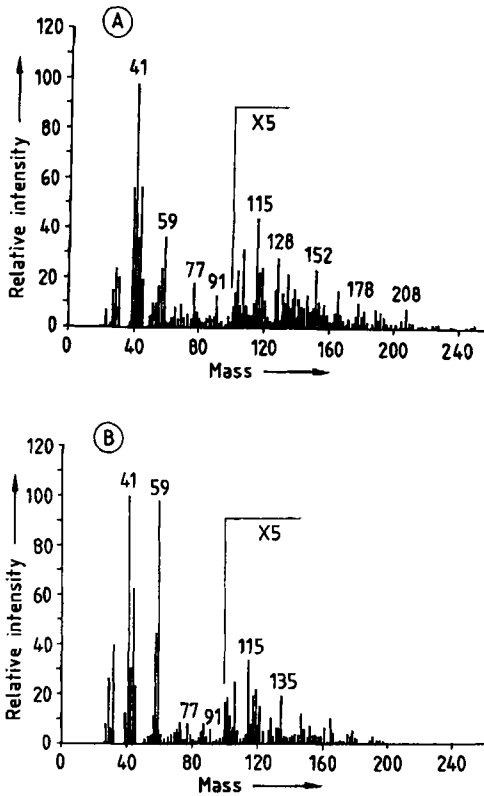
Figure 55. Positive SSIMS spectra from each side of the metal–paint interface on a hot-dip galvanized steel surface after accelerated cyclic corrosion testing [187]
A) Metal; B) Paint

phating. Indeed, several interfaces of importance probably exist, and corrosion protection may be a function of conditions at all of them. To study the effects of the variables without having to wait for natural weathering processes to cause corrosion, normal procedure is to apply an accelerated cyclic corrosion test (CCT) that involves scribing through the paint to the bare metal followed by salt spray, drying, and exposure to high humidity.

SSIMS has been applied by VAN OOIJ et al. [187] to study of the interfaces in a paint–metal system; both quadrupole and TOF SSIMS were employed, as well as XPS. Figure 55from their paper shows the positive SSIMS spectra from the metal (Fig. 55 A) and paint (Fig. 55 B) sides of an interface exposed by soaking samples in N-methylpyrrolidone, which removes the paint by swelling but does not attack it chemically. The metal surface had been phosphate treated before paint application. Both spectra are identified

in terms of organic materials; in other words the failure was in an organic layer close to, but not at, the metal surface. This identification is supported by the observations that the mass peak at 52 amu, characteristic of chromate, was negligible and that in the negative spectra the intensity of PO_x^- ions was also low. Note that epoxy-related peaks at 59 and 135 amu were observed on both sides, as well as aromatic fragments at 77, 91, 115, 128, 152, and 178 amu, the latter being more intense on the metal side. When the spectra were compared with those from an unphosphated and painted metal surface that had suffered delamination, the characteristic epoxy peaks were not seen on the metal side, indicating that in the absence of the phosphate layer the adhesion of the epoxyurethane to the substrate was lost.

27.4.5.3. Polymers

Treatment of polymer surfaces to improve their properties with respect to wetting or water repulsion and to adhesion, is by now a standard procedure. The treatment is designed to change the chemistry of the outermost groups in the polymer chain, without affecting bulk polymer properties. Any study of the effects of treatment therefore requires a technique that is specific mostly to the outer atom layers, which is why SSIMS has been used extensively in this area.

One of the favored forms of surface treatment is that of plasma etching. Depending on the conditions of the plasma discharge (i.e., nature of discharge gas, gas pressure, rate of gas flow, discharge power and frequency, and substrate polarity), plasma etching is able to alter the chemical characteristics of the surface over a wide range. By its nature a plasma discharge is also both highly controllable and reproducible. The effects of plasma treatment on the surface of polytetrafluoroethylene (PTFE) have been studied by MORRA et al. [188] using combined SSIMS and XPS. They found, as evidenced in Figure 56, that not only did the plasma change the surface chemistry, but the surface emission fragments also changed the plasma discharge conditions temporarily. Thus the untreated (Fig. 56 A) and 15-min oxygen plasma treated (Fig. 56 C) positive-ion spectra are virtually identical, whereas the 0.5-min oxygen plasma treated spectrum (Fig. 56 B) is quite different. The initial breakup of the PTFE surface produced hydrocarbon fragments which reacted with the oxygen to produce water vapor, thus changing the local nature of the discharge
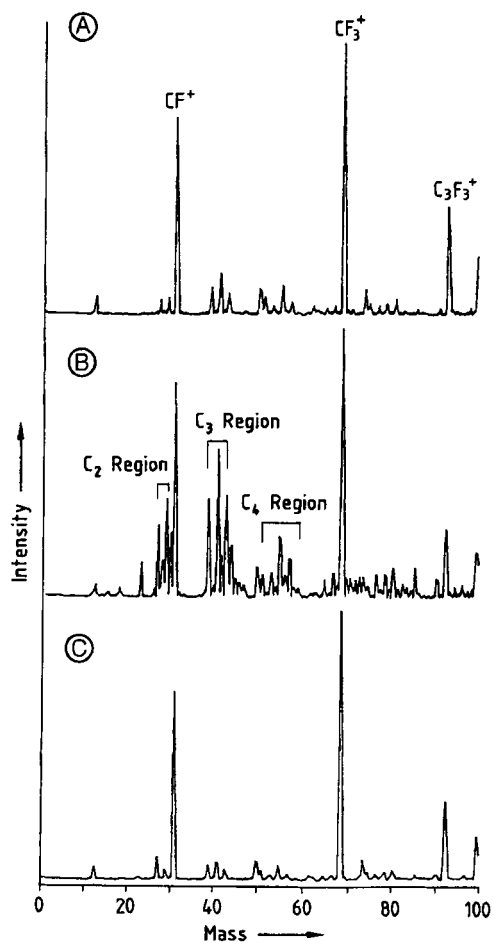
Figure 56. Positive SSIMS spectra from PTFE [188]
A) Untreated; B) 0.5-min oxygen plasma treatment;
C) 15-min oxygen plasma treatment

gas. The important deduction to be made is that one of the vital parameters in plasma treatment is the length of time in contact with the plasma; in this case, too long a contact time resulted in no surface changes.

Other SSIMS studies of polymer surfaces have included perfluorinated polyether [189], low-density polyethylene [190], poly(ethylene terephthalate) [191], and the oxidation of polyetheretherketone [192].

27.4.5.4. Surface Reactions

Because of the inherently destructive nature of ion bombardment, the use of SSIMS alone in study

of the reactions of surfaces with gases and vapors must be viewed with caution, but in combination with other surface techniques it can provide valuable additional information. The parallel techniques are most often XPS, TDS, and LEED, and the complementary information required from SSIMS normally refers to the nature of molecules on surfaces and with which, if any, other atoms they are combined.

A typical SSIMS spectrum of an organic molecule adsorbed on a surface is that of thiophene on ruthenium at 95 K, shown in Figure 57, from the study of Cocco and Tatarchuk [193]. The exposure was only 0.5 Langmuir (i.e., 5×10^{-7} torr $\cdot$ sec $\triangleq 37$ Pa $\cdot$ s) and the principal positive ion peaks are those from ruthenium, consisting of a series of seven isotopic peaks around 102 amu. However, ruthenium–thiophene complex fragments are found at ca. 186 and 160 amu, each showing the same complicated isotopic pattern, indicating that even at 95 K interaction between the metal and the thiophene had occurred. In addition, thiophene and protonated thiophene peaks are observed at 84 and 85 amu, respectively, with the implication that no dissociation of the thiophene had taken place. The smaller masses are those of hydrocarbon fragments of different chain lengths.

SSIMS has also been used in the study of the adsorption of propene on ruthenium [194], the decomposition of ammonia on silicon [195], and the decomposition of methanethiol on nickel[196].

Recently SSIMS has been applied in the TOF–SSIMS imaging mode to the study of very thin layers of organic materials [197], polymeric insulating materials [198], and carbon fiber and composite fracture surfaces [199]. In these studies a spatial resolution of ca. 200 nm in massresolved images was achieved.

27.5. Ion Scattering Spectroscopies (ISS and RBS)

Analysis of surfaces by ion scattering has a long history since several techniques are included according to the energy of the primary ion. They have been classified by Armour [200] as *low-energy ion scattering* (LEIS) for the range 100 eV to 10 keV, *medium-energy ion scattering* (MEIS) for the range 100 to 200 keV, and *high-energy ion scattering* (HEIS) for the range 1 to 2 MeV. However, LEIS is more often called *ion scattering*
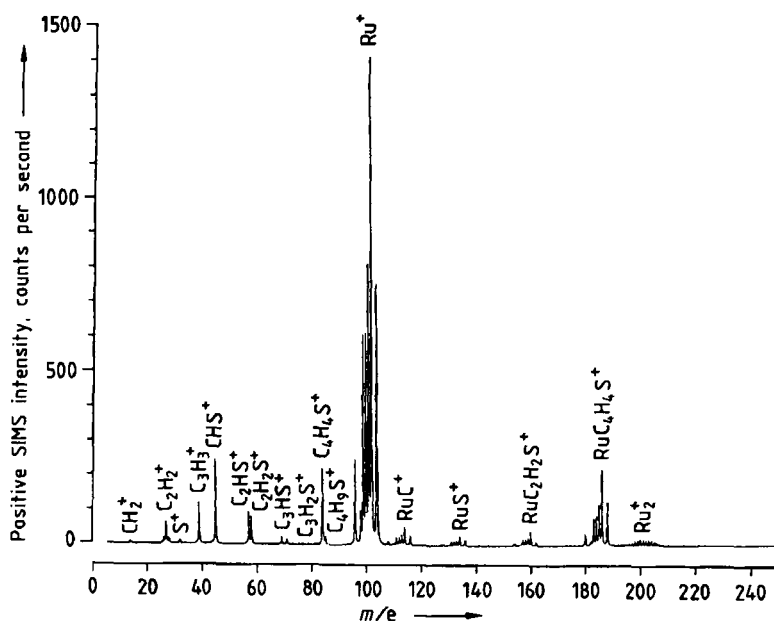
Figure 57. Positive SSIMS spectrum from Ru (0001) after exposure to 0.5 L thiophene at 95 K [193]

spectroscopy (ISS), whereas HEIS is always called *Rutherford backscattering spectroscopy* (RBS). ISS is invariably surface specific, while RBS is normally surface sensitive.

27.5.1. Ion Scattering Spectroscopy (ISS)

Although treated here as one of the major surface analytical techniques, ISS, like SSIMS, is used much less extensively than XPS or AES. Again, like SSIMS, it is often used in conjunction with other techniques and is valuable as a complement to the others for its two advantages of high surface specificity and ease of quantification.

27.5.1.1. Principles

A beam of inert gas ions, usually He^+ or Ne^+, of typical energy 0.5–4 keV and current 10^{-8}–10^{-9} A, is directed at a surface, and the energies of the ions of the same mass and charge that are scattered from the surface are analyzed. Because inert gas ions have a very strong electron affinity, the probability is high that on collision with a surface the majority will be neutralized. After two collisions the number of original ions left will be negligible. Thus analysis of scattered

ion energies will refer mostly to those ions that have suffered only one collision (i.e., from the outermost layer of atoms). Hence ISS is very surface specific. In addition, with primary ions of low mass and energy, the ion–atom interaction time is much shorter than the characteristic time of vibration of an atom in a solid, so that the collision can be approximated as that between the ion and a free atom. This means that binary collision (or "billiard-ball") theory can be applied, in which case the relationship between the energy E_1 after collision and the energy E_0 before collision, of the ion of mass M_1 with a surface atom of mass M_2, is

$$E_1/E_0 = \{1 + (M_2/M_1)\}^{-2} \left[\cos\theta \pm \left\{(M_2/M_1)^2 - \sin^2\theta\right\}^{1/2}\right]^2$$

(22)

where θ is the scattering angle (i.e., angle between incident and scattered beams).

If θ is chosen to be 90°, then Equation (22) reduces to the simple relationship

$$E_1/E_0 = (M_2 - M_1)/(M_2 + M_1)$$

(23)

Equation (22) shows that if the energy and mass of the incident ion, and the scattering angle, are all fixed, E_1 depends only on M_2, the mass of
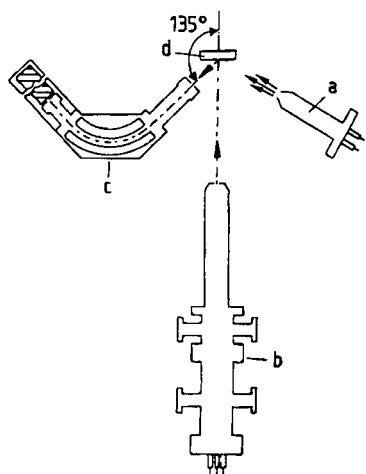
Figure 58. Experimental arrangement of ISS system used by MARTIN and NETTERFIELD [201]
a) Electron flood gun; b) Ion gun; c) ISS analyzer; d) Sample

the surface atom. Thus the energy spectrum measured in ISS can be converted directly into a mass spectrum, providing an elemental compositional analysis of the surface. Because only the mass of the surface atom is involved, no chemical state information is available.

Note from Equations (22) and (23) that as M_2 increases, the ratio E_1/E_0 tends to unity, which means that for a light ion such as He$^+$, heavy elements are difficult to resolve in the energy spectrum. ISS is therefore most suitable for the analysis of light elements on surfaces. The mass resolution can be improved by increasing the mass of the ion (e.g., by going to Ar$^+$ or Xe$^+$), but then problems of sputtering intrude. Another way of improving the resolution is by using scattering angles >90°, and in most ISS geometries the angle is $\approx 135°$.

27.5.1.2. Instrumentation

As with other surface-specific techniques, the vacuum requirements for ISS are stringent, vacua better than 10^{-8} Pa being needed. With the single monolayer specificity mentioned above, good UHV conditions are vital.

The basic instrumental requirements are an ion source, a sample target, and an energy analyzer. The source is normally a commercial ion gun, for example that shown in Figure 39, operated at a few kiloelectronvolts and with beam currents of He$^+$ or

Ne$^+$ of 1 – 100 nA. For scattered ion energy analysis, a standard CHA or CMA with reversed polarity can be used, or alternatively a special analyzer may be constructed. The basic arrangement is shown in Figure 58 from MARTIN and NETTERFIELD [201]; also shown is an auxiliary electron source used for charge neutralization by flooding the surface with low-energy electrons. In particularly sophisticated experiments the incident ion beam is mass-filtered before striking the target, and both target and analyzer are rotatable around the point of impact to allow angular resolution.

Similar to developments in SSIMS, significant improvements in ISS have been obtained by using *time-of-flight techniques*. The additional sensitivity available with a TOF system allows much lower incident ion current densities to be employed, thereby reducing the effects of beam damage to the surface. Such damage is of course negligible when using He$^+$, but if heavier ions such as Ar$^+$ are required for better mass resolution, the current density must be reduced. Another major advantage of a TOF energy analyzer is that it enables energy analysis not only of ions but of neutrals ejected from the surface, since the flight time is independent of charge state. This has led to the introduction of a structural technique called *noble-gas impact collision ion scattering spectroscopy* with neutral detection (NICISS), by NIEHUS [202].

27.5.1.3. Spectral and Structural Information

Spectra in ISS are recorded in the form of scattered intensity versus the ratio E_1/E_0, so that the range of the latter will be from 0 to 1. Some typical examples are shown in Figure 59, from BEUKEN and BERTRAND [203], for a set of Al$_2$O$_3$, MoS$_2$, and Co$_9$S$_8$ standards used by them in their study of cobalt – molybdenum sulfide catalysts. The conditions — He$^+$ at 2 keV and 15 nA, and a scattering angle of 138° — are also typical. Acquisition time per spectrum was about 5 min, and an electron flood gun was used for charge neutralization.

As seen in Figure 59, the E_1/E_0 ratio can be converted directly to mass and therefore to elemental identification. Notice also the disadvantage mentioned above, that as $E_1/E_0 \rightarrow 1$, the mass separation becomes poor; i.e., if cobalt ($M = 59$) and molybdenum ($M = 96$) had appeared in the same spectrum they would have only just been resolved
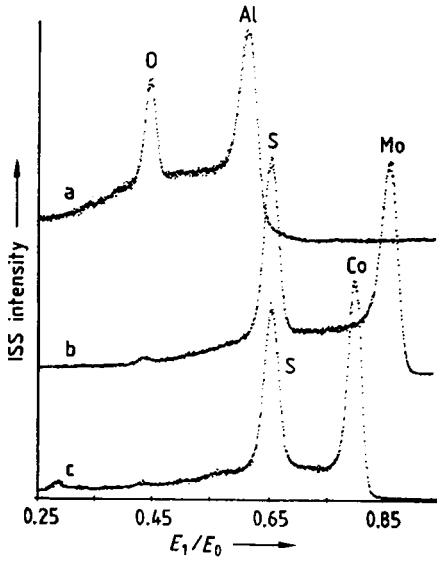
Figure 59. Examples of ISS spectra for a set of standard materials used in the study of cobalt – molybdenum sulfide catalysts [203]
a) Al_2O_3; b) MoS_2; c) Co_9S_8

Figure 60. Ion trajectories for a parallel beam of light ions hitting an atom [204]

under the conditions. However, the resolution for light elements is good.

In addition to elemental compositional information, ISS can also be used to provide information about surface structure, and in this application it is complementary to LEED. The reason for the structural possibility is indicated schematically in Figure 60, which shows the ion trajectories calculated by VAN DER VEEN [204] for a parallel beam of ions hitting an atom at the surface. Most ions are scattered in the forward direction, with only a few making an interaction strong enough to be scattered back toward the detector. Thus a "shadow cone" arises behind the atom, R being the radius of the cone at the position of the next atom in the row. Since R is of the same order as interatomic spacing, atoms in a row lying immediately behind the scattering atom are "invisible" to the primary beam. S is the so-called impact parameter, which depends on ion energy and distance of nearest approach between ion and atom.

If the experimental geometry is varied, by varying either the angle of primary ion incidence or the target azimuthal angle, while the scattering angle is kept constant, the shadow cone is effectively swept through the rows of atoms behind a surface atom. The scattered intensity will then

vary periodically according to whether the scattering is mostly from surface atoms or whether at certain angles contributions to the scattering come from atoms further along a row. If the near-surface crystallinity is good, then the variations in angular intensity can be used to derive or confirm a surface structure. The technique is particularly valuable when using scattering angles close to 180°, since the problem of multiple scattering (i.e., successive scattering off two or more atoms) is minimized, and the dependence of the scattered intensity on the angle of incidence is a direct representation of the shape of the shadow cone. This variation of ISS is called *impact collision ion scattering spectroscopy* (ICISS). Figure 61 from NIEHUS and COMSA [205] compares the ICISS measurements from a Cu (110) crystal surface before (Fig. 61 A) and after adsorption of oxygen (Fig. 61 B). Various models had been proposed for the structure of the oxygen-reacted surface, but only the "missing row" model shown on the lower right can explain the observed angular dependence.

27.5.1.4. Quantification and Depth Information

If the scattering process is considered as a simple binary collision, the intensity I_A^+ of primary ions scattered from atoms of type A in the surface is given by

$$I_A^+ = I_O^1 N_A \sigma_A (1 - P) S \cdot \Delta\Omega \cdot T \tag{24}$$

where I_O^+ is the intensity of the primary beam; N_A the density of scattering atoms A; σ_A the scattering cross section of atoms A; P the neutralization probability of a primary ion; S the shadowing factor; $\Delta\Omega$ the acceptance angle of the analyzer; and T the transmission function of the analyzer.
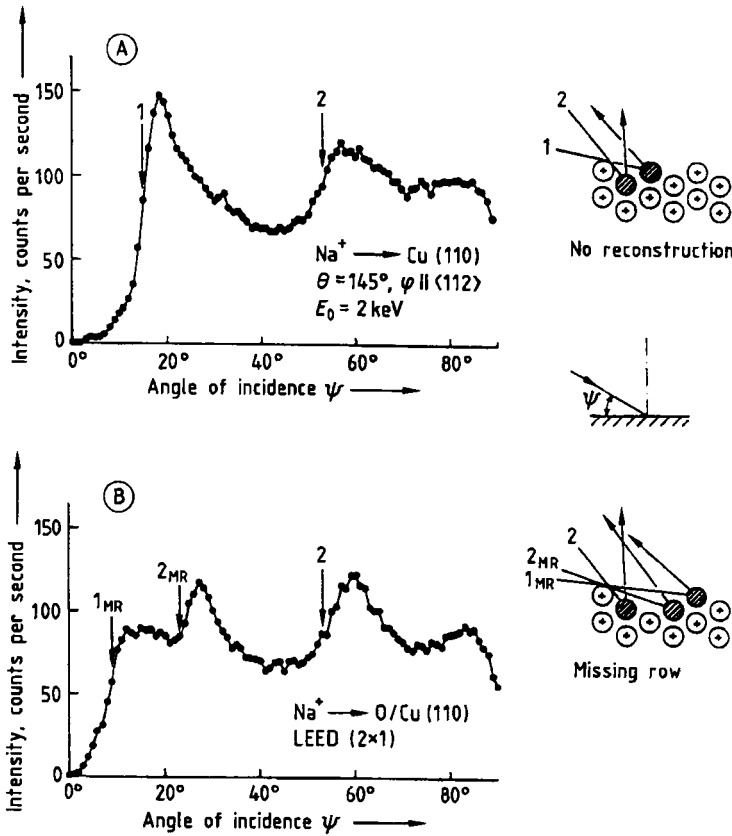
Figure 61. Impact collision ISS of Na^+ scattering from a Cu (110) surface before (A) and after (B) surface reconstruction due to oxygen adsorption, as a function of angle of incidence [205]

In an analysis, the primary ion current, the experimental geometry, and the transmission function are kept constant, so that to relate Γ_A^+ to N_A requires knowledge of σ_A and P.

Scattering Cross Section. To evaluate σ_A, information is needed about the physical nature of the interaction between the ion and the atom. In the ion energy range used in ISS the interaction can be considered in terms of purely repulsive potential (i.e., a Coulomb repulsion with inclusion of some screening, since some electronic screening occurs during the interaction). The general form of the interaction potential is thus

$$V(r) = \frac{Z_1 Z_2 e^2}{r} \, \varphi(r/a) \tag{25}$$

where Z_1, Z_2 are the atomic numbers of ion and scattering atom; r is the ion – atom distance; and a the characteristic screening length.

The most widely used screening function is that due to MOLIÈRE [206] using Thomas – Fermi statistics, which is approximated by

$$\varphi(r/a) = \varphi(x) = 0.35 \exp(-0.3x)$$
$$+ 0.55 \exp(-1.2x) + 0.1 \exp(-6x) \tag{26}$$

with

$$a = 0.4685 \left[Z_1^{2/3} + Z_2^{2/3} \right]^{-1/2} \tag{27}$$

Thus $V(r)$ can be computed and hence the socalled scattering integral
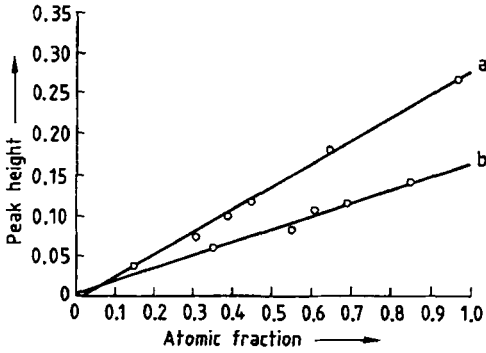
Figure 62. Calibration of the ISS peak heights of gold and silver against the known bulk compositions of a series of Au–Ag alloy standards [210]
a) Au; b) Ag

$$\psi = \pi - 2p \int\limits_{r_{\min}}^{\infty} dr/r^2 \left\{ 1 - (p^2/r^2) - V(r)/E \right\}^{1/2} \quad (28)$$

where $\psi = \theta + \sin^{-1}\{(\sin\theta)/A\}$, θ being the scattering angle, and $A = M_2/M_1$; p is the impact parameter; and $E = \{M_2/(M_1 + M_2)\}E_0$.

From Equation (28) can be calculated the differential scattering cross section for scattering into unit solid angle, which is

$$d\sigma(\theta)/d\Omega = (p/\sin\theta)(dp/d\theta) \quad (29)$$

and hence the cross section.

Neutralization Probability. The neutralization probability P in Equation (24) is not quite so straightforward to compute. As stated earlier, neutralization is a highly probable event, because the approaching ion represents a deep potential well compared to the work function of a solid. Based on the theory of neutralization by Auger electron ejection developed by HAGSTRUM [207], the probability of an ion escaping neutralization has the form

$$(1 - P) = \exp\left\{ \left(\frac{1}{V_{IT}} + \frac{1}{V_{ST}} \right) V_c \right\} \quad (30)$$

where V_{IT}, V_{ST} are the components of velocities of initial and scattered ions, respectively, normal to the surface; and V_c is the characteristic velocity.

V_c is dependent on the electron transition rate, which in turn depends, among other things, on the chemical state of the surface. It is approximated by

$$V_c = \int\limits_{r_0}^{\infty} R(r)\,dr \quad (31)$$

where $R(r)$ is the transition rate and r_0 the distance of closest approach of ion to surface.

V_c is thus a complex quantity, and difficult to estimate, which makes $(1 - P)$ difficult to calculate. However, most estimates indicate that $(1 - P) \approx 10^{-3}$.

Quantification. As with the other surface analytical techniques, in principle one can therefore quantify by calculation, and indeed in a few cases, such as the quantitative study of potassium segregation to the V_6O_{13} (001) surface carried out by DE GRYSE et al. [208], such an approach has been successful. In general, however, the uncertainties in some of the quantities in Equation (24) make such a calculation too difficult, and quantification is based on standards. Thus, BRONGERSMA et al. [209], in their study of surface enrichment of copper in a Cu–Ni alloy, calibrated the ISS intensities against pure copper and nickel targets, whereas NELSON [210] used gold and silver elemental standards in ISS analyses of a range of Au–Ag alloys. He found a linear relationship between the surface concentrations of the alloys after prolonged sputtering to equilibrium composition, and the bulk concentrations measured by an electron microprobe, as shown in Figure 62. BEUKEN and BERTRAND [203] reduced each of the spectra from the standard materials in Figure 59 to components consisting of Gaussian peaks and inelastic tails, and then added the components in different proportions to synthesize the observed spectra from their sulfided Co–Mo catalysts. The result of such synthesis when applied to the ISS spectrum from the catalyst after heating to 800 °C is given in Figure 63. From such data processing the relative concentrations of the metallic elements on the catalyst surface after various heat treatments could be derived very accurately.

Depth Profiling. Unlike SSIMS, ISS is not inherently destructive under normal conditions of operation (i.e., He⁺ at a few kiloelectronvolts and 1–10 nA) since the sputtering rate is negligible. If information in depth is required, then surface erosion must be introduced, as for XPS and AES. One obvious way of performing depth profiling is to use the same primary ion for both ISS and sputtering, but at quite different power densities for the two operations. Thus ROSSI et al.
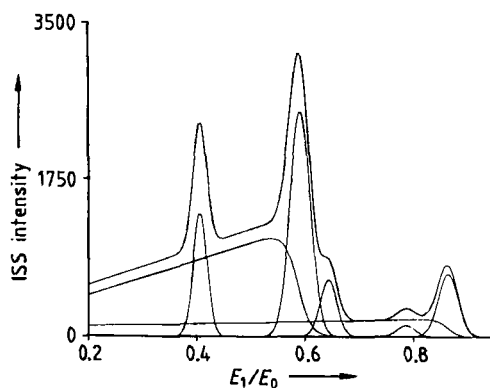
Figure 63. Curve synthesis method applied to the ISS spectra from the catalyst $CoMoS - \gamma - Al_2O_3$ after heat treatment at 800 °C [203]

[211] used Ne^+ at 45 nA and 3 keV (i.e., 1.35×10^{-4} W) for ISS, and at 300 nA and 5 keV (i.e., 1.5×10^{-3} W) for sputtering, in their study of passive film formation on Fe53 – Ni. Figure 64 shows the Fe – Ni regions of ISS spectra obtained by them after successively longer sputtering times. Inelastic backgrounds have been subtracted from the spectra, and the peaks have been resolved into iron and nickel contributions. After 5 min sputtering, the iron contribution is greatly enhanced, indicating an Fe-rich layer within the passive film, about 1 nm below the surface. On prolonged sputtering the bulk composition is reached.

Another method of depth profiling with ISS, used by KANG et al. [212], is to bombard the surface simultaneously with He^+ and Ar^+ ions. ISS is performed using He^+, while the surface is eroded with Ar^+. Because the gases were mixed together in the same ion source, a duoplasmatron, the energies of measurement and sputtering were the same, in the range 1.5 – 2.0 keV. According to the mixing ratio used, the relative currents of He^+ and Ar^+ could be varied.

The much smaller sampling depth of ISS, of the order of one atomic layer, compared to those of XPS and AES, implies that the depth resolution in ISS during profiling will be better. This improved depth resolution has been demonstrated by HAEUSSLER [213] in comparisons of ISS, XPS, and AES profiles of the same materials.
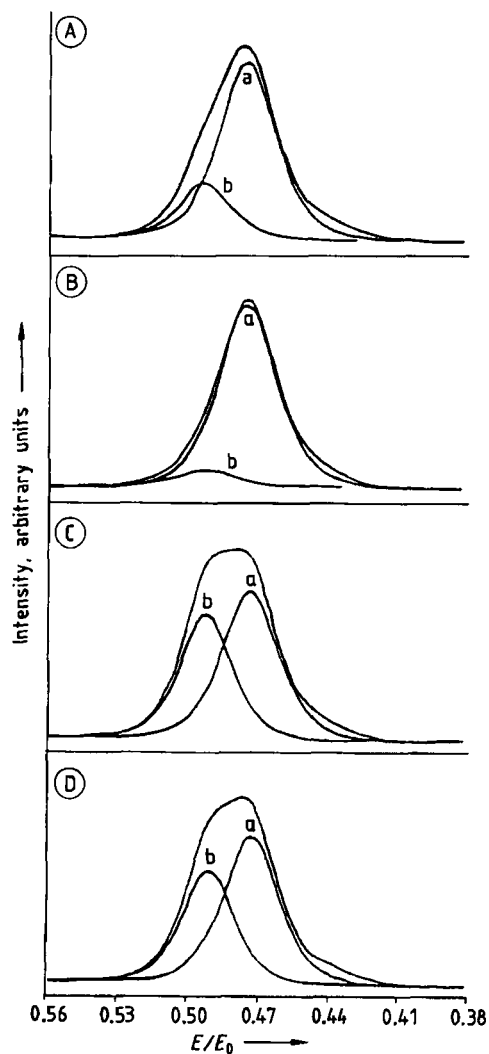
Figure 64. Sequence of ISS spectra from an Fe53 – Ni47 alloy passivated in 1 mol/L NaOH ($t_p = 5$ min, $E_p = 0.44$ V), during depth profiling [211]
The peaks have been resolved into iron and nickel contributions after subtraction of an inelastic background
A) 1 min; B) 5 min; C) 15 min; D) 30 – 120 min; a) Iron; b) Nickel

27.5.1.5. Applications

Because of its high surface specificity relative to XPS and AES, most published applications of ISS have been concerned with basic experiments on clean surfaces, usually of single crystals. The negligible sputtering rate when using low-energy He^+ also means that gas – solid interactions can be studied without worry about alteration of the surface during measurement.
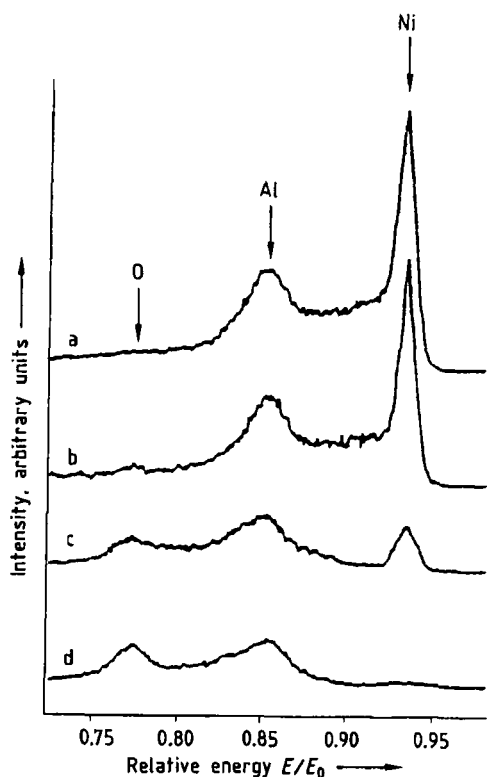
Figure 65. ISS spectra from Ni₃Al (001) after exposure to oxygen at 2×10⁻⁵ Pa and 700 °C [214]
a) Clean; b) 1-s Exposure; c) 15-s Exposure; d) 100-s Exposure

Figure 66. ISS spectra from a tungsten surface containing oxygen and barium, using a mixture of He^+ and Ne^+ to analyze for oxygen as well as separate barium and tungsten [215]
×4 = amplification factor

Typical of recent basic studies has been that of Shen et al. [214] of oxygen adsorption on Ni₃Al (001). The intermetallic compound Ni₃Al is known to have excellent oxidation resistance in air at high temperature, forming an adherent aluminum oxide film, so information about the initial stages of oxidation is of interest. The Ni₃Al (001) surface was cleaned by Ne^+ ion bombardment, and its stability as a function of temperature checked by ISS, which showed that up to 900 °C the Ni – Al ratio remained unchanged. Figure 65 shows a selection of ISS spectra from the alloy surface as a function of oxygen exposure at 700 °C, using 2-keV He^+ at current densities between 2×10⁻⁸ and 8×10⁻⁸ A/cm⁻². The oxygen peak increases, as expected, and both the aluminum and the nickel peaks decrease. However, the nickel peak intensity decreases at a much faster rate than that of aluminum, and the peak eventually disappears. Plots of normalized nickel and oxygen peak intensities indicated that the nickel peak disappeared when one

monolayer of Al_2O_3 was formed, confirming that the protective mechanism operates in the earliest stages of interaction. The authors also used ISS structurally to establish that the growing oxide had no order and was not in coherence with the underlying crystal.

A more technological application to which ISS is particularly suited is that of the surface composition of dispenser cathodes. These cathodes are believed to achieve optimum operation when an approximately monolayer coverage of barium and oxygen exists at the surface. When using XPS and AES, establishing that any barium signal comes only from the outer atomic layer of the surface is not easy, but the single-layer sensitivity of ISS should be able to provide that information. A problem in applying ISS to the cathode system is that the active barium–oxygen layer is supported on a tungsten substrate, and that both barium and tungsten are heavy elements whose resolution using He^+ would be difficult. On the other hand, oxygen must also be analyzed. The problem was solved by Marrian et al. [215] by performing ISS with a mixture of He^+ and Ne^+, so that oxygen would be analyzable by the He^+, while the use of Ne^+ would help separate barium and tungsten. A spectrum from their paper is given in Figure 66, showing that the analyses using the mixture did in fact
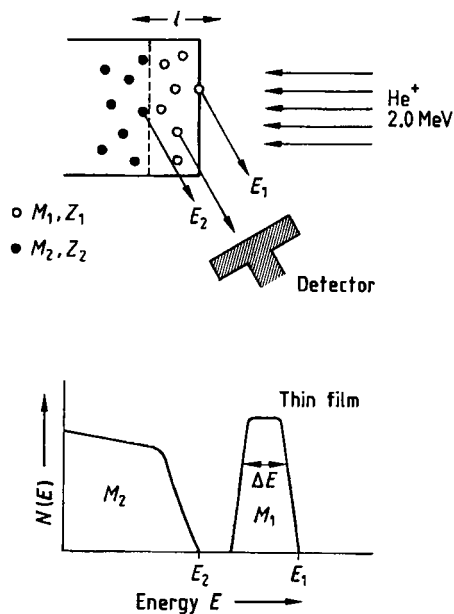
Figure 67. Illustration [219] of the principle of Rutherford backscattering spectroscopy (RBS) from a solid consisting of a thin film of thickness l composed of atoms of mass M_1 and atomic number Z_1 situated on a substrate of atoms of mass M_2 and atomic number Z_2

provide the analytical information needed. The authors were able to establish the existence of fractional monolayer coverage of barium and oxygen on the cathode surface.

Other recent applications of ISS have included the study of thin ZrO_2 films [201], a comparison of various techniques for cleaning InP substrates before epitaxial growth [216], analysis of conversion coatings on aluminum alloys [217], and study of alkali-induced sintering of an iron oxide monolayer on platinum [218].

27.5.2. Rutherford Backscattering Spectroscopy (RBS)

RBS has been in use since the 1960s for bulk materials analysis, and increasingly since the 1970s for thin-film analysis, particularly in the semiconductor field. The overall number and range of applications are enormous, but those dealing specifically with surface problems are less numerous. Nevertheless RBS ranks along with SSIMS and ISS as one of the major surface analytical techniques.

27.5.2.1. Principles

As in ISS, elemental compositional analysis is achieved by measuring the energy losses suffered by light primary ions in single binary collisions. Although in ISS the primary species is usually He^+, in RBS either H^+ or He^+ is used exclusively, and at very much higher energies, typically 1–2 MeV. Consequently, due to the penetration of the primary ion, both outer atomic layers and subsurface layers of a material are accessible; the surface sensitivity comes when the material in the outer few monolayers consists of heavy elements.

The energy of the backscattered ion, and hence elemental analysis, is governed by the same relationship (Eq. 22) already described for ISS. However, in RBS the ion penetrates the target and may be backscattered by the target atoms at any point along its path. Because of its very high energy, the ion can lose energy by collisional interactions with electrons and plasmons along either or both incoming and outgoing paths, which are straight lines. Thus, instead of the single peak observed by scattering from a particular element in the first atom layer in an ISS spectrum, the additional losses in energy cause a much broadened peak to be observed in an RBS spectrum. In RBS the energy spectrum is in fact a superposition of mass (i.e., elemental) analysis and depth analysis.

The principles have been illustrated elegantly by van Oostrom [219], as shown in Figure 67. There a thin film consisting of atoms of mass M_1, atomic number Z_1, and thickness l is situated on a substrate of atoms of mass M_2 and atomic number Z_2. The scattered ion energies from M_1 and M_2 are E_1 and E_2, respectively. If $M_1 > M_2$, then the energy $E_1 > E_2$, in which case the peak corresponding to M_1 is well separated from M_2, and the peak width ΔE is related directly to l. Because the atoms M_1 are in the surface, the threshold E_1 will be given by Equation (22) as in ISS, but since the atoms M_2 are below the thin film the threshold E_2 may be lower than that given by Equation (22) if energy losses occur in the film. The energy spectrum produced by scattering from atoms M_2 will start at the threshold E_2 and then increase into a very broad band since M_2 forms the substrate of effectively infinite depth. If $M_1 < M_2$, the peak due to M_1 merges with the broad band from M_2, and may or may not appear as a much smaller feature on top of the band.

The above makes clear that for surface analysis by RBS, conditions must be such that $M_1 \gg M_2$ for
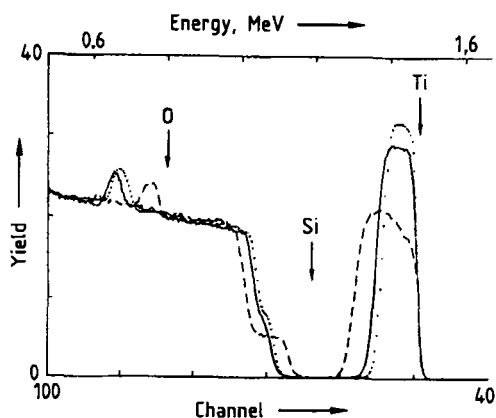
Figure 68. RBS spectra obtained from a sandwich structure of 110-nm titanium on top of 90 nm SiO_2 on a silicon substrate [221]
... =Before heating; ⎯ =After heating to 575 °C;
- - - - =After heating to 800 °C

the peak due to M_1 to be completely resolved. Under these circumstances, the sensitivity of RBS is comparable to AES and XPS (i.e., of the order of 10^{-5} atomic layers). As in ISS, no chemical information is available.

27.5.2.2. Instrumentation

He^+ or H^+ ions with energies in the megaelectronvolt region are produced in particle accelerators. These may take the form of a tandem electrostatic Van de Graaff accelerator; a cyclotron, which uses electromagnetic confinement to a circular orbit; or a linear accelerator, in which the ion is accelerated linearly across gaps between drift tubes by electromagnetic fields. The reader is referred to [220] for greater detail about accelerator technology and about accelerator-based methods of analysis. Accelerators normally produce ions in several charge states, so that the required ion is chosen by magnetic selection and often by velocity selection as well. The purified ion beam passes immediately into a UHV environment via differentially pumped apertures, since the vacuum environment in most accelerators is not that of UHV. For general analysis, in which channeling of the beam along particular crystal directions is not present, ions strike the specimen at normal incidence. When structural information is sought, however, in the same manner as in ISS, the specimen may be rotated about the point of ion impact

to vary the angle of incidence about channeling directions.

Detection of backscattered ions is normally performed by a silicon surface barrier detector at high scattering angles for increased surface sensitivity, as indicated in Figure 67. The detector position is often variable about the specimen position to allow various scattering angles to be chosen.

The ion beams used in RBS can be focused to small spots if required, although conventional analysis would normally be performed over an area of ≈ 1 mm^2. Spot sizes of $5-10$ μm can be obtained by suitable lens systems, but scanning at such high energies is too difficult. Selected area analyses using small spot beams is carried out by specimen manipulation.

27.5.2.3. Spectral and Structural Information

An RBS spectrum, like that of an ISS spectrum, consists of the energy distribution of backscattered primary ions. Unlike ISS, however, the energies of the backscattered ions are governed not only by the mass of the scattering atom but also by the depth of that atom.

An example of the nature of the compositional information obtainable from an RBS spectrum is based on Figure 68, from the work of KUIPER et al. [221]. They grew a 90-nm-thick film of SiO_2 on a silicon substrate, then deposited a 110-nm-thick film of titanium on top of the SiO_2, and heated the sandwich to various temperatures. In Figure 68 the dotted line is the spectrum recorded before heating; the solid line, after heating to 575 °C for 1 h; and the dashed line, after heating to 800 °C for 1 h. The vertical arrows are positioned at the backscattered energies expected for atoms of the indicated elements if they were in the surface layer.

For the titanium film, some titanium atoms always exist in the outer layer, but on heating to successively higher temperatures, titanium is found at progressively greater depths (i.e., it is redistributing). The width of the peak before heating corresponds exactly to the thickness of the deposited film. In the case of silicon, at no stage do any silicon atoms appear in the outer layer, although after heating to 800 °C the silicon edge is near the surface. The double edge of the silicon signal is due to the fact that, after heating to 800 °C, silicon is in two chemical states, one as a silicide in combination with titanium, and the other in the underlying SiO_2, which has reduced in thickness to 14 nm. After 800 °C heating, the ox-
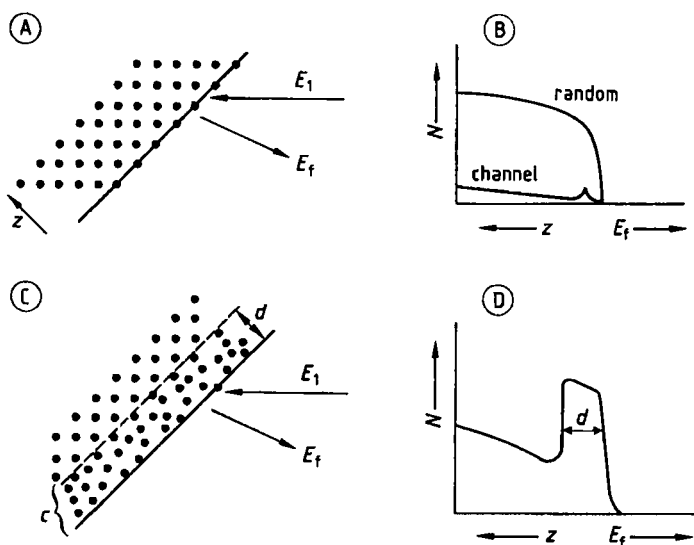
Figure 69. Illustration of RBS expected according to direction of incident beam [222]
A) Beam along a channeling (open) direction; B) Spectra expected from either nonchanneling (random) or channeling (open) directions; C) Beam incident on an amorphous layer of thickness d on surface; D) Spectrum expected from amorphous layer

ygen atoms have mostly moved nearer to the surface, with little change in their distribution after 575 °C treatment.

For relatively simple situations such as that of Figure 68, a spectrum simulation program can be used, in which the adjustable parameters are the stoichiometries of the likely compounds and their distribution with depth. Thus after 575 °C heating, the simulation gave the result that the surface consisted of titanium on top of a layer of $TiO_{0.23}$ overlaying 72 nm of SiO_2, with negligible silicide. After 800 °C, however, the fit was to a layer of Ti_2O_3 on top of the silicide Ti_5Si_3, covering the remaining 14 nm of SiO_2. The results agreed with depth profiling by AES performed at the same time.

RBS can be used to detect and measure the thickness of an amorphous layer on an otherwise crystalline substrate. The method is illustrated schematically in Figure 69, from WERNER [222]. If the incident ion beam is directed onto the crystal in an "open" or channeling direction (Fig. 69 A), it will penetrate more deeply into the crystal than if it is aligned along a nonchanneling or "random" direction. Thus the overall backscattered intensity in the channeling direction will be much lower than in the random direction, and in addition the small surface peak arising from collisions at the surface will be visible, as indicated in schematic

spectra in Figure 69 B. Should the surface be covered with an amorphous layer (Fig. 69 C), no channeling is possible in the layer, so that more collisions will take place with atoms near the surface, and the intensity of the backscattered signal from the surface will increase (for the same incident beam direction), as shown schematically in Figure 69 D.

An application of this method of obtaining surface structural information has been demonstrated by BHATTACHARYA and PRONKO [223]. Implantation in compound semiconductors for production of the required electrical characteristics is usually accompanied by crystalline damage, and in particular a surface amorphous layer may be formed. Figure 70 from their paper shows the results of the RBS channeling measurements on GaAs implanted with silicon before and after annealing at 950 °C. The upper spectrum (a) is taken with the ion beam in a nonchanneling or "random" direction, and has the expected structureless shape. When the beam was aligned with a channeling direction, the spectrum before annealing (b) was also of high intensity and revealed a broad peak near the edge, while after annealing the intensity dropped substantially and a much narrower near-edge peak remained (c). The broad peak in the unannealed state showed that the surface had become amorphous to a thickness of
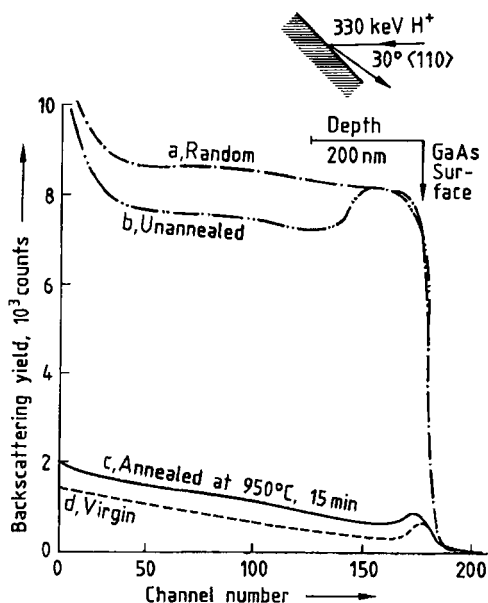
Figure 70. RBS spectra from GaAs implanted with silicon before and after annealing at 950 °C (120 keV Si$^+$–GaAs, dose: 5×10^{15} cm^{-2}) [223]
Uppermost spectrum is taken in a nonchanneling direction; the others are in channeling directions

≈ 140 nm, while the spectrum after annealing resembled closely that of the crystalline unimplanted crystal (d). Further structural information was obtained by recording the backscattered intensity as a function of angle around the GaAs [100] and [110] crystal directions, from which it was possible to calculate that about 70 % of the implanted silicon occupied substitutional lattice sites.

27.5.2.4. Quantification

The intensity of the current of high-energy ions backscattered from surface atoms of type A is very similar to Equation (24) for ISS, except that of course the neutralization probability is not relevant. Equation (24) then becomes for RBS

$$I_A^- = I_o^- N_A \sigma_A S \cdot \Delta\Omega \cdot T \tag{32}$$

where the symbols have the same meanings as before. Here σ can be derived from the Rutherford differential scattering cross section dσ given by

$$d\sigma = \frac{Z_1 Z_2 e^2}{4E^2} \cdot \frac{\cos\theta - \left\{1 - [(M_1/M_2)\sin\theta]^2\right\}^{1/2}}{\sin^4\theta\left\{1 - [(M_1/M_2)\sin\theta]^2\right\}^{1/2}} \tag{33}$$

θ being the scattering angle. For a particular primary ion and fixed experimental conditions, the differential cross section is thus a function only of the mass and atomic number of the scattering atom A, and can be calculated. Thus in principle, all the terms in Equation (32), except the required N_A, are known. In practice, the effective solid angle $\Delta\Omega$ of the analyzer is not well known, and using standards against which to calibrate the peak in the RBS spectrum arising from surface atoms (the so-called surface peak) is more common than attempting to derive calculated compositions. Such a peak can be seen clearly in the lower spectra of Figure 70.

Calibration is achieved by either implanting into or depositing on a light element such as silicon, a known amount of a much heavier element. Then if N_S is the known density of the standard element, one can write

$$N_A = (\sigma_S/\sigma_A) \cdot (I_A^+/I_S^+) N_S \tag{34}$$

given the same experimental conditions of measurement on the standard and the unknown. Then, in terms of the usual relative sensitivity factors, Equation (34) can be written as

$$N_A = R(A{:}S)I_A^+ \tag{35}$$

The accuracy depends on being able to measure the areas of the surface peaks in both standard and unknown with sufficient accuracy, which is often easy. Accuracies in RBS under favorable conditions can be better than 1 %.

27.5.2.5. Applications

Since RBS has operated as a major compositional and structural technique for bulk analysis for about 30 years, the number of published applications in that area is now enormous. Its applications in the field of specifically surface studies have not been so numerous, but they are increasing in number with the advent of improved vacuum technology. Many have been of the type already discussed in Section 27.5.2.3, in which the reaction or interdiffusion of one or more heavy metals on a semiconductor surface has been studied.

Typical of such studies is that of Nava et al. [224] on the interaction of an Ni–Pt alloy with crystalline silicon. The system is a particularly favorable one since both nickel and platinum are much heavier than silicon, and platinum is in turn much heavier than nickel. Alloy films about
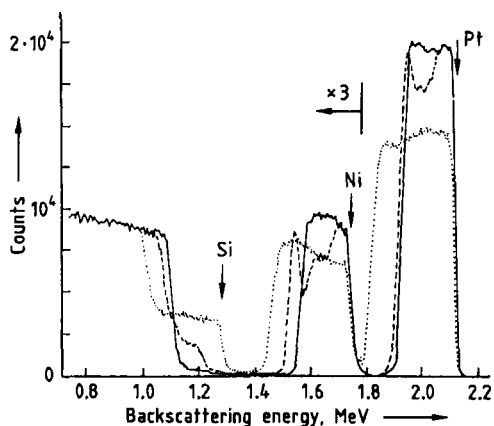
Figure 71. RBS spectra of a 145-nm-thick Ni – Pt alloy film on silicon before and after annealing [224]
— = Before annealing; - - - = Annealed at 400 °C; . . . = Annealed at 700 °C

145 nm thick were deposited on silicon by code-position of nickel and platinum in approximately equal amounts. Annealing of the film and substrate was then carried out at progressively increasing temperature. Figure 71 shows the RBS spectra before annealing (solid line), after 30 min at 400 °C (dashed line), and after 30 min at 700 °C (dotted line). As before, the vertical arrows indicate the energy positions of the backscattered edges to be expected if the indicated elements are in the surface layer.

As shown, the as-deposited film contained both platinum and nickel atoms in the surface layer, and covered the silicon completely, since no silicon atoms appeared at the surface layer energy position. The sharpness of the trailing edges of the platinum and nickel peaks, as well as that of the leading silicon edge, showed that the interface between film and substrate was itself sharp. The film composition, based on relative peak area under the platinum and nickel peaks, was $Ni_{55}Pt_{45}$. Annealing at 400 °C causes little movement of the platinum, but segregation of the nickel into the interface, coupled with movement of silicon in the opposite direction (i.e., also into the interface). The compound formed there is identified as NiSi from the silicon step height. At 700 °C platinum has also shifted considerably, along with further movement of nickel, until they are in fact completely intermixed. The silicon edge has now moved up to the surface position, indicating that it is mixed with both platinum and nickel. Careful

quantification establishes that both PtSi and NiSi have been formed in approximately equal quantities. Confirmation of the formation of these compounds came from X-ray diffraction. Thus the alloy – silicon reaction occurred first by the formation of NiSi, with a Pt-rich layer left above it, followed at slightly higher temperature by the reaction of platinum with diffusing silicon, and finally at the highest temperatures by complete mixing of PtSi and NiSi.

Similar metal – Si reaction studies have been performed by DE NIJS and VAN SILFHOUT [225] on Ti – Si, by FASTOW et al. [226] on pulsed ion beam irradiation of Ni – Cr on silicon, by ELLWANGER et al. [227] of $Pt_{0.4}Ni_{0.6}$ on silicon, and by OBATA et al. [228] of TiC on carbon and molybdenum.

RBS has also been used to characterize palladium and tin catalysts on polyetherimide surfaces [229], titanium nitride thin films [230], silicon oxynitride films [231], and silicon nitride films [232], and to study the laser mixing of Cu – Au – Cu and Cu – W – Cu thin alloy films on Si_3N_4 substrates [233], and the annealing behavior of GaAs after implantation with selenium [234].

27.6. Scanning Tunneling Methods (STM, STS, AFM)

Scanning tunneling microscopy (STM) was invented by BINNIG and ROHRER in 1983 [235], for which they received the Nobel Prize. The ability to produce images related directly to topography on an atomic scale was a major step forward for surface science, which until then had had to deduce surface atomic structure from techniques such as LEED and ISS. Since the appearance of STM, many sister techniques using similar principles have been developed for the study either of special materials, of surfaces in nonvacuum ambients, or of the surface distribution of effects other than purely topographical.

27.6.1. Principles

In *scanning tunneling microscopy* a metal tip is brought so close to a conducting surface (of the order of 0.5 – 2.0 nm) that an electron tunneling current becomes measurable at applied voltages in the range 2 mV to 2 V. The tip is then scanned over the surface. The tunneling current is monitored while scanning and used in either of two
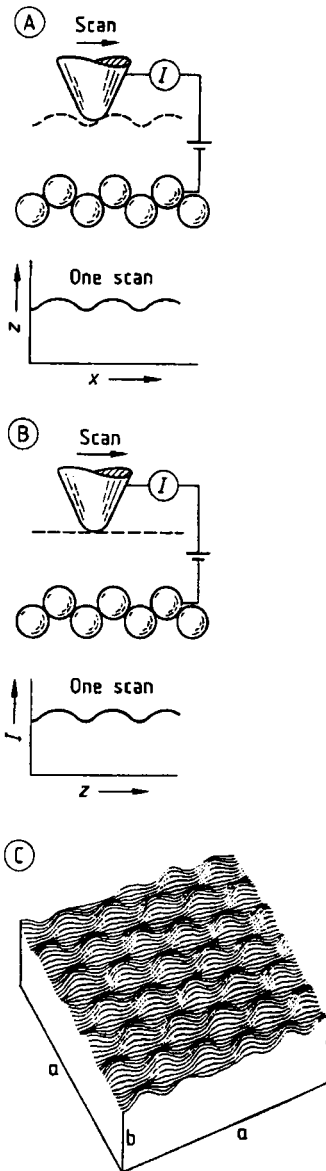
Figure 72. Illustration of modes used in STM [236], [237]
A) Constant current mode; B) Constant height mode;
C) Surface topographical-spectroscopic image produced by
scanning along progressive incremental distances in the *y*
direction
a) 1.3 nm; b) 0.5 nm

ways: in one, a feedback circuit changes the height
of the tip above the surface in order to maintain the
current constant (*constant current mode*); in the
other, the height of the tip is kept constant while
variations in current are recorded (*constant height*

mode). In the simplest, one-dimensional view, the
dependence of tunneling current I_T on applied
voltage V_T and tip–surface separation s is given
by the approximation

$$I_T \approx V_T \exp\left(-A\varphi^{1/2}s\right) \tag{36}$$

where A is a constant and φ is the effective local
work function. Equation (36) shows that for a
typical work function value of about 4.5 eV, the
current changes by about an order of magnitude
for each change in the separation of 0.1 nm, when
close to the surface. Therefore in the constant
current mode of operation the tip follows the sur-
face contours by rising or falling, whereas in the
constant height mode the current rises or falls. In
either case an image related to surface topography
is obtained by constructing a map from line scans
displaced progressively in one direction by equal
amounts. The principle is illustrated schematically
in Figure 72, taken from Hansma and Tersoff
[236] and Hansma et al. [237].

For constant local work function φ, Equation
(36) predicts a linear dependence of I_T on V_T, but
this is obeyed only at very low values of V_T, since
the equation is approximated from a more com-
plex and rigorous expression. At higher V_T, the
one-dimensional tunneling model on which Equa-
tion (36) is based breaks down, and the three-di-
mensional electronic structure of the surface re-
gion (i.e., the local band structure) must be con-
sidered. The tunneling current must then be calcu-
lated by first-order perturbation theory: if the as-
sumptions are made that the tip can be replaced by
a point and that the tip wave functions are local-
ized (i.e., a constant density of surface states), then
the current is given by

$$I_T \sim \sum_s \|\psi_s(r)\|^2 \delta(E_s - E_F) \tag{37}$$

where ψ_s is the wave function of surface state s; E_s
the energy associated with ψ_s; and E_F the Fermi
energy. The quantity on the right of Equation (37)
is simply the surface local density of states
(LDOS). Thus the tunneling current is propor-
tional to the surface LDOS at the position of the
point. For a fixed applied voltage V_T, the image in
STM is therefore not just a function of topograph-
ical variations, but is essentially an image of con-
stant surface LDOS. If the tip is now positioned
over a particular point on the surface, and the
current recorded as a function of V_T, the $I-V$
dependence will mirror the LDOS, that is, peaks
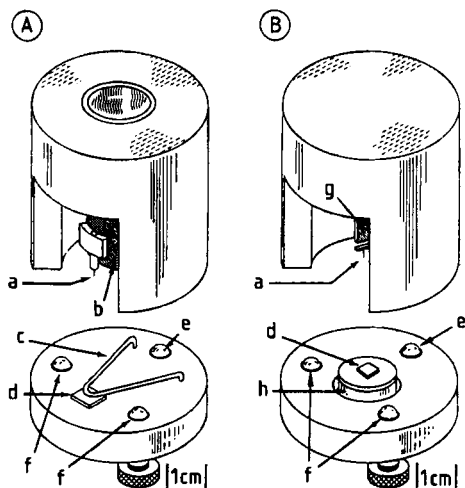
Figure 73. Schematics of the designs for an STM and an AFM used by HANSMA et al. [237]
A) STM; B) AFM
a) Tip; b) Single-tube *xyz* piezo; c) Sample clip; d) Sample; e) Fine advance screw; f) Coarse advance screws; g) Spring deflector sensor; h) *xyz* Piezo

will occur near saddle points in the surface band structure, along with features due to localized states (e.g., so-called dangling bonds). The $I-V$ dependence is therefore providing a tunneling spectroscopy, and according to whether $V_T < 0$ or $V_T > 0$, occupied or unoccupied states, respectively, are explored. Once a spectral feature has been found at a particular value of V_T, if the applied voltage is set to that value and the tip is scanned as before, a two-dimensional distribution of that feature is obtained. This is known as *scanning tunneling spectroscopy* (STS).

Because STM and STS are basically the same, in that topography and electronic distribution at the surface both contribute in varying proportions to the image that is actually seen, other methods of recording pure topography have been sought. In addition, STM (STS) requires that the surface be conducting, whereas the topography of many nonconducting surfaces would be of interest. In 1986, BINNIG et al. [238] calculated that a cantilever could be made with a spring constant lower than that of the equivalent spring between atoms. Thus the vibrational frequencies of atoms in a molecule or solid combined with the atomic mass lead to interatomic spring constants of about 10 N/m, compared with spring constants of 0.5 – 1 N/m in narrow strips of soft metal such as aluminum. A cantilever with a very fine tip virtually in contact

with the surface, with actual applied forces in the range of $10^{-13} - 10^{-6}$ N, and a means of sensing and monitoring the tip deflection, will therefore measure tip – surface interactions arising from van der Waals, electrostatic, magnetic, and other weak forces. This is known as *atomic force microscopy* (AFM). It does not depend on the conducting nature of the surface, and the information is purely topographical; a vacuum environment is not necessary, and indeed its application in a variety of media has led to a number of variants of the technique.

27.6.2. Instrumentation

All *scanning tunneling microscopes* have been constructed as variations on the schematic shown in Figure 72. Although several commercial instruments are available, the basic construction is relatively simple, so that many individual designs have appeared in the literature. An example is shown in Figure 73 A, from HANSMA et. al. [237]. The sample (d) is fixed to a table mounted on three advance screws (e, f). The tip (a) is attached to the end of a piezoelectric translator (b), the tripod arrangement of Figure 72 being replaced by a segmented tube since the latter is more rigid and therefore less sensitive to vibration. The sample is brought near the tip first by manual operation of the coarse advance screws while the tip is viewed through a microscope, until the separation is about 10 μm. Then a stepping motor drives the fine advance screw to bring the sample more slowly toward the tip, the final approach being monitored via the tunneling current itself. Since the ends of the coarse advance screws form the pivotal axis for the fine screw, the motion of the latter is reduced by a factor of 10, allowing very fine approach adjustment.

The tip is scanned linearly across the surface by ramped voltage applied to the *x* piezosegment, with progressive shifts in the *y* direction to create a rastered pattern of movement. Distance to the sample surface is controlled by a voltage applied to the *z* segment, the voltage being determined in a feedback circuit by the tunneling current. Images are accumulated by multiple scans, each scan consisting of the voltage applied to the *z* segment as a function of that applied to the *x*, and displaced successively by *y* shifts.

Fabrication of tips does not require any difficult technology. The reason for this is that any fine point fabricated by mechanical or electrochemical
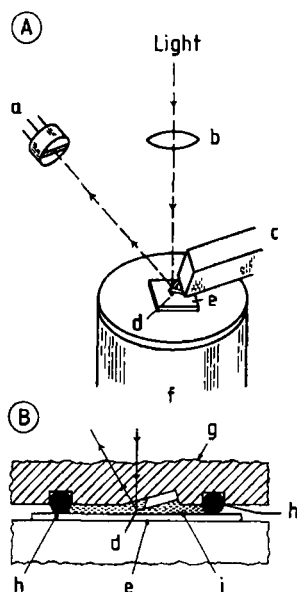
Figure 74. Cantilever deflection sensing used in AFM [239]
A) Principal scheme; B) Example of use in a liquid cell
a) Photodiode; b) Lens; c) Cantilever; d) Tip; e) Sample;
f) *xyz* Translator; g) Plexiglas; h) O-ring; i) Liquid

means will not be smooth on an atomic scale, but will have many "microtips" of near-atomic dimensions. When the tip is brought close enough to the surface, the tunneling current automatically selects the microtip that is nearest. Occasionally tips compete, or undesirable whiskers may be present, in which case an increase in voltage beyond the tunneling range will usually rectify the situation by removal of one or the other.

The other constructional requirement is rigidity. Low-frequency vibrations such as those found in buildings must not reach the tip – sample region; so, as indicated in Figure 73, components must be robust and rigid in order that such vibrations are damped.

In an *atomic force microscope* (an example is given in Fig. 73 B), the deflection of the cantilever spring is normally sensed by reflection of a laser beam from a mirror on the back of the cantilever into a two-segment photodiode, as shown in Figure 74 from WEISENHORN et al. [239]. The sample is positioned on a piezoelectric translator, and the signal from the photodiode is sent to a feedback circuit controlling the voltage applied to the z piezoelement, which moves the sample up or down in response to small changes in cantilever deflection. In this way an almost constant force

can be maintained on the tip, which is important when imaging vulnerable biological material. For the latter the surface is scanned at the smallest force compatible with tip contact. Rastering to produce images is performed by the x and y piezoelements of the translator.

Since the AFM does not use tunneling currents to sense the deflection of the cantilever and does not require a vacuum environment for operation, it has been used very successfully under water and other liquids. Figure 74 B shows the type of liquid cell employed by HANSMA and coworkers [239] – [243]. The tip with its mirror is attached to the underside of a strip of poly-(methyl methacrylate) that also carries a circular groove for a small O-ring. Once the sample has made contact with the O-ring, the cell is sealed, but the O-ring material is sufficiently flexible to allow movement of the sample under the tip in the usual way. A recent modification [244] of the design allows the fluid in the cell to be circulated in order to simulate flow conditions, while inserted electrodes allow cyclic voltammetry to be performed.

27.6.3. Spatial and Spectroscopic Information

Since the inception of STM, and latterly of AFM, a large number of elegant and often beautiful images have appeared in the literature, increasingly in color to emphasize the contrast. Although atomic resolution seems to have been achieved frequently with *STM*, care must be taken in the interpretation of images because of the overlap of topographical and electronic distribution variations. Thus in Figure 75, from EBERT et al. [245], imaging of the InP (110) surface reveals the positions either of the phosphorus atoms (A) or the indium atoms (B), depending on the applied voltage. In (A), at $V_T = -2.6$ V, the occupied state density is mapped, and in (B), at $V_T = +2.2$ V, the unoccupied state density. In both cases the resolution is atomic, but the spatial localization of the states is quite different.

With *AFM*, the images are unambiguous, provided the tip is very sharp so that multiple microtip effects are absent. Several materials have been used as tests of resolution, a favorite being freshly cleaved, highly oriented pyrolytic graphite, in which the average distance between carbon atoms is 0.14 nm. Calculations by GOULD et al. [246] show that every surface carbon atom should be imaged with equal intensity, and Figure 76 from

(a)   (b)

[001]

[ī10]
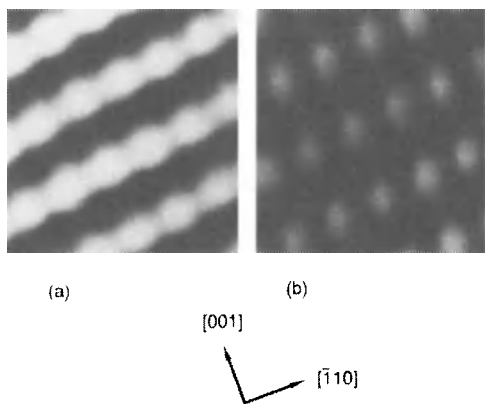
Figure 75. STM images of the InP (110) surface over an area of 2×2 nm² [245]
A) Tunneling voltage –2.6 V; B) Tunneling voltage +2.2 V. Images at the two voltages reveal the occupied and unoccupied state densities, respectively

Figure 76. AFM image of graphite in air [247]
Image size 1.17 nm. Carbon atoms can be seen as darker regions around central holes

GOULD et al. [247] confirms this prediction. The field of view in the figure is 1.17 nm across, and individual carbon atoms can be seen as alternate light and dark regions around each hole.

According to the applied voltage range, as indicated above for InP, STM, or rather, STS, can give information on both occupied and unoccupied surface states. The ST spectra over an extended range of V_T are therefore likely to be comparable with those from other techniques, although the comparison may not be exact because of the very localized nature of the tunneling that takes place in STS. Only if the bulk band structure is similar to the surface density of states (DOS) is agreement likely, and this has been found by KAISER and JAKLEVIC [248] for the clean Pd (111) surface. In Figure 77 from their paper, the $I-V$ spectrum from –2.4 to +2.4 V (Fig. 77 A) is compared with the experimentally determined band structure (Fig. 77 B) and the UPS spectrum (Fig. 77 C). The STS peaks at –1.4 and +1.0 V correspond to the positions of the d-band edge at about –1.2 eV and of the saddle point at +0.94 eV in B, while in the UPS spectrum (Fig. 77 C) surface-state peaks are found at about –0.5 and –2.1 eV, which match the peaks at –0.6 and –2.1 V. The minor feature in A near +0.5 V may also be a surface state. STS has also been compared with IPES [249] over the unoccupied range of surface states of graphite.

27.6.4. Applications

The ability of *AFM* to provide topographical images of all types of surface at spatial resolutions varying from the atomic to the micrometer scale, in air, vacuum, or under liquids, has led to a great variety of applications. They can be divided roughly into inorganic and organic (biological) applications.

An example of the way in which AFM has been applied to *inorganic material* surfaces is provided by the study of GRATZ et al. [250] on the dissolution of quartz in potassium hydroxide solution. Such a study forms part of a general attempt to understand the phenomena of the dissolution and growth of crystal surfaces, either of natural origin or grown synthetically. The authors etched optical-quality natural quartz for 4 h at 148 °C in KOH solution and then examined the surface with AFM, in this case in ethanol, which was found to keep the surface clean. Quartz is an insulator and could not therefore be studied by STM.

The etching created prismatic etch pits with smooth floors. Figure 78 shows some of the results. Image A, of area 1490×1490 nm², reveals the contours in the pit; the height separation from top to bottom of the pit is about 20 nm. The c and a axes of the quartz are indicated by the long and short arms of the cross, respectively. In image B, one quadrant of the pit is shown at higher magnification and in three dimensions, the field of
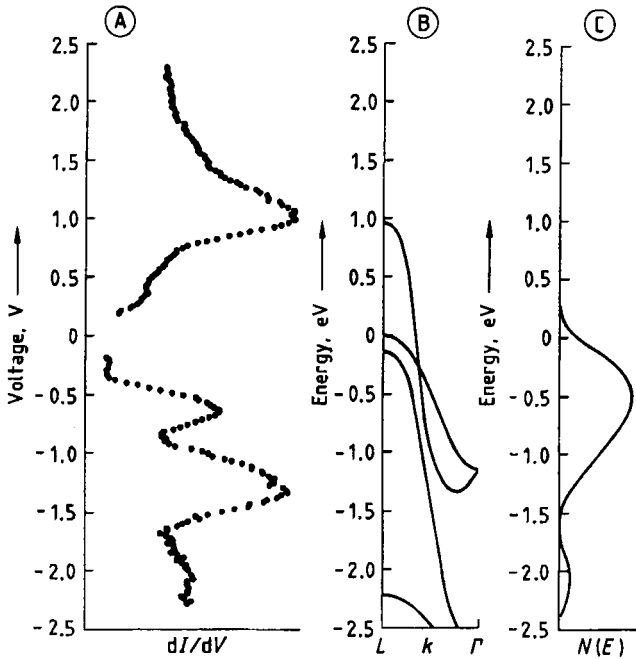
Figure 77. Comparison of
A) STS spectrum; B) Experimental band structure; and C) UPS spectrum, of palladium (111) [248]L, k, Γ = directions of symmetry in three-dimensional space

view being 650×650 nm^2. On the vertical scale the range is only 14 nm. The ledges visible in B are shown in C to be of only 0.7-nm height, to be atomically flat, and to be about 30 nm wide. Thus the etch pit can be described as an elongated amphitheatre, with a series of flat terraces of individual height only a few tenths of a nanometer. The results are said to be in agreement with the classical model of crystal dissolution.

Other recent inorganic material applications have included synthetic ultrafiltration membranes in air and in water [251], clay mineral surfaces [252]–[253], binding of molecules to zeolite surfaces [239], and electrochemical deposition of copper on gold in sulfate and perchlorate solutions [244].

Both AFM and STM have been able to image *biological materials* without apparent damage, which might be expected for AFM but is surprising for STM. In fact the imaging conditions used in STM must be kept as weak as possible to avoid damage. Thus, to obtain the image shown in Figure 79, from ARSCOTT and BLOOMFIELD [254], of a series of parallel fibers of Z-DNA, tunneling current and voltage levels were 1 – 2 nA and –100 to + 100 mV, respectively. The characteristic left-

handed twist of the Z-DNA helix is obvious in the image, and the measured helical periodicity of 4.19 nm is only 8 % less than the crystallographic value, indicating minimal contraction due to dehydration. The measurements were performed in air.

STM has also successfully imaged phosphorylase from muscle [255] and biomembranes [256]–[257], while AFM has produced images of proteins in a buffer solution [258], immunoglobulin [259]–[260], DNA [261], [252], and bacteria [263].

Although many surface analytical techniques, including some of those already described, can provide valuable information about the nature of surface reactions, the information is invariably averaged over a larger number of reaction sites. With STS, reactions and reactivity can be studied at selected individual, even atomic, surface sites, which is an exciting capability. The observations of WOLKOW and AVOURIS [264] on the reaction of NH_3 with the Si(111) surface illustrate the possibilities elegantly. Figure 80 from their paper shows a series of images of the Si (111)–(7×7) surface. In (A), the clean surface is imaged at + 0.8 V, revealing the characteristic hexagonal sur-
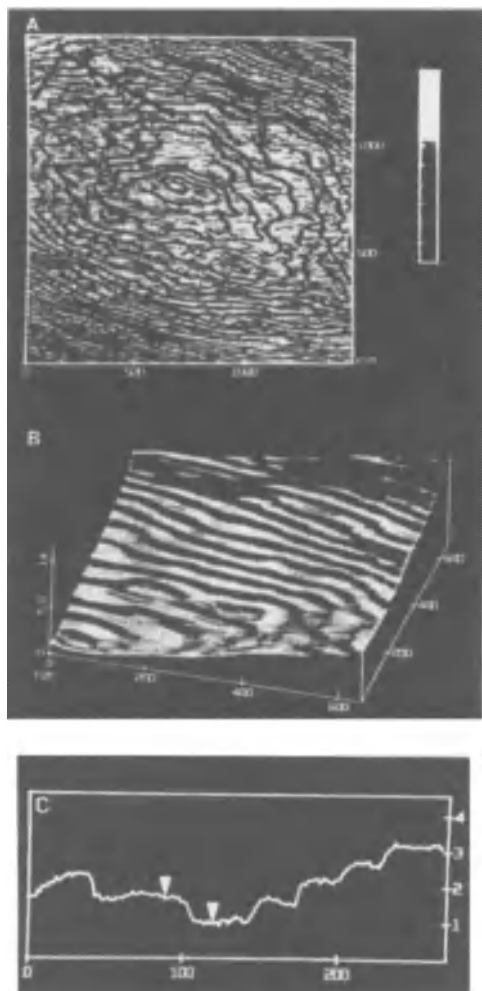
Figure 78. AFM images of an etch pit formed on quartz by etching in KOH at 148 °C [250]

The images were recorded under ethanol

A) Vertical view showing contours; B) Three-dimensional view of one pit quadrant; C) Profile across pit floor; height between cursors is 0.7 nm

Figure 79. STM image in air of parallel fibers of Z-DNA [254]

Tunneling current and voltage were kept very low to minimize damage

face structure, while in (B), the same tunneling voltage of +0.8 V is used, but after exposure to 1 Langmuir (i.e., 10^{-6} Torr · s or 1.33×10^{-4} Pa · s) of NH_3. In (B), about half the surface atoms seem to be missing. If the tunneling voltage is increased to +3.0 V, as in (C), the missing atoms reappear, but with different contrast from (A). The apparent disappearance is not real, but is an electronic effect due to the reaction of some of the surface atoms with NH_3; the basic structure is preserved, changes being those in the local DOS at certain atoms.

The accepted model for the (7×7) reconstruction of the Si (111) surface, that of TAKAYANAGI et al. [265], has silicon atoms in three special positions in the unit cell, those (in position A) underneath holes in a regular array in the outer layer, and others (B and C) present as adatoms (i.e., situated on top of the outer layer). Atoms B are adjacent to the holes at A; atoms C are in the center of the side of the unit cell. With STS the tip can be stationed over each of these atomic sites in turn, and the differences in local DOS recorded, both before and after reaction with NH_3. Figure 81 A shows the spectra from the clean surface, Figure 81 B, after the reaction. The quantity $(dI/dV)/(I/V)$ is proportional to the local DOS.

Spectra from the clean surface show a very strong band at ≈ -0.8 eV from atoms A, suggesting a fully occupied dangling bond surface state. Since atoms B and C have much weaker occupied bands, the implication is that charge transfer occurs from B and C to A, with more being contributed by C than B. Both B and C have stronger unoccupied bands, confirming the transfer. After reaction with NH_3, the strong occupied dangling bond state at A has vanished, and at B (the dashed line) the small occupied band has also been lost. Elimination of these states accounts for the apparent disappearance of reacted atoms from the image in Figure 80 B. The other spectra in Figure 81 B, the solid lines for B and C, are for atoms of those types not themselves reacted but adjacent to reacted A atoms. In both cases, but particularly for
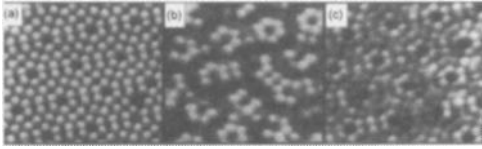
Figure 80. Series of STM images of the Si(111)–(7×7) surface [264]
A) Clean surface at 0.8 V; B) After exposure to 1 L of NH$_3$, also at 0.8 V; C) Same as B), but at 3.0 V

C, a transfer of intensity has occurred from the unoccupied to the occupied states, suggesting that during reaction a charge transfer takes place from A to C, and to a lesser extent B, allowing A to react with NH$_3$.

The above experiment demonstrates clearly how STS can follow changes in electronic distribution, and differences in reactivity, on an atomic scale — an unique capability. Although the Si (111) surface is probably one of the best candidates for such a demonstration, being atomically flat and possessing strong dangling bond states, the possibilities that are opened up by this technique in general are very clear.

STS has also been applied to a study of the unoccupied surface states of graphite [249], hydrogen-like image states on clean and oxygen-covered nickel and on gold epitaxed on silicon (111) [266], [267], the superconducting energy gap in Nb$_3$Sn [268], the electronic structure of the InP (110) surface [245], and copper phthalocyanine adsorbed on Cu (100) [269].

27.7. Other Surface Analytical Methods

27.7.1. Ultraviolet Photoelectron Spectroscopy (UPS)

In its principle of operation, UPS is similar to XPS, in that a surface is irradiated with photons and the energies of the ejected photoelectrons are analyzed. The physical relationship involved is the same as that of Equation (1) for XPS, but in UPS the energy $h\nu$ of the exciting photons is much lower since the photons are derived from a gaseous discharge that produces hard UV radiation. The gas normally used is helium, which, depending on pressure conditions in the discharge, will provide line sources of energy 21.21 eV (He I) or 40.82 eV (He II) with very narrow linewidths
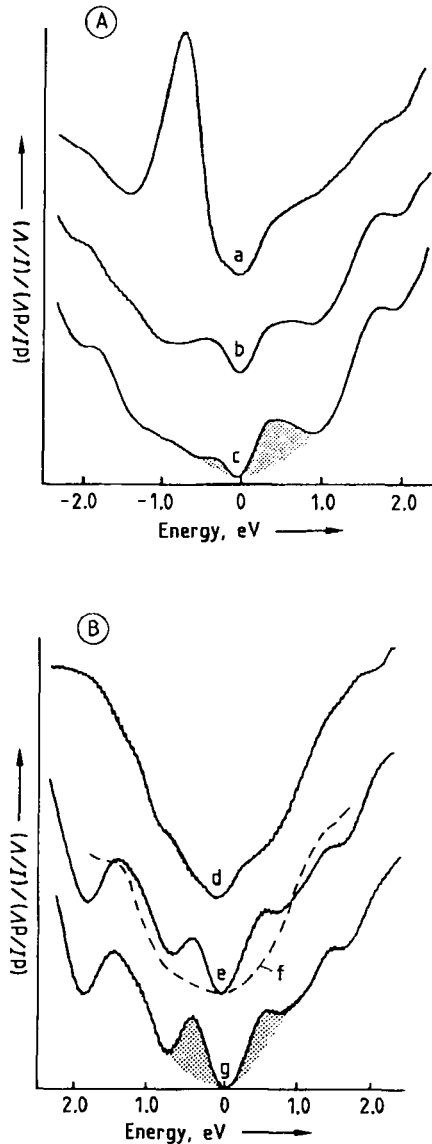
Figure 81. STS spectra recorded from the Si(111)–(7×7) surface [264]
A) In the clean condition; B) After reaction with 1 L of NH$_3$
a) Atom A; b) Atom B; c) Atom C; d) Atom A; e) Atom B not itself reacted but situated next to reacted atoms; f) Reacted atom B; g) Atom C not itself reacted but situated next to reacted atoms

(≈ 20 meV). Because of these low exciting energies, the binding energies E_B that appear in Equation (1) refer not to core levels, as in XPS, but to shallow valence band levels and to other shallow levels, such as those of adsorbates, near the valence band. For energy analysis, the CHA is again
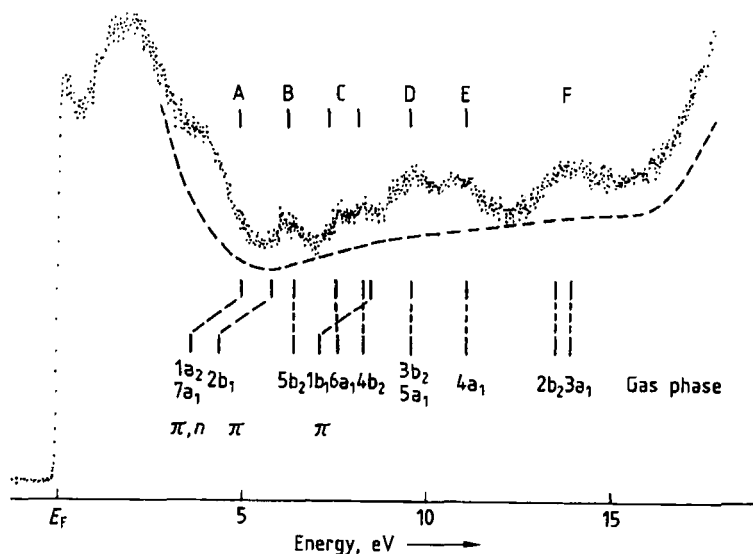
Figure 82. UPS He II spectrum from 5 Langmuirs pyridine adsorbed on Pd (111), with probable molecular orbital assignments
$(x = 60°, \theta = 50°)$ [270]
The dashed line is part of the spectrum from clean Pd (111)
A – F = designation of spectral position of peaks

used, but the pass energies employed for UPS are much lower than in XPS since better energy resolution is required.

UPS is not an analytical technique in the sense that it can provide quantified elemental concentrations on a surface as XPS and AES can, but it is a powerful tool in basic research and is another of the group of most frequently used techniques. It is used in two ways. In its *angle-integrated* form it has been employed very extensively in studying the interactions of gaseous molecules with surfaces, usually in combination with other techniques such as XPS, AES, LEED, and TDS. For complex molecules, interpretation of the UPS spectra during interaction is often possible in terms of the molecular orbital structure of the free molecule, as demonstrated in Figure 82 from NETZER and MACK [270]. The other way in which it has been used is in *angle-resolved UPS (ARUPS)*. In this form the energy analyzer, a CHA of small radius, is mounted in such a way mechanically that it can traverse the space around the sample in both polar and azimuthal directions, while the angle of photon incidence is kept constant. The angle of incidence is then changed by rotation of the sample. If the sample is a single crystal with a clean surface, the dispersion of spectral features (i.e., their movement along the energy axis and their increase or decrease in magnitude) can be used

theoretically to plot the band structure in the surface region. The changes in surface-associated band structure during reaction with gaseous molecules can then be followed, leading to additional information about the nature of the reaction.

27.7.2. Light-Spectroscopic Methods (RAIRS, SERS, Ellipsometry, SHG, IBSCA, GDOS)

The light-spectroscopic techniques have one thing in common: information is provided by them in the form of light, visible or IR, that has either had its spectroscopic properties changed by reflection from a surface or is emitted from atoms in excited states ejected from the surface. Four of the techniques (RAIRS, SERS, ellipsometry, SHG) also use light as the incident probe, the other two (IBSCA, GDOS) use incident ions.

Reflection – Absorption Infrared Spectroscopy. In RAIRS, IR radiation is reflected from a plane surface through an adsorbed layer, the reflected light losing intensity at those frequencies at which the light frequency coincides with a vibrational mode at the surface. The vibrations may be those within the adsorbed species itself, or they may arise from interaction with surface atoms.
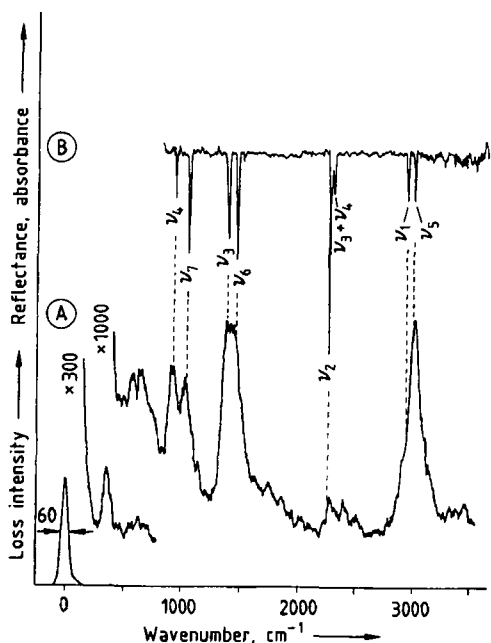
Figure 83. Comparison of HREELS and RAIRS vibrational spectra of CH_3CN adsorbed on Cu (111) at 100 K [271] A) HREELS; B) RAIRS

The substrate surface must be a good reflector. The incident IR beam is plane polarized and is arranged to strike the surface near grazing incidence, typically 80–82°; both plane polarization and grazing incidence are required theoretically. For adequate signal-to-noise ratio in the reflected signal, the absorption cross section of the adsorbed layer at the characteristic frequencies must be high enough to be able to use good energy resolution.

Most earlier research using RAIRS was carried out with dispersive spectrometers to analyze the reflected light, but Fourier transform methods have largely taken over now, leading to the acronym FT-RAIRS. The advantages of using FT methods are, firstly, that either much improved signal-to-noise ratios can be obtained for the same acquisition time or data can be acquired faster for the same signal-to-noise ratio; and, secondly, that all spectral features are recorded simultaneously.

FT-RAIRS is closely related to high-resolution electron energy loss spectroscopy (HREELS) (see Section 27.7.3), and they are often used together in the same system. The two techniques are complementary, since FT-RAIRS has high energy resolution (≈ 2 cm^{-1}) and speed of acquisition, but

its effective spectral lower limit is ≈ 800 cm^{-1} due to the transmission characteristics of the IR transparent optical components, while HREELS can cover the entire spectral range down to 0 cm^{-1} but has inferior energy resolution ($\approx 20-60$ cm^{-1}). A comparison of spectra obtained by the two techniques is shown in Figure 83, from ERLEY [271], for the adsorption of CH_3CN on Cu (111) at 100 K. The bands labeled ν_3, ν_6, ν_1, and ν_5 can be resolved by FT-RAIRS but not by HREELS, but HREELS can record vibrational bands at wavenumbers below the FT-RAIRS cutoff.

Surface-Enhanced Raman Scattering. Conventional Raman spectroscopy is beginning to be used more often in the study of very thin surface films, with the availability of improved detectors, but it still suffers from low intensity in the scattered signal. The discovery by FLEISCHMANN et al. [272] of the *surface enhancement effect* has enabled the Raman scattering effect to be applied more widely, even though restrictions exist. When the surfaces of certain metals, particularly silver, copper, and gold, are roughened on an atomic or near-atomic scale, the intensity of the scattered light is enhanced by factors of $10^5 - 10^6$. Roughening is usually performed electrochemically and is subject to some variation. Various theories have been put forward for the effect, without general agreement as yet. Despite the restrictions on the nature of the substrate, SERS has found very many applications, some of which have been carried out in UHV (e.g., the reactions of N_2 and CO_2 on silver, copper, and potassium surfaces at 40 K) [273]. For a thorough review of the physics of the SERS phenomenon and a discussion of applications, see [274].

Ellipsometry is a method of measuring the thicknesses, and changes in thickness, of very thin films on surfaces, to very high sensitivity, of the order of 5×10^{-3} nm. Polarized light, of wavelength within the UV to IR range of the spectrum, is specularly reflected from a surface with one or more layers of refractive index different from that of the substrate. The latter must be flat and highly reflecting. After the reflection the light is no longer plane polarized but elliptically polarized. The parameters measured are tan ψ, the amplitude ratio of the resolved components of the electric vector of the reflected light parallel to and perpendicular to the plane of incidence, and cos Δ, the phase difference of the two components. *Fixed-*

photon-energy ellipsometry can provide only a measure of the thickness, but if the elliptical polarization parameters are measured as a function of the light energy and the angle of incidence, the technique is known as *spectroscopic ellipsometry*. The observed dispersion with energy can be related to the dielectric loss function; since the energy dependence of loss functions varies from one material to another, the contributions of different surface layers can in principle be distinguished. Ellipsometry in either form is therefore not a compositional analytical technique, but it can provide valuable additional physical information when used with one of the compositional techniques such as AES or XPS.

Optical second harmonic generation (SHG) is another light reflection technique of fairly recent origin that is being applied increasingly in UHV, although it can be used in almost any medium. Like ellipsometry, to which it is related, SHG is very sensitive to the interfacial regions between thin films and substrates, and can be used to study adsorption, molecular orientation, and surface structure. If a linearly polarized wave at a frequency ω is incident on a surface whose dielectric constant is different from that of the ambient medium, then a nonlinear source polarization is induced at the surface, which radiates at frequency 2ω in both reflected and transmitted directions. Inversion symmetry dictates that the second-order nonlinear susceptibility vanishes in the bulk of the solid medium, but at the surface the inversion symmetry is broken, leading to electric dipole contributions to the second-order nonlinearity. By using a laser pulsed at high frequency, time-resolved reactions at surfaces with time scales in the range $10^{-14} - 10^{-6}$ s can be studied. Enhancement of the SHG response can be achieved by tuning the laser to an electronic transition in the substrate or an optical transition of an adsorbate. The technique has recently been reviewed at length [275].

Ion Beam Spectrochemical Analysis. As remarked earlier, ion irradiation of a surface produces visible light photons as well as many other types of particle. The light emission arises from the deexcitation of ions and excited neutrals, and is characterized by a spectrum consisting of a series of sharp lines at wavelengths associated with the type of ionized or excited atom. Spectral analysis thus provides elemental identification by reference to standard emission wavelength tables.

The technique is called ion beam spectrochemical analysis (IBSCA) but has several other names as well. Typical operating conditions include bombardment by Ar^+ ions at energies in the range $2 - 20$ keV, with current densities of $10 - 100$ $\mu A/cm^2$; the light emitted just above the target surface is then focused onto the entrance slit of a scanning optical monochromator, which analyzes the light spectrum. Spatial resolution depends on the type of ion source being used, and can be of the order of $2 - 5$ μm with a suitable type such as a duoplasmatron. Because emission takes place from particles ejected from the surface layer, the surface specificity is similar to SSIMS (i.e., one atom layer). Quantification can in principle be carried out theoretically (e.g., based on the Andersen – Hinthorne LTE model [185] mentioned in Chapter 27.4) but is normally performed by relative intensity measurements, that is, by measuring the intensity of emission from a pure elemental standard under the same instrumental conditions of analysis as used for an unknown. Applications have included studies of the oxidation of chromium [276], the effect of segregation of carbon and sulfur to the surfaces of Fe – Cr – Mo crystals [277], the reaction between a thin PbO film and fused silica [278], and the protonation of lithium silicate glasses [279].

Glow Discharge Optical Spectroscopy. Unlike many of the other techniques described in this section, GDOS is widely used industrially and instruments are available commercially. In the great majority of applications, however, GDOS is used not as a surface-specific analytical technique, but as a rapid and quantifiable technique for determination of elemental concentration profiles through films often tens of micrometers thick. GDOS has certain similarities to IBSCA. The surface to be analyzed is made the cathode in a diode-type discharge cell, whose design originated with GRIMM [280], and argon is flowed through the cell at pressures of $10^3 - 10^4$ Pa. Application of a d.c. voltage of ≈ 1 kV between an anode and the sample cathode establishes a glow discharge. Positive ions produced in the discharge sputter the sample surface and erode it continuously, most of the sputtered material leaving the surface as neutral atoms. The emitted atoms pass into the discharge, in which they are excited by multiple collisions, and the excited states then decay by the emission of visible light, whose spectrum is analyzed in a multichannel spectrophotometer. According to the chosen operating conditions, the rate of erosion of
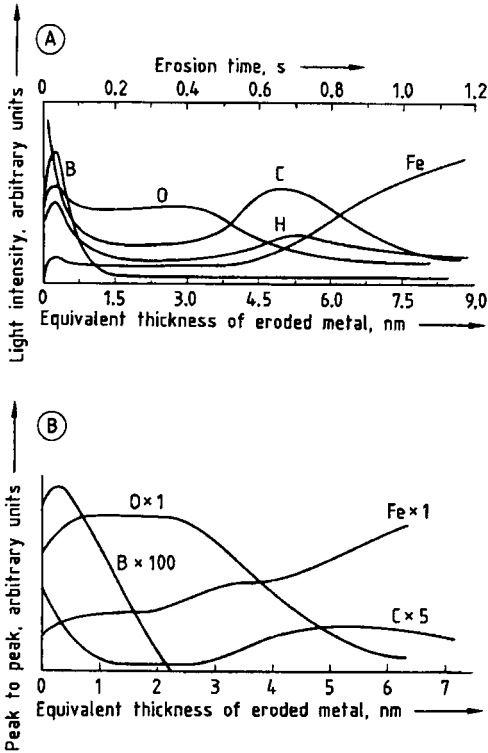
Figure 84. Comparison of depth profiles through a passive anodic film on iron [282]
A) GDOS; B) AES

the sample can be varied between 30 nm/min and 1 µm/min. The inherent sensitivity of the technique is high so that, if required, thin surface layers could be analyzed.

In arriving at a quantification of GDOS intensities, one must realize that as a result of the rapid uniform sputtering of the sample surface, both the surface composition (due to preferential sputtering) and the surface geometry (due to crater formation) are changing during the analysis. The latter change alters the glow discharge operating conditions since during sputtering the cathode surface is receding from the discharge region. BENGTSON [281] has shown how to obtain reasonable quantification on the basis of empirical expressions for the excitation function of the discharge and the sputtering rate in the discharge. If the intensity I_λ of an emission at wavelength λ from an atom of type A is measured, then the intensity is related to the concentration c_A of A by

$$I_\lambda = k_{\lambda A}\, c_A\, C\, i^2\, (V - V_0)^{x(\lambda, A)} \tag{38}$$

where $k_{\lambda A}$ is a constant containing instrumental and atomic factors; C a constant dependent on cell geometry; i the bombardment current; V the applied cell voltage; V_0 the discharge threshold voltage, also dependent on geometry; and $x(\lambda, A)$ a constant characteristic of the emission line.

Equation (38) can be used to measure the concentration of an element in an unknown material if the emission parameters from a standard are known, since by taking the ratio of currents from unknown and standard, some of the parameters in the expression disappear. For elements in low concentration in alloys or compounds the accuracy is good, typically ±5 %.

Although GDOS is normally applied as a very fast profiling technique, leading to the publication of many profile diagrams of films up to $\approx 20\,\mu m$ thick, its operating principles indicate that it is inherently surface-specific, and it can indeed be operated at a much lower erosion rate to study the near-surface region. GDOS also has an advantage over most of the other principal surface analytical techniques in that it can analyze for hydrogen. This ability and also its operation in a surface-specific manner are demonstrated in Figure 84, from BERNERON and CHARBONNIER [282], where the profiles by GDOS (A) and by AES (B) of an anodic passive film on iron are compared. The film was only $\approx 5\,nm$ thick. GDOS shows a pile-up of hydrogen at the film–metal interface and is able to detect boron (from the boric acid buffer) beyond the point at which AES cannot. The profiles are qualitatively similar. Note that GDOS gives purely elemental information, with no direct chemical information available.

Other near-surface applications have included the implantation profiles of phosphorus in iron and FeO [283]; TiN and copper coatings on steel [284]; chromate layers on aluminum [285]; and implantation profiles of phosphorus and boron in iron and nickel [286].

27.7.3. Electron Energy Loss Methods (ELS, CEELS, HREELS, IETS)

Energy Loss Spectroscopy and Core-Electron Energy Loss Spectroscopy. The techniques of ELS and CEELS can be discussed together because they use the same experimental arrangement, basically that of conventional AES. A beam of primary electrons interacts with electrons in surface

istic of the bulk structure and of the surface structure, the surface plasmon having theoretically an energy lower than that of the bulk by a factor of $2^{1/2}$. In *CEELS* the primary electron interacts with a core level, usually but not always leading to ionization, and the losses are in general much greater than in ELS. Thus in ELS the observed losses are normally within 30–50 eV of the primary energy, whereas in CEELS they may be several hundred electronvolts away.

Since all loss features are by definition associated with the primary energy, features characteristic of the surface can be compared to those of the bulk by variation of the primary energy (i.e., by effectively varying the electron escape depth). Such variation is standard practice in ELS, in which the range of primary energies used is typically 100–1000 eV. In CEELS, on the other hand, although the relation between primary energy and escape depth still applies, the primary energy must be high enough to ionize the core level with adequate efficiency, as discussed in Section 27.3.2.2 for AES. Thus for CEELS the primary energy range would be the same as for conventional AES (i.e., 2–5 keV).

Applications. ELS cannot be used to extract compositional information, but it has been used extensively in conjunction with other techniques, particularly AES and UPS, in basic studies of reactions occurring at clean or otherwise well-characterized surfaces. Changes in both plasmon and valence band losses occurring during reaction, particularly as a function of primary energy, provide useful information complementary to that offered by the other techniques. A typical example is shown in Figure 85, from the work of ROSINA et al. [287], for the interaction of H_2 with the clean Ce (001) surface. A primary energy of 102.9 eV was used for most of the measurements, to ensure maximum surface sensitivity, with a switch to 1000 eV occasionally to emphasize more bulk-related features. In the clean surface spectrum (curve e) the surface plasmon at 8.2 eV is dominant, with a shoulder at ≈ 12.5 eV due to the bulk plasmon and another shoulder at ≈ 4 eV probably arising from an intrinsic surface state. On hydrogenation, two additional features at 1.6 and 3.4 eV appear simultaneously with suppression of the 4-eV shoulder. The new features increase with H_2 exposure, and become dominant at 900 L and a primary energy of 102.9 eV. When the primary energy is changed to 1000 eV, the bulk plasmon is seen to have shifted to ≈ 14 eV from ≈ 12.5 eV. Both the features at 1.6 and 3.4 eV and the new

Figure 85. ELS spectra at two primary energies (E_p) recorded during interaction of hydrogen with the Ce (001) surface [287]
a) +900 L H_2; b) +900 L H_2; c) +600 L H_2; d) +300 L H_2; e) Clean

atoms, and the energy losses suffered by the primary electrons are measured by recording the appropriate part of the secondary electron spectrum in an energy analyzer. In *ELS* the losses are relatively small being due to the energies required to stimulate electronic transitions within the valence band of the solid, and collective oscillations of the conduction electrons (plasmon oscillations). For each solid the plasmons have energies character-

position of the bulk plasmon are characteristic of the formation of CeH_2. The increase in the 1.6-eV loss relative to that at 3.4 eV in going to the higher primary energy is indicative of a hydride phase increasing in hydrogen concentration with depth. This particular example is also important because it illustrates that ELS can give information about surface interactions with H_2 that neither XPS nor AES can give.

The many other recent applications of ELS include the interaction of oxygen with titanium [288], [289]; of oxygen [290] and of hydrogen [291] with Co – Ti; of oxygen with tin [292]; of pyrazine with silver [293]; and of ytterbium films with aluminum [294].

CEELS has been applied more to the study of core-electron excitation to localized final atomic states than to excitation to the continuum (i.e., complete ionization). Core-level ionization thresholds can in principle be used to obtain an elemental composition, but the spectral features are weak and the analysis is performed far better by XPS. On the other hand, where shallow, highly localized final states exist, CEELS has been able to provide useful information. Such localized states are plentiful throughout the lanthanide rare-earth series, in some cases being very close to the valence band. Figure 86, from STRASSER et al. [295], shows some examples of the fine structure to be found in CEELS above the main $4d$ excitation threshold, in the lighter lanthanides. The structure arises from strong coupling between the $4d$ holes left after excitation and the $4f$ electrons or holes, the $4f$ states being highly localized. Where the $4f$ states are close to the valence band, their occupancy is dependent on the atomic valence, and valence changes can be monitored exactly by recording the $4d$ CEELS spectra during treatments such as oxidation that usually alter the valence. This analytical variation has been explored exhaustively by NETZER and coworkers [296] – [298].

High-Resolution Electron Energy Loss Spectroscopy. HREELS has already been mentioned in the description of RAIRS as being a vibrational spectroscopy. In this technique, electrons of very low primary energy, typically 3 – 5 eV, are directed at a surface carrying adsorbed molecules. As the electrons approach closely enough, they interact with the molecular field and are reflected (i.e., scattered) having lost discrete amounts of energy corresponding to characteristic vibrational frequencies, either within the

molecules or between the molecules and the surface. Analysis of the scattered electron energy distribution then reveals the energy losses, and hence the vibrational frequencies. If the molecules are attached to the surface only loosely (*physisorption*), many of the characteristic vibrations within a molecule are similar to those observable in the gas phase. These consist of various stretching and deformation modes of vibration. In addition, other modes appear as a result of attachment to the surface. In the gas phase a molecule can translate and rotate freely, but if one end is fixed to a surface it can no longer do so, and therefore certain translational and rotational modes of vibration become observable that cannot be observed in the free molecule. Should the molecule be attached strongly to the surface (*chemisorption*), perhaps with the loss of reactive groups, and certainly with a transfer of electronic charge, then the characteristic free-molecule stretching vibrations can become so distorted that they no longer resemble those seen in the gas phase. Thus the HREELS spectrum from a complex polyatomic molecule adsorbed on a surface can itself be very complex and contain a great deal of information concerning the way in which the molecule has reacted with the surface and also with neighboring molecules. The orientation of the molecule can often be deduced as well, particularly when HREELS is used with other techniques such as LEED, UPS, and TDS.

Because the energy widths of the characteristic vibrational energies are small, < 0.1 meV, for optimum energy resolution one must achieve as narrow an energy spread in the incident electron beam as possible. In all HREELS equipment, therefore, extreme measures are taken to reduce the spread in a variety of ways. Most importantly, the incident electrons are energy selected in a monochromator that is basically an electron energy analyzer used in reverse (see, for example, [299]). Monochromators properly constructed can reduce the energy spread in the incident beam to 2 – 3 meV, but to maintain that width in the beam at the sample surface requires the complete elimination of stray magnetic and electrostatic fields around the sample. Thus very careful screening of the entire HREELS analytical arrangement is necessary. Many experiments are also carried out at low temperature to reduce the thermal broadening of characteristic vibrations.

The above description indicates that the applications of HREELS have been entirely basic and concerned with the interactions of molecules with clean or well-characterized surfaces. A very typi-
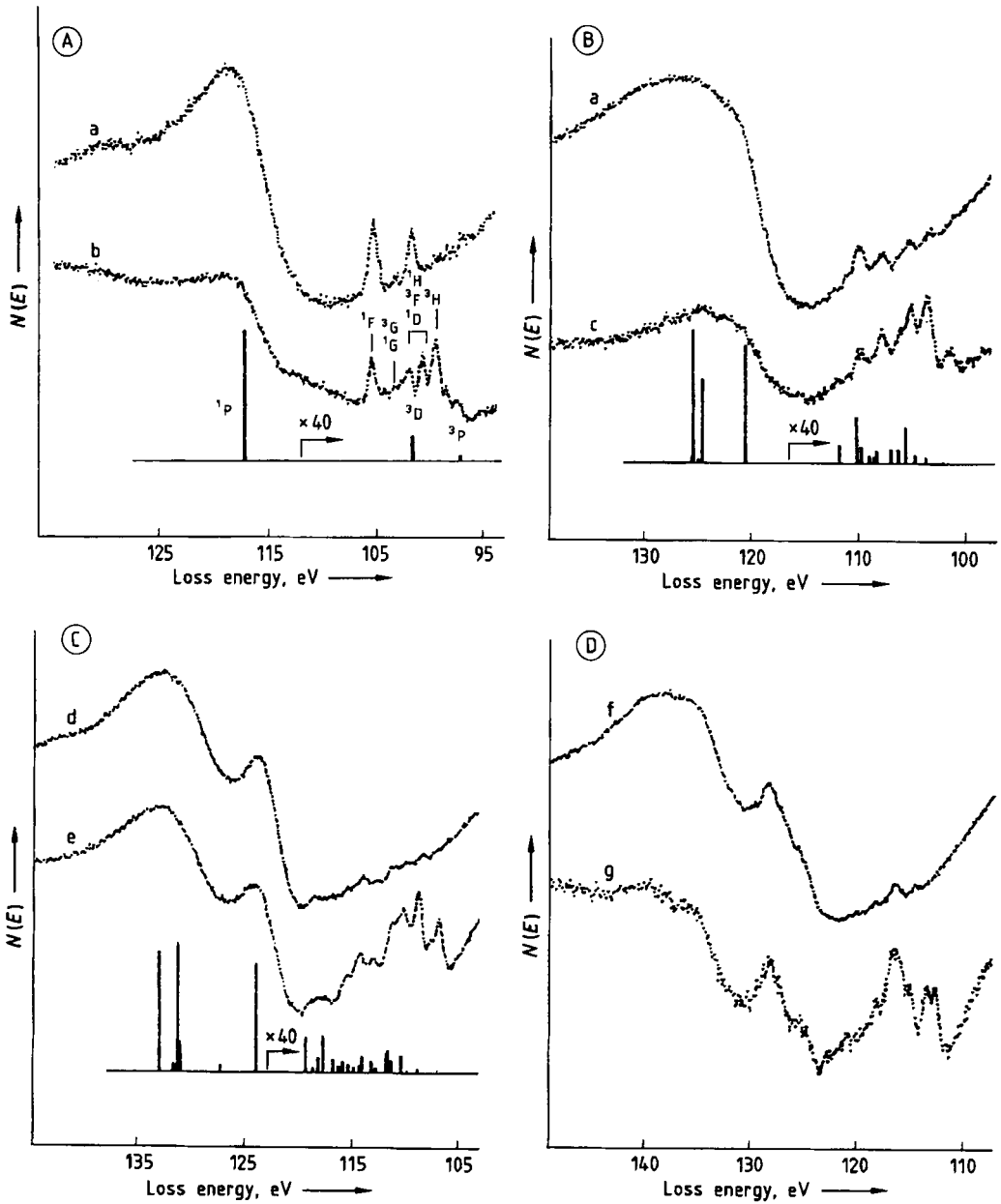
Figure 86. The $4d \rightarrow 4f$ CEELS spectra of the light lanthanides, at two primary energies [295]
A) Lanthanum: B) Cerium: C) Praeseodymium: D) Neodymium
a) $E_p = 1595$ eV: b) 315 eV: c) 350 eV: d) 1750 eV; e) 325 eV: f) 1725 eV: g) 290 eV
^{1}P, ^{1}F, etc: are final-state spectroscopic terms labeling the loss peaks

cal example of the HREELS spectra observed during increasing coverage of an adsorbed molecule is shown in Figure 87, from BARTKE et al. [300], for the adsorption of NO on Nb (100) at 20 K. The energy loss scales are expressed both in millielectronvolts and reciprocal centimeters for convenience, and the energy resolution, as governed by the width of the primary beam, is ≤ 4 meV.
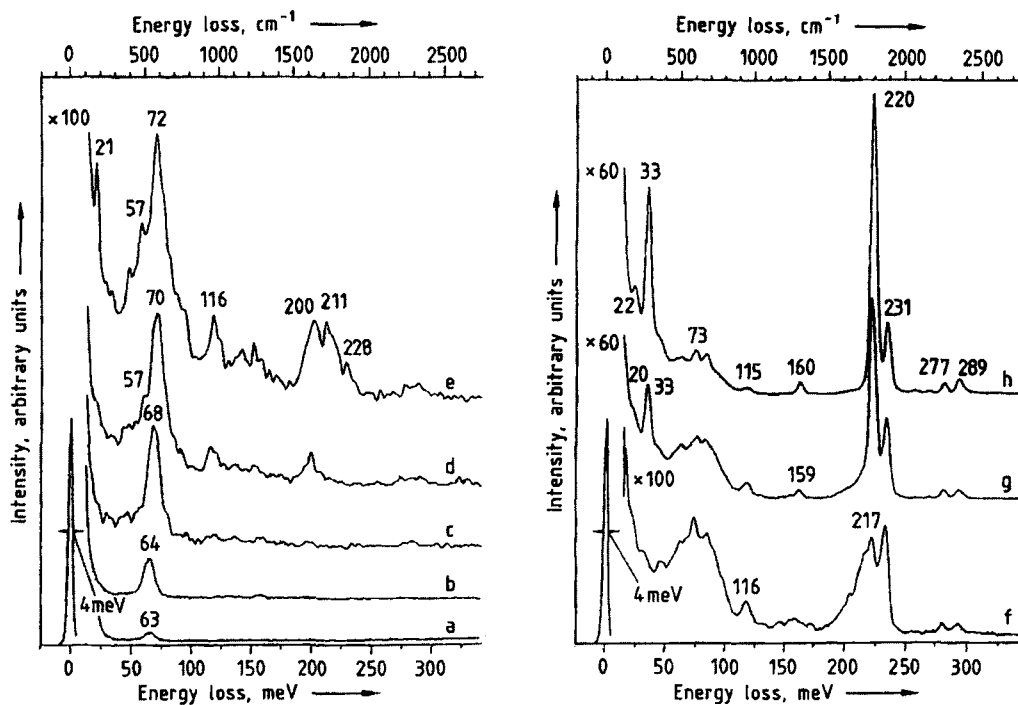
Figure 87. HREELS spectra recorded during the adsorption of NO on Nb (110) at 20 K (primary energy 2.5 eV, $I_0 = I_s = 52°$) [300]
Intensities are normalized to the elastic peak, and spectra were taken in the specular reflection mode
a) 0 L; b) 0.25 L; c) 0.5 L; d) 1 L; e) 2 L; f) 4 L; g) 8 L; h) 16 L; L = Langmuir (10^{-6} Torr · s)

The loss spectrum from the "clean" surface (curve a) is featureless apart from a tiny peak at 63 meV due to atomic oxygen, undetectable by AES. Initial adsorption of NO causes the peak to grow and others to appear; with increasing coverage the oxygen peak (arising from lattice Nb–O vibrations) shifts to 72 meV. The new peaks at 57 and 116 meV are ascribed to atomic nitrogen bonded in a "bridged" position (i.e., across two Nb atoms) and to atomic oxygen bonded on top of a niobium atom, respectively. Further new peaks at 200, 211, and 228 meV, after 2 Langmuirs adsorption, are associated with stretching vibrations within the NO molecule adsorbed on top of a niobium atom. The slight differences in loss energy are explained by different angles of adsorption of the NO (i.e., tilted or bent). With further NO adsorption (curves f–h) the NO stretching vibration shifts to 220 meV and becomes very strong, with a simultaneous decrease in the atomic vibration, indicating a preponderance of molecular adsorption. Additional peaks at 231 meV and at 22 and 33 meV are attributed

to vibrations within an ON–NO dimer on the surface. Thus, with correct interpretation, a few spectra such as those in Figure 87 can provide a large amount of information on the course of a reaction and the nature of the entities actually on the surface.

HREELS has also been used recently to study the adsorption systems 2,3-dimethyl-2-butene and 2-butyne on Ni (111) [301]; straight-chain alcohols on Ag (100) [302]; CO and potassium on Cu (111) [303]; O_2 on cleaved and sputtered graphite [304]; CO on copper-covered Ni (111) [305]; O_2 on Fe (100) [306]; and CO on Ru ($10\overline{1}0$) [307]. Many of these systems are models for catalytic reactions of technological interest.

Inelastic Electron Tunneling Spectroscopy. IETS is unique in that it is entirely surface specific, but does not require a vacuum environment and has almost never been performed in UHV. Its principle is relatively simple. If two metals are separated by a thin (≈ 3 nm) insulating layer, and a voltage is applied across them, electrons

can tunnel from one to the other through the insulator, and if no energy is lost by the electrons, the process is known as elastic tunneling. However, should discrete impurity states occur in the interface between either metal and the insulator, or molecules exist in such an interface that possess characteristic vibrational energies, then the tunneling electron can give up some of this energy either to the state or to the vibrational mode, before reaching the other metal. This is called inelastic tunneling. Obviously, the applied voltage must be greater than the state or vibrational energies. If the current across the metal – insulator – metal sandwich is recorded as a function of applied voltage, the current increases as the threshold for each state or vibrational mode is crossed. The increases in current are in fact very small, and for improved detectability, the current is double differentiated with respect to voltage, thereby providing in effect a vibrational spectrum that can be compared directly with free-molecule IR and Raman spectra.

The metal – insulator – metal sandwich is known as a tunnel junction, and its preparation is all-important. The standard junction consists of an aluminum strip ca. 60 – 80 nm thick and 0.5 – 1.0 mm wide deposited in a very good vacuum onto scrupulously clean glass or ceramic, the surface of the strip then being oxidized either thermally or by glow discharge. The resultant oxide layer is extremely uniform and about 3 nm thick. Introduction of adsorbed molecules onto the oxide layer, or "doping" as it is called, is then effected either by immersion in a solution followed by spinning to remove excess fluid, or by coating from the gas phase. The final stage in preparation is deposition of the second metal (invariably lead) of the sandwich; this deposition is carried out in a second vacuum system (i.e., not the one used for aluminum deposition), the final thickness of lead being about 300 nm, and the width of the lead strip being the same as that of the aluminum strip. The reason for using lead is that all IETS measurements are carried out at liquid-helium temperature, 4.2 K, to optimize the energy resolution by reducing the contribution of thermal broadening to the linewidth, and of course lead is superconducting at that temperature.

Normally not one, but several, junctions are fabricated simultaneously, since the likelihood of junction failure always exists. After fabrication the electrical connections are made to the metal electrodes, the junctions are dropped into liquid helium, and measurements are commenced at once.

Although insulators other than aluminum oxide have been tried, aluminum is still used almost universally because it is easy to evaporate and forms a limiting oxide layer of high uniformity. To be restricted therefore to adsorption of molecules on aluminum oxide may seem like a disadvantage of the technique, but aluminum oxide is very important in many technical fields. Many catalysts are supported on alumina in various forms, as are sensors, and in addition the properties of the oxide film on aluminum metal are of the greatest interest in adhesion and protection.

Penetration of moisture into an adhesion bond to aluminum is one of the reasons for ultimate failure of such a bond, and the problem is minimized by treatment of the aluminum oxide surface with a "hydration inhibitor" before adhesion. Some of the most successful inhibitors have been aminophosphonates, the suggestion being that P – O – Al bonds are formed that prevent the ingress of moisture. The interaction of various phosphorus acids with alumina has been studied by RAMSIER et al. [308]. Figure 88 compares the IETS spectra (spectra b) with the IR absorption spectra (spectra a) for phosphonic acid (A) and phosphinic acid (B). In both cases the strong IR peaks at 979 and 1185 cm^{-1} for phosphinic acid and at 946 and 1179 cm^{-1} for phosphonic acid, characteristic of P – OH and P=O stretching vibrations, respectively, are completely absent in the IETS spectra. These absences indicate that the acids are adsorbed on the alumina in resonance-stabilized forms reached via a condensation mechanism in which a water molecule is formed, which then disperses into solution.

Other applications of IETS have included complexation reactions of $CoBr_2$ and $CoCl_2$ on silane-modified alumina [309], silylation of plasma-grown aluminum oxide [310], the effect of hydrogen on the growth of MgO [311], and adsorption of silane adhesion promoters on aluminum oxide [312].

27.7.4. Appearance Potential Methods (SXAPS, AEAPS, DAPS)

Appearance potential methods all depend on detecting the threshold for ionization of a shallow core level and the fine structure near the threshold, differing only in the way in which detection is performed. In all of them the primary electron energy is ramped upward from near zero to whatever is appropriate to the sample material, while
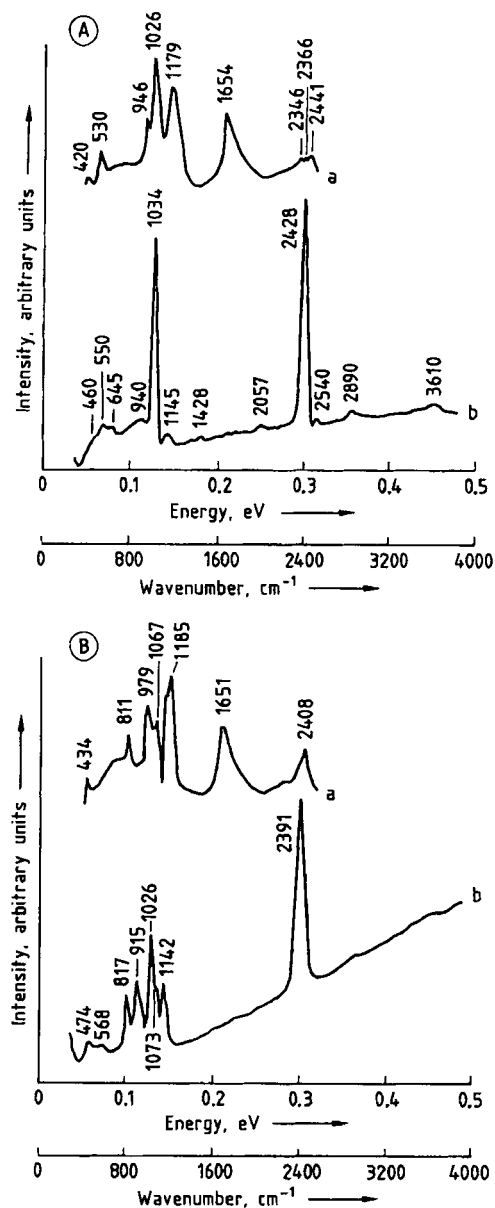
Figure 88. Comparison of IR absorption spectra in water with IETS spectra when adsorbed on aluminum oxide [308], of
A) Phosphonic acid; B) Phosphinic acid
a) IR spectrum; b) IETS spectrum

the primary current to the sample is kept constant. As the incident energy is increased, it passes through successive thresholds for the ionization of core levels of atoms in the surface. An ionized core level, as discussed earlier, can recombine by

the emission either of a characteristic X-ray photon or of an Auger electron.

Soft X-Ray Appearance Potential Spectroscopy. In SXAPS the X-ray photons emitted by the sample are detected, normally by letting them strike a photosensitive surface from which photoelectrons are collected, but also — with the advent of X-ray detectors of increased sensitivity — by direct detection. Above the X-ray emission threshold from a particular core level the excitation probability is a function of the densities of unoccupied electronic states. Since two electrons are involved, the incident and the excited, the shape of the spectral structure is proportional to the self-convolution of the unoccupied state densities.

Auger Electron Appearance Potential Spectroscopy. Due to the emission of an Auger electron as an alternative to soft X-ray photons, the total secondary electron yield will show an increase as an ionization threshold is crossed. It is the total secondary yield that is monitored in AEAPS, in a rather simple experimental arrangement. The secondary current arises not just from the Auger electrons themselves, but also from inelastic scattering of Auger electrons produced at greater depths below the surface. Since the yield change is basically a measure of the probability of excitation of a core-level electron to an empty state above the Fermi level, the fine structure above the threshold will be similar to that seen in SXAPS. One of the problems in AEAPS is that of a poorly behaved background, which means that SXAPS is preferred.

Disappearance Potential Spectroscopy. Crossing an ionization threshold means that electrons are lost from the primary beam as a result of ionization of a core hole. Thus if the reflected current of electrons at the primary energy, more usually termed the *elastically reflected current*, is monitored as a function of energy, it should show a sharp decrease as a threshold is crossed. This is the principle of operation of DAPS. It is in a sense the inverse of AEAPS, and indeed if spectra from the two techniques from the same surface are compared, they can be seen to be mirror images of each other. Background problems occur also in DAPS.

The principal advantages of AEAPS and DAPS over SXAPS is that they can be operated
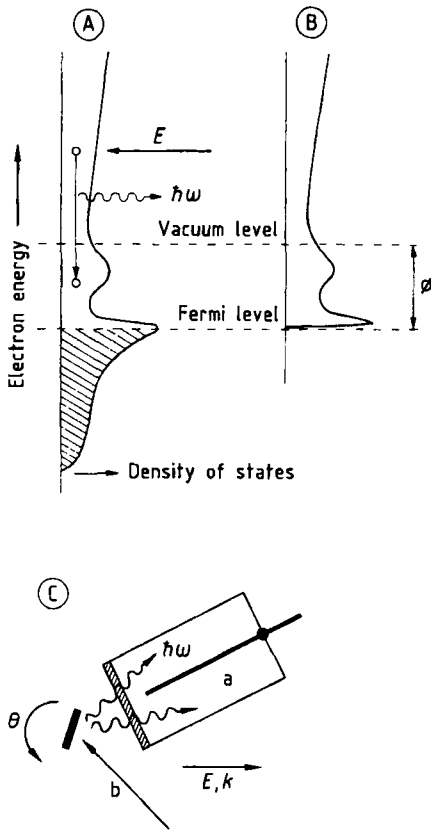
Figure 89. Schematic of the principles of IPES (A) and BIS
(B) [313]; C) Experimental arrangement
a) Geiger–Müller counter; b) Electron beam

at much lower primary electron currents, thus
causing less disturbance to any adsorbed species.

27.7.5. Inverse Photoemission (IPES, BIS)

Irradiation of a surface with electrons leads to
the emission not only of X-ray photons of energies
characteristic of the material, but also of a con-
tinuous background of photons called *brems-
strahlung radiation*. If a detector is set to detect
only those background photons of a particular
energy, and the primary electron energy is ramped
upward from zero, then the variations in photon
flux should mirror the variations in the densities of
unoccupied electron states. The process is called
inverse photoemission because it is clearly the
opposite of ordinary photoemission as in XPS
and UPS, and is illustrated in Figure 89 from
WOODRUFF et al. [313]. Although the name inverse

photoemission spectroscopy (IPES) should apply
to all forms of the inverse photoemission tech-
nique, in fact it is confined by usage to the form
in which the energy at which the photons are
detected is in the UV region, typically with
$hv = 9.7$ eV. If on the other hand a crystal mono-
chromator normally used to provide monochro-
matized X-rays for XPS is used in reverse, with
a detector placed where the electron source is in
Figure 3, then detection is at the X-ray energy
$hv = 1486.6$ eV, and the technique is called brems-
strahlung isochromat spectroscopy (BIS).

Since IPES maps the densities of unoccupied
states, it is related to other techniques that do the
same (i.e., STS and SXAPS). When used in con-
junction with a technique that maps the densities
of occupied surface states, such as UPS or ELS, a
continuous spectrum of state density from occu-
pied to unoccupied can be obtained. Just as in
UPS, in which angular resolution allows elucida-
tion of the three-dimensional occupied band struc-
ture, so in IPES angular resolution allows mapping
of the three-dimensional unoccupied band struc-
ture. This version is called *KRIPES* (i.e., *K-re-
solved IPES*).

The various versions of inverse photoemission
have been applied to purely basic problems, just as
with UPS and ELS. Recent typical examples in-
clude the bonding of H_2 on Ni (110) [314]; the
oxidation of a titanium thin film [315]; the effects
on H_2 chemisorption of a monolayer of nickel on
Cu (111) [316]; unoccupied electronic structure in
La_2CuO_4 [317] and in $Bi_2Sr_2CaCu_2O_8$ [318]; and
KRIPES studies of the Ge (113) (2×1) [319] and
$TiN_{0.83}$ (100) [320] surfaces.

27.7.6. Ion Excitation Methods (IAES, INS, MQS, SNMS, GDMS)

Ion (Excited) Auger Electron Spectroscopy.
Auger emission following creation of a core hole
by electron or photon irradiation has been de-
scribed under AES and XPS. Incident ions can
also create core holes, so that Auger emission as
a result of ion irradiation can also occur, giving
rise to the technique of IAES. The ions used are
normally noble-gas ions, but protons and α-par-
ticles have occasionally been used as well. In ad-
dition to the normal Auger features found in AES
and XPS spectra, peaks are found in IAES spectra
arising from Auger transitions apparently taking
place in atoms or clusters sputtered from the sur-
face. These peaks do not always coincide with

those found in gas-phase excitation, and others are found that are not present in gas-phase measurements. Because of the complexity of the spectra, IAES cannot be used directly as an analytical technique, but it is very useful in basic physics experiments that study Auger processes occurring in excited atoms. Good examples of ion-excited Auger spectra and their interpretation can be found in [321]–[327].

Ion Neutralization Spectroscopy. The technique of INS is probably the least used of all those described here, because of experimental difficulties, but it is also one of the physically most interesting. Ions of He^+ of a chosen low energy in the range $5-10\,eV$ approach a metal surface and within an interaction distance of a fraction of a nanometer form ion–atom pairs with the nearest surface atoms. The excited quasi molecule so formed can deexcite by Auger neutralization. If unfilled levels in the ion fall outside the range of filled levels of the solid, as is the case for He^+, then an Auger process can take place in which an electron from the valence band of the solid fills the core hole in the ion and the excess energy is given up to another valence electron, which is then ejected. Since either of the valence electrons can come from anywhere within the valence band, the observed energy spectrum reflects the local density of states at the solid surface, but is a self-convolution of the LDOS and of transition probabilities across the valence band. The technique of INS was originated by HAGSTRUM [328], who has given a very detailed exposition of the complex mathematical treatment necessary for deconvolution of the spectra [329], [330]. Because the technique of STS is also capable of extracting the LDOS, the results from INS and STS should be comparable, but such a comparison does not seem to have been attempted yet. HAGSTRUM has used INS extensively to study nickel, both the differences in the various low-index faces and the effects of adsorption of the chalcogen elements oxygen, sulfur, and selenium [331]–[333].

Metastable Quenching Spectroscopy. MQS is in a sense an extension of INS. Instead of He^+ ions, helium atoms in metastable states are used in the incident beam, at the same low energies. The excited singlet state $He^*(2^1S_1)$ has an energy of 20.62 eV and a lifetime of 2×10^{-2} s, and so is suitable. It is produced by expanding helium gas at high pressure through a nozzle into a cold cathode discharge sustained by combined electrostatic and magnetic fields. The high fields prevent ions and fast neutrals from leaving the discharge, and the beam is then nearly all $He^*(2^1S_1)$. As the metastable ion approaches the surface, either of two mechanisms can lead to deexcitation, both resulting in Auger emission similar to INS. If the excited level in the atom *can resonate with empty states at the Fermi level* of the surface, electron transfer from the atom to the surface can occur, leading to resonance ionization of the helium atom. Auger neutralization then takes place as in INS, and the resultant spectrum is again a self-convolution of the LDOS. If, however, the *excited level cannot resonate with empty surface states*, then direct Auger deexcitation can occur, in which the hole in the inner shell of the metastable helium atom is filled from a surface state of the sample, followed by ejection of the excited electron from the helium atom. The process is also called Penning ionization. In the latter process, only one electron is ejected, and the resultant spectrum is thus an unconvoluted reflection of the LDOS.

MQS was first demonstrated by CONRAD et al. [334], and the mechanisms involved in deexcitation have been discussed by CONRAD et al. [335] and by BOZSO et al. [336]. Even the brief description given above makes clear that during reaction at a surface, leading to changes in the surface electronic structure, the situation can easily change from resonance to nonresonance of the excited atomic level with surface states, in which case the deexcitation then switches from resonance ionization + Auger neutralization to Penning ionization. The resultant spectrum then changes in character from one similar to INS to one similar to UPS. Such a switch has been observed by BOZSO et al. [337]. Like INS, MQS is highly surface specific. Its applications have been entirely basic, for example, the adsorption of NH_3 on Ni (111) [338] and Ni (110) [339]; of CO, PF_3, NH_3, C_2H_2, and C_6H_6 on palladium and copper surfaces [340]; of CO on copper and on copper-covered ruthenium [341]; and of potassium on hydrogen-covered ruthenium [342].

Secondary Neutral Mass Spectrometry and Glow Discharge Mass Spectrometry. SNMS and GDMS are grouped together because of the close similarities in principle between them. In both techniques the surface to be analyzed is sputtered and the contribution to the sputtered flux of secondary neutrals is measured. They differ in the methods used to measure the secondary neutral flux. In *GDMS*, as in GDOS, atoms are sputtered
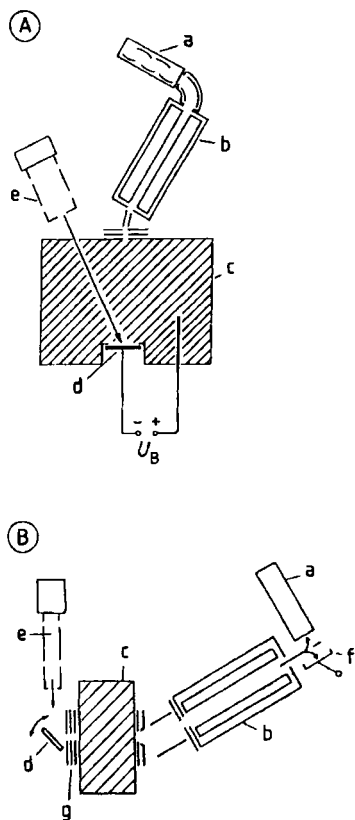
Figure 90. Schematics of designs for SNMS [344]
A) Direct bombardment mode via noble-gas ions in the plasma, or separate bombardment mode via rastered ion beam traversing the plasma: B) External bombardment mode via external ion gun
a) Multiplier; b) Quadrupole mass spectrometer; c) High-frequency plasma: d) Sample; e) Ion gun: f) Faraday cup; g) Electrical diaphragm

via a glow discharge in which the surface is the cathode, and on entering the discharge region, many of the atoms are ionized. The ions are then extracted from the discharge and passed via differential pumping stages into a region at much lower pressure where they can be mass analyzed. Either magnetic or quadrupole mass spectrometers are used for mass analysis. In *SNMS*, three different approaches to postionization of neutrals have been adopted. The first uses *electron impact ionization* (i.e., direct ionization of neutrals by an electron beam from a hot filament as they leave the surface). Secondary ions produced as a result of sputtering are deflected out of the secondary particle flux before they reach the postionization region. If the ion deflection and hot filament are

both switched off, the instrument can be used as a normal SSIMS analyzer. The second postionization approach is that of *plasma ionization* in a high-frequency, low-pressure gas discharge (i.e., under conditions very different from the higher pressure discharge used for GDMS and GDOS). Sputtering can be performed either by accelerating positive ions from the discharge to the sample surface or by an external ion gun. In either case, neutrals produced by sputtering pass into the plasma where some are ionized and are thus extracted directly into a quadrupole mass spectrometer. Again, with an external ion gun and without the plasma, SSIMS can be performed. This variation of SNMS has been pioneered by OECHSNER and coworkers [343], and Figure 90 shows schematically some of the experimental arrangements developed by OECHSNER [344]. Thirdly, postionization can be achieved by *high-intensity laser beams*, involving either resonant or nonresonant multiphoton ionization. In the resonant form, a tunable dye laser is tuned to exactly the correct wavelength to excite a chosen atom to an excited state; while in that state, interaction with a second photon leads to ejection of an electron (i.e., ionization). In the nonresonant form, excimer lasers of such high power are used that even if the laser frequency is a long way off resonance for many atoms, adequate ionization efficiency is achieved. The latter variations of SNMS have been given their own acronyms, surface analysis by resonance ionization of sputtered atoms (*SARISA*) and surface analysis by laser ionization (*SALI*), respectively.

Like SSIMS and GDOS, both SNMS and GDMS have been developed commercially, and GDMS has also had wide industrial application. Most of the applications have, like GDOS, consisted of rapid profiling of surface films many micrometers thick, with special emphasis on the depth distribution of trace elements, and would not be regarded as surface specific. Nevertheless, again like GDOS, GDMS is in principle a surface-specific technique and has occasionally been used [345] in an ultrahigh-vacuum system and at very low erosion rates.

One of the attractions of SNMS and GDMS is the relative ease and accuracy of quantification using standard materials and relative sensitivity factors. For most elements, particularly the metallic and semiconducting elements, the spread of sensitivity factors is less than an order of magnitude, comparable to the situation in XPS. In addition, the actual sensitivities are significantly

GaAs 20 ppm(wt) Si
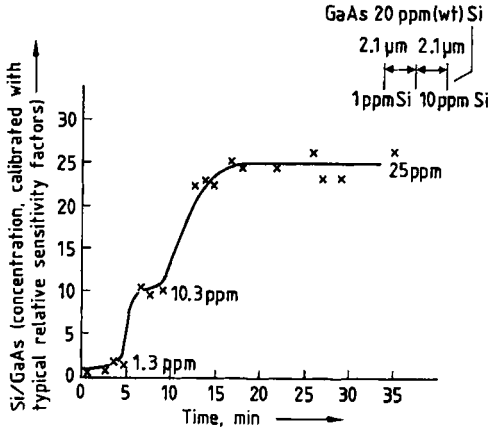
Figure 91. Profile by GDMS through two layers of silicon-doped GaAs grown epitaxially on bulk GaAs [346]

Figure 92. Profile by SNMS of multilayer Ni – Cr sandwich system, with layer thickness of 15 nm [347]
A) Raw intensity data; B) Time-dependent total sputter yield; C) Plot of concentration versus depth evaluated from A) and B)

higher than in XPS or AES, with analysis at the part-per-million level being routine and at the part-per-billion level being quite possible. A typical profile using GDMS to detect low concentrations is shown in Figure 91 from SANDERSON et al. [346], through two layers of silicon-doped GaAs grown epitaxially on bulk GaAs, the silicon levels being different in each layer and in the bulk. Each epilayer can be distinguished easily in terms of the silicon concentration at the part-per-million level, and the semiquantitatively derived silicon concentrations agree well with the amounts known to be incorporated during growth.

A demonstration of both the high depth resolution obtainable in SNMS and the quantification of the technique is given by Figure 92 from WUCHER et al. [347]. It shows a profile through a sandwich structure consisting of 39 alternate layers of chromium and nickel, each 15 nm thick, recorded using a commercial instrument with both bombardment and neutral postionization provided by a high-frequency plasma with an applied voltage of 515 V. Figure 92 A shows the raw intensity data; Figure 92 B, the measured total sputter yield, in which the chromium yield is markedly lower than that of nickel; and Figure 92 C, the intensities corrected for differing sputter yields and then quantified. The sharpnesses of the interlayer interfaces allowed the thicknesses of the individual layers to be determined with a standard deviation of only 10 %.

GDMS has been applied extensively in the steel industry for trace-element analysis and for profiling surface films and coatings. Some examples are given in [348], e.g., detection of uranium and thorium in concentrations of a few parts per billion in aluminum; of some 16 trace elements in concentrations of fractions of a part per billion in indium; of more than 25 impurity elements in a high-temperature alloy in concentrations ranging from 0.2 to 2000 ppm; and of 5 nonmetallic impurity contents in low-alloy steels. Good accuracy was achieved throughout. Applications of SNMS have included determination of hydrogen profiles in a-Si : H [349]; analysis of insulators such as SiO_2, Si_3N_4, PTFE, and glasses [350]; the study of aminoacid layers on metals [351]; depth profiles of silicides [352]; analysis of bulk polymers [353];
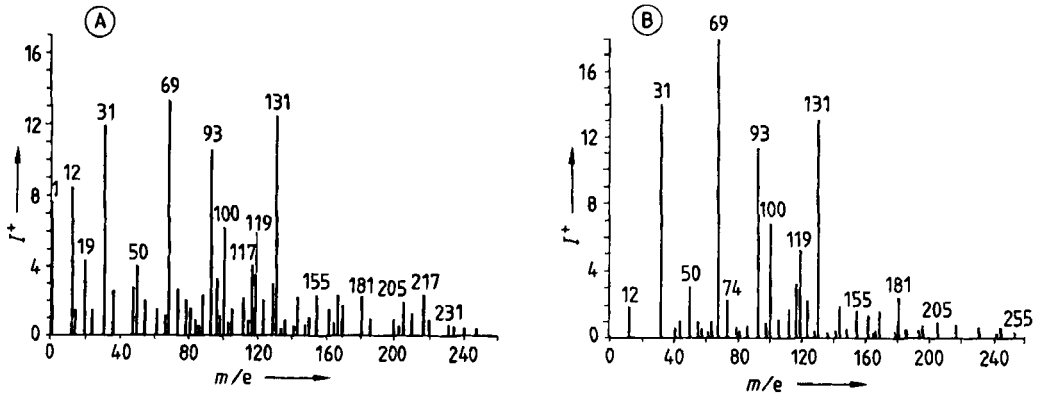
Figure 93. Comparison of SSIMS (A) and FABMS (B) positive secondary-ion spectra from PTFE [357]

thin-film systems [354]; and Cu – Al alloy [355] by SALI, and various metallic and nonmetallic materials by SARISA [356].

27.7.7. Neutral Excitation (FABMS)

Fast-atom bombardment mass spectrometry is very similar to SSIMS in practice, the only difference being that instead of using positively charged ions as the primary probe, a beam of energetic neutral atoms is used. Secondary ions are emitted as in SSIMS and analyzed in a mass spectrometer, usually of the quadrupole type. The beam of fast atoms is produced by passing a beam of ions through a charge-transfer cell, which consists of a small volume filled with argon to a pressure of about 10^2 Pa. Charge transfer occurs through resonance between fast argon ions and argon atoms with thermal energy, with efficiencies of 15 – 20 % if geometric and pressure conditions are optimized. Residual ions are removed by electrostatic deflection.

FABMS has two advantages over SSIMS, both arising from the use of neutral rather than charged particles. Firstly, little or no surface charging of insulating materials occurs, so that organics such as polymers can be analyzed without the need to employ auxiliary electron irradiation to neutralize surface charge. Secondly, the extent of beam damage to a surface, for the same particle flux, is much lower using FABMS than SSIMS, thus allowing materials such as inorganic compounds, glasses, and polymers to be analyzed with less worry about damage introducing ambiguity into the analysis. An example illustrating these advantages is given in Figure 93 from MICHAEL and STULIK [357].

There the SSIMS (Fig. 93 A) and the FABMS (Fig. 93 B) secondary positive ion spectra are compared for PTFE; the surface of the PTFE has to be coated with a gold film for SSIMS measurement but not for FABMS. More fragment ions occur in the SSIMS spectrum than in FABMS, and in the FABMS spectrum the size of the peak at $m/e = 12$ (C^+) relative to the principal peak at $m/e = 69$ (CF_3^+) is much smaller than in SSIMS, indicating much less damage of the polymer surface during FABMS analysis.

FABMS was developed as an alternative to SSIMS by VICKERMAN and coworkers; for a detailed description, see [358] – [362]. Applications have included analyses of tin and lead oxides [363], of poly(methyl methacrylate) – PTFE sandwich structures [364], of multicomponent glasses [358], of silica [358], and of phosphate- and chromate-treated aluminum [360].

27.7.8. Atom Probe Field-Ion Microscopy (APFIM, POSAP, PLAP)

The atom probe field-ion microscope (APFIM) and its subsequent developments, the *position-sensitive atom probe* (*POSAP*) and the *pulsed laser atom probe* (*PLAP*), have the ultimate sensitivity in compositional analysis (i.e., single atoms). The APFIM was developed by MÜLLER et al. [365], having grown out of *field-ion microscopy* (*FIM*) which also originated earlier with MÜLLER [366].

FIM is purely an imaging technique in which the specimen in the form of a needle with a very fine point (radius 10 – 100 nm) is at low temperature (liquid nitrogen or helium) and surrounded by a noble gas (He, Ne, or Ar) at $10^{-2} - 10^{-3}$ Pa. A
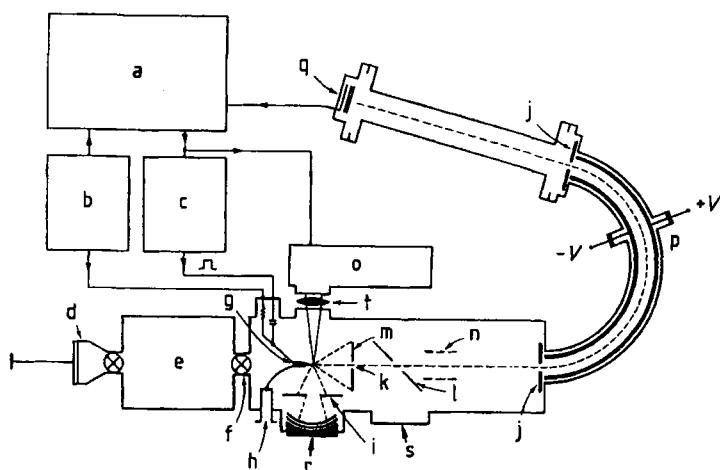
Figure 94. Energy-compensated time-of-flight mass spectrometer for APFIM measurements [367]
a) Computer-controlled fast digital CAMAC system; b) d.c. High voltage; c) High-voltage pulse; d) Airlock; e) Preparation chamber; f) Isolation value; g) Specimen; h) Cryostat; i) Variable iris; j) Herzog electrode; k) Probe aperture; l) Mirror; m) Channel plate; n) Einzel lens; o) Pulsed laser; p) Energy-compensating lens; q) Single-atom detector; r) Atom probe imaging viewpoint; s) FIM viewpoint; t) Focusing lens

fluorescent screen or a microchannel plate is situated a few centimeters from the needle. A high positive voltage is applied to the needle, which causes noble-gas atoms approaching the needle to be ionized over points of local field enhancement (i.e., over prominent atoms at the surface); the ions are then repelled from these points and travel in straight lines to the screen or plate, where an atomic image of the needle tip is formed. This is called a field-ion image, and the voltage at which the image is optimized is called the *best imaging voltage*. The latter is in the range 5 – 20 kV.

Atom Probe Field-Ion Microscopy. If the voltage is raised further above the imaging voltage, the cohesive energy that binds atoms to the surface can be exceeded, and atoms can be removed by what is termed *field evaporation*. The evaporation field required is a function of the sample material and the crystallographic orientation of the needle. The removal of atoms is the basis of APFIM and its daughter techniques. If the evaporation field is pulsed with very short ($\leq$10 ns) voltage pulses of extremely sharp rise times ($\approx$ 1 ns), and the time taken for a field-evaporated atom to travel from the tip to a detector is measured, then the mass-to-charge ratio of the ionized atom can be established; i.e., elemental identification of the individual atom is possible. The region on the tip from which atoms are removed is selected by rotating or

tilting the tip until the desired region, as viewed in the FIM image, falls over an aperture of about 2-mm diameter in the fluorescent screen. Most field-evaporated atoms will strike the screen, but those from the selected area will pass through the aperture into the time-of-flight mass spectrometer. For an aperture of 2 mm, the area analyzed on the tip is ca. 2-nm in diameter.

A modern instrument using an energy-compensated TOF mass spectrometer is shown in Figure 94 from MILLER [367]. Atoms may also be removed from a tip by focusing a pulsed laser onto it, as an alternative to field evaporation, leading to the variation of APFIM called the *pulsed laser atom probe (PLAP)*.

Position-Sensitive Atom Probe. Pulsed operation of the APFIM leads to analysis of all atoms within a volume consisting of a cylinder of ca. 2-nm diameter along the axis of the tip and aperture, with single atomic layer depth resolution but no indication of just where within that volume any particular atom originated. The advent of position-sensitive detectors has allowed APFIM to be extended so that three-dimensional compositional variations within the analyzed volume can be determined. The development has been pioneered by SMITH and coworkers [368], [369], and is called position-sensitive atom probe (POSAP). The aperture and single-ion detector of APFIM are replaced by a wide-angle double channel plate, with

a position-sensitive anode just behind the plate. Field- or laser-evaporated ions strike the channel plate, releasing an electron cascade, which is accelerated toward the anode. The impact position is located by the division of electric charge between three wedge and-strip electrodes. From this position the point of origin of the ion on the tip surface can be determined, since the ion trajectories are radial. Thus after many evaporation pulses, leading to removal of a volume of material from the tip, both the identities and the positions of all atoms within that volume can be mapped in three dimensions. Because an evaporated volume may contain many thousands of atoms, the data collection and handling capabilities must be particularly sophisticated.

POSAP is the only technique available for identifying and locating precipitates, second phases, particles, and interfaces on an atomic scale, and has therefore had considerable application in metallurgical and semiconductor problems. A typical example of a POSAP three-dimensional analysis is given by CEREZO [370] for the ferrite phase of an aged duplex stainless steel CF3 containing 21 % chromium and 9 % nickel as well as manganese, silicon, molybdenum, and carbon. The Fe – Cr system has a spinodal region in which it is thermodynamically unstable with respect to small compositional fluctuations, leading to chromium-rich and iron-rich phases forming a complex interconnected morphology. Figure 95 from CEREZO shows a series of sections by POSAP, each image being about 15 nm in diameter, with a vertical separation between sections of about 3 nm. The images are in chromium, on a gray scale, with white representing ≤ 5 at % Cr, and black ≥ 50 at % Cr. The sequence shows clearly that the chromium-rich phase is interconnected in depth. From such a sequence, with the help of powerful data handling, a three-dimensional reconstruction can be derived, as shown in Figure 96, where the chromium isosurface for concentrations about halfway between the maximum and minimum values is illustrated. The view is the same as in Figure 95 (i.e., along the direction of analysis). The isosurface represents the approximate interface between the chromium-rich and chromium-depleted regions, and shows clearly the nature of the interconnected structure with an average width of only 2 – 3 nm.

POSAP has also been applied to the phase chemistry of AlNiCo 2 magnet alloys; the precipitation of copper in iron-based alloys; surface segregation in ceramic oxide superconductors; hetero-

structures in compound semiconductors [369]; and Pt – Rh catalysts and multiquantum well structures [371]. Applications of *APFIM* have included finely dispersed carbides in a 14Co – 10Ni – 2Cr – 1Mo high-strength steel [372]; the distribution of vanadium atoms in the ordered alloy Fe – Co – 2V [373]; the location of hafnium atoms in the ordered alloy Ni_3Al [367]; the formation of precipitates and segregation of elements in a nickel-based superalloy [374]; and metal – semiconductor and Si – SiO_2 interfaces [375].

27.7.9. Desorption Methods (ESD, ESDIAD, TDS)

Electron-Stimulated Desorption and Electron-Stimulated Desorption Ion Angular Distribution. Electron irradiation of a surface, particularly one covered with one or more adsorbed species, can give rise to many types of secondary particles, including positive and negative ions. In ESD and ESDIAD, the surface is irradiated with electrons of energies in the range of 100 – 1000 eV, and the ejected positive-ion currents of selected species are measured in a mass spectrometer. If the angular distribution of the secondary ions is also measured, either by display on a screen or by using position-sensitive detection, then electron-stimulated desorption (ESD) becomes electron-stimulated desorption ion angular distribution (ESDIAD).

The electron desorption techniques are not used, and probably cannot be used, for compositional analysis, but they provide valuable information on the nature of electronic interactions leading to the breaking of bonds and, in the angle-resolved form, on the geometry of surface molecules and the orientation of broken bonds. The primary electrons do not, at the energies employed, succeed in breaking molecular or surface-to-molecule bonds or in knocking ions out of the surface directly, but the process is one of initial electronic excitation. An electron is absorbed by a surface – adsorbate complex or an adsorbed molecule itself, leading to excitation to an excited state by a Franck – Condon process. If the excited state is antibonding and the molecule or radical is already far enough from the surface, desorption can occur. Since return to the ground state after excitation is a much more probable process, the cross sections for ion desorption are low, $10^{-20} – 10^{-23}$ cm². If core-level ionization is involved in the initial interaction with the incident
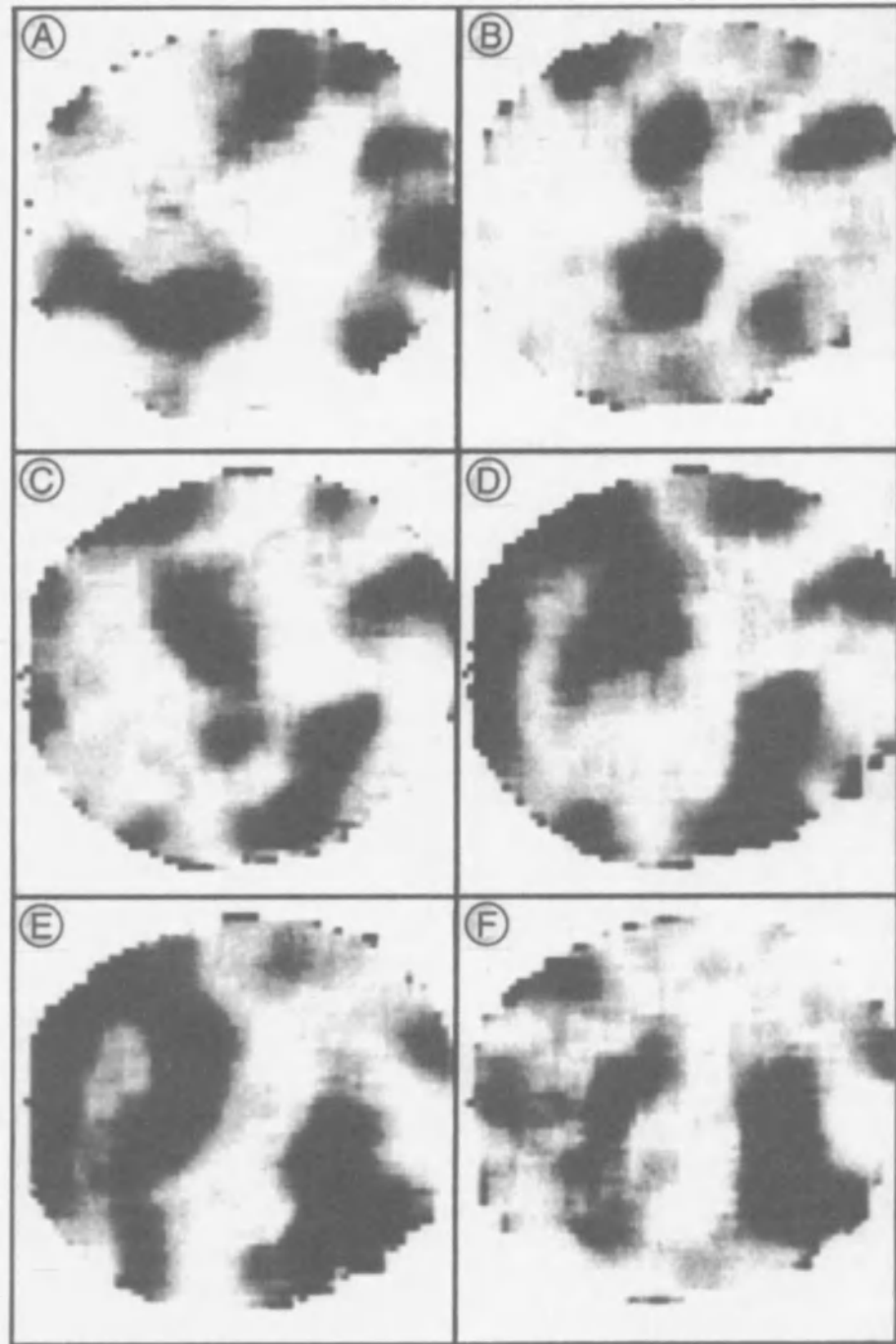
Figure 95. Sequence of images A – F obtained in POSAP by CEREZO [370] from the ferrite phase of an aged duplex stainless steel CF3

Each image is 15 nm in diameter, and the amount removed between each successive image corresponds to a vertical separation of ca. 4 nm

White areas represent ≤5 at % Cr, dark areas ≥50 at % Cr
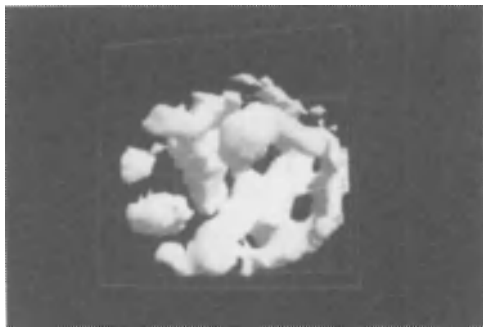
Figure 96. Three-dimensional reconstruction of data of Figure 95 [370]
The isosurface has been drawn through all points halfway between maximum and minimum chromium concentrations. The interconnected structure is Cr-rich

electron, rather than valence levels as in the Franck – Condon-type excitation, then a desorption mechanism based on Auger decay has been proposed by KNOTEK and FEIBELMAN [376]. The core level left behind after interatomic Auger decay creates a positive ion, which is then expelled by the repulsive Madelung potential.

The most common ions observed as a result of electron-stimulated desorption are atomic (e.g., H^+, O^+, F^+), but molecular ions such as OH^+, CO^+, H_2O^+, and CO_2^+ can also be found in significant quantities following adsorption of H_2O, CO, CO_2, etc. Substrate metallic ions have never been observed, which means that ESD is not applicable for surface compositional analysis. The most important application of ESD in the angularly resolved form ESDIAD is in determining the *structure and mode of adsorption* of adsorbed species. This derives from the fact that the ejection of positive ions in ESD is not isotropic. Instead the ions desorb along certain directions only, characterized by the orientation of the molecular bonds that are broken by electron excitation. This orientational dependence is exemplified in Figure 97 from MADEY [377], illustrating the H^+ ESDIAD patterns and suggested models for the adsorption of H_2O on Ag (110). Series A is for adsorption of H_2O on clean silver, B for adsorption on silver with a preadsorbed bromine layer, and C for adsorption on silver with a preadsorbed oxygen layer. The top row consists of three-dimensional perspective plots of the ion intensities; the second row shows two-dimensional intensity contour maps; and the bottom row shows schematic models. On clean silver, the random emission centered about the

surface normal indicates that the adsorbed H_2O is disordered locally, with no preferred orientation of the broken bonds. When bromine is present, the H_2O molecules are believed to form hydration shells around each bromine atom, leading to pronounced ordering and hence preferred bond orientation, which is reflected in the very distinct four-beam ESDIAD pattern seen in B. With preadsorbed oxygen, another type of ordering is found, caused by the dissociation of H_2O and reaction with the adsorbed oxygen to form adsorbed OH species. These OH radicals are thought to be tilted along (001) azimuths, giving rise to the intense two-beam pattern, with much weaker beams along the $(1\bar{1}0)$ azimuths, seen in C. The weaker beams are due to OH-stabilized H_2O in low concentration.

ESDIAD is obviously not a diffraction technique such as LEED, but it gives direct information about surface structure in real space. The sensitivity is to local bonding geometry, and long-range order is not necessary as in LEED. ESDIAD is especially sensitive to the orientation of hydrogen atoms in surface complexes, which is difficult to observe by any other technique.

Thermal desorption spectroscopy (TDS), sometimes called temperature-programmed desorption (TPD), is simple in principle. A gas or mixture of gases is allowed to adsorb on a clean metal foil for a chosen length of time; then, after the gas is pumped away, the foil is heated at a strictly linear rate to a high temperature, during which the current of a particular ion or group of ions is monitored as a function of temperature. The ion masses are selected in a quadrupole mass spectrometer. As the binding energy thresholds of the adsorbed species on the surface are crossed, peaks in the desorbed ion current appear at characteristic temperatures. From the characteristic temperatures and the shape of the desorption peak above the threshold, the activation energies for desorption can be obtained, along with information about the nature of the desorption process. The mass spectrum from the mass spectrometer, of course, provides information about the species that actually exist on the surface after adsorption.

Although simple in principle, experimental artifacts that are possible in TDS must be avoided. Thus, ions accepted by the mass spectrometer must originate only from the surface of the foil, and in addition, the temperature distribution across the foil should be uniform to avoid the overlapping of desorption processes occurring at different tem-
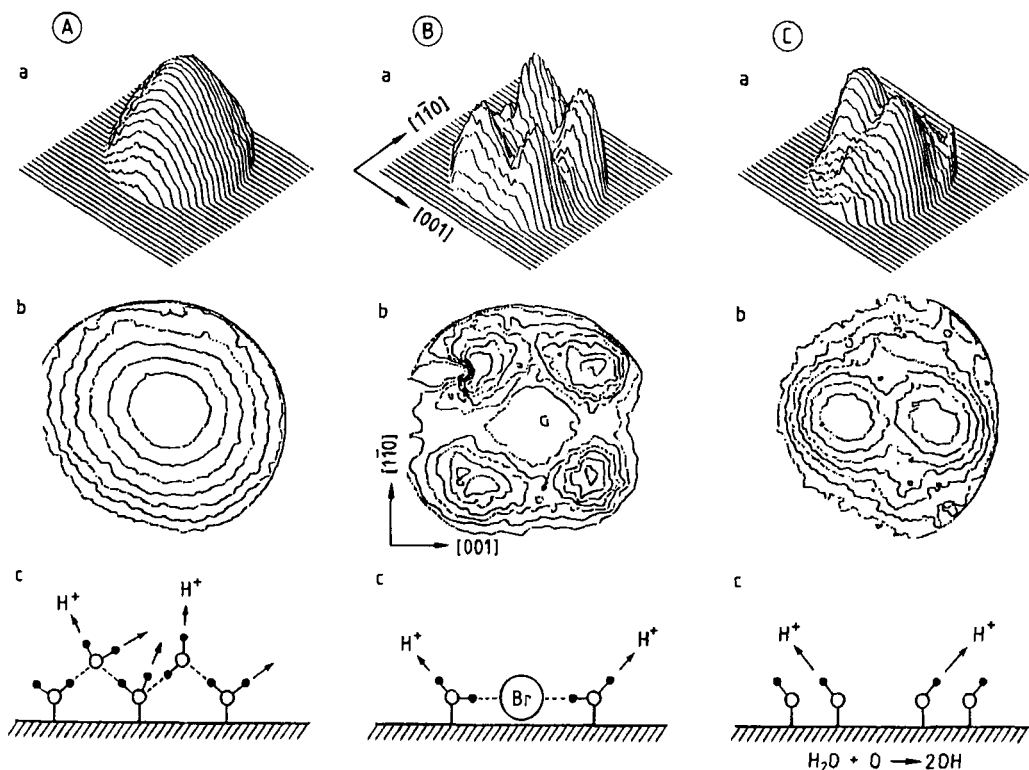
Figure 97. ESDIAD pattern in H^+ from Ag (100) [377]
A) Water-adsorbed; B) Water on preadsorbed Br; C) Water on preadsorbed O
a) Three-dimensional perspective plots; b) Two-dimensional contour maps; c) Schematic models

peratures. To ensure that these experimental requirements are met, the angle of acceptance into the spectrometer is restricted by placing a drift tube with an aperture between the foil and the spectrometer, and the foil itself is usually in the form of a long thin ribbon, from whose center section only are ions accepted. In addition, the heating rate must be sufficiently fast that the desorbed species accepted into the mass spectrometer is characteristic of the desorption process, but not so fast that the pressure rise around the foil is too great.

With correct experimental procedure, TDS is straightforward to use and has been applied extensively in basic experiments concerned with the nature of reactions between pure gases and clean solid surfaces. The majority of these applications have been catalysis related (i.e., performed on surfaces acting as models for catalysts), and in every case TDS has been used with other techniques such as UPS, ELS, AES, and LEED. To a certain extent it is quantifiable, in that the area under a desorption peak is proportional to the

number of ions of that species desorbed in that temperature range, but measurement of the area is not always easy if several processes overlap.

Figure 98 from KRÜGER and BENNDORF [378] shows some typical desorption spectra, following the adsorption of ethylene oxide on Ag (110) at 110 K at exposures increasing from 0.22 to 13.3 Langmuirs. The oxidation of ethylene to ethylene oxide is an industrial process for which silver has important catalytic properties. At very low coverage, only a single mass 29 (the major fragment from the cracking of C_2H_4O in the spectrometer ionization source) desorption peak is found at 170 K, which increases to saturation at about 3 Langmuir, and shifts down to 160 K. Beyond 3 Langmuir exposure, peaks b and c appear at lower temperatures, and c in particular continues to increase with exposure without reaching saturation. For the latter reason, peak c is obviously due to multilayer condensed ethylene oxide. Peaks a and b originate from the adsorption of ethylene oxide in two different sites on the Ag (110) surface, site a being filled first, followed
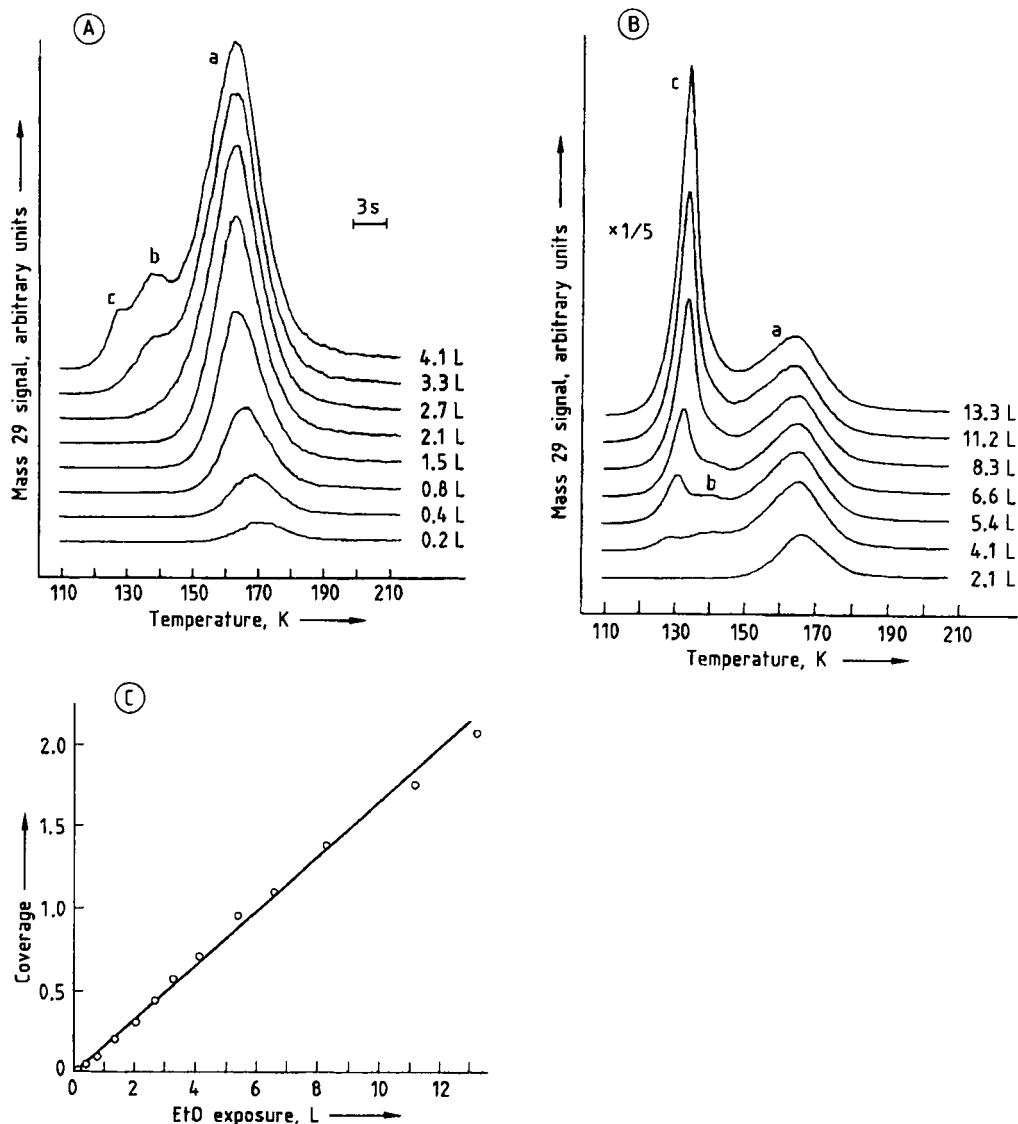
Figure 98. Desorption spectra following adsorption of ethylene oxide on silver at 110 K up to an exposure of 13.3 L [378]
A) 0.2–4.1 L; B) 2.1–13.3 L; C) Relation between surface coverage and exposure of ethylene oxide

by b when all a sites are occupied. Figure 98 C shows the total integrated area of all peaks as a function of exposure, a strictly linear dependence indicating that the occupancy of sites a and b and condensation into a multilayer all have equal probabilities. Data such as these, when taken with measurements from complementary techniques, are invaluable in establishing reaction mechanisms.

27.7.10. Diffraction Methods (LEED,RHEED) (see also →Structural Analysis by Diffraction)

Both LEED and RHEED are probes of surface structure that depend on the existence of long-range order of periodic structure in the surface. They do not provide compositional information and are always used along with other techniques such as AES that can analyze for composition.

LEED is probably the oldest surface-specific technique, having been discovered by DAVISSON and GERMER in 1927 [379].

Low-Energy Electron Diffraction. In LEED a collimated and monoenergetic beam of electrons of variable energies between about 20 and 1000 eV strikes the surface of a single crystal at normal incidence. An energy selector, almost universally consisting of a set of spherical grids centered on the specimen, repels all secondary electrons except those elastically scattered (i.e., those at the primary energy). The elastically scattered electrons are then accelerated to a spherical fluorescent screen, on which the diffraction pattern can be seen. Measurement of diffraction spot intensity can be made either with a photometer or with a movable Faraday cup.

Diffraction occurs due to the wave nature of electrons. The waves represented by the incident electron beam interfere as a result of reflection by a periodic crystal structure in such a way that the reflected beams are anisotropic in direction and are in fact concentrated in certain directions related directly to the crystal structure. Where the reflected beams are intercepted by the fluorescent screen, a spot pattern results. The purely geometric relationship between the observed pattern and the surface structure is governed by the well-known Bragg equation, which connects the directions of the interference maxima with the surface lattice spacing and the primary energy. The observed diffraction pattern is a direct representation of the reciprocal lattice of the surface. Since the electron wavelength is of the same order of magnitude as the surface lattice spacing, only a few diffraction spots appear at low primary energies (< 100 eV), but as the energy is increased, further spots appear and move with energy toward the normally deflected (i.e., undiffracted) beam. At the same time, the existing spots vary in intensity, appearing and disappearing on the fluorescent screen as they move with energy inward along radial directions.

The simple geometric, or kinematic, representation of the LEED pattern is valid only for scattering from those atoms actually in the surface (i.e., exposed to the incident electron beam). However, the incident beam has a finite depth of penetration, which of course increases with increasing energy; therefore the periodicity of the crystal normal to the surface must also be taken into account. This can be explored by recording the intensity of a particular diffraction spot as a func-

tion of primary electron energy (i.e., wavelength). The normally reflected beam will have intensity maxima due to the vertical periodicity, which are called primary Bragg peaks if the maxima are predicted by the Bragg equation, and secondary Bragg peaks if they are not. The latter are caused by multiple scattering effects, which can be accounted for only in a dynamic theory of scattering. The latter is complicated and can be applied only by using computer simulation, in which a feasible structure for the surface is guessed, from which diffraction intensities are calculated and compared with experiment. In this way a surface structure can be solved completely, but ambiguities cannot be eliminated in most cases.

Reflection High-Energy Electron Diffraction. The same diffraction mechanism is used in RHEED, but the experimental arrangement is quite different, and the observed pattern differs as well. Much higher primary electron energies are used than in LEED, in the range of 15–30 keV, and the primary beam is directed onto the crystal surface at near grazing incidence, at angles < 5°. As a result, the diffraction pattern is a mixture of spots and streaks. For such low angles of incidence the surface must be very smooth and flat, which is a limitation of the technique, but the requirement also means that any growth feature appearing above the original surface is easily identified since it produces its own characteristic diffraction pattern in transmission. In a variation of RHEED, the angle of the primary beam is varied backward and forward through a few degrees, which is called *rocking*, by analogy with the rocking patterns recorded in X-ray diffraction. Analysis of surface structure, with the assistance of sophisticated data processing, is less ambiguous when the rocking variation is used.

Applications. In virtually all basic work on interactions with, and reactions on, single-crystal surfaces over the last 50 years *LEED* has been employed to provide essential structural information. The structures proposed would in general be ambiguous, in the sense of the existence of several possibilities, if LEED were used on its own, but when the LEED data are taken in conjunction with those from other techniques applied at the same time, they have proved invaluable. The number of publications that have included LEED is probably greater than for any of the other techniques described here.

RHEED has also been used in basic studies, although to a far lesser extent because of its inherent restrictions, but in recent years it has found technological application as well. Semiconductor surfaces are usually flat, and epitaxial films grown on them are often flat as well. In those circumstances RHEED has proved to be a useful tool in the quality control of the structural condition of a semiconductor substrate and then of the perfection or otherwise of any films or layers grown or deposited on it. Thus each of the large UHV systems built for molecular beam epitaxy (MBE) contains a RHEED diagnostic system as one of the standard inspection techniques of the epilayer during and after growth.

27.8. Summary and Comparison of Techniques

Important surface analytical techniques are summarized and compared in Table 5.

27.9. Surface Analytical Equipment Suppliers

By now a large number of suppliers of surface analytical equipment can be found; a few are able to provide any combination of several techniques; the majority specialize in just one or two.

Supplier	Techniques		
Kratos Analytical Ltd., Barton Dock Road, Urmston, Manchester M31 2LD, United Kingdom	XPS SSIMS	AES UPS	SAM ISS
Ulvac Corporation, 10-3, 1-Chome Kyobashi, Chuo-Ku, Tokyo 104, Japan	AES	SSIMS	LEED
Riber, 133–137, Boulevard National, F-92503 Rueil Malmaison, B.P. 231, France	XPS SAM	AES	SSIMS
V. G. Microtech, Bellbrook Business Park, Bell Lane, Uckfield, East Sussex TN22 1BR, United Kingdom	XPS HREELS SNMS UPS AFM	AES SSIMS ISS BIS APFIM	SAM GDMS STM
Perkin Elmer, Physical Electronics Division, 6509 Flying Cloud Drive, Eden Prairie MN 55344, United States	XPS SSIMS	AES UPS	SAM
SPECS GmbH, Voltastrasse 5, 13355 Berlin, Germany	SNMS	SSIMS	
Scienta Instruments AB, Seminariegatan 33H, S-752 28 Uppsala, Sweden	XPS		
Omicron Vakuumphysik GmbH, Idsteinerstrasse 78, 65232 Taunusstein, Germany	STM	AFM	LEED
Oxford Applied Resarch Ltd., Crawley Mill, Witney, Oxon. OX8 5TJ, United Kingdom	RHEED		

Supplier	Techniques		
Staib Instrumente, Obere Hauptstrasse 45, 85354 Freising, Germany	RHEED		
Princeton Research Instruments, Inc., P.O. Box 1174, Princeton NJ 08542, United States	LEED		
National Electrostatics Corp., Graber Road, Box 310, Middleton WI 53562, United States	RBS		
J. A. Woollam Co., 650 J Street, Suite 39, Lincoln NE 68508, United States	ELL		
Rudolph Research, One Rudolph Road, Box 1000, Flanders NJ 07836, United States	ELL		
Sub Monolayer Science, 486 Ellis Street, Mountain View CA 94043, United States	SSIMS		
McAllister Technical Services, West 280 Prairie Avenue, Coeur d'Alene ID 83814, United States	STM		
Park Scientific Instruments, 1171 Borregas Avenue, Sunnyvale CA 94089, United States	STM	AFM	
Topometric, 1505 Wyatt Drive, Santa Clara CA 95054, United States	STM	AFM	
Digital Instruments, 6780 Cortona Drive, Santa Barbara CA 93117, United States	STM	AFM	
Burleigh Instruments Inc., Burleigh Park, Fishers NY 14453, United States	STM		
LK Technologies Inc., 3910 Roll Avenue, Bloomington IN 47403, United States	STM		
W. A. Technology Ltd., Chesterton Mills, French's Road, Cambridge CB4 3NP, United Kingdom	STM	AFM	LEED

Table 5. Summary and comparison of the more important techniques of surface analysis

Technique	Primary probe	Secondary particle(s)	Quantity measured	Elemental range	Type of information	Depth of information	Lateral resolution	Sensitivity, at %	Ease of quantification	Insulator analysis	Destructive	UHV environment
XPS	X-ray photons (Al or Mg K_α)	e^- (photo.)	E	all except H, He	elemental, chemical	4–8 monolayers	1 mm (standard), 5 µm (imaging)	10^{-4}	good	yes	hardly at all	yes
XAES	X-ray photons (Al or Mg K_α)	e^- (Auger)	E	all except H, He	Auger parameter	4–8 monolayers	1 mm	10^{-4}	N/A*	yes	hardly at all	yes
AES	e^- (5–10 keV)	e^- (Auger)	E	all except H, He	elemental	2–6 monolayers	0.5 µm	10^{-4}	moderate	difficult	often	yes
SAM	e^- (20–50 keV)	e^- (Auger)	E	all except H, He	elemental	2–6 monolayers	30 nm	10^{-4}	moderate	no	often	yes
SSIMS	n^+ (Ar$^+$, Ga$^+$)	n^+, n^-, clusters	M	all	chemical	1–2 monolayers	1 mm (standard), 50 nm (LMIS)	$10^{-2} - 10^{-6}$	poor	difficult	yes	yes
ISS	n^+ (He$^+$, Ne$^+$)	n^+ (He$^+$, Ne$^+$)	E	all except H, He	elemental, structural	1–2 monolayers	≈1 mm	10^{-4}	good	yes	not with He$^+$, possibly with Ne$^+$	yes
RBS	n^+ (H$^+$, He$^+$)	n^+ (H$^+$, He$^+$)	E	high Z on low Z	elemental, depth distribution	1 monolayer–2 µm	5 µm–1 mm	10^{-6}	good	yes	no	no
UPS	UV photons (He I, He II)	e^- (photo.)	E	(adsorbed molecules)	chemical (valence band)	2–10 monolayers	≈2 mm	10^{-3}	N/A*	yes	no	yes
RAIRS	IR photons	IR photons (reflected)	wavenumber	(adsorbed molecules)	chemical (vibrational)	1–2 monolayers	≈1 mm	10^{-4}	N/A*	no	no	no
SERS	laser photons	photons (scattered)	wavenumber	restricted	chemical (vibrational)	10–100 nm	≈0.5 mm	10^{-6}	N/A*	no	no	yes
GDOS	n^+ (Ar$^+$)	$n^* \rightarrow h\nu$	wavelength	all	elemental (profile)	10 nm–5 µm (profile)	none	10^{-6}	good	yes	yes	no
ELS	e^- (50–2000 eV)	e^- (loss)	E	(adsorbed molecules)	chemical	1–6 monolayers	≈0.5 mm	10^{-3}	N/A*	yes	no	yes
HREELS	e^- (2–5 eV)	e^- (loss)	E	(adsorbed molecules)	chemical (vibrational)	1 monolayer	≈1 mm	10^{-3}	N/A*	yes	no	yes

Table 5. (continued)

Technique	Primary probe	Secondary particle(s)	Quantity measured	Elemental range	Type of information	Depth of information	Lateral resolution	Sensitivity, at %	Ease of quantification	Insulator analysis	Destructive	UHV environment
IETS	(high field)	e^- (loss)	I (tunneling)	(adsorbed molecules)	chemical (vibrational)	1 monolayer	N/A*	10^{-3}	moderate	no	no	no
IPES	e^- (ramped)	photons (UV or X-ray)	photon flux	all, except H, He	chemical (unoccupied states)	1–6 monolayers	0.5–5 μm	10^{-3}	N/A*	yes	possibly	yes
SNMS	n^+ (Ar^+)	n^0	M	all	elemental	1–2 monolayers	≈1 mm	$10^{-2}–10^{-6}$	moderate	yes	yes	yes
GDMS	n^+ (Ar^+)	$n^* \rightarrow n^+$	M	all	elemental	10 nm–5 μm (profile)	none	10^{-6}	good	yes	yes	no
APFIM	(high field) (laser)	n^+	M	all	elemental	1 monolayer	2 nm	(1 atom)	very good	no	yes	yes
POSAP	(high field)	n^+	M	all	elemental	1–100 monolayers (profile)	2–20 nm	(1 atom)	very good	no	yes	yes
STM, STS	(high field)	e^-	I (tunneling)	all	topographic, chemical	1 monolayer	1 atom	(1 atom)	N/A*	no	possibly	yes
AFM	mechanical force		mechanical deflection	all	topographic	1 monolayer	1 atom	(1 atom)	N/A*	yes	no	no
LEED	e^- (200–1000 eV)	e^- (elastic)	I (diffracted)	all	structural	1–4 monolayers	≈0.5 mm	10^{-3}	N/A*	no	no	yes
RHEED	e^- (15–30 keV)	e^- (elastic)	I (diffracted)	all	structural	1–30 monolayers	30 nm	10^{-3}	N/A*	no	possibly	yes
TDS	T (ramped)	n^+, n^-	I (desorbed)	(adsorbed molecules)	chemical	1–2 monolayers	≈2–3 mm	10^{-4}	moderate	no	yes	yes

* N/A = not applicable.

27.10. References

[1] J. C. Rivière: *Surface Analytical Techniques.* Oxford University Press, Oxford 1990.

[2] M. P. Seah, D. Briggs in D. Briggs, M. P. Seah (eds.): *Practical Surface Analysis,* 2nd ed., **vol. 1,** J. Wiley & Sons, Chichester 1990, p. 13.

[3] D. M. Hercules, S. H. Hercules, *J. Chem. Educ.* **61** (1984) 402–409.

[4] M. J. Higatsberger, *Adv. Electron. Electron. Phys.* **56** (1981) 291–358.

[5] H. W. Werner, *Mikrochim. Acta Suppl.* **8** (1979) 25–50.

[6] A. E. Morgan, H. W. Werner, *Phys. Scr.* **18** (1978) 451–463.

[7] J. Tousset, *Le Vide* **33** (1978) 201–211.

[8] G. B. Larrabee, *Scanning Electron Micros.* part 1 1977, 639–650.

[9] A. Benninghoven, *Appl. Phys.* **1** (1973) 3–16.

[10] K. Siegbahn et al.: *ESCA: Atomic, Molecular, and Solid State Structure Studied by Means of Electron Spectroscopy,* Almqvist and Wiksells, Uppsala 1967.

[11] P. Auger, *J. Phys. Radium* **6** (1925) 205–209.

[12] M. P. Seah, W. A. Dench, *SIA Surf. Interface Anal.* **1** (1979) 2–11.

[13] M. P. Seah in J. M. Walls (ed.): *Methods of Surface Analysis,* Cambridge University Press, Cambridge 1989.

[14] P. W. Palmberg, *J. Electron Spectrosc. Relat. Phenom.* **5** (1974) 691–703;*J. Vac. Sci. Technol.* **12** (1975) 379–384.

[15] P. Coxon, J. Krizek, M. Humpherson, I. R. M. Wardell, *J. Electron Spectrosc. Relat. Phenom.* **51/52** (1980) 821–836.

[16] A. Barrie, *Chem. Phys. Lett.* **19** (1973) 109–113.

[17] M. Scrocco, *Chem. Phys. Lett.* **63** (1979) 52–56.

[18] M. Campagna, G. K. Wertheim, Y. Baer in L. Ley, M. Cardona (eds.): *Photoemission in Solids II,* Springer Verlag, Berlin 1979, chap. 4.

[19] G. K. Wertheim, D. N. E. Buchanan, N. V. Smith, M. M. Traum, *Phys. Lett. A* **49 A** (1974) 191–192.

[20] J. H. Scofield, *J. Electron. Spectrosc. Relat. Phenom.* **8** (1976) 129–137.

[21] R. F. Reilman, A. Msezane, S. T. Manson, *J. Electron Spectrosc. Relat. Phenom.* **8** (1976) 389–394.

[22] M. P. Seah in D. Briggs, M. P. Seah (eds.): *Practical Surface Analysis.* 2nd ed., **vol. I,** J. Wiley & Sons, Chichester 1990, p. 222.

[23] D. Briggs, M. P. Seah (eds.): *Practical Surface Analysis,* 2nd ed., **vol. I,** J. Wiley & Sons, Chichester 1990, p. 635.

[24] P. Brüesch, K. Müller, A. Atrens, H. Neff, *Appl. Phys. A* **A 38** (1985) 1–18.

[25] S. Hoffmann in D. Briggs, M. P. Seah (eds.): *Practical Surface Analysis,* 2nd ed., **vol. I,** J. Wiley & Sons, Chichester 1990, p. 143.

[26] C. D. Wagner, *Faraday Discuss. Chem. Soc.* **60** (1975) 291–300.

[27] C. D. Wagner, L. H. Gale, R. H. Raymond, *Anal. Chem.* **51** (1979) 466–482.

[28] C. D. Wagner in D. Briggs, M. P. Seah (eds.): *Practical Surface Analysis.* 2nd ed., **vol. I,** J. Wiley & Sons, Chichester 1990, p. 595.

[29] J. M. Walls (ed.): *Methods of Surface Analysis,* Cambridge University Press, Cambridge 1989.

[30] D. P. Woodruff, T. A. Delchar: *Modern Techniques of Surface Analysis.* Cambridge University Press, Cambridge 1988.

[31] J. F. Watts: *An Introduction to Surface Analysis by Electron Spectroscopy,* Oxford University Press, Oxford 1990.

[32] G. Ertl, J. Küppers: *Low Energy Electrons and Surface Chemistry,* 2nd ed., VCH, Weinheim 1985.

[33] D. Briggs (ed.): *Handbook of X-Ray and Ultraviolet Photoelectron Spectroscopy,* Heyden, London 1977.

[34] J. M. López Nieto, V. Cortés Corberán, J. L. G. Fierro, *SIA Surf. Interface Anal.* **17** (1991) 940–946.

[35] T. A. Patterson, J. C. Carver, D. E. Leyden, D. M. Hercules, *J. Phys. Chem.* **80** (1976) 1700–1708.

[36] G. Morea et al., *J. Chem. Soc. Faraday Trans.* **85** (1989) 3861–3870.

[37] A. Cimino, D. Gazzoli, M. Inversi, M. Valigi, *SIA Surf. Interface Anal.* **10** (1987) 194–201.

[38] A. Katrib et al., *Surf. Sci.* **189/190** (1987) 886–893.

[39] K. Balakrishnan, J. Schwank, *J. Catal.* **127** (1991) 287–306.

[40] B. L. Gustafson, P. S. Wehner, *Appl. Surf. Sci.* **52** (1991) 261–270.

[41] Yong-Xi Li, J. M. Stencel, B. H. Davis, *Appl. Catal.* **64** (1990) 71–81.

[42] G. Moretti, *SIA Surf. Interface Anal.* **17** (1991) 745–750.

[43] M. A. Baltanas, J. H. Onuferko, S. T. McMillan, J. R. Katzer, *J. Phys. Chem.* **91** (1987) 3772–3774.

[44] V. Di Castro, C. Furlani, M. Gargano, M. Rossi, *Appl. Surf. Sci.* **28** (1987) 270–278.

[45] V. Lindner, H. Papp, *Appl. Surf. Sci.* **32** (1988) 75–92.

[46] M. C. Burrell, G. A. Smith, J. J. Chera, *SIA Surf. Interface Anal.* **11** (1988) 160–164.

[47] D. L. Cocke, M. S. Owens, R. B. Wright, *Langmuir* **4** (1988) 1311–1318.

[48] J. L. G. Fierro, J. M. Palacios, F. Tomas, *SIA Surf. Interface Anal.* **13** (1988) 25–32.

[49] C. S. Kuivila, J. B. Butt, D. C. Stair, *Appl. Surf. Sci.* **32** (1988) 99–121.

[50] F. Lange, H. Schmelz, H. Knözinger, *J. Electron Spectrosc. Relat. Phenom.* **57** (1991) 307–315.

[51] J. Z. Shyu, K. Otto, *Appl. Surf. Sci.* **32** (1988) 246–252.

[52] K. Asami, K. Hashimoto, *Langmuir* **3** (1987) 897–904.

[53] P. Marcus, I. Olefjord, *Corros. Sci.* **28** (1988) 589–602.

[54] Shin-Puu Jeng, P. H. Holloway, C. D. Batich, *Surf. Sci.* **227** (1990) 278–290.

[55] P. J. Bremer et al., *SIA Surf. Interface Anal.* **17** (1991) 762–772.

[56] M.-P. Lahalle, R. Guillaumont, G. C. Allen, *J. Chem. Soc. Faraday Trans.* **86** (1990) 2641–2644.

[57] B.-P. Zhang et al., *Corros. Sci.* **33** (1992) 103–112.

[58] H. Yoshioka et al., *Corros. Sci.* **32** (1991) 313–325.

[59] R. B. Diegle et al., *J. Electrochem. Soc.* **135** (1988) 1085–1092.

[60] C. G. Pantano, R. K. Brow, *J. Am. Ceram. Soc.* **71** (1988) 577–581.

[61] Chin-An Chang, Yong-Kil Kim, A. G. Schrott, *J. Vac. Sci. Technol. A* **8** (1990) 3304–3309.

[62] F. J. Boerio, D. J. Ondrus, *J. Adhes.* **22** (1987) 1–12.

[63] J. N. Andersen et al., *J. Phys. Condens. Matter* **1** (1989) 7309–7313.

[64] A. A. Galuska, *Appl. Surf. Sci.* **40** (1989) 19–31, 33–40, 41–51.

[65] W. J. van Ooij, A. Sabata, A. D. Appelhans, *SIA Surf. Interface Anal.* **17** (1991) 403–420.

[66] Y.-H. Kim, Y. S. Chaug, N. J. Chou, J. Kim, *J. Vac. Sci. Technol. A* **5** (1987) 2890–2893.

[67] P. Steiner et al., *Z. Phys. B* **67** (1987) 497–502.

[68] C. Ziegler et al., *Fresenius J. Anal. Chem.* **341** (1991) 308–313.

[69] Y. Hwu et al., *Appl. Phys. Lett.* **59** (1991) 979–981.

[70] Y. Hwu et al., *Appl. Phys. Lett.* **57** (1990) 2139–2141.

[71] Q. Y. Ma et al., *J. Vac. Sci. Technol. A* **9** (1991) 390–393.

[72] R. P. Vasquez, M. C. Foote, B. D. Hunt, L. Bajuk, *J. Vac. Sci. Technol. A* **9** (1991) 570–573.

[73] H. Behner, G. Gieres, B. Sipos, *Fresenius J. Anal. Chem.* **341** (1991) 301–307.

[74] W. A. M. Aarnink, J. Gao, H. Rogalla, A. van Silfhout, *J. Less Common Met.* **164/165** (1990) 321–328.

[75] C. C. Chang et al., *J. Appl. Phys.* **67** (1990) 7483–7487.

[76] G. Frank, C. Ziegler, W. Göpel, *Phys. Rev. B. Condens. Matter* **43** (1991) 2828–2834.

[77] S. Tanaka et al., *Jpn. J. Appl. Phys.* **30** (1991) L1458–L1461.

[78] S. Kohiki et al., *Appl. Phys. A* **A50** (1990) 509–514.

[79] S. Kohiki et al., *J. Appl. Phys.* **68** (1990) 1229–1232.

[80] M. A. Sobolewski, S. Semancik, E. S. Hellman, E. H. Hartford, *J. Vac. Sci. Technol. A* **9** (1991) 2716–2720.

[81] G. Subramayam, F. Radpour, V. J. Kapoor, G. H. Lemon, *J. Appl. Phys.* **68** (1990) 1157–1163.

[82] K. Tanaka, H. Takaki, K. Koyama, S. Noguchi, *Jpn. J. Appl. Phys.* **29** (1990) 1658–1663.

[83] H. Yamanaka et al., *Jpn. J. Appl. Phys.* **30** (1991) 645–649.

[84] T. Suzuki et al., *Phys. Rev. B Condens. Matter* **42** (1990) 4263–4271.

[85] A. Fujimori et al., *Phys. Rev. B Condens. Matter* **42** (1990) 325–328.

[86] M. Klauda et al., *Physica C* **173** (1991) 109–116.

[87] K. Okada, Y. Seino, A. Kotani, *J. Phys. Soc. Jpn.* **60** (1991) 1040–1050.

[88] P. Adler, H. Buchkremer-Hermanns, A. Simon, *Z. Phys. B Condens. Matter* **81** (1990) 355–363.

[89] K. H. Young, E. J. Smith, M. M. Eddy, T. W. James, *Appl. Surf. Sci.* **52** (1991) 85–89.

[90] S. Gokhale et al., *Surf. Sci.* **257** (1991) 157–166.

[91] N. Aoto, E. Ikawa, N. Endo, Y. Kurogi, *Surf. Sci.* **234** (1990) 121–126.

[92] M. Niwano et al., *J. Vac. Sci. Technol. A* **9** (1991) 195–200.

[93] J. Onsgaard, M. Christiansen, F. Ørskov, P. J. Godowski, *Surf. Sci.* **247** (1991) 208–214.

[94] M. Kawasaki, G. J. Vandentop, M. Salmeron, G. A. Somorjai, *Surf. Sci.* **227** (1990) 261–267.

[95] M. Alonso et al., *J. Vac. Sci. Technol. B* **8** (1990) 955–963.

[96] M. Tang et al., *J. Vac. Sci. Technol. B* **8** (1990) 705–709.

[97] R. Klauser, M. Oshima, H. Sugahara, Y. Murata, *Surf. Sci.* **242** (1991) 319–323.

[98] U. K. Kirner, K. D. Schierbaum, W. Göpel, *Fresenius J. Anal. Chem.* **341** (1991) 416–420.

[99] F. S. Ohuchi, M. Kohyama, *J. Am. Ceram. Soc.* **74** (1991) 1163–1187.

[100] N. D. Shinn, C. H. F. Peden, K. L. Tsang, P. J. Berlowitz, *Phys. Scr.* **41** (1990) 607–611.

[101] S. Harel, J.-M. Mariot, E. Beauprez, C. F. Hague, *Surf. Coat. Technol.* **45** (1991) 309–315.

[102] G. Beamson, A. Bunn, D. Briggs, *SIA Surf. Interface Anal.* **17** (1991) 105–115.

[103] W. F. Stickle, J. F. Moulder, *J. Vac. Sci. Technol A* **9** (1991) 1441–1446.

[104] A. Naves de Brito et al., *SIA Surf. Interface Anal.* **17** (1991) 94–104.

[105] A. Chikoti, B. D. Ratner, D. Briggs, *Chem. Mater.* **3** (1991) 51–61.

[106] W. Pamler, F. Bell, L. Mühlhoff, H. J. Barth, *Mater. Sci. Eng. A* **139** (1991) 364–371.

[107] L. L. Cao, Y. M. Sun, L. Q. Zheng, *Wear* **140** (1990) 345–357.

[108] S. Noël, L. Boyer, C. Bodin, *J. Vac. Sci. Technol. A* **9** (1991) 32–38.

[109] A. Osaka, Y.-H. Wang, Y. Miura, T. Tsugaru, *J. Mater. Sci.* **26** (1991) 2778–2782.

[110] S. M. Mukhopadhyay, S. H. Garofalini, *J. Non Cryst. Solids* **126** (1990) 202–208.

[111] E.-T. Kang, D. E. Day, *J. Non Cryst. Solids* **126** (1990) 141–150.

[112] Z. Hussain, E. E. Khawaja, *Phys. Scr.* **41** (1990) 939–943.

[113] R. K. Brow, Y. B. Peng, D. E. Day, *J. Non Cryst. Solids* **126** (1990) 231–238.

[114] B. A. van Hassel, A. J. Burggraaf, *Appl. Phys. A* **A52** (1991) 410–417.

[115] P. Prieto, L. Galán, J. M. Sanz, *Surf. Sci.* **251/252** (1991) 701–705.

[116] K. Oyoshi, T. Tagami, S. Tanaka, *J. Appl. Phys.* **68** (1990) 3653–3660.

[117] Z.-H. Lu, A. Yelon, *Phys. Rev. B Condens. Matter* **41** (1990) 3284–3286.

[118] A. Carnera et al., *J. Non Cryst. Solids* **125** (1990) 293–301.

[119] R. W. Bernstein, J. K. Grepstad, *J. Appl. Phys.* **68** (1990) 4811–4815.

[120] A. Ermolieff et al., *Semicond. Sci. Tech.* **6** (1991) 98–102.

[121] S. V. Hattangady et al., *J. Appl. Phys.* **68** (1990) 1233–1236.

[122] Maria Faur et al., *SIA Surf. Interface Anal.* **15** (1990) 641–650.

[123] H. Hasegawa et al., *J. Vac. Sci. Technol. B* **8** (1990) 867–873.

[124] L. A. Harris, *J. Appl. Phys.* **39** (1968) 1419–1427, 1428–1431.

[125] R. E. Weber, W. T. Peria, *J. Appl. Phys.* **38** (1967) 4355–4358.

[126] P. W. Palmberg, G. K. Bohm, J. C. Tracy, *Appl. Phys. Lett.* **15** (1969) 254–255.

[127] E. Casnati, A. Tartari, C. Baraldi, *J. Phys. B* **15** (1982) 155–167.

[128] D. Landolt, H. J. Mathieu, *Oberfläche Surf.* **21** (1980) 8–15.

[129] L. E. Davis et al.: *Handbook of Auger Electron Spectroscopy*, Physical Electronic Ind., Minnesota 1976.

[130] (a) E. Kny, *J. Vac. Sci. Technol.* **17** (1980) 658–660.(b) J. Kleefeld, L. L. Levinson, *Thin Solid Films* **64** (1979) 389–393.(c) M. A. Smith, L. L. Levinson, *Phys. Rev. B Solid State* **16** (1977) 1365–1369.

[131] M. Gryzinski, *Phys. Rev.* **138** (1965) A305–321.

[132] S. Ichimura, R. Shimizu, *Surf. Sci.* **112** (1981) 386–408.

[133] R. Payling, *J. Electron Spectros. Relat. Phenom.* **36** (1985) 99–104.

[134] S. Mroczkowski, D. Lichtman, *J. Vac. Sci. Technol. A* **3** (1985) 1860–1865.

[135] T. Sato et al., *SIA Surf. Interface Anal.* **14** (1989) 787–793.

[136] M. P. Seah, *SIA Surf. Interface Anal.* **9** (1986) 85–98.
[137] R. Hamminger, G. Grathwohl, F. Thümmler, *J. Mater. Sci.* **18** (1983) 353–364.
[138] H. Goretzki, *Fresenius Z. Anal. Chem.* **329** (1987) 180–189.
[139] T. H. Chuang, *Mat. Sci. Eng.* **A 141** (1991) 169–178.
[140] D. J. Nettleship, R. K. Wild, *SIA Surf. Interface Anal.* **16** (1990) 552–558.
[141] C. Setti, *Mikrochim. Acta* **1** (1987) 437–444.
[142] K. Sasaki, A. Noya, T. Umezawa, *Jpn. J. Appl. Phys.* **29** (1990) 1043–1047.
[143] W. Pamler, A. Kohlhase, *SIA Surf. Interface Anal.* **14** (1989) 289–294.
[144] M. Procop, H. Raidt, B. Sandow, *Phys. Status Solidi A* **99** (1987) 573–580.
[145] A. Reif, P. Streubel, A. Meisel, D. Zeissig, *Phys. Status Solidi A* **122** (1990) 331–340.
[146] S. Stein et al., *Fresenius J. Anal. Chem.* **341** (1991) 66–69.
[147] A. Manara, V. Sistori, *SIA Surf. Interface Anal.* **15** (1990) 457–462.
[148] J. Shin, K. N. Geib, C. W. Wilmsen, Z. Lilliental-Weber, *J. Vac. Sci. Technol. A* **8** (1990) 1894–1898.
[149] S. V. Hattangady et al., *Mater. Res. Soc. Symp. Proc.* **165** (1990) 221-226.
[150] A. Hupfer, J. Albrecht, D. Dietrich, *Fresenius J. Anal. Chem.* **341** (1991) 439–444.
[151] F. Bertran et al., *Surf. Sci. Lett.* **245** (1991) L163–L169.
[152] X. Wallart et al., *J. Appl. Phys.* **69** (1991) 8168–8176.
[153] V. M. Bermudez, *Surf. Sci. Lett.* **230** (1990) L155–L161.
[154] H. Roux, N. Boutaoui, M. Tholomier, *Thin Solid Films* **172** (1989) 141–148.
[155] B. J. Burrow, A. E. Morgan, R. C. Ellwanger, *J. Vac. Sci. Technol. A* **4** (1986) 2463–2469.
[156] K. Machida, M. Enyo, I. Toyoshima, *Thin Solid Films* **161** (1988) L91–L95.
[157] K. Sasaki, A. Noya, T. Umezawa, *Jpn. J. Appl. Phys.* **27** (1988) 1190–1192.
[158] A. Gauzzi, H. J. Mathieu, J. H. James, B. Kellett, *Vacuum* **41** (1990) 870–874.
[159] W. A. M. Aarnink et al., *Appl. Surf. Sci.* **47** (1991) 195–203.
[160] Q. X. Jia, W. A. Anderson, *Appl. Phys. Lett.* **57** (1990) 304–306.
[161] J. Colino, J. L. Sacedon, L. Del Olmo, J. L. Vincent, *J. Vac. Sci. Technol. A* **8** (1990) 4021–4025.
[162] Udayan De, S. Natarajan, E. W. Seibt, *Physica C* **183** (1991) 83–89.
[163] E. W. Seibt, A. Zalar, *Mater. Lett.* **11** (1991) 1–5.
[164] C. Uebing, *Surf. Sci.* **225** (1990) 97–106.
[165] F. Reniers, M. Jardinier-Offergeld, F. Bouillon, *SIA Surf. Interface Anal.* **17** (1991) 343–351.
[166] S. Nakanishi et al., *Surf. Sci. Lett.* **247** (1991) L215–L220.
[167] M. J. van Staden, J. P. Roux, *Appl. Surf. Sci.* **44** (1990) 271–277.
[168] R. F. Visser, J. P. Roux, *Appl. Surf. Sci.* **51** (1991) 115–124.
[169] W. Arabczyk, H.-J. Müssig, F. Storbeck, *Surf. Sci.* **251/252** (1991) 804–808.
[170] A. Schönborn, H. Bubert, E. H. de Kaat, *Fresenius J. Anal. Chem.* **341** (1991) 241–244.
[171] J.-R. Chen et al., *Appl. Surf. Sci.* **33/34** (1988) 212–219.
[172] S. K. Sharma, S. Hofmann, *Appl. Surf. Sci.* **51** (1991) 139–155.
[173] S. Badrinarayanan, A. B. Mandale, S. Sinha, *Surf. Coat. Technol.* **34** (1988) 133–139.
[174] R. Hauert, J. Patscheider, R. Zehringer, M. Tobler, *Thin Solid Films* **206** (1991) 330–334.
[175] A. Benninghoven, *Z. Phys.* **220** (1969) 159–180.
[176] J.-C. Pivin, *J. Mater. Sci.* **18** (1983) 1267–1290.
[177] P. Sigmund in N. H. Tolk, J. C. Tully, W. Heiland, C. W. White (eds.): *Inelastic Ion-Surface Collisions*, Academic Press, New York 1977.
[178] P. D. Prewett, D. K. Jefferies, *J. Phys. D* **13** (1980) 1747–1755.
[179] A. R. Krauss, D. M. Gruen, *Appl. Phys.* **14** (1977) 89–97.
[180] P. Steffens et al., *J. Vac. Sci. Technol. A* **3** (1985) 1322–1325.
[181] A. J. Eccles, J. C. Vickerman, *J. Vac. Sci. Technol. A* **7** (1989) 234–244.
[182] A. Benninghoven, K.-H. Müller, M. Schemmer, P. Beckman, *Proc. Int. Conf. Solid Surf.* **3rd** (1977) 1063–1066.
[183] I. V. Bletsos, D. M. Hercules, A. Benninghoven, D. Greifendorf in A. Benninghoven, R. J. Colton, D. S. Simons, H. W. Werner (eds.): *SIMS V*, Springer Verlag, Berlin 1986.
[184] P. H. McBreen, S. Moore, A. Adnot, D. Roy, *Surf. Sci.* **194** (1988) L112–L118.
[185] C. A. Andersen, J. R. Hinthorne, *Anal. Chem.* **45** (1973) 1421–1438.
[186] J. A. Treverton et al., *SIA Surf. Interface Anal.* **15** (1990) 369–376.
[187] W. J. van Ooij, A. Sabata, A. D. Appelhans, *SIA Surf. Interface Anal.* **17** (1991) 403–420.
[188] M. Morra, E. Occhiello, F. Garbassi, *SIA Surf. Interface Anal.* **16** (1990) 412–417.
[189] D. E. Fowler, R. D. Johnson, D. van Leyen, A. Benninghoven, *SIA Surf. Interface Anal.* **17** (1991) 125–136.
[190] G. J. Leggett, D. Briggs, J. C. Vickerman, *SIA Surf. Interface Anal.* **17** (1991) 737–744.
[191] W. D. Ramsden, *SIA Surf. Interface Anal.* **17** (1991) 793–802.
[192] D. J. Pawson et al., *SIA Surf. Interface Anal.* **18** (1992) 13–22.
[193] R. A. Cocco, B. J. Tatarchuk, *Surf. Sci.* **218** (1989) 127–146.
[194] B. H. Sakakini et al., *Surf. Sci.* **271** (1992) 227–236.
[195] X.-L. Zhou, C. R. Flores, J. M. White, *Surf. Sci. Lett.* **268** (1992) L267–L273.
[196] M. E. Castro, J. M. White, *Surf. Sci.* **257** (1991) 22–32.
[197] D. Briggs, M. J. Hearn, *SIA Surf. Interface Anal.* **13** (1988) 181–185.
[198] D. Briggs et al., *SIA Surf. Interface Anal.* **15** (1990) 62–65.
[199] M. J. Hearn, D. Briggs, *SIA Surf. Interface Anal.* **17** (1991) 421–429.
[200] D. G. Armour, *Vacuum* **31** (1981) 417–428.
[201] P. J. Martin, R. P. Netterfield, *SIA Surf. Interface Anal.* **10** (1987) 13–16.
[202] H. Niehus, *Appl. Phys. A* **53** (1991) 388–402.
[203] J.-M. Beuken, P. Bertrand, *Surf. Sci.* **162** (1985) 329–336.
[204] J. F. van der Veen, *Surf. Sci. Rep.* **5** (1985) 199–287.
[205] H. Niehus, G. Comsa, *Surf. Sci.* **140** (1984) 18–30.
[206] G. Molière, *Z. Naturforsch.* **2 A** (1947) 133–143.

[207] H. D. Hagstrum, *Phys. Rev.* **123** (1961) 758–765.

[208] R. De Gryse, J. Landuyt, L. Vandenbroucke, J. Vennik, *SIA Surf. Interface Anal.* **4** (1982) 168–173.

[209] H. H. Brongersma, M. J. Sparnaay, T. M. Buck, *Surf. Sci.* **71** (1978) 657–666.

[210] G. C. Nelson, *J. Vac. Sci. Technol.* **13** (1976) 974–975.

[211] A. Rossi, C. Calinski, H.-W. Hoppe, H.-H. Strehblow, *SIA Surf. Interface Anal.* **18** (1992) 269–276.

[212] H. J. Kang, R. Shimizu, T. Okutani, *Surf. Sci.* **116** (1982) L 173–L 178.

[213] E. N. Haeussler, *SIA Surf. Interface Anal.* **2** (1980) 134–139.

[214] Y. G. Shen, D. J. O'Connor, R. J. MacDonald, *SIA Surf. Interface Anal.* **17** (1991) 903–910.

[215] C. R. K. Marrian, A. Shih, G. A. Haas, *Appl. Surf. Sci.* **24** (1985) 372–390.

[216] S. J. Hoekje, G. B. Hoflund, *Thin Solid Films* **197** (1991) 367–380.

[217] H. Puderbach, *Fresenius Z. Anal. Chem.* **314** (1983) 260–264.

[218] G. H. Vurens, D. R. Strongin, M. Salmeron, G. A. Somorjai, *Surf. Sci.* **199** (1988) L 387–L 393.

[219] A. van Oostrom, *Vacuum* **34** (1984) 881–892.

[220] T. W. Conlon, *Contemp. Phys.* **26** (1985) 521–558.

[221] A. E. T. Kuiper, M. F. C. Willemsen, J. C. Barbour, *Appl. Surf. Sci.* **35** (1988–1989) 186–198.

[222] H. W. Werner in H. Oechsner (ed.): *Thin Film and Depth Profile Analysis,* Springer Verlag, Berlin 1984.

[223] R. S. Bhattacharya, P. P. Pronko, *Appl. Phys. Lett.* **40** (1982) 890–892.

[224] F. Nava et al., *Thin Solid Films* **89** (1982) 381–385.

[225] J. M. M. de Nijs, A. van Silfhout, *Appl. Surf. Sci.* **40** (1990) 349–358.

[226] R. Fastow et al., *J. Vac. Sci. Technol. A* **5** (1987) 164–168.

[227] R. C. Ellwanger, A. E. Morgan, W. T. Stacy, Y. Tamminga, *J. Vac. Sci. Technol. B* **1** (1983) 533–539.

[228] T. Obata et al., *Thin Solid Films* **87** (1982) 207–214.

[229] M. C. Burrell, G. A. Smith, J. J. Chera, *SIA Surf. Interface Anal.* **11** (1988) 160–164.

[230] H. Z. Wu et al., *Thin Solid Films* **191** (1990) 55–67.

[231] A. E. T. Kuiper, S. W. Koo, F. H. P. M. Habraken, Y. Tamminga, *J. Vac. Sci. Technol. B* **1** (1983) 62–66.

[232] F. H. P. M. Habraken et al., *J. Appl. Phys.* **53** (1982) 404–415.

[233] Z. L. Wang, J. F. M. Westendorp, F. W. Saris: "Ion Beam Modification of Materials," in B. Biasse, G. Destefanis, J. P. Gailliard (eds.): *Proc. 3rd Int. Conf.,* Grenoble 1982.

[234] R. S. Bhattacharya et al., *J. Phys. Chem. Solids* **44** (1983) 61–69.

[235] G. Binnig, H. Rohrer, *Surf. Sci.* **126** (1983) 236–244.

[236] P. K. Hansma, J. Tersoff, *J. Appl. Phys.* **61** (1987) R 1–R 23.

[237] P. K. Hansma, V. B. Elings, O. Marti, C. E. Bracker, *Science (Washington D.C.)* **242** (1988) 209–216.

[238] G. Binnig, C. F. Quate, C. Gerber, *Phys. Rev. Lett.* **56** (1986) 930–933.

[239] A. L. Weisenhorn et al., *Science (Washington D.C.)* **247** (1990) 1330–1333.

[240] B. Drake et al., *Science (Washington D.C.)* **243** (1989) 1586–1589.

[241] H. G. Hansma et al., *Clin. Chem. (Winston Salem N.C.)* **37** (1991) 1497–1501.

[242] S. Manne, H.-J. Butt, S. A. C. Gould, P. K. Hansma, *Appl. Phys. Lett.* **56** (1990) 1758–1759.

[243] H.-J. Butt et al., *J. Struct. Biol.* **105** (1990) 54–61.

[244] S. Manne et al., *Science (Washington D.C.)* **251** (1991) 183–186.

[245] Ph. Ebert, G. Cox, U. Poppe, K. Urban, *Surf. Sci.* **271** (1992) 587–595.

[246] S. A. C. Gould, K. Burke, P. K. Hansma, *Phys. Rev. B Condens. Matter* **40** (1989) 5363–5366.

[247] S. A. Gould et al., *Ultramicroscopy* **33** (1990) 93–98.

[248] W. J. Kaiser, R. C. Jaklevic, *IBM J. Res. Dev.* **30** (1986) 411–416.

[249] B. Reihl, J. K. Gimzewski, J. M. Nicholls, E. Tosatti, *Phys. Rev. B Condens. Matter* **33** (1986) 5770-5773.

[250] A. J. Gratz, S. Manne, P. K. Hansma, *Science (Washington D.C.)* **251** (1991) 1343–1346.

[251] P. Dietz et al., *Ultramicroscopy* **35** (1991) 155–159.

[252] H. Hartman et al., *Clays Clay Miner.* **38** (1990) 337–342.

[253] H. Lindgreen et al., *Am. Mineral.* **76** (1991) 1218–1222.

[254] P. G. Arscott, V. A. Bloomfield, *Ultramicroscopy* **33** (1990) 127–131.

[255] R. D. Edstrom et al., *Ultramicroscopy* **33** (1990) 99–106.

[256] K. A. Fisher et al., *Ultramicroscopy* **33** (1990) 117–126.

[257] J. A. N. Zasadzinski et al., *Science (Washington D.C.)* **239** (1988) 1013–1015.

[258] A. L. Weisenhorn et al., *Biophys. J.* **58** (1990) 1251–1258.

[259] H. G. Hansma et al., *Clin. Chem. (Winston Salem N.C.)* **37** (1991) 1497–1501.

[260] J. N. Lin et al., *Langmuir* **6** (1990) 509–511.

[261] H. G. Hansma et al., *J. Vac. Sci. Technol. B* **9** (1991) 1282–1284.

[262] A. L. Weisenhorn et al., *Langmuir* **7** (1991) 8–12.

[263] H.-J. Butt et al., *J. Struct. Biol.* **105** (1990) 54–61.

[264] R. Wolkow, Ph. Avouris, *Phys. Rev. Lett.* **60** (1988) 1049–1052.

[265] K. Takayanagi, Y. Tanishiro, M. Takahashi, S. Takahashi, *J. Vac. Sci. Technol. A* **3** (1985) 1502–1506.

[266] G. Binnig et al., *Phys. Rev. Lett.* **55** (1985) 991–994.

[267] A. Baratoff et al., *Surf. Sci.* **168** (1986) 734–743.

[268] S. A. Elrod, A. L. de Lozanne, C. R. Quate, *Appl. Phys. Lett.* **45** (1986) 1240–1242.

[269] P. Sautet, C. Joachim, *Surf. Sci.* **271** (1992) 387–394.

[270] F. P. Netzer, J.-U. Mack, *Chem. Phys. Lett.* **95** (1983) 492–496.

[271] W. Erley, *J. Electron Spectros. Relat. Phenom.* **44** (1987) 65–78.

[272] M. Fleischmann, P. J. Hendra, A. J. McQuillan, *Chem. Phys. Lett.* **26** (1976) 163–174.

[273] W. Akemann, A. Otto, *Surf. Sci.* **272** (1992) 211–219.

[274] A. Otto, I. Mrozek, H. Grabhorn, W. Akemann, *J. Phys. Condens. Matter* **4** (1992) 1143–1212.

[275] G. L. Richmond, J. M. Robinson, V. L. Shannon, *Prog. Surf. Sci.* **28** (1988) 1–70.

[276] R. J. MacDonald, P. J. Martin, *Surf. Sci.* **67** (1977) 237–250.

[277] J. C. Hamilton, R. J. Anderson, *Surf. Sci.* **149** (1985) 81–92.

[278] H. Bach, F. G. K. Baucke, *J. Am. Ceram. Soc.* **65** (1982) 527–533.

[279] F. G. K. Baucke, H. Bach, *J. Am. Ceram. Soc.* **65** (1982) 534–539.

[280] W. Grimm, *Naturwissenschaften* **54** (1967) 586; *Spectrochim. Acta Part B* **23 B** (1968) 443–448.

[281] A. Bengtson. *Spectrochim. Acta Part B* **40 B** (1985) 631–639.

[282] R. Berneron, J. C. Charbonnier, *SIA Surf. Interface Anal.* **3** (1981) 134–141.

[283] J. Pons-Corbeau, *SIA Surf. Interface Anal.* **7** (1985) 169–176.

[284] A. Bengtson, L. Danielson, *Thin Solid Films* **124** (1985) 231–236.

[285] A. Quentmeier et al., *Mikrochimica Acta Suppl.* **11** (1985) 89–102.

[286] J. Takadoum et al., *SIA Surf. Interface Anal.* **6** (1984) 174–183.

[287] G. Rosina, E. Bertel, F. P. Netzer, *Phys. Rev. B Condens. Matter* **34** (1986) 5746–5753.

[288] B. M. Biwer, S. L. Bernasek, *Surf. Sci.* **167** (1986) 207–230.

[289] E. Bertel, R. Stockbauer, T. E. Madey, *Surf. Sci.* **141** (1984) 355–387.

[290] J. C. Rivière, F. P. Netzer, G. Rosina, *SIA Surf. Interface Anal.* **18** (1992) 333–344.

[291] F. P. Netzer, J. C. Rivière, G. Rosina, *Phys. Rev. B* **38** (1988) 7453–7460.

[292] C. M. Stander, *Appl. Surf. Sci.* **16** (1983) 463–468.

[293] A. Otto, B. Reihl, *Surf. Sci.* **178** (1986) 635–645.

[294] J. Onsgaard, I. Chorkendorff, O. Ellegaard, O. Sørensen, *Surf. Sci.* **138** (1984) 148–158.

[295] G. Strasser, G. Rosina, J. A. D. Matthew, F. P. Netzer, *J. Phys. F Met. Phys.* **15** (1985) 739–751.

[296] F. P. Netzer, G. Strasser, J. A. D. Matthew, *Solid State Commun.* **45** (1983) 171–174.

[297] G. Strasser, F. P. Netzer, *J. Vac. Sci. Technol. A* **2** (1984) 826–830.

[298] G. Strasser, E. Bertel, J. A. D. Matthew, F. P. Netzer in P. Wachter, H. Boppart (eds.): *Valence Instabilities*, North Holland Publ., Amsterdam 1982.

[299] M. Nishijima, S. Masuda, H. Kobayashi, M. Onchi, *Rev. Sci. Instrum.* **53** (1982) 790–796.

[300] T. U. Bartke, R. Franchy, H. Ibach, *Surf. Sci.* **272** (1992) 299–305.

[301] A. Fricke, H. Graupner, L. Hammer, K. Müller, *Surf. Sci.* **272** (1992) 182–188.

[302] Q. Dai, A. J. Gellman, *Surf. Sci.* **257** (1991) 103–112.

[303] S. Bao et al., *Surf. Sci.* **271** (1992) 513–518.

[304] M. J. Nowakowski, J. M. Vohs, D. A. Bonnell, *Surf. Sci. Lett.* **271** (1992) L351–L356.

[305] X. H. Feng et al., *J. Chem. Phys.* **90** (1989) 7516–7522.

[306] J.-P. Lu, M. R. Albert, S. L. Bernasek, D. J. Dwyer, *Surf. Sci.* **215** (1989) 348–362.

[307] G. Lauth, T. Solomun, W. Hirschwald, K. Christmann, *Surf. Sci.* **210** (1989) 201–224.

[308] R. D. Ramsier, P. N. Henriksen, A. N. Gent, *Surf. Sci.* **203** (1988) 72–88.

[309] K. W. Hipps, U. Mazur, *Surf. Sci.* **207** (1989) 385–400.

[310] P. N. T. van Velzen, M. C. Raas, *Surf. Sci.* **161** (1985) L605–L613.

[311] M. C. Gallagher, Y. B. Ning, J. G. Adler, *Phys. Rev. B Condens. Matter* **36** (1987) 6651–6656.

[312] D. M. Brewis et al., *SIA Surf. Interface Anal.* **6** (1984) 40–45.

[313] D. P. Woodruff, P. D. Johnson, N. V. Smith, *J. Vac. Sci. Technol. A* **1** (1983) 1104–1110.

[314] G. Rangelov, N. Memmel, E. Bertel, V. Dose, *Surf. Sci.* **236** (1990) 250–258.

[315] R. Konishi, S. Ikeda, T. Osaki, H. Sasakura, *Jpn. J. Appl. Phys.* **29** (1990) 1805–1806.

[316] K. H. Frank, R. Dudde, H. J. Sagner, W. Eberhardt, *Phys. Rev. B Condens. Matter* **39** (1989) 940–948.

[317] N. B. Brookes et al., *Phys. Rev. B Condens. Matter* **39** (1989) 2736–2739.

[318] T. Watanabe et al., *Physica C* **176** (1991) 274–278.

[319] J. M. Nicholls, B. Reihl, *Surf. Sci.* **218** (1989) 237–245.

[320] P. L. Wincott, K. L. Håkansson, L. I. Johansson, *Surf. Sci.* **211/212** (1989) 404–413.

[321] E. W. Thomas, *Vacuum* **34** (1984) 1031–1044.

[322] R. Whaley, E. W. Thomas, *J. Appl. Phys.* **56** (1984) 1505–1513.

[323] R. A. Baragiola, L. Nair, T. E. Madey, *Nucl. Inst. Methods Phys. Res.* **B 58** (1991) 322–327.

[324] R. H. Milne, E. A. Maydell, D. J. Fabian, *Appl. Phys. A* **A 52** (1991) 197–202.

[325] H. Brenten, H. Müller, D. Kruse, V. Kempter, *Nucl. Inst. Methods Phys. Res.* **B 58** (1991) 328–332.

[326] C. Benazeth, N. Benazeth, M. Hou, *Surf. Sci.* **151** (1985) L137–L143.

[327] M. Nègre, J. Mischler, N. Benazeth, *Surf. Sci.* **157** (1985) 436–450.

[328] H. D. Hagstrum, *Phys. Rev.* **150** (1966) 495–515.

[329] H. D. Hagstrum, G. E. Becker, *Phys. Rev. B Solid State* **4** (1971) 4187–4202.

[330] H. D. Hagstrum, G. E. Becker, *Phys. Rev. B Solid State* **8** (1973) 1592–1603.

[331] H. D. Hagstrum, G. E. Becker, *J. Chem. Phys.* **54** (1971) 1015–1032.

[332] G. E. Becker, H. D. Hagstrum, *Surf. Sci.* **30** (1972) 505–524.

[333] H. D. Hagstrum, G. E. Becker, *J. Vac. Sci. Technol.* **14** (1977) 369–371.

[334] H. Conrad et al., *Phys. Rev. Lett.* **42** (1979) 1082–1086.

[335] H. Conrad et al., *Surf. Sci.* **100** (1980) L461–L466.

[336] F. Bozso et al., *J. Chem. Phys.* **78** (1983) 4256–4269.

[337] F. Bozso et al., *Surf. Sci.* **136** (1984) 257–266.

[338] F. Bozso et al., *Surf. Sci.* **138** (1984) 488–504.

[339] L. Lee et al., *Surf. Sci.* **165** (1986) L95–L105.

[340] W. Sesselmann et al., *Surf. Sci.* **146** (1984) 17–42.

[341] G. Rocker, H. Tochihara, R. M. Martin, H. Metiu, *Surf. Sci.* **181** (1987) 509–529.

[342] G. H. Rocker et al., *Surf. Sci.* **208** (1989) 205–220.

[343] H. Oechsner, E. Stumpe, *Appl. Phys.* **14** (1977) 43–47.

[344] H. Oechsner, *Nucl. Inst. Methods Phys. Res.* **B 33** (1988) 918–925.

[345] J. W. Coburn, E. Taglauer, E. Kay, *J. Appl. Phys.* **45** (1974) 1779–1786.

[346] N. E. Sanderson et al., *Microchim. Acta* **1** (1987) 275–290.

[347] A. Wucher, H. Oechsner, F. Novak, *Thin Solid Films* **174** (1989) 133–137.

[348] P. M. Charalambous, *Steel Res.* **5** (1987) 197–203.

[349] J. Sopka, H. Oechsner, *J. Non Cryst. Solids* **114** (1989) 208–210.

[350] H. Oechsner, *Scanning Microscopy* **3** (1989) 411–418.

[351] A. Benninghoven, M. Kempken, P. Klüsener, *Surf. Sci.* **206** (1988) L927–L933.

[352] P. Beckmann, M. Kopnarski, H. Oechsner, *Mikrochim. Acta Suppl.* **11** (1985) 79–88.

[353] U. Schühle, J. B. Pallix, C. H. Becker, *J. Vac. Sci. Technol. A* **6** (1988) 936–940.

[354] J. B. Pallix, C. H. Becker, N. Newman, *J. Vac. Sci. Technol. A* **6** (1988) 1049–1052.

[355] E. Kawatoh, R. Shimizu, *Jpn. J. Appl. Phys.* **30** (1991) 832–836.

[356] D. M. Gruen et al., *Nucl. Inst. Methods Phys. Res.* **B 58** (1991) 505–511.

[357] R. Michael, D. Stulik, *App. Surf. Sci.* **28** (1987) 367–381.

[358] D. J. Surman, J. A. van den Berg, J. C. Vickerman, *SIA Surf. Interface Anal.* **4** (1982) 160–167.

[359] A. Brown, J. C. Vickerman, *Analyst (London)* **109** (1984) 851–857.

[360] A. Brown, J. C. Vickerman, *SIA Surf. Interface Anal.* **6** (1984) 1–14.

[361] A. Brown, J. A. van den Berg, J. C. Vickerman, *Spectrochim. Acta Part B* **40 B** (1985) 871–877.

[362] A. Brown, *Eur. Spectros. News* **81** (1988) 13–21.

[363] W. Unger, W. Pritzkow, *Int. J. Mass. Spec. Ion Process.* **75** (1987) 15–26.

[364] R. Michael, D. Stulik, *Appl. Surf. Sci.* **28** (1987) 53–62.

[365] E. W. Müller, J. A. Panitz, S. B. McLane, *Rev. Sci. Instrum.* **39** (1968) 83–86.

[366] E. W. Müller, *Z. Phys.* **131** (1951) 136–142.

[367] M. K. Miller, *Int. Mat. Rev.* **32** (1987) 221–240.

[368] A. Cerezo, T. J. Godfrey, G. D. W. Smith, *Rev. Sci. Instrum.* **59** (1988) 862–866.

[369] A. Cerezo et al., *J. Micros. (Oxford)* **154** (1989) 215–225.

[370] A. Cerezo, *Vacuum* **42** (1991) 605–611.

[371] A. Cerezo et al., *EMSA Bull.* **20** (1990) 77–83.

[372] Li Chang, G. D. W. Smith, G. B. Olson, *J. Phys. Colloq.* **C 2, Suppl. 3, 47** (1986) 265–275.

[373] S. S. Brenner, M. K. Miller, *J. Met.* **35** (1983) 54–63.

[374] G. D. W. Smith, *Metals Handbook,* **10** (1986) 9th ed., 583–602.

[375] A. Cerezo, G. D. W. Smith, *Materials Forum* **10** (1987) 104–116.

[376] M. L. Knotek, P. J. Feibelman, *Phys. Rev. Lett.* **40** (1978) 964–967.

[377] T. E. Madey, *Science (Washington D.C.)* **234** (1986) 316–322.

[378] B. Krüger, C. Bendorff, *Surf. Sci.* **178** (1986) 704–715.

[379] C. J. Davisson, L. H. Germer, *Phys. Rev.* **30** (1927) 705–740.

28. Chemical and Biochemical Sensors

KARL CAMMANN, Institut für Chemo- und Biosensorik, Münster, Germany (Chap. 28, 28.4. 28.5. Sections 28.2.2. 28.2.3.1.3, 28.2.3.1.4, 28.2.3.4, 28.2.4, 28.2.5)

BERND ROSS, Institut für Chemo- und Biosensorik, Münster, Germany (Sections. 28.2.3, 28.2.3.1, 28.2.3.1.2)

ANDREAS KATERKAMP, Institut für Chemo- und Biosensorik, Münster, Germany (Sections 28.2.3.2, 28.3.3.2)

JÖRG REINBOLD, Institut für Chemo- und Biosensorik, Münster, Germany (Sections 28.2.2.2, 28.2.3.3)

BERND GRÜNDIG, SensLab GmbH - Bioelektrochemische Sensoren, Leipzig, Germany (Sections 28.3.1 – 28.3.2.3, 28.3.5)

REINHARD RENNEBERG, Hong Kong University of Science & Technology - Department of Chemistry, Hong Kong, PR China (Sections 28.3.1, 28.3.2.4 – 28.3.4)

28.1. Introduction to the Field of Sensors and Actuators

When a person suffers the loss of one of the five senses it is recognized as a serious obstacle to carrying out everyday activities. In the case of blindness and deafness the expression "disabled" is even used as an indication of the severity of the loss, just as with those whose limbs (legs or arms) or other crucial body components are incapable of functioning properly. In the animate world generally, sensing and then acting or reacting is closely linked with the very essence of life, where input from some type of sensory system always precedes intelligent action. From the earliest days of mankind it was the availability of a powerful "coping mechanism" that constituted the driving force leading first to intelligent action and later to the unbiased accumulation of knowledge.

Any intellectual information structure capable of producing theories is dependent upon a reliable input of information. Imperfect observations, or perceptual processes subject to frequent errors or some bias , will never provide a satisfactory basis for working theories subject to verification by others. In this respect analytical chemistry can be regarded as one of the cognitive means available to the chemist interested in ascertaining the composition of the material world and the distribution and arrangement of atoms and molecules within a particular sample. Since both the nature and the three-dimensional arrangement of atoms in space determine the characteristics of matter, reliable information with respect to both is crucial for understanding the whole of the material world. This observation is equally valid for the inorganic, the organic, and also the living world. A high level of structured information is essential to the work of most natural scientists. The ability to describe in detail a complex chemical process (e.g., the combustion of fuel in an engine or waste in a municipal incinerator, the synthesis of minerals deep within the earth, or the course of biochemical reactions taking place in living systems) is crucially dependent on the quality of the information available. High technology is impossible without detailed information on the state of the system in question, as has been amply demonstrated in semiconductor technology, where knowledge of the purity of the materials and the state of cleanliness in the production plant is absolutely essential.

Complex and especially dynamic systems involving chemical processes are often described in terms of simplified models. The more chemical information is available, the better will be the model — and consequently the more complete the understanding and control over the system. Chemical analysis constitutes the primary source of information for characterizing the various states of such a system.

As a result of a tremendous increase in the demand for analytical chemical information, the number of analyses conducted for acquiring information about particular samples is increasing dramatically. It is thought that worldwide more than 10^9 analyses are now performed every day — with a double-digit growth rate. The latter because quality control has become a worldwide topic, agreed upon in many international trade contracts. Many complex systems are governed by time-dependent parameters, which makes it essential that as much information as possible be acquired within a particular period of time so that the various reactions taking place within the system can be followed, and to provide insight into the dynamics of the mechanisms involved. Devices designed to permit continuous measurement of a quantity are called *sensors*. Various types of sensors have been developed for monitoring many important *physical* parameters, including temperature, pressure, speed, flow, etc. The number of potential interferences is rather limited in these cases, so it is not surprising that the corresponding sensors offer a level of reliability considerably higher than that for sensors that measure *chemical* quantities (e.g., the concentration of a certain species, the activity of an ion, or the partial pressure of a gaseous material).

There are various major fields of application for chemical sensors: environmental studies, quality control of chemically produced compounds and processed food and biomedical analysis, especially for medical diagnostics. In the field of environmental analysis there is a great demand for the type of continuous monitoring that only a sensor can provide, since the relevant parameter for toxicological risk assessment is always the *dose* (i.e., a concentration multiplied by an exposure time). The study of environmental chemistry also depends upon a continuous data output that provides not only baseline levels but also a reliable record of concentration outbursts. *Data networks* can be used to accumulate multiple determinations of a particular analyte regardless of whether these are acquired simultaneously or at different times within a three-dimensional sampling space (which may have dimensions on the order of miles),

thereby providing information about concentration gradients and other factors essential for pinpointing the location of an emission source. The inevitable exponential increase in the number of chemical analyses associated with this need for maximum information threatens to create serious financial and also ecological problems, since most classical analytical chemical techniques themselves produce a certain amount of chemical waste.

Our society is characterized by a population with an average age that is continuously increasing. This has stimulated the widespread biomedical application of chemical sensors for monitoring the health of the elderly without confining them to hospitals. Sensors for determining the concentrations of such key body substances as glucose, creatinine, or cholesterol could prove vitally important in controlling and maintaining personal health, at the same time enhancing the quality of the patient's life. The same principle underlies sensor applications in intensive care wards. Sensors also offer the potential for significant cost reductions relative to conventional laboratory tests, and they can even improve the quality of the surveillance because continuous monitoring can lower the risk of complications. For example, the monitoring of blood potassium levels can give early warning of the steady increase that often precedes an embolism, providing sufficient time for clinical countermeasures.

It should be noted at the outset, however, that the development costs associated with a chemical sensor are typically many orders of magnitude higher than those for a sensor measuring a physical quantity. A satisfactory return on investment therefore presupposes either mass production in the case of low-cost sensors or special applications that would justify an expensive sensor.

Another important field of application for chemical sensors is *process control.* Here the sensor is expected to deliver some crucial signal to the actuators (valves, pumps, etc.) that control the actual process. Fully automated process control is feasible with certain types of feedback circuits. Since key chemical parameters in many chemical or biochemical processes are not subject to sufficiently accurate direct determination by the human senses, or even via the detour of a physical parameter, sensor technology becomes the key to automatic process and quality control. Often the performance of an entire industrial process depends on the quality and reliability of the sensors employed. Successful adaptation of a batch process to *flow-through reactor production technology*

is nearly impossible without chemical sensors for monitoring the input, the product quality, and also — an increasingly important consideration — the waste. *Municipal incinerators* would certainly be more acceptable if they offered sensor-based control of SO_2, NO_x, and, even more important, the most hazardous organic compounds, such as dioxines. Compared with the high construction costs for a plant complete with effective exhaust management, an extra investment in sensor research can certainly be justified.

These few examples illustrate the increasing importance of chemical sensor development as a key technology in various closed-looped sensor/actuator-based process control applications. Rapid, comprehensive, and reliable information regarding the chemical state of a system is indispensable in many high technology fields; for this reason there is also an increasing need for intelligent or "smart" sensors capable of controlling their own performance automatically.

Finally, there is an urgent and growing demand for analytical chemists and engineers skilled in producing reliable chemical sensors delivering results that are not only reproducible but also accurate (free of systematic errors attributable to interferences from the various sample matrices).

28.2. Chemical Sensors [1]–[9]

28.2.1. Introduction

Until the introduction in the 1980s of the lambda-(λ) probe [10]–[12] for oxygen measurement in the exhaust systems of cars employing catalytic convertors (see Section 28.2.2.3), the prototype for the most successfully applied chemical sensor was certainly the electrochemical pH electrode, introduced over 50 years ago and discussed in detail in Section 28.2.3.1.1. This device can be used to illustrate the identification and/or further optimization of all the critical and essential features characterizing sensor performance, including reversibility, selectivity, sensitivity, linearity, dynamic working range, response time, and fouling conditions.

Electrochemical electrodes for chemical analysis have actually been known for over 110 years (the Nernst equation was published as early as 1889), and they can therefore be considered as the forerunners of chemical sensors. In general, a chemical sensor is a small device placed directly

| Receptor (recognition system) | Transducer | Electronics | Output |
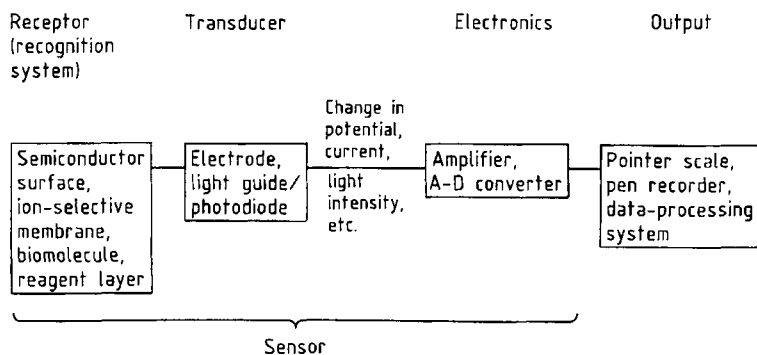
Figure 1. Schematic arrangement of a typical chemical or biochemical sensor [13]

on or in a sample and designed to produce an electrical signal that can be correlated in a specific way with the concentration of the *analyte,* the substance to be determined. If the analyte concentration varies in the sample under investigation, the sensor should faithfully follow the variation in both directions; i.e., it should function in a reversible way, and it should deliver analytical data at a rate greater than the rate of change of the system (quasi-reversible case). From the standpoint of practical applications, complete and fast reversibility is relatively unimportant so long as measurements (even readings of such as from a dosimeter) are acquired with sufficient rapidity compared with changes in the analyte concentration that are to be followed. Another restriction arises

from the demand for a reagent-free measurement: an extremely selective or specific cooperation between the analyte atom, ion, or molecule and the surface of a receptor-like sensor element must then be the sole basis for gaining information about the analyte concentration.

The general construction principle of a chemical sensor is illustrated in Figure 1. The most important part is the *sensing element* (receptor) at which the molecular or ionic recognition process takes place, since this defines the overall selectivity of the entire sensor. The sensing-element surface is sometimes covered by an additional surface layer acting as a protective device to improve the lifetime and/or dynamic range of the sensor, or to prevent interfering substances from reaching the sensing-element surface. The analyte recognition process takes place either at the surface of the sensing element or in the bulk of the material, leading to a concentration-dependent change in some physical property that can be

transformed into an electrical signal by the appropriate transducer. The transducer cannot itself improve the selectivity (exceptions are transducer like electrochemical electrodes which intrinsically deliver an analyte proportional signal), but it is responsible for the sensitivity of the sensor, and it functions together with the sensing element to establish a dynamic concentration range. The electrical signal of the transducer is usually amplified by a device positioned close to the transducer or even integrated into it. If extensive electrical wiring is required between the sensor and the readout – control unit, the electrical signal might also be digitized as one way of minimizing noise pick-up during signal transmission. Several chemical signals produced by multiple sensing elements, each connected to its own specific transducer, can be transmitted via modern multiplexing techniques, which require only a two-wire electrical connection to the control unit also delivering the electrical power ("Bus"-technique in German).

If a biochemical mechanism (mostly enzymatic catalysis, immuno-chemical reaction or complementary DNA hybridization) is used in the molecular recognition step, the sensor is called a *biosensor.* Systems of this type are dealt with in Chapter 28.3.

Key features of chemical sensors subject to specification are, in order of decreasing importance:

1) Selectivity
2) Limit of detection (LOD) for the analyte
3) Accuracy (trueness — a new English term — more important than reproducibility)
4) Sensitivity (slope of the calibration curve)
5) Dynamic response range
6) Stability

7) Response time
8) Reliability (maintenance-free working time)
9) Lifetime

Selectivity. The *selectivity* of a chemical sensor toward the analyte is often expressed in terms of a dimension that compares the concentration of the corresponding interfering substance with an analyte concentration that produces the same sensor signal. This factor is obtained by dividing the sensitivity of the sensing device toward the corresponding interfering substance (= slope of the calibration curve; see below) by the sensitivity for the analyte. Typical selectivities range from 10^{-3} to $< 10^{-12}$ the last extremely small value stands for the sodium error of classical pH-glass-electrodes. In order to estimate the error attributable to limited selectivity the concentration of the interferent must be multiplied by the selectivity coefficient. If, for example, a sodium-selective glass electrode shows a typical selectivity coefficient toward potassium ions on the order of 10^{-3}, a thousand-fold excess of potassium- over sodium-ion concentration would produce an equivalent sensor signal, therefore doubling the reading and leading to an error of 100 %. It should be noted, however, that such selectivities are themselves concentration- and matrix-dependent and often depends on the way they are determined. This means that near the limit of detection of a particular sensor an interfering species will disrupt the measurement of an analyte to a greater extent than it will at higher concentrations. Therefore, response factors can rarely be used for correction purposes.

The selectivity is the most important parameter associated with a chemical sensor since it largely determines the trueness of the analytical method. Trueness (or *accuracy*) is more than simply precision, repeatability, or reproducibility: it represents agreement with the true content of the sample. This is the concentration found independent of the method, the time, the analyst, the laboratory, the country, etc. and is demanded in international trade agreements as comparability. All the other parameters mentioned are subject to systematic error, which can only be recognized if an analyte is determined by more than one method. Since selectivity is always limited, all chemical sensors are prone to report higher concentrations than a sample actually contains. In the case of environmental analysis this positive error can be regarded as a safety margin if relevant interferents are also present. In such a situation the sensor acts as a probe for establishing when a sample should

be taken to be analyzed in the laboratory. The reproducibility of chemical sensor measurements is often well below 5 % relative.

However, it should also be noted that sensing elements are subject to poisoning, resulting in a diminished response for the analyte. Sensor poisons are substances that either cover the sensing surfaces, thereby influencing the analyte recognition process or, in the case of catalytic sensors, decrease the activity of the catalyst. Since these fouling effects are not linear, predictable, or mathematically additive, great care is necessary whenever data from a sensor array are treated by the methods of mathematical data management (e.g., pattern recognition, neural networks, etc.). The influence of sensor poisoning can be completely different and unpredictable for every single sensing element. There is no simple theory that permits the transfer of poisoning effects with respect to one sensing element to other parts of a sensing array as a way of avoiding recalibration. Thus, any attempt to overcome insufficient selectivity by the use of a sensor array with multiple sensor elements displaying slightly differing selectivities for the analyte and the interferents, followed by the application of chemometrics to correct for selectivity errors, is very dangerous in the context of unknown and variable sample matrices. Unfortunately, this is usually the situation in environmental analysis. In the learning and/or calibration phase of sensor-array implementation it is essential that the sample matrix be understood. In the case of a known matrix and a closed system, and if the number of possible interferents is > 5, the number of calibration steps (involving different mixtures of the interfering compounds, selected to cover all possible concentration and mixture variations) required to provide reliable information on the 5-dimensional response space of the sensor array for a $5 + N$ array is extremely large (usually several hundred). With real samples the time necessary for calibrating the sensors might exceed the period of stability. Since with open environmental systems the type and number of interferents is not known, any attempt to correct for inadequate sensor selectivity by use of a sensor array increases the risk of systematic errors. This assertion is valid for every type of mathematical data treatment (pattern recognition, neural networks, etc.) for samples of unknown and/or variable matrices. However, for trace analysis in the sub-ppm range all analytical methods (with the exception of isotope dilution) are subject to systematic errors, as has been demonstrated with in-

terlaboratory comparison tests using even the most sophisticated techniques [14]. Therefore one should not expect an accuracy of better than 1 – 5 % relative from sensor measurements.

Limit of Detection (LOD) for the Analyte. In analytical chemistry the *limit of detection* (LOD) is exceeded when the signal of the analyte reaches at least three times the general noise level for the reading. No quantitative measurement is possible at this low concentration; only the presence of the analyte can be assumed with a high probability (> 99 %) in this concentration range. The range of quantification ends at ten times the LOD. Therefore, the LOD of an analytical instrument must always be ten times smaller than the lowest concentration to be quantified. It is worth mentioning in this context that LOD values reported by most instrument manufacturers disregard the influence of interfering compounds. Furthermore, the actual LOD is often worse than published values with simple mono-analyte solutions when an analyte must be determined within its characteristic matrix (e.g., a pollution site, urine, or blood).

In some cases with extremely sensitive sensors a zero-point calibration might be difficult to perform, because traces of the analyte would always be present or might easily be carried into the calibration process by solvents or reagents. The lowest *measurable* level is then called the *blank value*. The LOD is then defined as three times the standard deviation of the blank value, expressed in concentration units (not in the units of the signal, as is sometimes improperly suggested).

Normally the LOD becomes worse as a sensor ages.

Sensitivity. As mentioned above, the *sensitivity* of a sensor is defined by the signal it generates, expressed in the concentration units of the substance measured. This corresponds to the slope of the corresponding calibration curve when the substance is the analyte, or to the so-called response curve for interferents. With some sensors the sensitivity rises to a maximum during the device's lifetime. A check of the sensitivity is therefore a valuable quality-assessment step. Intelligent sensors are expected to carry out such checks automatically in the course of routine performance tests. Since in most cases the sensitivity depends on such other parameters as the sample matrix, temperature, pressure, and humidity, certain precautionary measurements are necessary to ensure

that all these parameters remain constant both during calibration and in the analysis of real samples.

Dynamic Response Range. The *dynamic response range* is the concentration range over which a calibration curve can be described by a single mathematical equation. A potentiometric sensing device follows a logarithmic relationship, while amperometric and most other electrochemical sensors display linear relationships. Both types of signal-to-concentration relationship are possible with optical sensors; in absorption measurements it is the absorbance with its logarithmic base that is the determining factor, whereas a fluorescence measurement can be described by a linearized function. The broader the concentration range subject to measurement with a given sensing system, the less important are dilution or enrichment steps during sample preparation. The measurement range is limited by the LOD at the low concentration end, and by saturation effects at the highest levels. Modern computer facilities make it possible also to use the nonlinear portion of a calibration curve as a way of saving sample preparation time, but analytical chemists are very reluctant to follow this course because experience has shown that the initial and final parts of a calibration curve are generally most subject to influence by disturbances (electronic and chemical). A good chemical sensor should function over at least one or two concentration decades. Sensors with excellent performance characteristics include the lambda oxygen probe and the glass pH electrode, both of which cover analyte concentration ranges exceeding twelve decades !

Stability. Several types of signal variation are associated with sensors. If the signal is found to vary slowly in two directions it is unlikely to be regarded as having acceptable stability and reproducibility, especially since it would probably not be subject to electronic correction.

Output variation in a single direction is called *drift*. A steadily drifting signal can be caused by a drifting zero point (if no analyte is present) and/or by changes in sensor sensitivity (i.e., changes in the slope of the calibration curve). Drift in the sensor zero-concentration signal (*zero-point drift*) can be corrected by comparison with a signal produced by a sensor from the same production batch that has been immersed in an analyte-free solution.

When drift observed in a sensor signal is attributable exclusively to zero-point drift, and there

has been no change in the sensitivity, a one-point calibration is permissible: that is, either the instrumental zero point must itself be adjusted prior to the measurement, or else the sensor signal for a particular calibration mixture must be recorded. Assuming there is no change in the slope of the calibration curve, this procedure is equivalent to effecting a parallel shift in the calibration curve. If the sensor is routinely stored in a calibrating environment between measurements then such a one-point calibration becomes particularly easy to perform.

Special attention must be directed toward the physical and chemical mechanisms that cause a sensor signal to drift. Differential techniques are satisfactory only if the effect to be compensated has the same absolute magnitude and the same sign with respect to both devices. A consideration often neglected in potentiometry is proper functioning of the reference electrode. Several authors [15]–[17] have suggested that in ion-selective potentiometry a "blank membrane" electrode (one with the same ion-selective membrane composition as the measuring electrode but without the selectivity-inducing electroactive compound) would create the same phase-boundary potential for all interfering compounds as the measuring electrode itself. This is definitely not the case, however. At the measuring electrode, analyte ions establish a relatively high exchange-current density, which fixes the boundary potential. The effect of a particular interfering ion will depend on its unique current–voltage characteristics at such an interface. Therefore, interfering ions may establish different phase-boundary potentials depending on the particular mixed potential situations present at the membranes in question, which means that a differential measuring technique will almost certainly fail to provide proper compensation.

If both the zero point of a sensor and its sensitivity change with time, a two-point calibration is necessary. The combination of a changing slope for the calibration curve and a change in the intercept is generally unfavorable, but this is often the situation with catalytic gas sensors. If the slope of the calibration curve also depends on the sample matrix, only standard addition techniques will help.

The stability of a chemical sensor is usually subject to a significant aging process. In the course of aging most sensors lose some of their selectivity, sensitivity, and stability. Some sensors, like the glass pH electrode, can be rejuvenated, while others must be replaced if certain specifications are no longer fulfilled.

Response Time. The *response time* is not defined in an exact way. Some manufacturers prefer to specify the time interval over which a signal reaches 90 % of its final value after a ten-fold concentration increase, while authors sometimes prefer the 95 % or even 99 % level. The particular percentage value chosen represents a pragmatic decision, since most *signal–time* curves follow an exponential increase of the form:

$$\text{signal} \times \left(1 - e^{-kt}\right)$$

where the true final value is unknown and/or will never be reached in a mathematical sense. Therefore, specification in terms of the time constant k is clearer. This corresponds to the time required for a signal to reach about 67 % of its final value, which can be ascertained without waiting for a final reading: as the slope of the curve ln (signal) vs. time. If the response is constant and independent of the sample matrix, this equation can be used to calculate a final reading immediately after the sensor has been introduced into the sample.

Typical *response times for chemical sensors* are in the range of seconds, but some biosensors require several minutes to reach a final reading. At the other extreme, certain thin-film strontium titanate oxygen sensors show response times in the millisecond range.

The response time for a sensor is generally greater for a decreasing analyte concentration than for an increasing concentration. This effect is more pronounced in liquids than in gaseous samples. Both the surface roughness of the sensor and/or the dead volume of the measuring cell have some influence on the response time. Small cracks in the walls of a measuring cell can function as analyte reservoirs and diminish the rate of analyte dilution.

With certain sensors the response time can also depend on the sample matrix. In the presence of strongly interfering substances the response time for a chemical sensor might increase as a result of an increase in the time required to reach final equilibrium.

Reliability. There are various kinds of reliabilities. One involves the degree of trust that can be placed in an analytical result delivered by a particular chemical sensor. Another is a function of the real time during which the sensor actually

performs satisfactorily without a breakdown and/or need for repair.

There is no way to judge fairly the analytical reliability of a chemical sensor, since this depends strongly upon the expert ability of the analyst to choose a suitable sensor and a suitable sample preparation routine in order to circumvent predictable problems.

With respect to the first generation of sensors, the ion-selective electrodes, the experience of many users has led to the following "reliability hit list", with the most reliable devices cited first: glass pH electrodes, followed by fluoride, sodium glass, valinomycin-based potassium, sulfide, and iodide electrodes, and culminating in the electrodes for divalent ions. In a general way, sometimes very personal judgements regarding reliability have always been linked to a particular sensor's specificity toward the analyte in question.

Reliability can be improved considerably if analytical conclusions are based on measurements obtained by different methods. For example, it is possible to determine sodium either with an appropriate glass-membrane electrode or with a neutral carrier-based membrane electrode. If both give the same analytical results, the probability of a systematic error is very low. Likewise, potassium might be measured using selective membrane electrodes with different carrier molecules. This is of course different from analysis based on a sensor array, in which the selectivities of the individual sensors differ only slightly, and different also in the sense that no correction procedure is invoked in case two results are found to disagree. Use of the Nernst–Nikolsky equation (see Section 28.2.3.1.1), together with the introduction of additional electrodes for determining the main interferents, has been shown to be an effective way of correcting errors in the laboratory with synthetic samples. Since selectivity numbers (i.e., parameters like the selectivity coefficient) are influenced by the sample matrix through its ionic strength, content of surface-active or lipophilic compounds, or interfering ions, it is dubious whether data from real environmental samples should be corrected by this method.

The length of time over which a chemical sensor can be expected to function reliably can be remarkably great (in the range of years), as in the case of glass-membrane or solid-state membrane electrodes, the lambda probe (Section 28.2.2.3), and the Taguchi gas sensor (see Section 28.2.2.2). On the other hand, a biosensor that depends upon a cascade of enzymes to produce an analyte signal will usually have a short span of proper functioning (a few days only). Potentiometric ion-selective membrane electrodes and optical chemosensors have lifetimes of several months. In the case of biosensors it should be noted that anything that changes the quaternary space-orientation of the recognition biomolecules will destroy the proper functioning of the sensor. Enzymes can be influenced by such factors as pH, certain heavy-metal ions, certain inhibitors, and high temperature (resulting in denaturation).

Lifetime. As already mentioned above, some rugged solid-state sensors have *lifetimes* of several years. Certain sensors can also be regenerated when their function begins to deteriorate. The shortest lifetimes are exhibited by biosensors. Ion-selective electrodes and optical sensors based on membrane-bound recognition molecules often lose their ability to function by a leaching-out effect. In optical sensors the photobleaching effect may also reduce the lifetime to less than a year. On the other hand, amperometric cells work well for many years, albeit with restricted selectivities.

Problems associated with inadequate lifetimes are best overcome with mass-produced miniaturized replacement sensors based on inexpensive materials. Minimizing replacement costs may well represent the future of biosensors. Installation of a new sensor to replace one that is worn-out can also circumvent surface fouling, interfering layers of proteins, and certain drift and poisoning problems. The integrated optical system IOS developed at the ICB Muenster and described in Section 28.3.3.2 has shown for the first time how to work with pre-calibrated low-cost immuno-chips with long storage capability.

Comparison of Sensor Data with that Obtained by Traditional Analytical Methods. It is extremely difficult to compare sensor data with traditional data; indeed, generalizations of this type are rarely possible anywhere in the field of analytical chemistry. Sensor developers are often confronted with the customer's tendency to consider use of a sensor only if all else has failed. This means that the most adverse conditions imaginable are sometimes proposed for the application of a chemical sensor. It is also not fair to compare a device costing a few dollars with the most expensive and sophisticated instrumentation available, nor is it appropriate to compare the performance of a chemical sensor with techniques involving time-consuming separations. In this case only

the corresponding detector should be compared with the sensor. Sometimes the very simple combination of a selective chemical sensor with an appropriate separation technique is the most effective way to obtain the redundant data that offer the highest reliability.

Chemical sensors are generally superior to simple photometric devices because they are more selective or faster (as in the case of optical sensors based on the photometric method), more flexible, more economical, and better adapted to continuous sensing. The latter advantage can also be achieved by traditional means via flow-through measuring cells, but this leads to a waste of material and sample, and also to the production of chemical waste. Reagent-free chemical sensors show their greatest advantages in continuous-monitoring applications, in some of which they are called upon to fulfill a control function without necessarily reporting an analytical result, as in the case of the lambda probe described below (Section 28.2.2.3).

28.2.2. Molecular Recognition Processes and Corresponding Selectivities

In any sensing element the functions to be fulfilled include sampling, sample preparation, separation, identification, and detection. Therefore, successful performance in these tasks largely determines the quality of the chemical sensor as a whole. Selective recognition of an analyte ion or molecule is not an easy task considering that today there are more than five million known compounds and a real sample may contain hundreds of potentially interfering compounds. Recognition can be accomplished only on the basis of unique characteristics of the analyte in question. Several different recognition processes are relevant to the field of chemical sensors, ranging from energy differences (in spectroscopy) to thermodynamically determined variables (in electrochemistry), including kinetic parameters (in catalytic processes). The most specific interactions are those in which the form and the spatial arrangement of the various atoms in a molecule play an important role. This is especially true with biosensors based on the complementary (lock-and-key) principle. Here the analyte molecule and its counterpart have exactly complementary geometrical shapes and come so close together that they interact on the basis electrostatic interaction or with weak van der Waals interaction forces. Apart from biomole-

cules, supramolecular chemistry has been given increased attention [18], [19].

28.2.2.1. Catalytic Processes in Calorimetric Devices

Pellistors are chemical sensors for detecting gaseous compounds that can be oxidized by oxygen. A catalyst is required in this case because the activation energy for splitting a doubly-bonded oxygen molecule into more reactive atoms is too high for the instantaneous "burning" of oxidizable molecules. In most cases platinum is the catalyst of first choice because of its inertness. The principle underlying oxidizable-gas sensors involves catalytic burning of the gaseous analyte, which leads to the production of heat that can in turn be sensed by various temperature-sensitive transducers. Often what is actually measured is the increase in electrical resistance of a metal wire heated to an elevated temperature (ca. $300-400\,°C$) by the current flowing through it. However, it is also possible to use more sensitive semiconductor devices or even thermopiles in order to register temperature changes of $< 10^{-4}\,°C$.

In order to understand the selectivity displayed by a pellistor toward various flammable compounds it is necessary to consider the elementary steps in the corresponding catalytic oxidation of analyte at the catalyst surface. Since this is in fact a surface reaction, various adsorption processes play a dominant role. First, oxygen must be adsorbed and chemisorbed, permitting the oxygen double bond to be weakened by the catalyst. Then the species to be oxidized must also be adsorbed onto the same catalyst surface where it can subsequently react with the activated oxygen atoms. The process of adsorption may follow one of two known types of adsorption isotherms, as reflected in the calibration curves for these devices. An equilibrium consisting of adsorption, catalyzed reaction, and desorption of the oxidized product leads to a constant signal at a constant analyte concentration. The sensor response function is influenced by any change in the type or number of active surface sites, since this in turn affects both the adsorption processes and the catalytic efficiency. Likewise, compounds that have an influence on any of the relevant equilibria will also alter the calibration parameters. Especially problematic are strongly adsorbed oxidation products, which lower the turnover rate of analyte molecules and thus the sensitivity of the gas sensor.

Given the sequence of events that must occur when an organic molecule is oxidized by atmospheric oxygen, thereby delivering the heat that is actually to be measured, one can readily understand the importance of changes in the adsorption and desorption equilibria. The selectivity observed with such a calorimetric device arises not because some gas reactions are associated with larger enthalpy changes than others, but rather because those gas molecules that exhibit the most rapid adsorption and desorption kinetics are associated with the highest turnover rates. The latter of course depend on molecular-specific heats of ad- and desorption, which have a major influence on the overall reaction kinetics. A change in these specific heats always results in a corresponding change in sensor selectivity. Consequently, any change in the catalyst material, its physical form, or its distribution within the mostly ceramic pellet will in turn alter the gas selectivity and sensitivity. The same is true for variations in the working temperature. On the other hand, nonreacting compounds can also influence a sensor's response and thereby the calibration function if they in some way affect one of the relevant equilibria and/or the catalytic power of the catalyst. Catalyst poisons such as hydrogen sulfide or organic silicon compounds show a strongly detrimental effect on sensor response.

Summarizing the selectivity characteristics of these calorimetric gas sensors, any gas will be subject to detection if it can be catalytically oxidized with a high turnover rate by atmospheric oxygen at elevated temperature. From a kinetic point of view, smaller molecules like carbon monoxide or methane are favored. In contrast to biosensors there is here no precise molecular "tight-fit" recognition of the geometrical form of the analyte. Thus, the selectivity of this type of gas sensor is rather limited. Differences in the often reaction-rate controlled adsorption and desorption processes for different oxidizable gas molecules are not sufficiently large to allow selective detection of only a single compound. However, this is not necessarily a disadvantage in certain sensor applications, such as detecting the absence of explosive gases (especially important in the mining industry) or carbon monoxide in an automobile garage. In the latter case any positive error resulting from gasoline interference could in fact be regarded as providing a safety margin, since unburned gasoline should be absent from such locations as well.

28.2.2.2. Reactions at Semiconductor Surfaces and Interfaces Influencing Surface or Bulk Conductivities

Introduction. Since the early 1960s it has been known that the electrical conductance of certain semiconductor materials such as binary and ternary metal oxides (e.g., SnO_2 [20], ZnO [21], Fe_2O_3 [22]—all of which are n-semiconductor materials—and CuO or NiO [23]—p-semiconductors) depends on the adsorption of gases on their solid surfaces. The underlying principle here involves a transfer of electrons between the semiconductor surface and adsorbed gas molecules, together with charge transduction in the interior of the material. Typical gases detected by semiconducting devices include oxidizable substances such as hydrogen, hydrogen sulfide, carbon monoxide, and alkanes (SnO_2, ZnO, etc.), as well as reducible gases like chlorine, oxygen, and ozone (NiO, CuO).

In 1967 both SHAVER [24] and LOH [25] described effects achievable with oxide semiconductors modified by the addition of noble metals (e.g., Pt, Pd, Ir, Rh), and since that time the sensitivity and selectivity of semiconductor sensing devices has been significantly enhanced. Intense efforts in this direction, coupled with the further addition of metal oxides [26]–[28], resulted in widespread application of semiconductor gas sensors beginning in the 1970s.

One of the earliest SnO_2 sensors, designed by N. TAGUCHI, is referred to as the "Figaro sensor" (see below). Sensors of this type make it possible to detect as little as 0.2 ppm of an oxidizable compound such as carbon monoxide or methane [29]. Nevertheless, certain details of the associated sensing mechanism are still not fully understood theoretically. An important aim of current research is to overcome limitations of the present generation of sensors, especially instability, irreproducibility, and nonselectivity [30].

Construction and Characterization of Semiconductor Sensors. Semiconductor gas sensors are characterized by their simple construction. A schematic overview of the construction principle of a homogeneous semiconducting gas sensor is provided in Figure 2 A [31]. Sensor operation is based on a change in the surface resistance (or conductance) of an oxidic microcrystalline semiconductor in the presence of interacting gases. A time-dependent record reflecting transient expo-
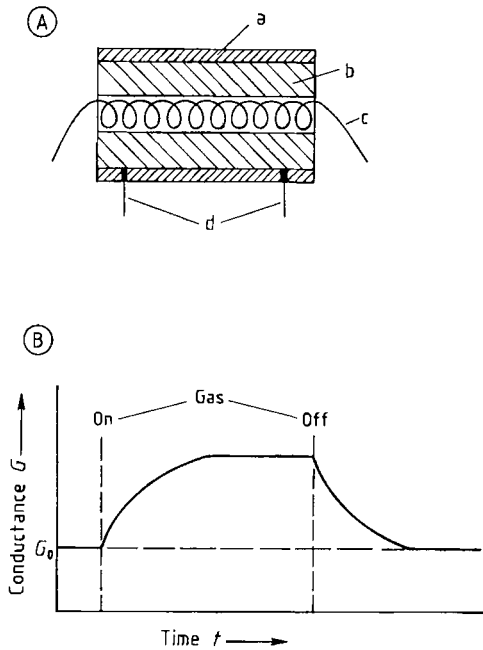
Figure 2. A) Schematic drawing of a homogeneous semi-conducting gas sensor; B) Result of transient exposure of the sensor to a gas that increases the conductance [31]
a) Semiconductor; b) Ceramic; c) Heater; d) Contacts
(with permission from Elsevier, Amsterdam)

Figure 3. Design characteristics of gas sensors (with permission from VDE-Verlag, Berlin)
A) Figaro Type TGS 813 [32]; B) SnO_2 thick-film type [13]
a) Stainless steel screen; b) Sensor element; c) Leads;
d) Heater; e) Epoxy resin; f) Nickel contact; g) Gold; h) Heat contact; i) SnO_2; j) Insulating layer; k) Heater; l) Substrate

sure of a sensor to a gas leading to an increase in conductance is illustrated in Figure 2 B. The sensitivity of a semiconductor sensor is strongly affected by its operating temperature, which is normally in the range 200 – 400 °C. Chemical regeneration of the oxide surface is possible by a reheating process.

A commercial sensor of the "Figaro" type (TGS 813) is shown in Figure 3 A; 400 000 such sensors had already been sold in 1988 [32].

Miniaturization leads to a more modern version of the SnO_2 sensor, normally prepared by thick-film techniques (Fig. 3 B) in which thin SnO_2 films, insulator layers (SiO_2), and integrated heating films are sputtered onto silicon substrates. This approach is compatible with high rates of heating and low-cost production [13].

Working Principles and Theory. The mechanisms responsible for semiconductor gas-sensor operation can be divided into two classes. The first class involves changes in bulk conductance (transducer function), while the second relies on changes in surface conductance (receptor func-

tion). The physical phenomena associated with these two mechanisms are shown schematically in Figure 4 A [33]; Figure 4 B addresses the same problem at the microstructural level [34].

The description of functional principles that follows relates directly to n-semiconductors, but its application to p-semiconductors is straightforward. In a first step, oxygen molecules from the air form a layer of more or less strongly adsorbed (chemisorbed/ionisorbed) oxygen molecules at the surface, resulting in a local excess of electrons. In other words, oxygen acts as a *surface acceptor*, binding electrons from the surface space – charge layer. With respect to the principal energy states (levels) of the electrons in the surface space – charge layer, ionisorption results in a decrease in the electron concentration and an increase in the electronic energy (Fig. 5) [32].

Subsequent reaction with reducing gases (e.g., CH_4 or CO) leads to an increase in charge density (and therefore an increase in conductivity), asso-
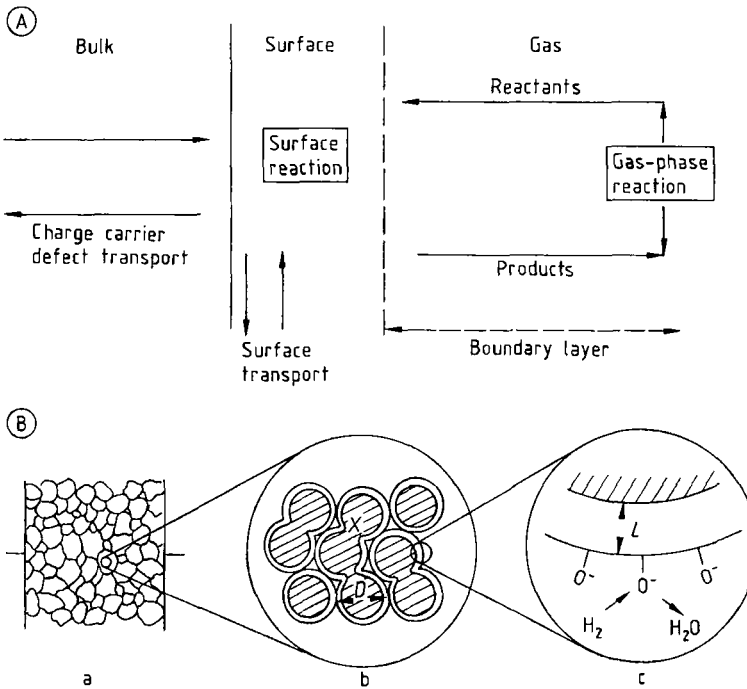
Figure 4. A) Illustration of the physical and chemical phenomena involved in the transduction of a change in the gaseous atmosphere into a change in conductance of a ceramic oxide (with permission from Adam Hilger, Bristol); B) Surface reactions (receptor function) may occur between adsorbed species and defects in the solid (transducer function), between different adsorbed species, or between gas molecules and adsorbed species (or any or all of these together), and the surface reaction may involve a catalyst [33], [34] (with permission from Elsevier, Amsterdam)
a) Element (output resistance change); b) Microstructure (transducer function); c) Surface (receptor function)
D = Particle size; X = Neck size; L = Thickness of space – charge layer

ciated with three possible mechanisms [23], [31], [35]:

1) Adsorption of the reducing molecules as donors, causing electrons to be shifted into the conductance band of the oxide
2) Reaction of the reducing molecules with ionisorbed oxygen under conditions leading to the production of bound electrons
3) Reduction of oxidic oxygen by the reducing molecules, resulting in oxygen vacancies which act as donors, thereby increasing the conductance

Cases 1) and 2) alter the amount of charge stored in the *surface states*, and therefore the amount of charge of opposite sign in deeper parts of the region. For a theoretical derivation of the relationship between the conductance of a semiconducting oxide layer and the composition of the

gaseous surroundings, the following facts must be considered [30], [36]:

Oxygen must be present; that is, these sensors respond only to nonequilibrium gas mixtures containing both combustible gases (CO, hydrocarbons, H_2, etc.) and oxygen.

There exists a temperature of maximum response; that is, the relative change in conductivity upon introduction of a combustible gas increases with increasing temperature, but falls to zero at sufficiently high temperatures.

With respect to the relationship between conductivity (σ) and gas partial pressure, Equation (1) has been found to apply:

$$\sigma \sim p^{\beta} \tag{1}$$

where p is the partial pressure of the combustible gas and β is generally in the range $0.5 - 1.0$ depending on the mechanism.

Oxygen   Atomic hydrogen

Depletion   Enrichment
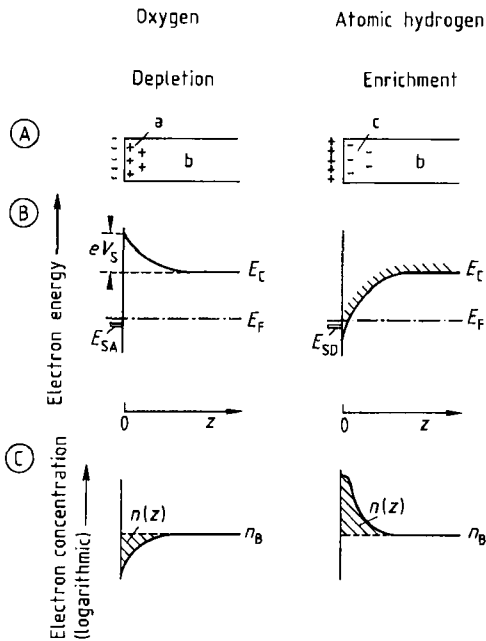
Figure 5. Space – charge layer at the surface of ZnO after exposure to oxygen and hydrogen (with permission from VDE-Verlag, Berlin)
A) Vacuum; B) Band structure in the vicinity of the conducting edge; C) Concentration n(z) of electrons in the conductance band [32]
a) Ionized donors; b) Crystal; c) Electrons
E_{SA} = Acceptor level; E_{SD} = Donor level; n_B = Bulk electron density; z = Distance from surface; eV_S = Surface energy barrier height; E_C = Conductance band edge; E_F = Fermi level

This microscopic model is limited to idealized thin films (10 – 100 nm of, for example, SnO_2). Here the gas – solid interactions can be described in terms of the electronic surface states, which are related to energy levels in the band gap model (Fig. 5). Other surface phenomena involved are lattice defect points, trace impurities, and material segregations near the surface. Crystal dislocations and adsorbed atoms or molecules also play an important role in states at the gas – solid interface. As far as n-semiconductors are concerned, both adsorbed hydrogen and oxygen vacancies can function as surface donors, whereas ionisorbed oxygen is a surface acceptor. A surface potential difference develops as a consequence of negatively charged adsorbed oxygen (O_2^- and O^{2-}) and positively charged oxygen vacancies within the space – charge region below the surface layer. Increasing the amount of negatively charged oxygen at the surface causes the surface potential to increase up to the Fermi level (the highest level

occupied by electrons). This defines the surface potential and represents the surface state, but the precise magnitude of the potential is a function of the oxygen partial pressure. The distribution of the various adsorbed oxygen species depends on the temperature, and is influenced by the presence of hydrogen and other gaseous compounds. With respect to the two possible charged oxygen species, O_2^- and O^-, it can be assumed that only O^- is reactive, and that the rate of interconversion of the species is low compared to the rate of the surface combustion reaction, consistent with the following kinetic scheme [33], [37]:

$$O_2(g) \rightleftharpoons O_2(s) \rightleftharpoons O_2^- \rightleftharpoons O^- \rightleftharpoons O^{2-}$$

with O^- coupled to $2OH^-$ via H_2O, and O^- leading via R to RO.
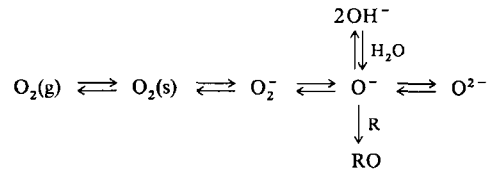
At low temperature the adsorbed species is mainly O_2^-, which is converted into O^- when the temperature is increased above 450 K.

If a combustible gas R reacts with the adsorbed oxygen species, a steady state occupancy Θ of the surface state is established, which is less than the equilibrium occupancy in air. The following mechanism can be assumed:

$$n' + 1/2\, O_2(g) \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} O^-(s)$$
$$R(g) + O^-(s) \xrightarrow{k_2} RO(g) + n'$$
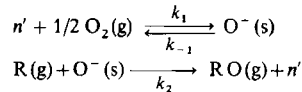
where n' denotes a conductance electron.

Necessary conditions for sensitivity with respect to the partial pressure p of the combustible gas are then:

$$k_{-1} << k_2 p(R) \quad \text{and} \quad k_1 p(O_2)^{1/2} << k_2 p(R) \tag{2}$$

in which case

$$\Theta = \frac{k_1 p(O_2)^{1/2}}{k_{-1} + k_1(O_2)^{1/2} + k_2 p(R)} \tag{3}$$

The third case described above is somewhat different: here the observed effect is related to a change in bulk conductance as a function of stoichiometric changes in oxygen activity in the interior of the crystal lattice. The change in conductivity can be described by the relationship [33]:
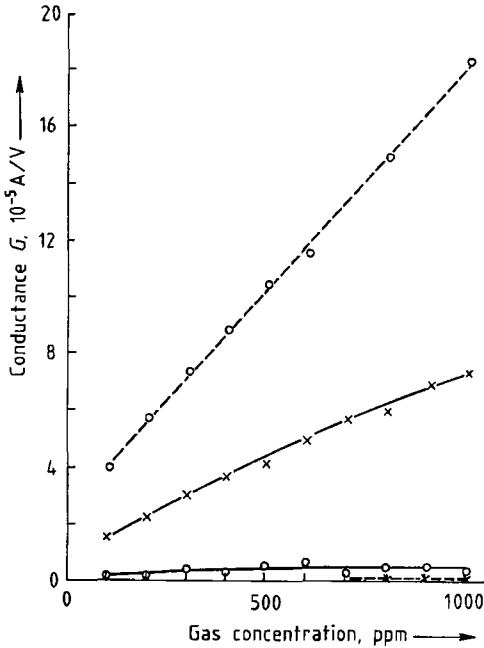
Figure 6. Sensitivity data from commercial SnO_2-based sensors, with sintered layers containing stabilizing additives like SiO_2 (conductance as a function of CO and CH_4 concentration in air); the Type 711 sensor probably operates at a lower temperature and incorporates different catalytically active additives than the Type 813 sensor [31], [38] (with permission from Elsevier, Amsterdam)
----- Sensor Type 711; ——— Sensor Type 813;
x = CH_4; o = CO

$$\sigma = \sigma_0 \exp(E_A/kT) p(O_2)^{\frac{1}{n}} \qquad (4)$$

where σ_0 denotes the electrical conductivity in air, k denotes the Boltzmann constant, T is the temperature in Kelvin, E_A is an activation energy, and the sign and value of n depend on the nature of the point defects arising when oxygen is removed from the lattice. At low temperature ($< 500\,°C$) the rate-determining step is interfacial combustion, whereas oxygen vacancies within the lattice dominate at higher temperatures below ca. $1000\,°C$.

Thus, the observed overall conductance represents a combination of surface effects (both electronic and ionic) together with grain-boundary and volume-lattice effects. The principles elaborated above have been applied widely in the control of oxygen in combustion processes at high temperatures, as in automobile engines.
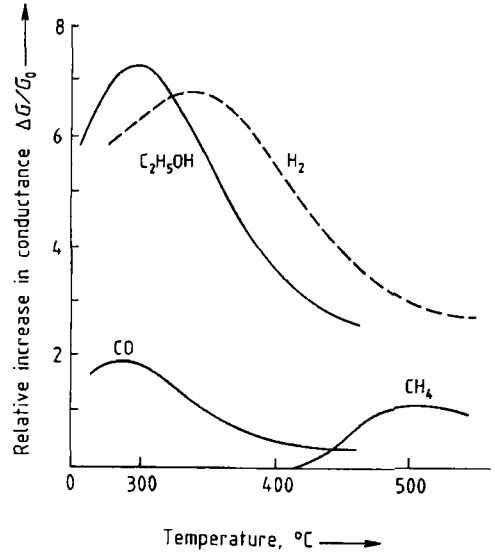
Figure 7. Sensor response to reducing gases (1000 ppm in air) as a function of temperature (sintered layer of SnO_2, thickness 0.05 mm, 0.05 wt % Sb; based on [39], [31]) (with permission from Elsevier, Amsterdam)

Parameters. Under a constant set of conditions (temperature, air, pressure, humidity, etc.) the ratio of the conductance in sample air to that in pure air (G/G_0) is proportional to the analyte concentration (Fig. 6). Apart from the concentration of adsorbed gases, other parameters, both physical and chemical, may also have an influence on the conductance and the optimum working conditions. Considering both components of the conductance change described above, the most important factors influencing conductance are:

1) The microstructure of the semiconducting particles and the surface composition, both of which are characterized by the crystallite size D, the grain-size distribution, and the coagulation structure, all subject to some control through the inclusion of additives.

2) The working temperature of the material, which has a unique optimum (maximum) value for a particular material composition, and must therefore be determined separately for every analyte gas and every substrate composition (Fig. 7).

Tin dioxide has become the favorite material for sensor applications of this type because of its simple preparation and high sensitivity at low operating temperature. Commercially available sensors are usually sintered $\geq 700\,°C$ to ensure sufficient
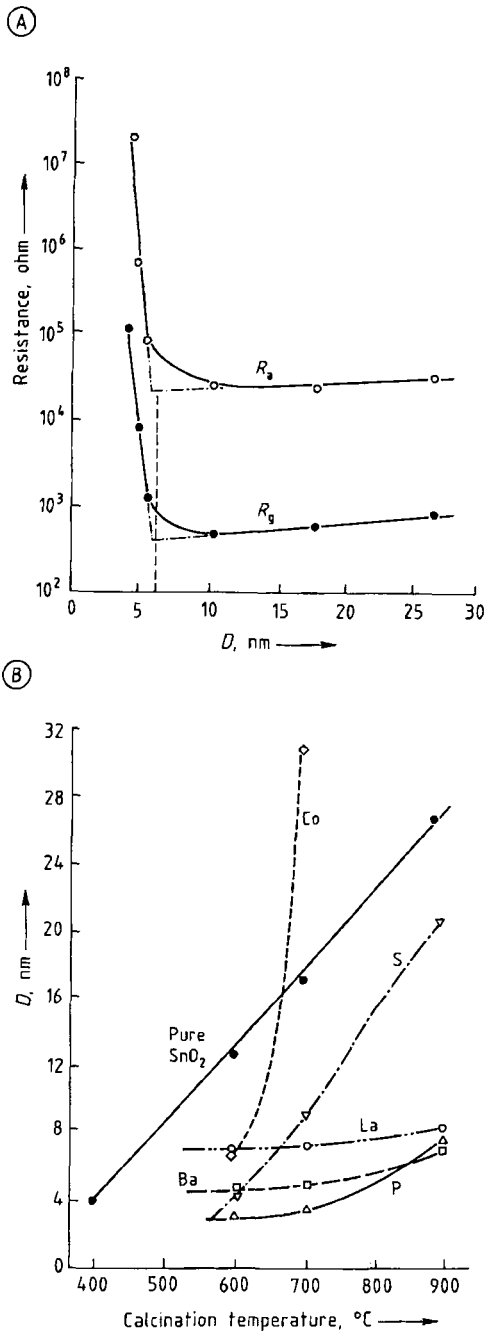
Figure 8. A) Influence of crystallite size (D) on electrical resistance for an SnO_2-based sensor in dry air (R_a) and in air containing 800 ppm H_2 (R_g) at 300 °C (elements sintered at 400 °C); B) Changes in SnO_2 crystallite size as a function of calcination temperature for pure SnO_2 samples and SnO_2 impregnated (5 %) with oxides of various foreign metals [34] (with permission from Elsevier, Amsterdam)

mechanical strength; the cross-sectional diameter of the SnO_2 particles is typically ≥ 20 nm. Unfortunately there is a sharp decrease in gas sensitivity (corresponding to a decrease in the resistance of the SnO_2) when the particle size D increases beyond about 6 nm (Fig. 8 A) [34]. Additives such as barium, phosphorus, or lanthanum oxides can be used to restrict thermal growth of the SnO_2 particles to a limit of ca. 10 nm (Fig. 8 B) [34].

Various approaches have been applied to interpreting the parameters influencing gas sensitivity. Models that relate gas sensitivity to grain size (D) and the geometry of contacts between the SnO_2 particles include the intergrain model and the neck model. An extended approach taking into account both models has been developed by YAMAZOE and MIURA [34], in which gas sensitivity is related to electron charge transfer within the microstructure of the polycrystalline elements (cf. Fig. 4 B).

Selectivity. The selectivity of all semiconductor gas sensors is very limited, but it can be optimized within narrow limits by choosing the best operating temperature and the most appropriate dopant. As in the case of pellistors, overall selectivity is controlled by the ultimate rate-determining step; i.e., the compound that shows the highest oxidation rate determines the selectivity. As is generally the case, the slowest step controls the overall reaction rate. In order to construct a sensor that selectively transforms only the analyte into another product, side reactions must be prevented between the compound that is actually detected (here oxygen) and compounds other than the analyte. This is nearly impossible to achieve in the case of a considerable excess of interfering compounds. Since the physicochemical phenomenon underlying heterogeneous catalysis is based on adsorption–desorption steps and subject to the kinetics of the heterogeneous chemical reaction itself, no known mechanism displays exceptional selectivity. It is not generally possible for a solid-state surface to recognize a single class of molecules because the forces involved in the ad- and desorption steps are rather strong, and are influenced only by overall molecular design and not by details of molecular structure, in contrast to the situation with biocatalysts. It must also be kept in mind that every compound can be considered as a potential interferant, since it might influence the heterogeneous rate constant for catalysis even if it does not itself react with oxygen. This also means that every catalyst poison, such as silicones or even dust blocking the active surface,
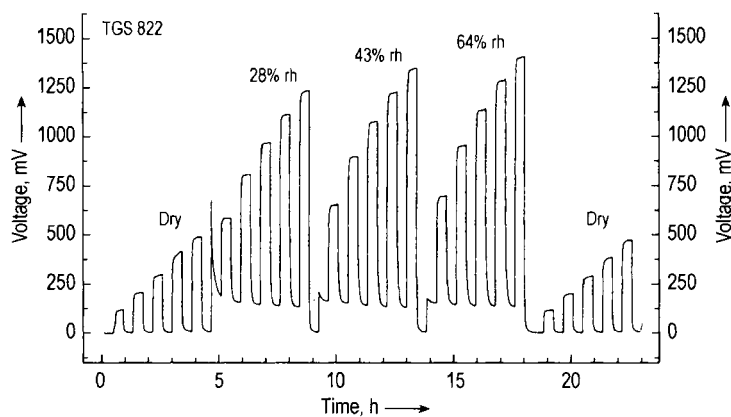
Figure 9. Effect of the relative air humidity on the response of an SnO_2 semiconductor sensor (Taguchi Type TGS 822) on the calibration cycle with ethanol as analyte. The ethanol concentrations were varied from 13, 26, 65 ppm and the relative humidity between dry, 23 %, 43 %, and 64 % rh.

Table 1. Humidity interference of SnO_2-based gas sensors. Analyte, ethanol; relative humidity rh; S = sensitivity /slope of calibration curve at 12.5 ppm ethanol; norm = factor of sensitivity increase in corresponding humid air (values in parenthesis)

Ethanol, ppm	rh, %	S, mV/ppm		
		TGS 800	TGS 813	TGS 822
13	0.00 ± 0.00	14.368 (1.00)	3.568 (1.00)	9.571 (1.00)
26	22.23 ± 0.42	57.752 (4.02)	13.850 (3.88)	47.423 (4.95)
39	42.13 ± 1.18	61.776 (4.30)	15.488 (4.34)	52.851 (5.52)
52	64.43 ± 0.93	63.932 (4.45)	16.801 (4.71)	54.275 (5.67)
65	0.00 ± 0.00	13.978 (0.97)	3.330 (0.93)	9.296 (0.97)

is likely to interfere in a measurement, which in turn limits the possibility of using chemometric methods to correct for errors caused by lack of selectivity.

Humidity effect. Humidity often influences both the zero point of a Taguchi sensor and the sensitivity, which makes corrections difficult as shown in Figure 9.

Table 1 demonstrates that this is the general behavior of semiconductor gas sensors. The slope of the calibration curve can vary by up to a factor of 5. In this table, three different Taguchi sensors (TGS 800, TGS 813, and TGS 822) were carefully compared [257] with respect to their humidity interference using ethanol as the analyte.

This behavior has to be taken into account if these devices are used in electronic noses!

28.2.2.3. Selective Ion Conductivities in Solid-State Materials

The lambda probe, based on selective oxygen ion diffusion through solid yttrium-doped ZrO_2 above 400 °C, is an extremely selective sensor for gaseous oxygen. Frequently used for measuring the residual oxygen content in exhaust gases from internal combustion engines, the lambda probe is primarily a potentiometric device that functions as a concentration cell in which gaseous oxygen is in equilibrium with lattice oxygen in the solid electrolyte (Fig. 10). A higher concentration of oxygen on one side of the membrane leads to a potential difference which, according to the Nernst equation, is proportional to the ratio of the two oxygen concentrations. Likewise the fluoride ion-selective electrode based on selective fluoride-ion diffusion through solid-state (crystalline) LaF_3 is an extremely selective sensor for fluoride ions in solution.

Selectivity in both of these cases is controlled by the size and the charge of the diffusing ion. The
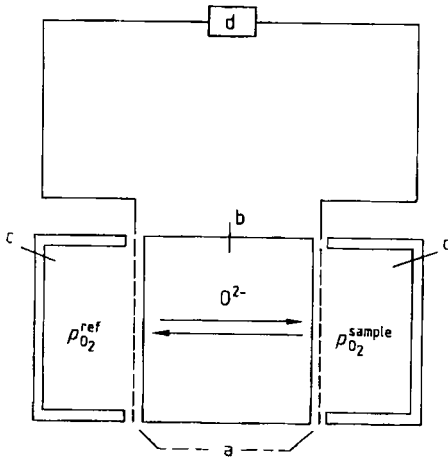
Figure 10. Schematic representation of a lambda probe operated as a potentiometric sensor
a) Porous electrode; b) ZrO_2, doped; c) Oxygen; d) Voltage measurement system

diffusion process takes place via discrete ion jumps to nonoccupied lattice positions and/or from and to interstitial positions. Ion selection here can be regarded as a filtering process, with the additional requirement for an appropriate ionic charge to prevent overall electroneutrality. Because the analyte molecules and ions involved are rather small, they can be separated quite readily from larger molecules or ions. Thus, the solid-state material functions like an ideal permselective membrane. Any solid state material characterized by exceptionally high ionic conductivity, such as Ag_2S (which has extremely high Ag^+-ion conductivity), will be a very good sensor for that ion. In the case of the ion-selective membrane electrodes used as Nernstian potentiometric sensors in solution, however, more detailed consideration is necessary (see Section 28.2.3.1.1).

Since the analyte recognition process is here rather straightforward, and because the filtering effect is accompanied by additional restrictions (appropriate charge and multiple straining), the selectivities obtained through this approach to measuring are extremely high. Many of the compounds that interfere do so not through direct participation in the solid-state diffusion process but rather by affecting the analyte concentration on the sensor surface by other means (e.g., reactions with the analyte, etc.). Thus, with the help of appropriate catalysts and high temperatures the λ-probe can also be used for the detection of other analytes that are oxidized at the surface and thereby reduce the amount of adsorbed oxygen.

28.2.2.4. Selective Adsorption – Distribution and Supramolecular Chemistry at Interfaces

Supramolecular chemistry has been a source of fascinating new molecular recognition processes. At the active center of a biocatalyst (enzyme) and/or in the binding region of a large antibody macromolecule, the recognition process is attributable mainly to a perfect bimolecular approach of complementary structures. Association or binding occurs only if the relatively weak intermolecular forces involved (dipole – dipole and van der Waals forces) are further supported by a uniquely favorable bimolecular approach that encourages interatomic interactions. A selective reaction of this type requires that the host molecule be carefully tailored so that it matches the guest molecule perfectly. Molecular modeling with attention to the activation energies for closest approach is an especially valuable tool in the design of synthetic host molecules for particular analytes. The analyte can be either a charged or an uncharged species. In the case of an uncharged molecule, however, selective binding of the analyte to the host molecule must be detected with the aid of a mass-sensitive transducer or by following the changes of certain optical parameters.

Further research in the field of supramolecular chemistry for sensors seems warranted given the prospect for improved selectivity based on custom-tailored synthetic guest molecules for the selective binding of particular host molecules (analytes). Even differentiation between enantiomers may be achievable. The latter could be demonstrated by employing known stationary phases used in gas chromatography for the separation of enantiomers also on the surface of mass sensitive transducers. During recent years it has been claimed that the technique of molecular imprinting polymers (MIP) is able to perform a similiar high selectivity. While this may be achievable in chromatographic separations due to the high repetition of multiple interface equilibria, it has not yet been convincingly shown for a chemical sensor with a single "hit and recognize" event. The rigidity of the polymer backbone of the imprint makes it highly unlikely that the same recognition power as flexible biomolecules with their "induced-fit" approach can be achieved.

28.2.2.5. Selective Charge-Transfer Processes at Ion-Selective Electrodes (Potentiometry)

An important aim of quantitative analytical chemistry is the selective detection of specific types of ions in aqueous solution. An electrode capable of selectively detecting one type of ion is called an *ion-selective electrode* (ISE). The operating principle of a particular ISE depends upon the nature of the interaction between ions in solution and the surface of that electrode. Consistent with thermodynamic distribution, the ion in question crosses the phase boundary and interacts with the electrode phase after first discarding its hydrate shell (new studies have shown that this may not always be the case in a quantitative sense, since water can be detected in ion-selective PVC membranes), which results in a net flow of ions across the boundary. Ideally, the counterion would not take part in this flow, so a charge separation would develop, causing the counterion to remain in the neighboring border area. This is almost the situation when there will be no net energy gain by this transfer due to some supramolecular binding, and a distribution because of the lipophilicity of counterions should be less pronounced. The resulting electrode potential (difference) between the interior of the electrode and the bulk of the sample solution can be described thermodynamically and kinetically — again, in the ideal case — by the Nernst equation.

The selectivity of such a potentiometric device is that of an ideal Nernstian sensor, the potential of which remains stable irrespective of the fact that some current may be flowing through the sensor interface. An ideal Nernstian sensor is characterized by a current – voltage curve that is very steep and nearly parallel to the current axis.

The selectivity of a potentiometric ion-selective interface is controlled largely by interfacial charge-transfer kinetics and not by the stability constant for reaction of the analyte ion with the electroactive compound (e.g., valinomycin in the case of potassium ion). The exchange-current density for the analyte ion relative to interfering ions determines the extent of potentiometric selectivity. Any factor that increases the former (conditioning of the electrode with the measuring ion, ion-pair formation by lipophilic counterions, etc.) leads to a better and more selective Nernstian sensor. The ideal electroactive compound apparently behaves like a selective charge-transfer catalyst for the analyte ion. Advantages that might perhaps be gained from tailored host molecules for accelerating the transfer of analyte ions into the surface of the electrode have not yet been fully investigated. Recognition of an analyte ion depends upon size and charge as well as the speed with which the ion loses its solvent sphere prior to entering the electrode phase. Therefore, the electroactive compounds most suitable for embedding in the membrane of an ion-selective electrode (usually consisting of PVC containing a plasticizer) are those that play an active role in successive replacement of the often firmly held hydrate shells of the analyte ions. Recent results in the author's institute clearly demonstrate that the thermodynamic complex stability constants determined in homogeneous solution (e.g. via NMR) do not mirror the potentiometric selectivity! A newly synthesized compound for sodium ions, as an example, turned out to yield a very selective PVC membrane for $Pb(OH)^+$ ions [40]. While in theory supramolecular recognition worked perfectly well in homogeneous solutions, at the interface a different mechanism obviously plays a dominant role.

28.2.2.6. Selective Electrochemical Reactions at Working Electrodes (Voltammetry and Amperometry)

Depending on the potential range at the working electrode, as many as about ten different electrode processes can be separated voltammetrically based on the typical resolution of a polarographic curve, which requires a difference voltage of about 200 mV for the half-wave potential and a concentration-dependent limiting current. Selectivity in the process is therefore rather limited, and depends on such thermodynamic parameters as the reaction enthalpy for reduction or oxidation at the electrode surface. The potential of the working electrode determines whether a particular compound (analyte) can be oxidized or reduced, and compounds with similar redox characteristics might interfere.

The problem of interference is of course greatest when extremely positive or negative working electrode potentials are applied at the working electrode, since at high potential the small differences between electrochemical reaction rates become negligible. The "resolution" of a typical current – voltage curve is diminished if the electrochemical reaction is kinetically hindered, resulting in so-called irreversible current steps that are less steep. Irreversible electroactive species require a greater overpotential before the diffusion-controlled limiting current plateau is reached,

therefore increasing the chance that an interfering compound will become electrochemically active. The limited resolution power of voltammetry can be improved by the use of chemically modified electrode surfaces. If a surface layer or a thin membrane is introduced between the phase boundary this additional layer will increase the selectivity if only the analyte can pass through this layer.

In certain biosensors an irreversible heterogeneous reaction is transformed into a homogeneous redox reaction via a very reversible redox system. The latter can be transformed back at the working electrode by a much lower overpotential, thereby reducing the chance of co-oxidation or co-reduction of interfering compounds. Since the limiting current is strictly proportional to the analyte concentration and also to the concentrations of interfering compounds, chemometric data treatment and/or differential measurements may be used to correct for errors.

28.2.2.7. Molecular Recognition Processes Based on Molecular Biological Principles

Metabolic biosensors are based on special enzyme-catalyzed reactions of the analyte. The recognition process, also known as the lock-and-key principle, is extremely selective, permitting differentiation even between chiral isomers of a molecule. Recognition in this case demands a perfect geometrical fit as well as the appropriate dipole and/or charge distribution to permit binding of the analyte molecule inside the generally much larger biomolecule (host). Since fitting is associated with all three dimensions, a very large biomolecule with its stabilized structure is more effective than the smaller host molecules commonly encountered in supramolecular chemistry. The selectivity of a specific biocatalytic reaction can be further improved by ensuring detection of only the transformed analyte or a stoichiometric partner molecule through a chemical transducer located behind the recognition layer. In a certain sense, this type of biosensor represents the highest possible level of selectivity, and comparable results would be difficult to achieve by other means.

Immunosensors based on mono- or polyclonal antibodies constitute the second best choice with respect to high selectivity. Depending on the amount of effort expended in screening and choosing the desired antibodies from among the millions of different antibody molecules available, the selectivity can be remarkably high. On the other hand, it is also possible to choose antibody molecules that bind only to a certain region (epitope) of the analyte molecule, leading to biomolecules capable of recognizing all members of some class of compounds sharing similar molecular structures. This type of sensor, though it cannot be calibrated, is very valuable for screening purposes used to detect the presence of an only certain class of substance . The application of antibody molecules has been considerably simplified by the large-scale preparation of monoclonal versions with precisely identical features. It has even become possible to construct immunosensor arrays based on monoclonal antibody molecules with exactly matched selectivities. Results obtained from such arrays must be evaluated by modern methods of pattern recognition.

In order to overcome the limited stability of large protein molecules, research is currently in progress to isolate only the binding region of the F_{ab} part (antibody-binding fragment) of an antibody molecule. Smaller molecules are of course likely to display somewhat more limited recognition ability, however, with selectivities approaching those of supramolecular host molecules synthesized in the traditional way. Recent research has also focussed on recombinant antibodies produced from large protein libraries without the need for any animal immunization.

28.2.3. Transducers for Molecular Recognition: Processes and Sensitivities

28.2.3.1. Electrochemical Sensors

Electrochemical sensors constitute the largest and oldest group of chemical sensors. Although many such devices have reached commercial maturity, others remain in various stages of development. Electrochemical sensors will be discussed here within the broadest possible framework, with electrochemistry interpreted simply as any interaction involving both electricity and chemistry. Sensors as diverse as enzyme electrodes, high-temperature oxide sensors, fuel cells, and surface-conductivity sensors will be included. Such sensors can be subdivided based on their mode of operation into three categories: *potentiometric* (measurement of voltage), *amperometric* (measurement of current), and *conductometric* (measurement of conductivity) [13], [41].

Electrochemistry implies the transfer of charge from an electrode to some other phase, which may be a solid or a liquid. During this process chemical

changes take place at the electrodes, and charge is conducted through the bulk of the sample phase. Both the electrode reactions themselves and/or the charge transport phenomenon can be modulated chemically to serve as the basis for a sensing process. Certain basic rules apply to all electrochemical sensors, the cardinal one being the requirement of a closed electrical circuit. This means that an electrochemical cell must consist of at least two electrodes, which can be regarded from a purely electrical point of view as a sensor electrode and a signal return.

Another important general characteristic of electrochemical sensors is that charge transport within the transducer portion of the sensor and/or inside the supporting instrumentation (which constitutes part of the overall circuit) is always electronic. On the other hand, charge transport in the sample under investigation may be electronic, ionic, or mixed. In the latter two cases electron transfer takes place at the electrode–sample interface, perhaps accompanied by electrolysis, and the corresponding mechanism becomes one of the most critical aspects of sensor performance.

The overall current – voltage relationship is complex, and it can vary as the conditions change. The relationship is affected primarily by the nature and concentration of the electroactive species, the electrode material, and the mode of mass transport. A total observed current can be analyzed in terms of its cathodic and anodic components. If the two currents are equal in magnitude but opposite in sign, there must be an *exchange current* passing through the electrode-surface/sample interface.

Both the charge-transfer resistance and the exchange-current density are critically important parameters in the operation of most electrochemical sensors, since they reflect the kinetics of an electrode reaction. These parameters are directly proportional to the electrode area, so the smaller the electrode the higher will be the resistance, all other parameters remaining unchanged. Therefore, it is the electrochemical process taking place at the smaller electrode that determines the absolute value of the current flowing through the entire circuit. The auxiliary electrode will begin to interfere only if its charge-transfer resistance becomes comparable to that of the working electrode. Microelectrodes represent one approach to avoiding such interferences (see Section 28.2.3.1.2).

On the other hand, in zero- current potentiometry the relative sizes of the two electrodes is immaterial. Acquiring useful information in this case

requires only that the potential of the working electrode be measured against a well-defined and stable potential from a reference electrode. Any foreign potential inadvertently present within the measuring circuit can contaminate the information, so it is mandatory that the reference electrode be placed as close to the working electrode as is practically possible. Thus, in an amperometric (or conductometric) measurement the source of information can be localized by choosing a small working electrode, whereas it cannot be localized with zero- current potentiometric measurements.

28.2.3.1.1. Self-Indicating Potentiometric Electrodes

Fundamental Considerations. Within the context of this chapter "potentiometry" is understood to mean the measurement of potential differences across an indicator electrode and a reference electrode under conditions of zero net electrical current. Such a measurement can be used either for determining an analyte ion directly (direct potentiometry) or for monitoring a titration (see below).

In recent years potentiometry has proven to be well suited to the routine analysis of a great number and variety of analytes [5], [40], [42]–[60]. Ion-selective electrodes (ISEs) are commercially available for many anions and cations, as indicated in Tables 4–7. Other analytes can be determined using ISEs in an indirect way. Chapter 28.3 deals with various types of potentiometric biosensors. A special advantage of ion-selective potentiometry is the possibility of carrying out measurements even in microliter volumes without any loss of analyte.

At present, most ISEs unfortunately do not provide absolute selectivity for a single ion. This fact requires that one have access to very detailed information regarding the nature of samples to be analyzed so that interferences can be either eliminated or otherwise taken into account.

The theory of (ion-selective) potentiometry has already been introduced in Section 28.2.2.5. Here it is necessary only to remind the reader of the Nernst equation and the Nernst–Nikolsky equation, both of which are used in the evaluation of potentiometric measurements.

The *Nernst equation* is valid only under ideal conditions (with no interfering ions, etc.):

$$E = E^0 + (RT/zF) \ln a_M \tag{5a}$$

$$= E^0 + (0.059/z) \log a_M \tag{5b}$$

The *Nernst–Nikolsky equation* on the other hand takes into consideration the influence of interfering ions on the potential of an ISE:

$$E = E^0 + (RT/z_M F)\ln \left[a_M + \sum K_{MI}(a_I)^{z_M/z_I} \right] \tag{6}$$

where

E = potential of the ISE, measured with zero net current
E^0 = standard potential of the ISE
R = gas constant ($8.314 \ J \ K^{-1} \ mol^{-1}$)
T = absolute temperature (K)
F = faraday constant ($96\,485$ C/mol)
a_M = activity of the ion to be measured
a_I = activity of an interfering ion
z_M, z_I = electrical charges of the measured and interfering ions
K_{MI} = selectivity coefficient

The value of K_{MI} depends both on the activity a_M and on the particular combination of analyte ion and interfering ion. Recently BAKKER [48] suggested another equation which described the EMF versus concentration behavior of an ISE with a higher accuracy. The selectivity coefficients are determined here without conditioning the electrode with the measured ion, which yields constant selectivity coefficients. The new equation describes the situation when a monovalent analyte ion I^+ is interfered by a divalent ion J^{2+} especially well, and *vice versa*:

$$E = E_I^0$$
$$+ \frac{RT}{F} \ln \left(\frac{a_I(IJ)}{2} + \frac{1}{2}\sqrt{a_I(IJ)^2 + 4a_J(IJ)(K_{IJ}^{pot})^2} \right) \tag{6a}$$

for $z_I = 2$ and $z_J = 1$

$$E = E_I^0$$
$$+ \frac{RT}{F} \ln \left(\sqrt{a_I(IJ) + \frac{1}{4}K_{IJ}^{pot}a_I(IJ)^2} + \sqrt{\frac{1}{4}K_{IJ}^{pot}a_I(IJ)^2} \right) \tag{6b}$$

HORVAI [49] developed another more general equation for ion-selective ion-exchange membranes, which also considers the concentration dependency of the selectivity coefficient:

$$E \cong c_I^{z} + \frac{\varrho}{\zeta_I \Phi} \ln \alpha_I(IJ) + \frac{\varrho}{\zeta_I \Phi} \varkappa_{IJ}\alpha_I(IJ)^{\zeta_J/\zeta_I}\alpha_J(IJ) \tag{6c}$$

One can see here, separated from each other, the response of the ISE towards the pure analyte ion I and (as the last term) the EMF deviation caused by the interfering Ion J. Note that there is no logarithm in the last term and also that the last term depends both on the interfering and primary ion concentrations.

Types of Ion-Selective Electrodes. Simple metal-ion electrodes such as those based on silver, gold, platinum, etc., will not be dealt with in this discussion since they are used mainly for indicating purposes in potentiometric titrations. Apart from these, the following types of ISEs can be distinguished:

1) Glass-membrane electrodes
2) Solid-membrane electrodes
3) Liquid-membrane electrodes (including PVC-membrane electrodes)
4) ISEs based on semiconductors (ion-selective field-effect transistors, etc.)

These four types will be the subject of more detailed consideration.

Glass-Membrane Electrodes [50]. The best-known glass-membrane electrode is the pH electrode. Its most important component is a thin glass membrane made from a sodium-rich type of glass. Depending on the intended application the membrane may take one of several geometric forms: spherical, conical, or flat. For applications in process streams, special pH electrodes have been devised that are resistant to high pressures, and sterilizable pH electrodes are available for biotechnological applications.

Examples of the different types of pH electrodes are illustrated in Figure 11. The *spherical* type is most frequently used for the direct measurement of pH or for acid–base monitoring. It is robust and appropriate for most routine applications. *Conical* electrodes can be used as "stick-in electrodes" for pH measurements in meat, bread, cheese, etc. A conical membrane can easily be cleaned, which is important if measurements are to be made in highly viscous or turbid media. A *flat* membrane facilitates pH measurements on surfaces, such as on human skin.

The pH electrode must be combined with a reference electrode [51] that provides a constant potential for completing the electrical circuit. A suitable reference system is frequently integrated into the body of the pH electrode itself. Combined electrodes of this type are particularly easy to
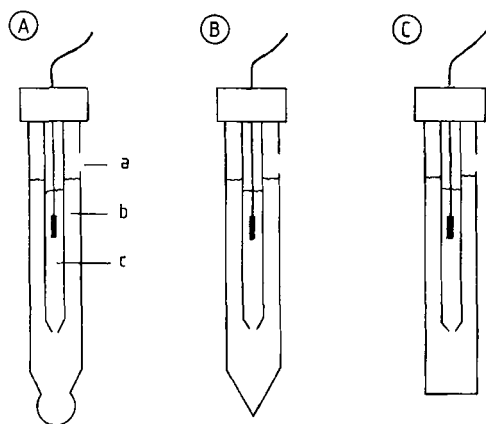
Figure 11. Various types of glass-membrane pH electrodes
(with permission from Springer-Verlag, Heidelberg)
A) Spherical membrane; B) Conical membrane; C) Flat
membrane
a) Fill hole; b) Inner filling solution (buffer); c) Reference
system

handle. pH Measurements at $>$ pH 10 by means of
glass electrodes are subject to interference by al-
kalimetal ions, especially Na^+. The selectivity co-
efficient K_{MI} for interference by the sodium ion is
10^{-13} in special lithium containing glasses with a
reduced alkali error. This means that a 10^{13}-fold
concentration of sodium is required to produce the
same potential as a given concentration of protons.

It is important that the surface of a glass mem-
brane be swollen before measurements are taken.
Prior to first use the electrode should therefore be
allowed to swell overnight in a 3 mol/L KCl so-
lution. For the same reason the electrodes should
also be stored in a solution of this type between
measurements and after measurements have been
conducted in nonaqueous solvents. Here, the
choice of a double junction reference electrode
with the same solvent as in the measuring solution
and added organic salt for sufficient conductivity
is important. In the case of endpoint detection a
platinum/ferrocene redox element without a liquid
junction could also be used.

Apart from the pH-selective glass-membrane
electrodes, other commercially available glass-
membrane electrodes are responsive to sodium
ions, although the selectivity of a pNa electrode
is much lower than that of a pH electrode. Approx-
imate selectivity coefficients K_{MI} in this case are
10^3 for interference by H^+ and 10^{-3} for interfer-
ence by Li^+, K^+, and Cs^+. However, an interfer-

ence by H^+ ions is seldom serious since by appro-
priate pH adjustment one can easily shift the H^+
ion concentration to a range where it does not
influence the electrode potential.

Solid-Membrane Electrodes. The sensing ele-
ment of a solid-membrane electrode consists of a
material showing ionic conductivity for the partic-
ular ions that are to be determined. Such a device
may be manufactured from a single crystal, as in
the well-known fluoride-selective electrode,
where the sensing element is a single crystal of
lanthanum fluoride doped with europium(II) flu-
oride to lower the electrical resistance. Alterna-
tively, a membrane can be produced by grinding
the crystalline sensor material together with an
appropriate additive (PTFE powder, silicone
rubber), followed by solidification either by appli-
cation of high pressure or by addition of a cross-
linking agent. It is thus essential that the crystal
particles stay in contact which each other. In ad-
dition to these so-called heterogeneous mem-
branes, the homogeneous membranes based on
silver sulfide are better known. They consist of
Ag_2S which is freshly precipitated, washed, dried,
and pressed into a pellet with a traditional KBr-IR
press. Since Ag_2S is rarely stoichiometric it shows
a high Ag^+ ion conductivity leading to a perfect
Nernstian-sensor for that ion. Because the standard
exchange current density at its suface is much
higher than on a silver metal surface (overpotential
due to crystallization polarization) it senses lower
Ag^+ ion concentrations much better. By incorpo-
rating about 30 weight % of a certain silver halide
into this pellet one obtains the corresponding hal-
ide sensitive membrane. Note, the same result
would be obtained if the appropriate silver halide
were to be added as a powder to the measuring
solution. That means that it is still the Ag^+ ion
which determines the potential via the correspond-
ing solubility product. If certain metal sulfides
such as CuS, PbS, or CdS are added the membrane
senses those metal ions via two solubility products
and an intermediate S^{2-} ion-concentration. Thus,
Ag^+ ions must be totally absent as well as com-
pounds oxidize S^{2-}.

The membrane is then fixed to the electrode
body. Electrical contact is established either by a
conducting adhesive or by an internal reference
system. Figure 12 illustrates schematically the
construction of two types of solid-membrane elec-
trodes. A detailed compilation of commercially
available solid-state electrodes is provided in Ta-
bles 4 – 7 at the end of this section.
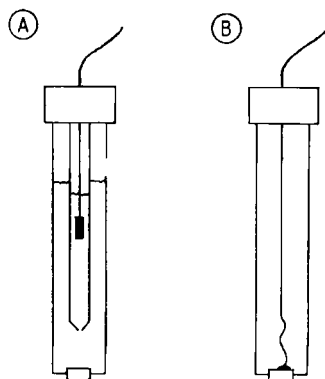
Figure 12. Two types of solid-membrane electrodes (with permission from Springer-Verlag, Heidelberg)
A) With internal filling solution and reference system;
B) With solid contact

Figure 13. Two types of liquid-membrane electrodes (with permission from Springer-Verlag, Heidelberg)
A) Porous-membrane electrode; B) PVC-membrane electrode
a) Reservoir for liquid ion carrier; b) Porous membrane;
c) Reference system; d) PVC membrane; e) Reference system

Generally speaking, interference must be anticipated from any ion that tends to react with the membrane material to form compounds less soluble than the membrane itself. Thus, the chloride electrode, which consists of a $AgCl/Ag_2S$ membrane, is poisoned by mercury(II) ions. Due to the solid state of the membrane the reaction rate is rather low, so a poisoned membrane can be repaired by carefully polishing it with an appropriate polishing powder (as described in instruction manuals furnished with the electrodes).

Liquid-Membrane Electrodes. The electroactive phase in this type of electrode consists of a liquid that has been stabilized mechanically. For this purpose the electroactive components may be dissolved in a highly viscous and apolar organic solvent such as one of the higher alcohols or paraffin. A porous material (e.g., commercial membrane filters such as Sartorius cellulose nitrate with pore diameters of about 0.45 μm), are initially made hydrophobic with, e.g., hexamethyl disilazane and then soaked with this solution and introduced in place of the solid membrane in an electrode of the type described above. More frequently, however, the liquid is not fixed in a porous material, but rather embedded into a high molecular mass polymer like PVC. In this case the plasticizer (normally present at a concentration of about 66 % by weight) acts as a solvent for the electroactive component (the concentration, which seldom exceeds 1 %!). Simplified construction schemes for these two types of electrodes are illustrated in Figure 13.

For applications like those discussed in Sections 28.2.2.4 and 28.2.2.5 the electroactive component of a liquid-membrane electrode would consist of a (charged) ion exchanger or an (uncharged) ion carrier. In both cases the electroactive component must have a rather high affinity for the ion to be determined, but at the same time display the greatest possible selectivity with respect to interfering ions. Table 2 lists ion exchangers and ion carriers frequently used in liquid-membrane ISEs. For an excellent and almost exhaustive compilation of ISEs based on this principle see the reviews of BAKKER, BÜHLMANN, and PRETSCH [59], [60].

The first part of these publications deals with the most recent theoretical treatment while the second part describes in detail nearly all known membrane compositions for nearly every ion of particular importance. It should be noted that PVC-based ISEs are especially easy to construct in the laboratory just before they are to be used. The Fluka company offers a large collection of electroactive compounds (including the most famous ones with their characteristic ETH number), the right PVC and plasticizer and also construction information. Recently reasons for the typical detection limits in the range of μM solutions have been found [61]. If the internal filling solution contains too high a concentration of the measured ion it will be distributed through the entire membrane and can be exchanged with interfering ions at the measuring surface and thereby produce a certain concentration in the sample solution. Thus, in case of trace determinations the inner concentration of the measured ion should be < μM.

Table 2. Example of ion exchangers and ion carriers frequently encountered in liquid-membrane electrodes

Ion	Electroactive compound
Cations	
K^+	valinomycine
NH_4^+	nonactine – monactine
Li^+	lithium carrier, e.g., ETH 1644 (Fluka)
Ca^{2+}	calcium salt of didecylphosphonic acid
Anions	
NO_3^-	$Ni(o\text{-phen})_3^{2+}$
ClO_4^-	$Fe(o\text{-phen})_3^{2+}$
BF_4^-	$Fe(o\text{-phen})_3^{2+}$

However, care is needed not to loose the internal measured ion by adsorption processes. Thus, an appropriate ion-buffer leading to a constant low concentration is recommended.

For anion determinations, most PVC-membrane based ISEs follow the so-called Hofmeister series, which is: the more lipophilic anion is always more favored by the organic membrane and therefore sensed better. However, there are exceptions to this general rule which are also most interesting from a theoretical point of view, especially when a supramolecular carrier is used and more than the typical one lipophilic salt addition is performed. Because of this, owing to the incorporation of a lipophilic cation and anion, a carbonate selective ISE could be constructed with decreased interference from salicylate by five orders of magnitude [62]. This addition surprisingly increases the charge transfer resistance of the ISE so that it can be measured besides the high bulk resistance in series. However, the ratio of the transfer resistances for the analyte and the interfering ions changes too, in favor of the measured ion!

ISEs Based On Semiconductors (ISFETs). Devices based on ion-selective field-effect transistors (ISFETs) are sufficiently distinctive to warrant separate treatment in Section 28.2.3.1.4.

Instrumentation. The discussion here of instrumentation required for potentiometric measurements will be limited to a brief overview of suitable reference electrodes [51]. The conventional reference electrode consists of an electrode of the second kind, like the Ag/AgCl/KCl system. Much effort has been devoted to the development of miniaturized reference systems suited to direct integration into ISFETs, but most systems of this type fail to fulfill all the requirements for a "true" reference electrode with a well-defined and stable

potential difference relative to the standard hydrogen electrode. The greatest problem is the short distances needed for the liquid junction, leading to an influx of unwanted compounds within too short a time period. Reference electrodes with gel-stabilized internal electrolytes are produced for applications in process analysis.

Table 3 provides an overview of the most frequently used reference electrodes.

Evaluation of Potentiometric Results. The aim of all methods for the evaluation of potentiometric data is the computation of substrate activities from voltage changes observed between an ISE and a reference electrode. Generally speaking, three different approaches are possible: (1) the direct potentiometric measurement of activities, (2) monitoring during the course of a titration, and (3) flow-injection analysis (FIA).

A determination by *direct potentiometric measurement* is accomplished either by calibrating the electrode with solutions of known concentration, or by using the techniques of standard addition or standard subtraction. Since calibration with standard solutions usually involves solutions of pure salts, thorough knowledge of the samples to be analyzed is a prerequisite to avoiding erroneous results due to interfering ions present in the sample matrix. Influences attributable to other substances present in the sample that do not interfere directly with the determination but produce changes in ionic strength can be minimized by adding a total ionic strength adjusting buffer ("TISAB solution") to both standard solutions and samples. Standard-addition or -subtraction methods should be invoked whenever it is not possible to make appropriate standard solutions (owing to a very complex matrix, for example). Figure 14 illustrates the evaluation of a standard-addition analysis.

Monitoring a titration by means of an ISE has the advantage over direct potentiometric measurement that the accuracy is determined largely by the titration reaction and not simply by the calibration function of the electrode. This is especially important for ISEs with changing or unknown slopes. For this reason the determination of a surfactant should always be carried out as a titration and not as a direct measurement.

Flow-Injection Analyses with ISEs are especially easy to carry out. If the ISE has a flat surface it can be connected to the liquid junction of a reference electrode via a small strip of filter paper. This paper strip is then held by a clamp and the carrier stream of the FIA system is applied just

Table 3. Common reference electrodes

Electrode	Potential vs. NHE	Useful temperature range, °C	Interferences
$Hg/Hg_2Cl_2/KCl$ (0.1 mol/L)	+0.334 V		
$Hg/Hg_2Cl_2/KCl$ (1.0 mol/L)	+0.280 V	15–70	complexing agents (CN^-), S^{2-}, strong oxidizing or reducing agents
$Hg/Hg_2Cl_2/KCl$ (satd.)	+0.241 V		complexing agents (CN^-), S^{2-}, strong oxidizing or reducing agents
$Ag/AgCl/KCl$ (0.1 mol/L)	+0.290 V		complexing agents (CN^-), S^{2-}, strong oxidizing or reducing agents
$Ag/AgCl/KCl$ (1.0 mol/L)	+0.222 V	15–110	complexing agents (CN^-), S^{2-}, strong oxidizing or reducing agents
$Ag/AgCl/KCl$ (satd.)	+0.197 V		complexing agents (CN^-), S^{2-}, strong oxidizing or reducing agents
$Hg/Hg_2SO_4/K_2SO_4$ (0.5 mol/L)	+0.682 V	15–70	complexing agents (CN^-), S^{2-}, strong oxidizing or reducing agents
$Hg/Hg_2SO_4/K_2SO_4$ (satd.)	+0.650 V		
$Tl(Hg)/TlCl/KCl$ (3.5 mol/L)	−0.575 V	15–120	S^{2-}, strong oxidizing agents; no interference from complexing agents
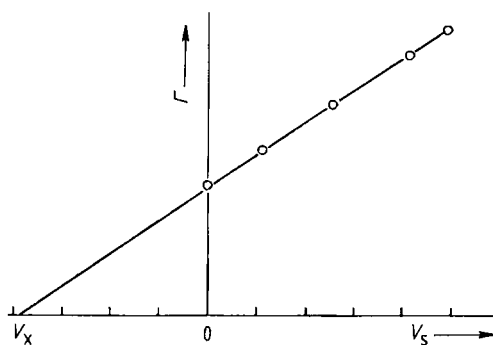
Figure 14. Plot for the evaluation of a standard-addition analysis
V_S = Volume of standard solution; V_x = Volume of analyte
$\Gamma = pE \cdot V_{total}$, with $pE = 10^{\Delta E \cdot (z/S)}$ (ΔE = Potential difference; z = Charge of the ion, including sign; S = Slope of the electrode, including sign)
The unknown concentration of the analyte c_x is
$c_x = (-c_s \cdot V_x)/V_{initial}$, where c_s = Concentration of the standard solution and $V_{initial}$ = Initial volume of the analyte

before the measuring ISE, passing over its surface and later over the reference electrode. New levels of ultra trace analysis can be reached — as already mentioned — by never conditioning the ISE membrane with higher concentrations of the measured ion either from the outside or from the inside! This automatically leads to a differential FIA operating mode in which the injected sample segment is first passed over one of the ISE surfaces and then with a matched retarding coil to the other side of the ISE membrane. Figure 15 shows results and the experimental set-up with two external reference electrodes, the ISE membrane and the redarding

coil which should direct the sample segment to the second site of the ISE-membrane exactly at the moment it leaves the first side [63]. In ultra trace analysis the carrier solution should contain only traces of the measured ion in order to get a stable voltage reading. Note that the injection volume in Figure 15 is in the µL range. Thus, such a simple and low cost approach allows the determination of pg amounts as it is demonstrated in Figure 16.

Commercially Available ISEs. Tables 4–7 provide an overview of ISEs commercially available at the time. ISEs described only in the literature, and which must therefore be selfconstructed, have not been included in the tables. Instructions for the preparation of PVC-membrane elecrodes can be found in the literature [4], [5], [15].

28.2.3.1.2. Voltammetric and Amperometric Cells

Information is obtained with this type of electrochemical sensor from either the combined current/potential – concentration relationship (voltammetry) or from the current – concentration relationship alone (amperometry).

A voltammetric measurement is accomplished by scanning the potential difference across an electrochemical cell containing a working electrode, a reference electrode and usually an auxiliary electrode separated from the sample solution with a diaphragma from one preset value to another and recording the cell current as a function of the applied potential. A curve so generated is known as a *voltammogram*. Amperometric meas-
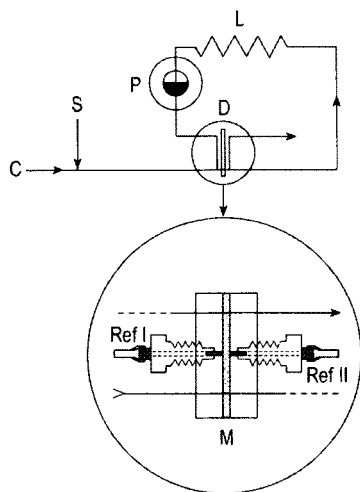
Figure 15. Differential flow-injection analysis with one single ISE membrane and two reference electrodes in an electrolytic closed loop.
An injector block used in liquid chromatography injects about 500 µL of sample into the carrier stream. The volume of the carrier stream loop from one side of the ISE membrane to the opposite side including the volume of both membrane sides is also approximately 500 µL. Thus, when the sample segment is fully washed out of the first membrane side the first front of it reaches the backside. Through this a positive and a negative analyte peak are obtained, which look like a second derivative of a peak.
C = carrier solution; S = sample (introduction with injector loop); P = pump; L = loop adapted to different sample volumes; D = detector; M = ISE membrane with two spacers with a flow channel on each side; Ref I and Ref II = commercial reference electrodes with threads

urements are made by recording the current flow in a cell at a single applied potential.

The essential operational function of a transducer for this purpose is the transfer of one or more electrons to or from the transducer surface (= working electrode). The flow of these electrons is what constitutes the output signal. Voltammetric and amperometric devices are also capable of conferring a degree of selectivity on the overall sensing process by the proper choice of the working electrode potential against a reference electrode.

Furthermore they are relatively simple: in the most elementary case, the transducer (electrochemical cell) consists of nothing more than two electrodes immersed in a suitable electrolyte. A more complex arrangement might involve the use of a three-electrode cell mentioned above, with the advantages that the reference half-cell is not disturbed (polarized) by the current flowing

and that the ixR drop can be compensated by such a set-up.

In spite of the inherent diagnostic advantages of voltammetry, a transducer based on this technique represents a rather cumbersome approach to sensing, mainly because of the electronic circuitry required to scan the applied potential, the time needed for a single scan, and the evaluation algorithm identifying the correct current peak to be measured. Accordingly, most sensing applications involve cells operating in an amperometric (fixed potential) mode.

Voltammetry. When a slowly changing potential is applied to an electrode immersed in an electrolyte solution containing a redox species, a current will be observed to flow as soon as the applied potential reaches a certain value. This current arises from a heterogeneous electron transfer between the electrode and the redox couple, resulting in either oxidation or reduction of the electroactive species. At a sufficiently oxidizing or reducing overpotential the magnitude of the current may become a function of mass transfer of the redox species to the electrode. In a well-stirred solution so-called quasi-stationary current – voltage curves with typical current steps and plateaus are obtained. The half-step (wave) potential is characteristic of the type of species being electrochemically oxidized or reduced, the height of the current steps represents the concentration of it. Without stirring, current peaks are obtained, since the diffusion towards the working electrode surface is no longer sufficient to maintain a constant current. Redox couples that give rise to symmetrical current peaks separated by $58/z$ mV (at 25 °C) with cyclic voltage ramps in unstirred solutions are frequently referred to as *reversible couples*. If the rate of electron transfer between the redox couple and the electrode is high compared with the rate of mass transfer, then the electrode reaction is reversible. Under these circumstances the concentration ratio for oxidized and reduced forms of the couple at the electrode surface is described by the Nernst equation [Equations (5a) and (5b)].

If electron transfer between the redox species and electrode is very slow (kinetical controlled) relative to the mass transport of solution species to the electrode, then the observed current will not be a function of mass transport. In this case, the low rate of electron transfer results in a concentration ratio of the two forms of the redox couple at the electrode surface that no longer conforms to the Nernst equation. Current – voltage curves of these
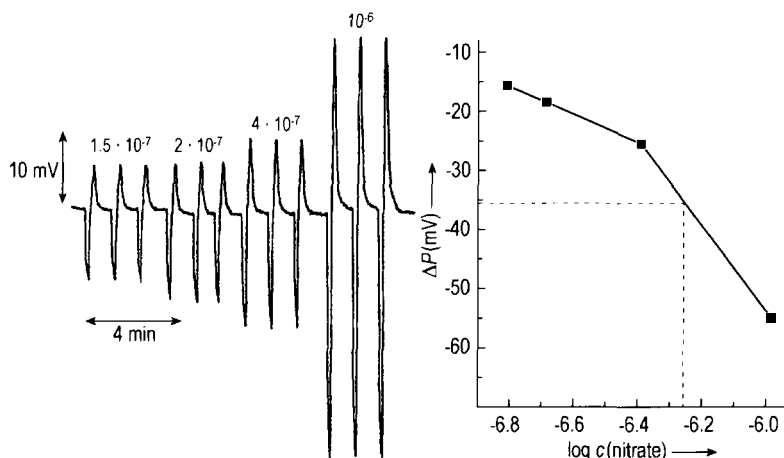
Figure 16. The calibration plot in the lower ppb range shown was constructed with peak-to-peak values resulting in a double Nernstian response. Because of the FIA detection the ISE membrane is never equilibrated with higher concentrations of the analyte ion.
Details: nitrate-selesctive ISE membrane: 1 % Tetradodecylammoniumbromide; 50 mol % of the first is potassium-tetrakis-(4-chlorophenyl)-borat; 49 % 2-.NPOE; 50 % PVC
Carrier solution: 10^{-3} M $Li_2SO_4 + 10^{-7}$ (!) M KNO_3

Table 4. Commercially available ISEs with selectivity for cations

ISE type	MT^a	Concentration range, mol/L ^{b}	Useful pH range ^{b}	Temp. range, °C ^{b}	Membrane resistance, $M\Omega^{b,c}$	Suppliers ^{d}	Interfering ions ^{b,c} in order of decreasing interference; miscellaneous remarks
pNa^+	G	$2 \cdot 10^{-6} - 1$	pH = pNa + 4	0 – 60	100 – 200	CIMORWΩ	H^+, K^+, NH_4^+, Li^+; Ag^+ must be absent
pK^+	P	$10^{-6} - 1$	2 – 12	0 – 40	10 – 20	CIMORWΩ	Rb^+, Cs^+, NH_4^+, Tl^+, H^+, Ag^+, $Tris^+$, Li^+, Na^+; cationic surfactants must be absent
pCa^{2+}	P	$5 \cdot 10^{-7} - 1$	2.5 – 11	0 – 40	1 – 4	CIMORWΩ	Pb^{2+}, Hg^{2+}, H^+, Sr^{2+}, Fe^{2+}, Cu^{2+}, Ni^{2+}, NH_4^+, Na^+, $Tris^+$, Li^+, K^+, Ba^{2+}, Zn^{2+}, Mg^{2+}; cationic surfactants must be absent
pCd^{2+}	S	$10^{-7} - 10^{-1}$	2 – 12	0 – 100	≤ 1	MOWΩ	Cu^{2+}, Hg^{2+}, Ag^+ must be absent; Fe^{2+}/Fe^{3+} interfere if present in concentrations $> c$ $[Cd^{2+}]$
pPb^{2+}	S	$10^{-6} - 10^{-1}$	4 – 7	0 – 100	≤ 1	MORWΩ	Cu^{2+}, Hg^{2+}, Ag^+ must be absent; Fe^{2+}/Fe^{3+} and Cd^{2+} interfere if present in concentrations $> c$ $[Pb^{2+}]$
pCu^{2+}	S	$10^{-8} - 10^{-1}$	2 – 12	0 – 80	≤ 1	CMORW	Hg^{2+}, Ag^+, S^{2-} must be absent; Cl^-, Br^-, Fe^{3+}, Cd^{2+} interfere if $c > c$ $[Cu^{2+}]$; Anions forming complexes with Cu^{2+} (Hal^-, OAc^-) interfere; this ISE is sensitive to light, so all measurements should be carried out in the dark
pAg^+	S	$10^{-7} - 10^{-1}$	2 – 12	0 – 80	≤ 1	CIMORWΩ	Hg_2^{2+}, Hg^{2+}, proteins
pNH_4^+	P	$< 10^{-6} - 1$	4 – 7	0 – 50	≤ 10	IR	cationic surfactants must be absent; interferences from K^+, Rb^+, H^+, Cs^+, Na^+/Li^+, Sr^{2+}, Ba^{2+}, Mg^{2+}

$^{a-h}$ See Table 7.

Table 5. Commercially available ISEs with selectivity for anions

ISE type	MT[a]	Concentration range, mol/L[b]	Useful pH range[b]	Temp. range, °C[b]	Membrane resistance MΩ[b,c]	Suppliers[d]	Interfering ions[b,e] in order of decreasing interference; miscellaneous remarks
pF⁻	S	10^{-6} – sat. solu.	5 – 11	0 – 100	≈ 0.2/≤1	CIMORWΩ	pH < 5, pH > 11
pCl⁻	S	$5 \cdot 10^{-5}$ – 1	2 – 12	0 – 50	≤ 0.1	CIMORWΩ	CN^-, I^-, Br^-, $S_2O_3^{2-}$, NH_3, OH^-, metal ions forming complexes with Cl^-; S^{2-} must be absent
pBr⁻	S	$5 \cdot 10^{-6}$ – 1	2 – 12	0 – 50	≤ 0.1	CIMORWΩ	CN^-, I^-, $S_2O_3^{2-}$, NH_3, Cl^-, metal ions forming complexes with Br^-; S^{2-} must be absent
pI⁻	S	10^{-7} – >1	2 – 12	0 – 80	≤ 0.1	CIMORWΩ	S^{2-}, CN^-, Br^-, Cl^-, $S_2O_3^{2-}$, metal ions forming complexes with I^-
pCN⁻	S	< 10^{-6} – 10^{-2}	11 – 13	0 – 80	≤ 30/≤0.5	CIMORWΩ	S^{2-} must be absent; interferences from I^-, CrO_4^{2-}, $S_2O_3^{2-}$, Br^-, Cl^-, metal ions forming complexes with CN^-
pSCN⁻	S	$5 \cdot 10^{-6}$ – 1	2 – 12	0 – 50	≤ 0.1	MOW	$S^{2-} \approx I^-$, Br^-, CN^-, $S_2O_3^{2-}$, Cl^-, OH^-
pClO₄⁻	P	$7 \cdot 10^{-6}$ – 1		0 – 40		O	not sorted: I^-, NO_3^-, Br^-, ClO_3^-, CN^-, NO_2^-, HCO_3^-, CO_3^{2-}, Cl^-, $H_xPO_4^{x-3}$, OAc^-, F^-, SO_4^{2-}
pNO₃⁻	P	10^{-5} – 1	3 – 12	0 – 50	1	CIMORWΩ	anionic surfactants must be absent; interference from $SCN^- \approx MnO_4^-$, $NO_2^- \approx ClO_4$, I^-, Br^-, HCO_3^-, $F^- \approx SO_4^{2-}$
pBF₄⁻	P	$7 \cdot 10^{-6}$ – 1	2.5 – 11	0 – 40		MOW	ClO_4^-, I^-, ClO_3^-, CN^-, $Br^- \approx NO_2^-$, NO_3^-, Cl^-, PO_4^{3-}, OAc^-, F^-, SO_4^{2-}
pS²⁻	S	10^{-6} – 10^{-1}	12 – 14	0 – 80		CIMORWΩ	Hg_2^{2+}, Hg^{2+}, proteins

$^{a-h}$ See Table 7.

Table 6. Commercially available ISEs with selectivity for gases

ISE type	MT[a]	Concentration range, mol/L[b]	Useful pH range[b]	Temp. range, °C[b]	Membrane resistance, MΩ[b,c]	Suppliers[d]	Interfering ions[b,e] in order of decreasing interference; miscellaneous remarks
pNH₃	M	< 10^{-6} – $5 \cdot 10^{-2}$	alkaline soln.	0 – 50		IMOWΩ	all types of detergents and wetting agents must be absent; interferences from volatile amines
pCO₂	M	$7 \cdot 10^{-6}$ – $2 \cdot 10^{-2}$	acidic soln.	0 – 50		IOΩ	all types of detergents and wetting agents must be absent; interferences from SO_2, NO_x, H_2S
pNO$_x$	M	10^{-6} – $5 \cdot 10^{-3}$	acidic soln.	0 – 50		IO	all types of detergents and wetting agents must be absent; interferences from SO_2, CO_2, volatile carbonic acids
pCl₂	M	10^{-7} – $3 \cdot 10^{-4}$		0 – 50		O	oxidizing agents

$^{a-h}$ See Table 7.

Table 7. Commercially available ISEs with selectivity for miscellaneous species

ISE type	M-T^a	Concentra-tion range, mol/L^b	Useful pH range^b	Temp. range, °C^b	Membrane resistance, M$\Omega^{b,c}$	Suppliers^d	Interfering ions^{b,e} in order of decreasing interference; miscellaneous remarks
Hard^f	P	$7 \cdot 10^{-6} - 1$		$0-50$		O	not sorted: Na^+, Cu^{2+}, Zn^{2+}, Fe^{2+}, Ni^{2+}, Sr^{2+}, Ba^{2+}, K^+
Surf^g	P	$\approx 10^{-6; h}$	$1-13$	$0-40$		MO	oppositely charged bulky ions interfere with the titration reaction; high concentrations of nonionic surfactants may interfere

^{a} MT = Membrane type: G = Glass-membrane electrode; P = PVC-membrane electrode; S = Solid-state membrane electrode; M = gas-permeable membrane electrode. Glass and solid-state membranes are more resistant to certain organic solvents (acetone, methanol, benzene, dioxane, etc.). PVC-membrane electrodes must not be allowed to come into contact with any organic solvent, because this dramatically reduces the lifetime of the electrode.
^{b} Mean values as provided by the suppliers.
^{c} Only if different suppliers provide very different values for membrane electrical resistance are two values listed in the table.
^{d} This list was compiled in Spring 1993 and should not be regarded as complete; C = Ciba Corning; I = Ingold Messtechnik; M = Metrohm; O = Orion; R = Radiometer; W = WTW; Ω = Omega.
^{e} A very detailed compilation of selectivity coefficients is available in [23].
^{f} Hard: water hardness; i.e., divalent cations.
^{g} Surf = ionic surfactants.
^{h} This type of ISE should be used only for the monitoring of titrations: the value cited refers to the mimimum titratable concentration of an anionic surfactant (pure solution).

so-called irreversible processes show the corresponding current peaks more to be separated and of different heights and are of limited analytical utility.

An intermediate situation arises when the electron-transfer and mass-transport rates are comparable. Such quasi-reversible electrode reactions are quite common, and their analytical utility depends to a large extent on careful control of the mass-transfer rate in the electroanalytical method used for their study.

Instruments suitable for voltammetry and amperometry consist of three basic components: a wave-form generator, some form of potential control, and an electrochemical cell. Modern electroanalytical systems employ a three-electrode arrangement for the electrochemical cell. A device called a *potentiostat* is used to maintain a programmed or fixed potential difference between the two current-carrying electrodes (the working electrode and the auxiliary electrode) relative to a third electrode (reference electrode), the function of which is to provide a fixed potential reference in the cell [64], [65].
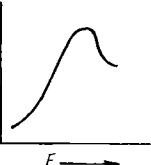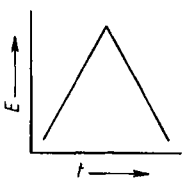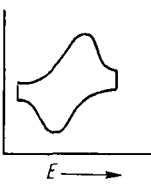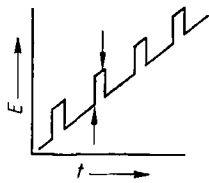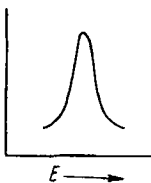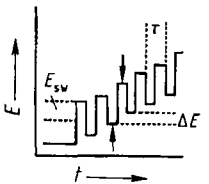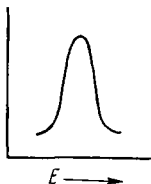
Both potentiostats and waveform generators have benefited substantially in recent years from the introduction of operational amplifiers, as well as from the availability of desktop computers. Electrochemical cell design is now wellestablished both for static systems and for detectors situated in flowing reagent streams.

Despite the fact that solid electrodes have been in use for electroanalytical purposes for some time and are well documented, selection of an appropriate electrode material and its preparation before use remains an area of considerable interest and great controversy among electroanalytical chemists [66]. Solid electrodes for analytical use should always be prepared in a rigorous and reproducible manner. This generally involves polishing the electrode physically with successively finer grades of carborundum or diamond paste. The polished electrode should then be rinsed thoroughly to remove as completely as possible all traces of the polishing materials.

Many different applied potential waveforms have been employed for analytical purposes, but only a few are of mainstream significance [67], [68]. These are presented in Table 8.* → = Point at which the current is measured; τ = Portion of the staircase during which pulses of width $\tau/2$ are superimposed; E_{sw} = Square-wave amplitude.

In *linear-sweep* voltammetry the current is recorded while the potential of the working electrode is swept from one selected limit to another at a rate between 1 mV/s and 1000 mV/s. In *cyclic* voltammetry, two linear sweeps are recorded, as shown in Table 8. The distinguishing feature of cyclic voltammetry is that electrogenerated species formed in the forward sweep are subject to the reverse electrochemical reaction in the return sweep [69]. Figure 17 shows a typical cyclic vol-

Table 8. Voltammetric techniques (adapted from [26])

Technique	Applied potential*	Measured signal	Comments
Linear-sweep voltammetry			used mainly for diagnostic purposes; current measured throughout the scan
Cyclic voltammetry			reversal of the potential at the end of the forward scan makes this a powerful diagnostic tool; current is measured throughout the scan
Differential-pulse voltammetry			the displayed signal is the difference between the current just prior to and at the end of the applied pulse; low scan rates, but good limits of detection
Square-wave voltammetry			the displayed signal is the difference between currents measured on the forward and reverse pulses; rapid scans and good detection limits

tammetric curve together with the important parameters. In both methods, as the potential is swept to the electroactive region of the redox couple, the current response rises to a peak value before decaying. This decay is caused by depletion of the electroactive species in the zone close to the electrode surface, which means that the diffusion zone spreads out further into the bulk solution.

Pulse and square-wave voltammetry [70] are much better candidates for incorporation into a chemical sensor than linear-sweep or cyclic voltammetry, primarily because of the possibility of discriminating between faradaic and capacitive currents. When a potential pulse is applied to an electrode, the capacitive current that flows is proportional to the magnitude of the pulse, and it decays exponentially with time. The faradaic current, on the other hand, decays according to the square root of time. Figure 18 illustrates the expected decreases in faradaic and capacitive currents, showing that the capacitive current decreases faster than the faradaic one. This gives the opportunity of sampling the current only after the capacitive one can be neglected.

Proper selection of the measuring time permits a signal-to-noise ratio to be improved dramatically. This characteristic is exploited in normal-pulse, differential-pulse, and square-wave voltammetry. In the first of these techniques it represents the only mechanism for decreasing the effect of capacitive current. Further elimination of the capacitive current can be achieved in differential-pulse voltammetry by limiting the duration of the applied pulse and by subtracting the current observed immediately prior to the imposition of the pulse. Increased rejection of the charging cur-
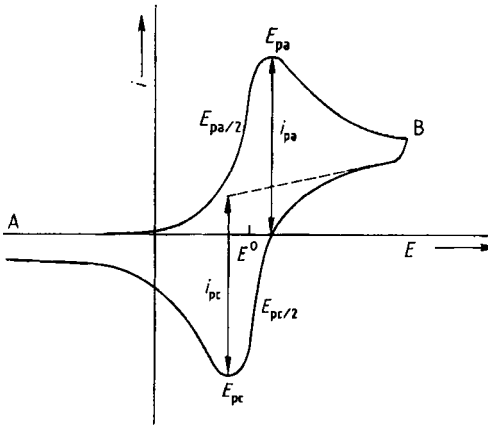
Figure 17. Typical potential vs. current (E vs. i) curve from a cyclic voltammetric experiment with a reversible redox couple
A) Starting potential; B) Reverse potential
E^0 = Standard potential; E_{pa} = Anodic peak potential; E_{pc} = Cathodic peak potential; $E_{pa/2}$ = Anodic half-wave potential; $E_{pc/2}$ = Cathodic half-wave potential; i_{pc} = Cathodic peak current; i_{pa} = Anodic peak current

Figure 18. Decrease in faradaic and capacitive current with elapsed time from the imposition of a potential pulse
$i^\bullet$ = Faradaic current; $i^\sim$ = Capacitive current

rent leads to improved detection limits; indeed, this variant of voltammetry permits concentrations as low as $10^{-8} - 10^{-7}$ mol/L to be measured quite readily. In order to avoid problems associated with the electroactive species, the delay between application of successive pulses must be approximately one-half second. This in turn imposes a limit on the scan rate in differential-pulse voltammetry, thereby decreasing its usefulness in sensor applications.

The technique of square-wave voltammetry offers greater promise as a voltammetric method for probing selective chemistry because of the high rate at which the corresponding scan can be executed. The analytical signal in this technique constitutes a difference between the current for the forward pulse and the current for the reverse pulse. Because of the large amplitude of the square wave in a reversible reduction, a reduced electroactive species formed at the electrode during the forward pulse is reoxidized by the reverse pulse. Consequently, the sensitivity of the method is enhanced relative to differential-pulse voltammetry.

This element of speed is crucial to square-wave voltammetry, because, like all voltammetric techniques based on pulse waveforms, the measured current is proportional to $t^{-1/2}$. However, in contrast to the other pulsed voltammetric techniques, square-wave voltammetry causes very little of the depletion that gives rise to distortion of the cur-

rent – voltage waveform. Accordingly, square-wave voltammetry is uniquely capable of benefiting from high scan rates. A typical compromise frequency for the waveform is 200 Hz. Coupled with a dE value of 10 mV, this gives rise to a scan rate of 2 V/s. The scan rate is limited by a concomitant increase in the capacitive current. At a solid electrode, the square-wave method is no better than pulsed-voltammetric techniques with respect to rejection of the capacitive current.

Another significant advantage of square-wave technology is the possibility of rejecting a wide range of background currents. With respect to capacitive currents this is achieved by the subtraction of two currents in a manner analogous to differential-pulse voltammetry. Slowly varying capacitive currents that arise when surface groups reorganize on certain types of solid electrodes are also eliminated provided the rate of variance is sufficiently low.

Amperometry. Amperometry has traditionally been concerned with maintaining a fixed potential between two electrodes, but pulsed techniques have recently attracted considerable attention as well. The applied potential at which current measurements are made is usually selected to correspond to the mass-transport-limited portion of the corresponding voltammetric scan. Table 9 summarizes the various amperometric methods.

Theoretically, the current obtained in a quiescent solution at a conventional set of electrodes gradually decays to zero according to the Cottrell equation [71]:

Table 9. Amperometric methods (adapted from [26])

Method	Applied potential	Measured signal	Comments
Chronoamperometry			the electrode potential is pulsed to a region in which the analyte is electroactive; a decaying current reflects the growth of the diffusion layer
Pulsed amperometry			the electrode potential is pulsed briefly to a region in which the analyte is electroactive; between pulses the diffusion layer may be eliminated by forced or natural convection
Pulsed amperometric detection			electrode conditioning, analyte sorption, and catalytic electrooxidation are all promoted by the use of this waveform; the current is measured only during the last part of the cycle (region *A*)

$$i_j = nFAC(D/\pi t)^{1/2} \qquad (7)$$

where t is the elapsed time from application of the potential pulse.

The observed decrease in current is due to a slow spread of the diffusion layer out into the bulk solution, with a concomitant decrease in the concentration gradient. In practice, this process continues for ca. 100 s, after which random convection processes in the solution take over, putting an end to further movement of the diffusion layer. Waiting nearly two minutes to obtain a steady-state current is not a particularly attractive alternative. Accordingly, in amperometric measurements for sensor applications the spread of the diffusion layer is controlled by invoking one or more of the following mechanisms:

1) *Convective Diffusion.* From a practical point of view, convective diffusion can be achieved in two ways: by moving the solution relative to the electrode, or moving the electrode relative to the solution. Of the systems developed to move the electrode, only one merits consideration here, the *rotating disc electrode* (RDE),

but there are few possibilities for realizing this electrode in a sensor.

2) *Imposition of a Physical Barrier in the Form of a Membrane.* There are four primary advantages to covering the electrodes of an amperometric device with a membrane permeable only to the analyte: (a) poisoning of the electrodes by electroactive or surface active species is limited; (b) the resolution of the system is enhanced, because extraneous electroactive species that might otherwise undergo electron transfer at the electrode are excluded; (c) the composition of the electrolyte occupying the space between the membrane and the electrodes remains constant; and (d) the membrane forms a physical barrier to prevent the diffusion layer from spreading into the bulk solution.

The limiting current obtained at a membrane-covered amperometric device is a function of time, reaching in due course a steady-state value. The current–time transient cannot be described by a single equation, because different transport mechanisms operate during this time interval. For a typical membrane-covered amperometric oxygen detector with an electro-
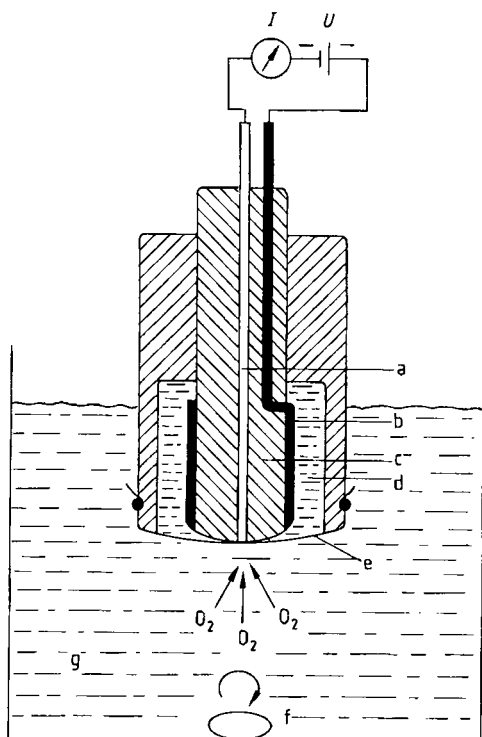
Figure 19. Schematic diagram of a Clark-type oxygen sensor
a) Cathode; b) Anode; c) Insulator; d) Electrolyte; e) Membrane; f) Stirrer; g) Sample for analysis

lyte thickness of 10 μm, a membrane thickness of 20 μm, an electrode radius > 2 mm, and a diffusion coefficient for the analyte two orders of magnitude lower in the membrane than in the electrolyte solution, it can be shown that the current – time transient can be described by three equations. For times > 20 s after the imposition of the potential pulse to the electroactive region, the limiting current is given by

$$i_1 = nFAC(P_m/d) \tag{8}$$

where P_m is the permeability coefficient of the membrane and d is its thickness.

An important consequence of this equation is that the limiting current is a function of the permeability of the membrane. Unfortunately, membrane properties such as permeability are themselves a function of time and ambient conditions, particularly temperature. Accordingly, there are advantages to operating membrane-covered detectors away from this steady-state condition. For a typical detector, the

limiting current < 100 ms after imposition of the measurement potential pulse is given by

$$i_1 = nFAC(K_b/K_0)(D_e/\pi t)^{1/2} \tag{9}$$

where t is elapsed time since imposition of the measuring potential, D_e is the diffusion coefficient for the analyte in the electrolyte, and K_b is the distribution coefficient for the analyte at the internal electrolyte – membrane interface. The ratio K_b/K_0 expresses the salting-out effect of the electrolyte solution.

In effect this equation describes the diffusion-limited current arising within the electrolyte layer. In other words, the diffusion layer has not had sufficient time to spread to the membrane, and so the limiting current is completely independent of the membrane. This has the advantage that alterations in membrane characteristics, as well as alterations in ambient conditions that might affect membrane performance, have no effect whatsoever on the limiting current. One of the factors to which this applies is stirring. Under steady-state conditions, stirring of some type is necessary to prevent changes that would otherwise be caused by the diffusion layer moving out into the bulk solution, although the rate of such stirring is not critical. When transient measurements are made, no stirring whatsoever is required. The true situation regarding transient measurements with amperometric devices is actually not quite so simple as depicted here; for example, there is also a need to take into account contributions from the capacitive current.

The classic amperometric membrane-covered device is the Clark oxygen sensor [72], which consists of a platinum working electrode, an Ag/AgCl reference electrode, a KCl electrolyte solution, and an oxygen-permeable membrane made of teflon or silicone rubber. Figure 19 is a schematic diagram of such a sensor. Dissolved oxygen passes through the permeable membrane, after which it is reduced at the working electrode. The current flow is proportional to the amount of oxygen present in the solution. The following reactions take place at the surface of the working electrode:

$$O_2 + 2H_2O + 2e \longrightarrow [H_2O_2] + 2OH^-$$

$$[H_2O_2] + 2e \longrightarrow 2OH^-$$
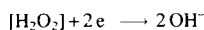
A separate polarographic wave corresponding to the reduction of hydrogen peroxide is observed only with a mercury cathode, which produces two waves for the reduction of oxygen. At other electrodes (Pt, Au, C) only one four-electron wave is obtained, corresponding to complete reduction to four hydroxide ions [73]. The potential E of the working electrode (at the diffusion-limiting current plateau) usually lies between -0.6 and -0.9 V vs. Ag/AgCl.

Progress in semiconductor technology during recent years has made it possible to miniaturize the classical oxygen sensor [74]–[77] for use as a transducer in several chemical and biochemical sensors.

3) *Pulsing the Electrode from a Region of Electroinactivity to a Region of Transport-Controlled Electroactivity, after which the Potential is Restored to the Electroinactive Region.* For electrochemical sensors without membranes, applying a potential pulse from a region of electroinactivity to a region of electroactivity has several advantages. For example, the duty cycle of the pulse can be arranged in such a way that the electrode spends the majority of its time at the electroinactive potential. Under these circumstances electrode reactions give rise to products that poison the electrode, but these products are of minimal significance because they are produced at a very low rate. Moreover, natural processes of convection are able to redistribute the electroactive species during the time that the electrode is at the rest potential, thus limiting the spread of the diffusion layer into the bulk solution. The drawback to this technique is the appearance of a capacitive current. Limiting the size of the pulse decreases the contribution of this current, but a small pulse from a rest potential to the electroactive potential is achievable only for reversible electrode reactions. The alternative is to allow sufficient time for the capacitive current to decay away between imposition of the potential pulse and measurement of the current.

4) *Pulsed Amperometric Detection* [78]. This is a relatively new technique that has made possible the direct amperometric determination of many compounds formerly considered unsuitable for analysis in this way. The corresponding waveform is illustrated in Table 9. This rather complex waveform serves several purposes. On platinum, for example, and at the extreme anodic potential, surface-confined oxidation products that tend to poison the electrode are desorbed. This process occurs concurrently with the formation of PtO_2 on the surface. At the cathodic potential the surface oxide is subsequently reduced, and a clean platinum surface is regenerated. In some cases, sorption of the analyte occurs at this cathodic potential, so the analytical signal is recorded at an intermediate potential. Several different mechanisms can give rise to this signal. Electroactive species that are not adsorbed at the electrode may undergo conventional oxidative electrode reactions. Alternatively, electroactive species that are adsorbed at the electrode during the cathodic swing of the cleaning cycle may undergo either conventional oxidative desorption or oxidative desorption catalyzed by an oxide species on the surface of the electrode. For many compounds the latter reaction is much more favorable from a thermodynamic point of view than oxidation in the absence of a catalyst. This type of reaction occurs most readily when both the analyte and the oxide are confined to the surface. Finally, adsorbed electroinactive analytes may also be detected by virtue of their ability to block the electrode sites at which the PtOH formation occurs. Consequently, any faradaic current arising from the oxidation of platinum is diminished in the presence of adsorbed electroinactive analyte.

5) *Microelectrodes.* Electroanalytical applications using microelectrodes have become more common in recent years [79]–[83] as silicon technology has opened new possibilities for constructing electrodes with micrometer and submicrometer dimensions [84], [85]. The rate of electrolysis at such electrodes is approximately equivalent to the rate of diffusion. Linear-sweep voltammetry in this case gives rise to the same S-shaped current–voltage curves observed with rotating disc or dropping mercury electrodes. The time-dependent current arising at a microelectrode is given by the equation for the chronoamperometric current at a conventional electrode under spherical diffusion conditions:

$$i_t = nFACD\left[\left\{1/(tD\pi)^{1/2}\right\} + \{1/r\}\right] \tag{10}$$

where r is the electrode radius. Indeed, a sphere or hemisphere represents a good model for the diffusion zone that surrounds such an

electrode, and because of enhanced mass transport to the electrode, a steady-state current is achieved very rapidly once a potential pulse has been applied to the electrode.

Although the diffusion zone is large in this case relative to the electrode surface, it is small compared to the zone surrounding a conventional electrode. This small size means that the faradaic current associated with a microelectrode is relatively immune to the effects of convection in the bulk solution, and the current in a flowing stream is independent of the rate of flow. Decreased effects due to capacitance and resistance at microelectrodes, coupled with high mass-transport rates, make it possible for electrochemical measurements to be obtained in this way in cells containing highly resistive solutions. In fact, provided sufficient free carrier is available to charge the double layer, microelectrodes can be used for direct electrochemical measurements in the gas phase [86]. One drawback of employing working electrodes with diameters less than about 5 μm are the small currents which will flow through such small interfaces. Current measurements in the pA range are prone to electromagnetic interferences (e.g., AC pick-up) and thus, the electric noise level can be a limiting factor. This problem can be solved by employing an array of microelectrodes all at the same potential and separated from each other by about ten times their diameter. Such arrays have been constructed and have shown all the benefits of single microelectrodes [87], [88].

28.2.3.1.3. Conductance Devices

Conductance is used as a sensor signal in two different areas of sensor technology. On one hand it is possible to determine the electrolytic conductance of electrolytes, which permits conclusions to be reached on a wide range of electrolyte properties. Moreover, partial pressures of gaseous compounds can be determined in the gas phase by examining their influence on the electronic conductance of such materials as semiconducting oxides [89]–[92].

Conductometric Sensors for Monitoring Electrolytic Conductance. *Introduction.* Electrolytic conductance serves as a useful signal of the electrical conductivity of mainly aqueous electrolyte solutions, and *conductometry* has been devel-

oped into an electrochemical analytical method based on measuring the conductance in electrolyte solutions. All ions in such a solution act as current carriers as a function of their mobility, charge, and concentration. For this reason conductivity measurements are inherently nonselective with respect to any system to which a changing voltage difference has been applied and the diffusion current is measured. Despite this disadvantage, conductivity measurements are indispensable sources of information, especially for monitoring the chemical purity of water samples. In pollution control, conductometric data provide a reliable measure of the total content of ionic pollutants, which is sometimes all that is required. Advantages of conductivity detectors include their universal applicability, simplicity, low price, and wide effective concentration range. Conductivity measurement is one of the most commonly used electrochemical techniques in the control of industrial chemical processes. Conductivity detectors are also employed in ion chromatography and HPLC as a way of detecting separated ions in the eluate from a column [93]. Furthermore, conductometric analyzers are used in the monitoring of atmospheric pollutants such as acidic or alkaline gases (e.g., HCl, CO_2, SO_2, SO_3, NH_3) after these have been transferred as ions into aqueous solution [94], [95]. Conductivity measurements are nondestructive with respect to the sample, and they are always advantageous when electrolytes must be determined in a medium of low self-conductivity.

Definitions. The electrolytic conductivity σ of a solution is defined as the conductance G of the electrolyte between two electrodes, each with an area A, separated by a distance d. The conductivity σ is then given by

$$\sigma = G \cdot d/A \qquad (11)$$

The traditional unit of electrolytic conductivity, Siemens per centimeter (S/cm), has recently been replaced by S/m [96], but the new unit has not yet been accepted widely by the instrument-producing industry, nor by the users of conductivity measurement devices. The English-language literature has historically made use of the "mho" (reciprocal ohm), which is another way of expressing the conductance G. The quotient d/A (cm^{-1}) is defined as the cell constant k, and depends on the geometry of the cell. The electric field is never homogeneous over the cross section of a real cell, so it is not in fact possible to calculate a cell constant k directly from geometric dimensions; it
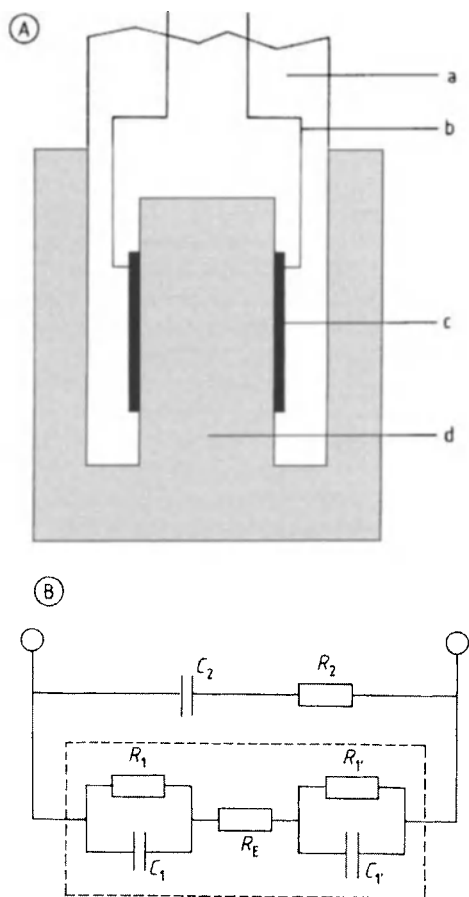
Figure 20. Depiction of a two-electrode cell (A) together with the corresponding equivalent circuit (B)
The dashed rectangle includes all network components involving the electrolyte in contact with the electrodes.
a) Cell body; b) Leads; c) Electrode; d) Sample
R_E = Electrolyte resistance; R_1, $R_{1'}$ = Charge-transfer resistances at the interface electrode – electrolyte; C_1, $C_{1'}$ = Capacitance of the electrochemical double layer; R_2, C_2 = Shunt circuit in the cell

must instead be determined with calibration solutions of known conductivity [97].

The temperature dependence of σ can be deduced from ionic mobilities. This dependence is expressed in terms of the temperature coefficient $\alpha = (d\sigma/dT)(1/\sigma)$, which generally has a positive value between 1 and 3 % per degree Celsius. Knowledge of α in turn makes it possible to calculate conductivities at temperatures other than the temperature of measurement. Whenever the need arises to determine the conductivity of a solution with a very low electrolyte concentration it is essential also to take into account the self-

dissociation of pure water, which has a conductivity $\sigma_{25} = 0.05483$ µS/cm.

Measurement Principles. The equipment used for measuring conductance usually consists of a conductivity sensor, a measuring attachment, and a temperature-compensation unit. Conductance values are obtained by establishing either the current flow between a set of electrodes subjected to a constant AC voltage, or the current induced in the secondary coil of a transformer connected to the primary coil by the electrolyte of interest. Capacitive coupled cells have been developed as well, but because of their expensive construction they are no longer of interest [98]–[101].

Two-Electrode Cells. Two-electrode cells are equipped with a pair of electrodes across which both the voltage and the current can be measured. Use of an AC voltage source is a convenient way of avoiding polarization effects at the electrode/electrolyte interface and electrolyte decomposition. Figure 20 illustrates a typical construction together with the corresponding equivalent circuit. The advantage of this design is the simplicity of both the cell and the measuring attachment.

Disadvantages are associated with the need for measuring both the current and the voltage at a single electrode, which may lead to electrode polarization and a measured conductivity lower than the true value. The resistance contributed by a cell in contact with a sample must also be taken into consideration in conjunction with other resistance and capacitance values. At high conductivities ($\sigma > 1$ mS/cm) the resistance of the connecting wires is another significant factor, and electrode reactions can cause polarization of the electrode – electrolyte interface. These problems can be overcome most effectively with an AC voltage oscillating at 1 – 5 kHz and electrode materials with minimum polarizability due to a large surface area, such as platinum covered with a layer of platinum black. At low conductivities ($\sigma < 10$ µS/cm) measured conductivity values may be too high because of capacitive – ohmic shunts caused by the cell body or the connecting wires ("Parker effect"). Contamination by adsorbed ions is also a problem with porous electrodes. The measurement frequency should therefore be kept as low as possible, and the electrodes should be smooth. Resistive precipitates at the electrode surface are sometimes a cause of artificially low conductivity.

For the reasons mentioned above, the "cell constant" is in fact constant only over a limited conductivity range. Suitable values for the cell

constant depend on the anticipated level of conductivity:

Expected conductivity	Cell constant
< 0.1 µS/cm – 1 mS/cm	0.1 cm⁻¹
1 µS/cm – 100 mS/cm	1.0 cm⁻¹
10 µS/cm – 1 S/cm	10 cm⁻¹

Four-Electrode Cells. The four-electrode cell was developed as a way of overcoming the problems inherent in two-electrode cells. Two current electrodes in this case generate an electric field, and two separate voltage electrodes are provided for measuring the voltage drop with the aid of a high-impedance amplifier. This design is insensitive to polarization effects, and it also avoids interference from insulating films or precipitates because these are formed outside the zone of measurement. The result is a single system that can be used to measure any conductivity > 1 µS/cm.

Inductively Coupled Cells. The principle underlying inductively coupled conductivity cells is very simple. Two coils of a transformer are shielded from each other magnetically such that the only coupling loop is formed by an insulating tube filled with the sample. An AC voltage in the primary coil induces an AC current in the sample, the magnitude of which depends on the dimensions of the loop and the conductivity of the sample. The voltage induced in the second coil depends only on the input voltage of the primary coil and the conductance of the sample. Typical working frequencies are $80 - 200$ Hz. An advantage of these cells is that they permit contactless measurement. Very high conductivities (>1 S/cm) can be measured without polarization errors because of the contactless decoupling of the signal. Even thin resistive films within the tube fail to falsify the measured conductivity. Another advantage associated with cells of this type is the possibility of encapsulating the entire device in a polymer offering high chemical resistance. Inductively coupled conductivity cells are thus of special interest for measurements in highly conducting and aggressive media [102].

Ion-Selective Conductometric Microsensors. The greatest problem in miniaturizing potentiometry was the lack of a longer working reference electrode. The development of ion-selective conductometric microsensors mainly for cations such as H^+, Li^+, Na^+, K^+, NH_4^+, Ca^{2+}, Cd^{2+}, Pb^{2+}, Hg^{2+}, and Ag^+ thus far in the ICB Muenster overcomes this problem [103]. In principle, they are partly based on the ion-selective optodes introduced by the late Prof. Simon who showed that the measured ion can also be selectively extracted into an ionophore-based membrane. In order to maintain electroneutrality in the bulk of the membrane a co-immobilized indicator dye looses a proton which passes into the sample solution and thereby changes the color. The influx of the measured cation, which will be complexed by the neutral ionophore and leads to an even larger cation complex in the PVC membrane, is electrically compensated by an outflux of protons that dissociate from a neutral lipophilic acid. This exchange of protons from the membrane against the measured cation will cause the membrane conductivity to change. This can easily be followed by a planar interdigital electrode array onto which a corresponding membrane is drop-coated. In order to keep the lipophilic acid in the membrane in its neutral state and in order to control the interfacial ion-exchange processes in a reproducible way, all measurements must be performed in a buffer of an appropriately chosen and constant pH. Figure 21 shows the measuring set-up and Figure 22 some results of a valinomycin based potassium-selective membrane. The length A of the interdigital electrode fingers is only about 1 mm and α the width of a finger and spacing lies in the region of about 10 µm. The membrane thickness is δ in the range $0.5 - 5$ µm. The membrane cocktail — similar to the one in case of ISE membranes — can be brought onto the interdigital electrode array by dip- or drop-coating. U_{in} an alternating voltage in the range $10 - 100$ mV and is applied via the corresponding output of a lock-in amplifier. U_{out} at the load resistor is fed to the input of the lock-in amplifier. The in-phase component of the complex impedance is measured and evaluated. In general, the characteristics of those bulk conductive microsensors resemble those of the corresponding ISE membranes with slight variations in the selectivity pattern. The advantage is that the accuracy of their measurements does not drop with increasing charge of the measured cation. With a membrane thickness in the region of 1 µm the response time is in the second range. A difference compared with ISE membranes lies in the increased concentration of ionophore that is added up to about 5 % by weight. These ion-selective conductometric microsensors do not need any reference electrode and can be scaled down to the µm range. The lifetime depends on the lipophilicity of the ionophore and plasticizer. Normally they work for up to several weeks and show a relatively small drift. The further advantages are that they are easily
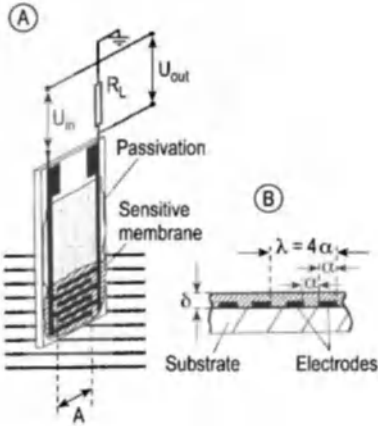
Figure 21. Schematic drawing of the construction and measuring circuit of ion-selective conductometric microsensors not requiring any reference electrodea) General view; b) Cross section
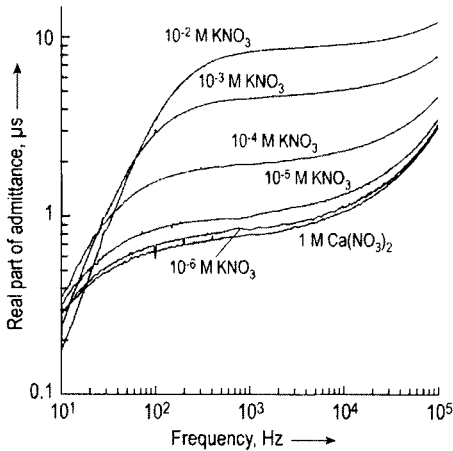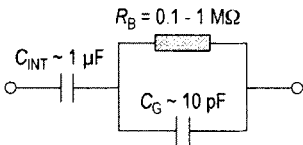
Figure 22. Simplified electronic equivalent circuit (measured is R_B = bulk membrane resistance) and impedance spectra of the real part (admittance) in a 1 M $Ca(NO_3)_2$ (!) supporting electrolyte and increasing K^+ concentrations with a valinomycin-based potassium conductometric microsensor. Note that even μM amounts of K^+ ions in a well conducting electrolyte solution can be sensed
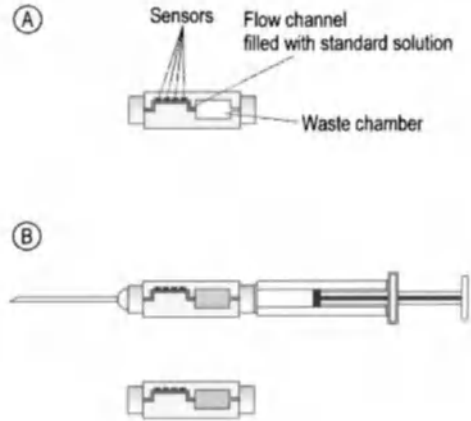
Figure 23. Measuring principle of a cartridge system using microsensors for the case of medical "point of care" blood electrolyte measurements
A) Storage of the cartridge with a standard solution; recall of data: measuring of the standard solution for plausibility and quality controls.
B) Drawing of the blood sample; displacement of standard; disconnection of the cartridge; insertion of the cartridge into the analyzer; measuring of the undiluted blood sample.

for which the Nernst equation only allows a small mV change per decade concentration change. Despite the fact that their selectivity pattern is more similar to the one of the corresponding optodes they lack the drawback of the latter with respect of chromo- or fluorophore bleaching. Furthermore, in case of light absorption measurements a compromise between sensitivity (depending on the thickness) and response time must be made.

Because of the excellent sensitivity, ease of fabrication, and storage capabilities these microsensors are ideally suited for medical bedside monitoring or "point of care" devices. Figure 23 shows a novel cartridge developed at the ICB Muenster for blood electrolyte measurements. In comparison with existing devices no open blood transfer with a certain risk of infection is needed. In another special version of the cartridge, a valve allows the continuous uptake of blood in the syringe needing only the first 100 μL for the electrolyte. The content of the syringe can then be analyzed as usual. The evaluation is performed by putting the cartridge in a hand-held readout unit. Note, prior to blood sampling the same cartridge has to be checked by the same readout unit for quality control reasons and re-calibration.

Conductometric Sensors for Monitoring Gaseous Compounds. *Introduction.* Gaseous com-

produced by CMOS technology and open also the way to the detection of higher charged analyte ions

pounds in the gas phase are detected with chemically sensitive devices that respond electrically to changes in conductance. BRATTAIN and BARDEEN (using germanium, in 1952) [104] and HEILAND (zinc oxide, 1954) [105] were the first to observe modulation of the electrical conductance of a solid-state device by gases from the surrounding atmosphere. The most commonly used material for conductometric solid-state gas sensors is now tin dioxide [106]. As noted previously (Section 28.2.2.2) commercial sensors based on polycrystalline SnO_2 ("Taguchi sensors") [107] are produced by companies such as Figaro Inc., for a wide range of industrial and domestic applications, including gas monitors, leak detectors, and alarm systems for toxic or inflammable gases. Apart from SnO_2, an n-semiconductor, many other inorganic semiconducting oxides can serve as sensitive materials in gas sensors, including TiO_2, ZnO, and Fe_2O_3 with various additives and dopants [13], [108]. Certain organic semiconductors such as phthalocyanines with appropriate central metal atoms, or polymers like polypyrrole or polysiloxane, also show sensitivity to gaseous compounds. These sensors are often referred to as homogeneous or heterogeneous semiconductors, or as dielectric sensors. Detailed lists of materials sensitive to detectable gaseous compounds together with the corresponding operating temperatures are provided in [89]–[92], [108]. The design principle upon which such sensors are based is quite simple: a sensitive semiconducting sample is placed between two metallic contacts, and the combination is equipped with a resistance heater capable of maintaining the appropriate working temperature, usually between 100 °C and 700 °C. Semiconductor sensors have dimensions in the range of millimeters, and are easily fabricated by either thick- or thin-film technology [109].

Heterogeneous Semiconductor Sensors. The response of a heterogeneous semiconductor sensor is based on changes in surface conductance as mentioned in the introduction , mainly due to changes in the free electron concentration within a thin surface layer of the semiconductor as a consequence of charge exchange with adsorbed species accompanying chemisorption or heterogeneous reactions [33] (see Section 28.2.2.2). The most commonly used semiconducting materials are the two very stable oxides SnO_2 [110], [111] and ZnO [112]. Because the conductance change is a surface phenomenon, the ratio of surface to bulk material should be made as large as possible (using thin layers) to maximize the effect. For

sintered polycrystalline materials the grain size should also be as small as possible.

At a given temperature the number of gas molecules adsorbed at a semiconductor surface — hence the conductance — correlates exponentially with the partial pressure of the gas. Oxidizing gaseous compounds such as oxygen, nitrogen dioxide, and chlorine cause a resistance increase by extracting electrons from the conduction band of an n-type semiconductor, while reducing gases like hydrogen, carbon monoxide, methane, hydrocarbons, or alcohols lead to a resistance decrease since they are themselves oxidized by the adsorbed oxygen . Humidity decreases the resistance of an n-type semiconductor, causing considerable interference.

Response times for sensors of this type are usually in the range of seconds or minutes. For some gases it has been found that catalytic effects at the semiconductor surface can have a significant impact on the response. Thus, sensor properties such as response and selectivity can be improved by adding catalytically active metals like palladium, rhodium, silver, or platinum. This doping of the semiconductor surface is one important way of modifying the selectivity of a sensor [113]–[115], but another possibility for adjusting sensor properties for a particular application is careful selection of the working temperature [116].

The influence of temperature on the selectivity of a sensor depends on the fact that each gas has its own *characteristic temperature:* a temperature at which the reversible relationship between gas concentration and the concentration of electrons at the semiconductor surface (and therefore the conductance) reaches a maximum [110]. Working temperatures for SnO_2-based sensors usually lie between 200 °C and 450 °C depending on the application.

Various techniques are used for the fabrication of semiconductor sensors. Conductance sensors from structurized sintered polycrystalline ceramics can be produced by thick- or thin-film technology. Chemically sensitive materials in the form of single crystals or whiskers can be attached to electrodes by thin- or thick-film techniques as well. Mass production of sensors requires that the resulting devices be characterized by a defined level of conductance. For example, the conductance of polycrystalline SnO_2 can be adjusted by subsequent thermal treatment > 800 °C under a controlled partial pressure of oxygen. Another approach to defined conductance involves doping the semiconductor with antimony or fluorine. The reproducibility and stability of a sensor *signal*

is generally diminished by irreversible changes in the defect structure of the semiconductor surface.

Homogeneous Semiconductor Sensors. The effectiveness of this type of semiconductor sensor depends on changes in bulk conductance. The stoichiometry and concentration of bulk point defects, and thus the conductance of such materials, is a function of the partial pressure of a gaseous analyte. Conductance changes arise through reversible changes in small deviations from ideal stoichiometry, which in turn influence the concentrations of bulk point defects and free electrons [33]. The deviation from ideal stoichiometric composition is usually very small; for example, in $Cu_{2-x}O$ the deficit x in the Cu ratio is on the order of 0.001. Since conductance and deviation from ideal stoichiometry are proportional, a measured conductance can be used to determine a corresponding partial pressure. Regarding the kinetics of these sensors, an important factor is diffusion of a compound like oxygen or the corresponding lattice defects that compensate for nonideal stoichiometry within the bulk of the semiconductor. Any gaseous reaction involving point defects at the surface of the semiconductor must also be appropriately fast. Both requirements can be met through the combination of high temperature and a catalytic coating (e.g., Pt).

Most bulk-conductance sensors consist of semiconducting oxides the conductance of which changes with the partial pressure of oxygen. They are used primarily for monitoring oxygen partial pressures in the air or in combustion gases. Titanium dioxide (TiO_2) is one of the most commonly used homogeneous sensor materials for the determination of oxygen partial pressures [117]. At 900 °C and an oxygen partial pressure of 0.1 MPa, TiO_2 has precisely the stoichiometric composition. At lower oxygen partial pressures TiO_2 has an oxygen deficit x expressed by the formula TiO_{2-x}. This oxygen deficit specifies the number of unoccupied oxygen sites and the corresponding excess of trivalent titanium in the lattice. Both defects act as electron donors, thereby contributing to the concentration of free electrons in the conductance band of the semiconductor. The oxygen deficit, and thus the conductance, varies inversely with the partial pressure of oxygen in a characteristic way. Thus, the conductance σ_e of an n-type semiconducting oxide like TiO_{2-x}, CeO_{2-x}, or Nb_2O_{5-x} is given by the equation:

$$\sigma_e = \text{const}(T) \cdot p(O_2)^{-1/m}. \tag{12}$$

where const (T) is a temperature-dependent constant and m usually varies between 4 and 8 depending on the nature of the disorder in the oxide structure.

On the other hand, p-type semiconducting oxides like $Cu_{2-x}O$, $Ni_{1-x}O$, and $Co_{1-x}O$ have a deficit x of metal relative to ideal stoichiometry. This deficit increases with increasing oxygen partial pressure. Ionic lattice defects occurring in oxides of this type act as electron acceptors. In this case the concentration of the defects, and therefore the conductance, increases with increasing oxygen partial pressure according to the expression:

$$\sigma_e = \text{const}(T) \cdot p(O_2)^{1/m} \tag{13}$$

Here again, m usually varies between 4 and 8.

The strong exponential dependence of const (T) on temperature is a great disadvantage with homogeneous semiconductor sensors, one that can lead to large measurement errors. Nevertheless, appropriate sensor design makes it possible to compensate for the temperature dependence. In the case of a TiO_2 combustion-gas sensor, for example, the conductance of an oxygen-sensitive TiO_2 sample can be compared with that of a second sample displaying the same response to temperature but coated in such a way as to eliminate its sensitivity to gas composition. The conductance difference between the two samples is then a function only of the gas composition, not of fluctuations in the temperature.

An important requirement for any metal oxide under consideration for use as a homogeneous semiconducting sensor is that it display a sufficiently high bulk diffusion coefficient $\tilde{D}$ for the conductance-determining species. Response times t for oxide layers with a thickness y can be estimated from the Einstein equation:

$$y^2 = 2\tilde{D} \cdot t \tag{14}$$

For example, a Cu_2O layer with a thickness of 1 μm reaches a response time of ca. 1 ms, corresponding to a diffusion coefficient of ca. 10^{-5} cm^2 s^{-1}, only upon heating to 700–800 °C. The diffusion coefficient can be altered by doping, but it also depends on the extent of deviation from ideal stoichiometry.

A number of oxides (e.g., TiO_2, CeO_2, Nb_2O_5, CoO) or titanates with a perovskite structure (Ba/Sr/CaTiO$_3$) have been successfully tested for application as extremely fast responding oxygen sensors [118]–[120] allowing individual cylinder

control in automobile engines. Most of these investigations involved determination of the oxygen partial pressure in exhaust gases from combustion processes at partial pressures between 10 and 10^{-18} kPa. At low oxygen partial pressures and high temperature, certain oxides (e.g., Cu_2O, CoO, NiO) are unsuitable because of their reduction to pure metal. Sensors based on TiO_2 ceramics are by far the most highly developed devices. These consist of porous sintered ceramic materials with a platinum surface doping, and are used to determine air – fuel ratios in the exhaust gases of combustion engines.

Apart from TiO_2, a number of other semiconducting oxides might be useful as sensor materials for special applications over limited ranges of oxygen partial pressure, as in monitoring the level of oxygen in inert gases. There is also a good possibility of optimizing various ternary and higher oxide combinations for sensor application by suitable choice of composition with respect to stability, diffusion, and surface properties [121] – [123]. Thus, barium and strontium ferrates show conductances with considerably less temperature dependence relative to TiO_2 or CoO.

Organic Conductometric Sensors. For several years, certain organic semiconducting polymers have attracted attention in the search for new chemical sensors [124]. The possibility of utilizing conductivity changes induced by gases adsorbed on organic materials is interesting for a number of reasons. For example, it might permit the detection of very low levels of atmospheric pollutants. Moreover, organic materials are much more easily modified than inorganic materials with respect to such characteristics as sensitivity, working temperature, and selectivity. The greatest disadvantage of organic materials in gas detection is that they are usually very poor conductors, and conductivity measurements would be correspondingly difficult. Also, organic materials are thermally unstable, so it is often impossible to use them at temperatures at which gas – solid interactions proceed rapidly and reversibly.

Phthalocyanines constitute one group of organic materials with thermal stability >400 °C, and they are usually considered to have semiconducting rather than insulating properties. Thin layers of phthalocyanine complexes based on metals like lead [125] and copper [126] have been particularly carefully studied as potential gas sensors [127]. The conductance of these films is a very sensitive function of gases with electron-acceptor properties, including nitrogen dioxide, chlorine,

and iodine. Films of phthalocyanines with various central metal atoms can be deposited on substrates with a planar interdigital electrode array by sublimation or screen printing. Changing the nature of the central metal atom makes it possible to adjust the sensitivity and selectivity of a phthalocyanine film in accordance with the proposed application [128] – [133].

Another organic semiconductor with gas-sensitive properties is polypyrrole. Chemiresistors with ambient-temperature gas sensitivity to both electron-donating and electron-accepting gases have been achieved using thin polypyrrole films prepared by electrochemical polymerization on electrodes. Sensitivities have been investigated toward such gases as nitrogen dioxide [134], ammonia [135], and methanol [136], [137]. A wide variety of polymers of this type is available, including substituted polypyrroles, polythiophenes, polyindoles, and polyanilines. This is leading the way to combining more of these sensors in sensor arrays and using chemometrics to evaluate unknown smells (electronic noses).

Humidity Sensors. Humidity sensors are devices that respond to changes in water-vapor pressure by a change in electrical resistance, capacitance, or a combination of the two as reflected in the impedance. The sensitive material might be an electrolyte, an organic polymer, or a metal oxide subject to a change in ionic – electronic conductivity or capacitance with respect to humidity as a consequence of physical adsorption or absorption of water molecules. In general these devices offer a very cost-effective continuous or spot-check approach to monitoring humidity in air and other gases [138], [139].

Electrolyte humidity sensors. Electrolyte humidity sensors are based on an electrolyte solution (e.g., the hygroscopic LiCl) the ion conductivity of which changes as a result of evaporation or condensation of water in response to changes in the relative humidity of the surrounding atmosphere. The fabrication of these sensors is quite straightforward. A porous substrate or an organic binder is impregnated with the electrolyte in such a way as to prevent electrolyte from flowing out of the sensor even at very high humidity. Contact is maintained with platinum electrodes on both sides of the substrate. Even thin films of potassium metaphosphate or barium fluorite are suitable for humidity measurements. Despite their slow response and limited linear range these sensors have an

outstanding reputation due to good reproducibility, long-term stability, and low cost.

Ceramic Humidity Sensors. The first widely-used ceramic humidity sensor was the aluminum oxide sensor, in which aluminum is oxidized anodically in an acidic solution to form a porous oxide film, sometimes less than 1 μm thick, that displays the necessary hygroscopicity. Adsorption of water molecules causes a decrease in resistance and an increase in capacitance, which is reversed by subsequent desorption. Other materials are also used in humidity sensors, including ZnO with added Li_2O and V_2O_5, colloidal Fe_3O_4, α-Fe_2O_3, mixed oxides of TiO_2 and SnO_2, $MgCr_2O_4 - TiO_2$, $TiO_2 - V_2O_5$, $ZnCr_2O_4 - LiZnVO_4$, $MgAl_2O_3$, $MgFe_2O_4$, or $H_3PO_4 - ZrSiO_2$, as well as $PbCrO_4$ with added alkali oxide [140]. Sometimes the signal of such a sensor is subject to drift as a result of changes in the surface structure or contamination from other adsorbed gases. Nevertheless, reproducibility can be restored by heat cleaning.

The most important considerations with respect to sensor characteristics are surface properties (hydrophilic – hydrophobic), pore-size distribution, and electrical resistance. To ensure adequate sensitivity and response a sensor of this type should consist of a very thin film of porous ceramic with a porosity >30 %. The contacting electrodes may be interdigitated or porous sandwich-type structures made from noble metals (e.g., Pd, Pt, Au) and so constructed that they do not obstruct the pores of the oxide film. Humidity affects not only the resistance of a porous ceramic but also its capacitance by extending the surface area in contact with the electrodes. The high dielectric constant of adsorbed water molecules also plays an important role.

Polymeric Humidity Sensors. Certain polymers are capable of taking up water as a function of the relative humidity of the surrounding atmosphere, leading to changes in such mechanical and electrical properties as volume, resistance, and dielectric characteristics. A number of polymers have been used as sensitive materials in this regard, including ion-exchange resins, polymer electrolytes, hydrophilic polymers, and hygroscopic resins [141]. Hydrophilic polymers [e.g., poly(ethylene oxide) – sorbitol] swell by absorption of water. If such a polymer is filled (mixed) with carbon or metal powder of very small particle size, the resistance observed depends on the distance and the number of particles that are in contact with each other. Uptake of water leads to swelling of the polymer and disturbance of the ohmic contacts

between conducting particles, which in turn increases the resistance. For this reason the resistance correlates with humidity. Useful materials include cellulose esters and polyalcohols.

The resistance of polymers with dissociable groups (e.g., sulfonates, amines, amides, or even hydroxides) also depends on humidity. The resistance of such polymers in fact decreases with absorption of water because of ionic dissociation and an increase in ionic conduction.

The absorption of water by polymers is especially suited to the development of a capacitive humidity sensor based on the high dielectric constant of water, since capacitance depends on the area, the thickness, and the dielectric constant of a dielectric. Uptake of water causes an increase in the dielectric constant followed by an increase in capacitance. Polyimide is a suitable sensitive material, but so are certain inorganic ceramics such as Al_2O_3 [142] and low-density Ta_2O_5, which change their capacitance with the formation of water dipoles. Low-density Ta_2O_5 layers can be formed by anodic oxidation of sputtered tantalum films, whereas polyimide films are deposited by spin coating and subsequent polymerization.

Conducting Polymer-Based Gas Sensors. The above mentioned ion-selective conductometric microsensors can easily be transformed into gas sensors by omitting the ionophore and adding an organic salt for a certain ground conductivity instead. These PVC/plasticizer membranes are sensitive to reaction with certain gases that can enter the membrane phase and alter the conductivity of the organic ions either by increasing their dissociation from an ion-pair situation or by increasing their diffusional property due to membrane swelling. Figure 24 gives an example of when the salt and the plasticizer are present simultaneously. The selectivity for a certain gas can be modulated by changing the type and concentration of the latter. Therefore, much more variations are possible compared with other polymer-based gas sensors based only on their known gas chromatographic distribution behavior. A typical membrane composition is: about 50 % PVC, 30 – 35 % plasticizer, the remainder being the organic salt such as tetraalkylammonium halides. The compounds are as usual dissolved in tetrahydrofuran and then drop-coated on the interdigital electrode array with similar dimensions as in the above mentioned case of the conductometric ion-selective microsensors.

Gases that can be detected in the lower ppm range with response time of a few seconds are so far: tetrachloroethene, ethanol, benzene, toluol,
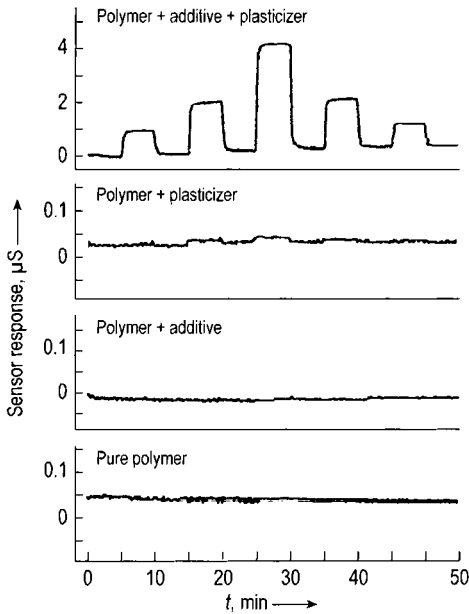
Figure 24. Variations of the sensor membrane composition. Sensor response against different concentrations of ethanol (500 ppm, 1000 ppm, 2000 ppm).
Polymer: polyisobutylene; plasticizer: ortho-nitrophenyloctylether; organic salt: tetraoctylammoniumbromide

Figure 25. Polymer membrane sensor response against tetrachloroethene. The membrane contains polyethylenoxide, 2-fluorophenyl-2-nitrophenylether, tetraoctylammoniumbromide
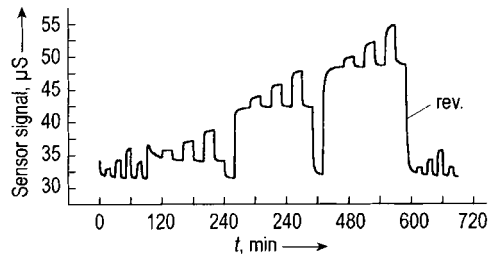
Figure 26. Humidity effect on a conductometric polymer gas sensor for ethanol.
Response toward: 350, 500, and 700 ppm ethanol at 10 %, 40 %, and 60 % relative humidity; membrane composition, 54 % polyepichlorhydrin, 36 % o-nitrophenyl-octyl-ether, 10 % tetradodecylammoniumbromide

and xylene. The measurement technique is exactly the same as with the conductometric ion sensors for application in liquids (Fig. 21). Compared with other gas-sensing devices the measurements are performed at room temperature. Thus, the energy consumption is very low, which will open the way to hand-held instruments. Unfortunately the optimization of the membrane composition for one gaseous analyte has yet to be on a purely empirical base. First studies at the ICB Muenster showed that the sensitivity and selectivity could be changed by many orders of magnitude [143], [144]. One example of a membrane with a certain selectivity towards tetrachloroethene to be deployed in the dry cleaning business is shown in Figure 25. Other examples are shown together with electronic nose developments.

Here the relevant limits of 25 ppm room concentration and 270 ppm in the cleaner before door is opened can easily be detected. Compared with the semiconductor gas sensors, humidity only changes the zero point for the measurements not the slope of the calibration curve. Figure 26 shows the humidity effect on a conductometric ethanol microsensor. Thus any placement of the sensor in

a location with similar humidity but zero analyte concentration will correct for that.

In order to solve the humidity problem of most gas sensors the ICB Muenster developed and instrument called "Air-Check" in which a short but efficient pre-sampling on Tenax material takes place. With this the sensitivity could be increased as needed, water no longer interferes and the selectivity could also be increased by analyzing the time course of the rapid thermal desorption step. The whole adsorption–desorption cycle can be performed automatically within less than 1 min!

28.2.3.1.4. Ion-Selective Field-Effect Transistors (ISFETs)

As described in Section 28.2.3.1.1, with conventional ion-selective electrodes (ISEs) the ion-
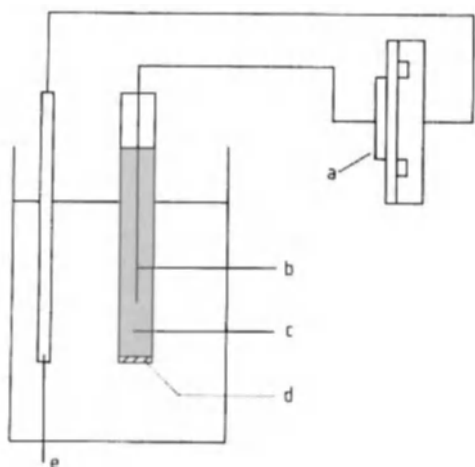
Figure 27. Schematic diagram of a membrane-electrode measuring circuit and cell assembly
a) MOSFET; b) Internal reference half cell; c) Internal filling solution; d) Ion-selective membrane; e) Reference electrode

Figure 28. Schematic diagram of an ISFET measuring circuit and cell assembly
a) Reference electrode; b) Ion-selective membrane; c) ISFET

selective membrane is placed between the analyte and the internal reference solution. An internal reference electrode makes electrical connection to the internal reference solution, and the measurement system is completed by a second reference electrode in contact with the analyte solution. In any modern pH meter the internal reference electrode is connected to the gate metal of a metaloxide semiconductor field-effect transistor (MOSFET), which acts as an impedance-converting device (Fig. 27). Normally, a conventional ISE functions very well. However, in certain applications problems can arise due to large size, high cost, and sensitive construction.

These problems might be resolved by the ionselective field-effect transistor (ISFET). In an ISFET the ion-selective membrane is placed directly on the gate insulator of the field-effect transistor; alternatively, the gate insulator itself, acting as a pH-selective membrane, may be exposed to the analyte solution (Fig. 28). If one compares an ISFET with the conventional ISE measurement system, the gate metal, connecting leads, and internal reference system have all been eliminated. An ISFET is a small, physically robust, fast potentiometric sensor, and it can be produced by microelectronic methods, with the future prospect of low-cost bulk production. More than one sensor can be placed within an area of a few square millimeters. The first ISFETs were described independently by BERGVELD [142] and MATSUO, ESASHI,

and INUMA [145]. The Institute of Microtechnology of the University of Neuchatel is very active in this field in recent days [146].

ISFET Operation. The ISFET can be regarded as a special type of MOSFET. To understand the operation of an ISFET it is therefore necessary first to understand the operation of a MOSFET.

An n-channel MOSFET (Fig. 29) consists of a p-type silicon substrate with two n-type diffusions, the source (c) and the drain (d). The structure is covered with an insulating layer (b), on top of which a gate electrode (a) is deposited over the area between the source and the drain. A voltage is applied to the gate that is positive with respect to both the source and the bulk material. This voltage produces a positive charge at the gate electrode and a negative charge at the surface of the silicon under the gate. The negative charge is formed by electrons, which are the minority carriers in the substrate, and this in turn creates a conducting channel between the source and the drain. The density of electrons in the channel, and thus the conductivity of the channel, is modulated by the applied gate voltage (V_G). If a constant voltage (V_D) is applied between the source and the drain a change in the conductivity of the channel leads to a change in current between source and drain. The drain current I_D in the unsaturated region ($V_D < V_G - V_T$) of a MOSFET is equal to:

$$I_D = \mu C_{ox} \frac{W}{L} \left[(V_G - V_T) V_D - \frac{1}{2} V_D^2 \right] \quad (15)$$

where
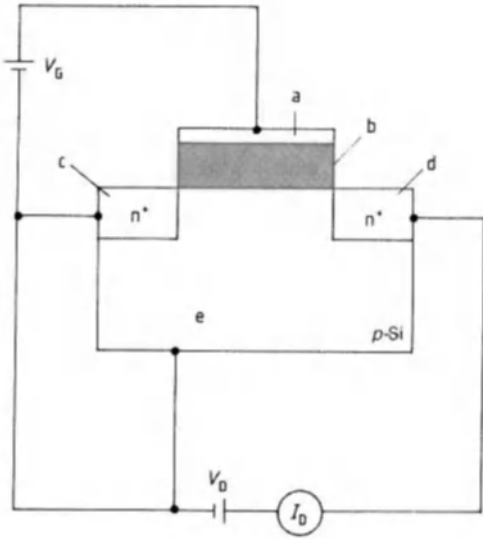
Figure 29. Schematic diagram of an n-channel MOSFET
a) Gate metal; b) Insulator; c) Source; d) Drain; e) Bulk
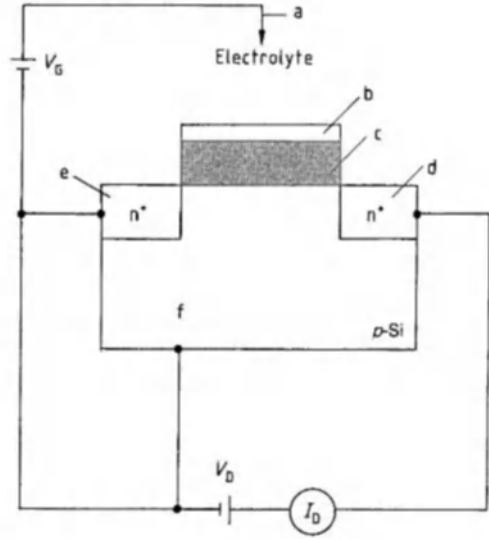
Figure 30. Schematic diagram of an ISFET
a) Reference electrode; b) Ion-selective membrane; c) Insulator; d) Drain; e) Source; f) Bulk

μ = Electron mobility in the channel
C_{ox} = Oxide capacitance per unit area
W/L = Ratio of channel width to length
V_T = Threshold voltage

A detailed description of the MOSFET is provided in [147].

In an ISFET (Fig. 30) the gate metal is replaced by a selective membrane and an electrolyte solution, or simply by the electrolyte solution, which is in contact with a reference electrode. The potential of the gate with respect to the substrate is the sum of the potentials of the various electronic and electrochemical processes that occur in the external path between the silicon substrate and the gate; i.e., the reference electrode half-cell potential E_{ref}, the solution – membrane or solution – insulator interfacial potential φ_M, and the metal – silicon contact potential. These voltages are added to the externally applied voltages V_g, and the resulting voltage has the same significance and function as that defined in the theory of MOSFET operation. V_{T*}, the threshold voltage of the ISFET, can be described as:

$$V_{T*} = V_T - E_{ref} - \varphi_M \tag{16}$$

From Equation (15) we obtain the following relationship:

$$I_D = \mu C_{ox} \frac{W}{L} \left[(V_G - V_T + E_{ref} - \varphi_M) V_D - \frac{1}{2} V_D^2 \right]. \tag{17}$$

φ_M, the solution – membrane or solution – insulator interfacial potential, depends on the activity of the ions in the analyte solution according to the Nernst – Nikolsky equation [Eq. (6), Section 28.2.3.1.1] for ISFETs with an additional membrane, or according to a similar equation (see below) for ISFETs with a solution – insulator interface.

If all parameters in Equation (17) apart from φ_M are kept constant, changes in I_D are caused exclusively by changes in the activity of ions in the analyte solution. Thus, in an ISFET, the drain current is modulated by the ionic activity in the analyte solution. A detailed description of ISFET operation is presented in [148].

pH-Sensitive ISFETs. Gate insulators such as Si_3N_4, Al_2O_3, and Ta_2O_3 have been found to be pH sensitive, so FETs in which bare gate insulators of this type are exposed to a solution respond to changes in pH. The pH response can be explained by the "site-binding theory." This model assumes that there are ionizable binding sites present at the surface of an insulator, and that

these can exhibit amphoteric behavior with respect to the potential-determining ion (i.e., H^+):

$$A-OH \rightleftharpoons A-O^- + H_S^+$$

$$A-OH + H_S^+ \rightleftharpoons A-OH_2^+$$
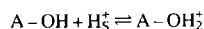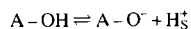
According to this model the pH dependence of φ_M is given by the following equation:

$$\varphi_M = 2.3 \frac{kT}{q} \left[\frac{\beta}{\beta+1} \right] \left(pH_{pzc} - pH \right) \qquad (18)$$

under the condition that $\beta \gg q\varphi_M/kT$. pH_{pzc} is the pH value for which $\varphi_M = 0$; β is a parameter that reflects the chemical sensitivity of the outer gate insulator, with a value dependent on the density of hydroxyl groups and the surface reactivity. Surfaces with a high β value show Nernstian behavior. For a detailed description of the site-binding model, see [148].

ISFETs Sensitive to Other Ions. Deposition of an ion-selective membrane on top of the gate insulator opens the way to measurement of ions other than H^+. Most of the sensitive materials for ISE applications described in Section 28.2.3.1.1 have been used in conjunction with IS-FETs, providing sensors covering a wide variety of species. Thus, ISFETs with solid membranes have been described with AgCl – AgBr membranes sensitive to Ag^+, Cl^-, and Br^- [149], for example, or with LaF_3 membranes sensitive to fluoride [150]. Ion-sensitive polymer-matrix membranes (liquid membranes with a polymer matrix) have also been used as sensitive membranes for ISFETs. The first experiments involved mainly PVC membranes [151], [152], but these membranes show poor adhesion and poor mechanical strength. To improve membrane adhesion, modified PVC was utilized as a matrix material [153]. The use of silicones as matrix materials has made it possible to prepare very durable ISFETs with polymer-matrix membranes [154]. Other reports describe the use of photopolymerized polymers as matrix materials [155], [156].

Problems. The commercial development of IS-FETs has been slow despite intensive research. From the standpoint of broad commercialization the ISFET is in competition with conventional ISEs with respect to price, performance, and reliability, and it is essential that some means be devised for producing ISFETs by commercially ac-

ceptable methods. Reaching this stage requires that several problems be solved, however.

ISFETs have so far been used primarily in conjunction with conventional macro reference electrodes. This fact greatly limits the potential benefits to be gained from miniaturization of the sensor. An optimal reference electrode for ISFETs should be miniaturizable and display long-term stability, and it should be subject to fabrication with microelectronic techniques. Attempts have been made to miniaturize macro reference electrodes [157], or to use modified surfaces with extremely low surface-site densities as reference FETs [158], but to date the above-mentioned requirements have not been met.

A further problem relates to encapsulation. The ion-sensitive gate area is exposed to the analyte solution, but all other parts of the sensor must be insulated from the solution. In most cases encapsulation is carried out by hand using epoxy resins, but it is impossible to produce ISFETs at low cost by this technique. Both photolithographic methods [159], [160] and an electrochemical method [161] have been proposed for the solution of this problem. One very promising development is the backside-contacted ISFET [162]. The electrical contacts in this case are protected from the analyte solution by means of O-rings, thereby circumventing the need for resins.

As noted previously, ISFETs selective to ions other than H^+ are prepared by deposition of an additional membrane on top of the gate insulator. If such ISFETs are to be commercialized this membrane must be deposited by a mass-production technique, which will be difficult to achieve with polymer membranes. One possible solution is the development of photolithographically patternable ion-sensitive membranes [163], but such membranes have so far been developed only for a few ions, and problems in the photolithographic deposition of more than one type of membrane on a single chip have not been resolved. The interface between an insulator and a sensitive membrane is also ill-defined in a thermodynamic sense, but this problem was solved by introducing a hydrogel between the sensing membrane and the insulator. Futhermore, ISFETs suffer from drift. The influence of this effect would be eliminated if the sensor were to be used in flow-injection analysis (FIA), so successful applications can be anticipated for ISFETs in the field of dynamic measurement [164].

Only if all the problems cited are successfully resolved will the ISFET live up to the promises
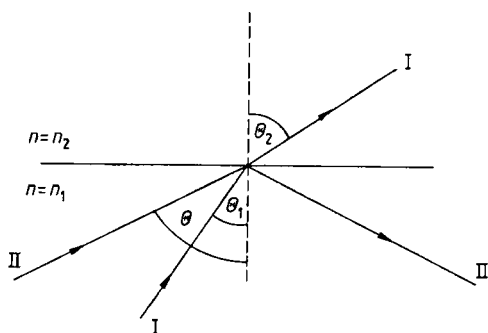
Figure 31. Refraction and total internal reflection
n = Refractive index

that have been held out for it for more than two decades: low cost, multifunctionality, and a robust and reliable construction.

28.2.3.2. Optical Sensors

28.2.3.2.1. Fiber-Optical Sensors

In recent years the use of optical fibers has become increasingly important in the field of chemical and biochemical sensors. This is clearly due to the fact that fiber-optical sensors combine the strengths of well-known bulk optical analytical methods with the unique advantages of optical fibers. More specifically:

1) An enormous technology base already exists for familiar optical methods of chemical analysis. There is probably no class of chemical analyte that has not at some time been the subject of optical determination through absorption or fluorescence spectroscopy. Methods also exist for labeling a target analyte with an appropriate chromophore or a fluorophore, a common practice in immunoassays.

2) Fiber-optic techniques offer increased sensitivity relative to conventional bulk optic approaches. For example, evanescent wave (EW) spectroscopy is significantly more sensitive than bulk attenuated total reflection spectroscopy. Fiber-optical EW spectroscopy is the only technique suitable for use with highly absorbing or scattering media.

3) The interrogating light in a fiber-optics system remains guided, so no coupling optics are required in the sensor region, and an all-fiber approach is feasible. This represents a sharp

contrast to bulk optics, where proper optical adjustment is extremely critical. Fiber optics also provides a good solution to the problem of wide separation between the sensing area and the detector, as in the in vivo monitoring of medical parameters.

4) Recent developments in optical communications and integrated optics have led to optical materials that permit significant miniaturization of the sensing and detecting devices.

5) Since the measurement signal is optical, electrical interference is avoided, and the absence of electrical connectors ensures safe operation in hazardous environments, such as those containing explosive vapors.

The light inside an optical fiber is guided by the principle of total internal reflection, as described by Snell's law and depicted in Figure 31. A ray of light striking the interface between two media with differing refractive indices n_1 and n_2 is refracted (ray I) according to the relationship $\sin(\Theta_1) \times n_1 = \sin(\Theta_2) \times n_2$. If the condition $n_1 > n_2$ is fulfilled, total internal reflection will be observed for several angles of incidence. Thus, at an angle close to the horizontal the incident beam will no longer be refracted into the second medium, but will instead be totally reflected internally at the interface (ray II in Fig. 31). The *critical angle of incidence* $\Theta_c = \arcsin(n_2/n_1)$, which defines the onset of total internal reflection, is determined by the refractive indices of the two media, and whenever $\Theta > \Theta_c$ light is guided inside the medium characterized by n_1.

The general structure of an optical fiber is illustrated in Figure 32. Light entering from the left is confined within the core, which is surrounded by a cladding of lower refractive index (n_{clad}). The whole arrangement is surrounded by a nonoptical jacket, usually made from plastic and designed to provide both stabilization and protection. The fibers themselves can be constructed from a wide variety of transparent materials with differing refractive indices, including various glasses, fused silica, plastics, and sapphire. Most commercially available fibers have glass cores with n_{core} in the vicinity of 1.48 and diameters ranging from $5-200$ μm; typical claddings have $n_{clad} < 1.48$ and thicknesses of $50-1000$ μm. An optical fiber is usually characterized by its numerical aperture $NA = (n_{core}^2 - n_{clad}^2)^{1/2}$. The value of NA determines the cone angle Θ_c within which incident light will become entrapped by the fiber, where $\Theta_c = \arccos(NA)$; $\Theta_{max} = \arcsin(NA/n_{env})$
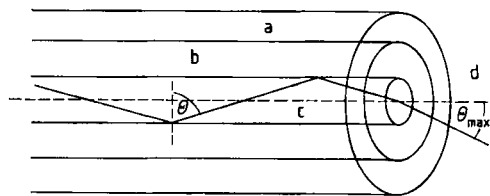
Figure 32. Structure of an optical fiber
a) Jacket; b) Cladding, $n = n_{clad}$; c) Core, $n = n_{core}$; d) Environment, $n = n_{env}$

for the cone angle Θ_{max} with which light leaves the fiber at its end (see Fig. 32).

The simple ray-optical description of light-guiding presented above fails to account for several phenomena important with respect to fiber-optical sensors. The most important of these is the *evanescent wave*. The intensity of light that is totally reflected internally does not in fact fall abruptly to zero at the core – cladding interface. Instead, the electromagnetic field intensity decays exponentially as a function of distance from the interface, thereby extending into the medium of lower refractive index. This field is called the *evanescent field*, and its penetration depth d is defined by the distance from the interface within which it decays to a fraction $1/e$ of its value at the interface. The angle of incidence Θ strongly influences the penetration depth. Values for d vary from $50 - 200$ nm, increasing as Θ decreases; d reaches its maximum when Θ is very close to the critical angle.

Fiber-optical sensors can be constructed in either of two fundamentally different optical configurations with respect to the sensing area. The first is a wave-guide-binding configuration, in which recognition occurs at the surface of the fiber core via the evanescent field. This is achieved by removing the jacket and cladding from the fiber and causing the analyte molecules to bind to the core surface, as in the example described below. Such an arrangement is described as an *evanescent wave sensor* or *intrinsic fiber-optical sensor*. The resulting sensitivity depends greatly on the value of NA [165].

Alternatively, the fiber-end configuration might be considered, in which analyte is released for recognition purposes into the illuminated space at the end of the fiber. In this case the fiber acts only as a light pipe, resulting in what is known as an *extrinsic fiber-optical sensor*.

Both configurations entail similar optical systems: a spectral light source for excitation (e.g., a
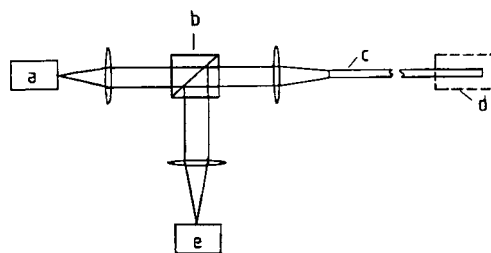
Figure 33. General configuration of an optical detection system for a fiber-optical sensor
a) Source; b) Beam splitter; c) Fiber; d) Sensing area; e) Detector
The excitation system consists of a xenon or tungsten lamp together with a monochromator or optical filter, an LED, or a laser, whereas the light detection system might be a monochromator or filter with a photomultiplier or photodiode.

xenon or tungsten lamp equipped with a monochromator or optical filter, an LED, or a laser) and a spectral light-detection device consisting of a monochromator or filter together with a photomultiplier or photodiode. A schematic diagram applicable to the optical systems for these two sensor configurations is presented in Figure 33.

More detailed information regarding the physical principles of fiber-optical chemical and biochemical sensors and the corresponding detection systems is available in [166] and [167].

Biochemical Sensors. Antibodies are widely used as recognition elements in fiber-optical biosensor applications. This is due to the high specificity of the binding reaction between an antibody and an analyte (antigen). To determine this reaction optically, labeling agents such as fluoresceine, rhodamine, or phycoerythrine are used, all of which show absorption and emission bands in the visible region of the spectrum. Linear amplification can be achieved by binding more than one label. Depending on the circumstances, either the antigen or the antibody might be tagged (see below).

The evanescent wave (intrinsic) configuration is most important for biosensors because it permits the antibody or antigen to be immobilized directly on the core surface, and it extends the path for possible interaction between the light and the recognition elements. Accordingly, only tagged molecules inside the evanescent field are subject to excitation. The amount of light energy absorbed, or the light emitted from these molecules and collected by the fiber, gives rise to the measurement signal. In contrast to the extrinsic sensor
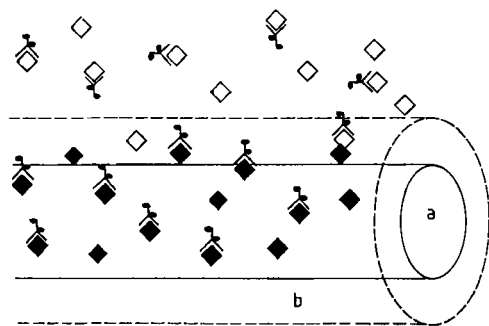
Figure 34. A fiber-optical immunosensor, in which labeled antibody molecules are gradually replaced by free analytes, causing the fluorescence signal to decrease
a) Core; b) Region of evanescent wave
◆ Fixed analyte; ◇ Free analyte;

Tagged antibody

approach, the intrinsic arrangement therefore reduces noise that would otherwise result from free labeled molecules or intrinsic fluorescence within the environment [167]. Despite these advantages, however, extrinsic biosensors have also been constructed, with binding devices placed in front of the fiber as a way of fixing the recognition elements. These devices can be either glass plates or transparent membranes. This offers the advantage of easy removal and replacement of the recognition elements [168].

Figure 34 provides an example of a biosensor of the intrinsic type. Here the antigen has been immobilized at the fiber core surface where it has been allowed to bind with a labeled antibody. Energy from the evanescent wave is absorbed by the labels, generating a detectable fluorescence signal. Addition of unlabeled analyte to the surrounding solution leads to competition for the binding sites. This permits some of the labeled antibody molecules to diffuse away from the surface, causing the fluorescence signal to decrease. The same general approach is also applicable to immobilized antibodies and tagged antigens [169].

Competition between a labeled antigen and an unlabeled analyte can also be effected in an external column, where the labeled antigen has previously been bound to an antibody. Once the competition reaches equilibrium any labeled antigen that has been released can be detected with antibody immobilized on a fiber. In this case the fluorescence signal would be expected to increase with an increased concentration of analyte.

Enzymes are in some cases also useful for recognition and labeling purposes. For example, glucose oxidase has been used to determine glucose on the basis of free oxygen that is consumed, where the disappearance of oxygen is measured with a fiber-optical oxygen subsensor [170]. Labeling of antibodies or antigens with urease or alkaline phosphatase and subsequent addition of a fluoro- or chromogenic substrate leads to detectable production of fluorophores or chromophores. The signal in this case is subject to multiple enhancement since it is proportional to the concentrations of both labeled antigen and antibody as well as to incubation time.

Chemical Sensors. Recognition with a *chemical sensor* is achieved via an analyte-specific reaction involving chemical compounds located inside the sensing area. The reaction must be one that is accompanied by changes in absorbance or fluorescence characteristics. Several examples are described below. A common physical arrangement utilizes a fiber-end approach and two single fibers or two fiber bundles, one for guiding light into the sensing area and the other to carry it away. The chemical reagents are confined within a transparent membrane or a small segment of the fiber itself, near its end and shielded by a membrane.

A particularly wide variety of fiber-optical sensors has been developed for measuring pH values, and many of these sensors are also used as subsensors for other analytes. Sensors for pH determination can be divided into two general types. The first relies on colorimetric acid – base indicators to produce changes in absorbance. For example, the dye phenol red can be copolymerized with acrylamide and bisacrylamide to yield dyed polyacrylamide microspheres. These are in turn packed into a length of cellulose dialysis tubing placed at the end of a pair of fibers that can serve as a vehicle for monitoring changes in absorbance [171]. Bromothymol blue immobilized on styrene – divinylbenzene copolymer has been similarly employed [172]. Alternatively, fluorometric acid – base indicators can be used to detect changes in fluorescence intensity that accompany protonation. Thus, one pH sensor was based on fluorescence quenching by H^+ of the excited state of immobilized fluoresceinamine [173]. 8-Hydroxypyrene-1,3,6-trisulfonate (HPTS) fixed on ethylene vinyl acetate has also been invoked as a fluorescence pH sensor [174].

Chemical sensors for other ions have been reported as well, including the halides and
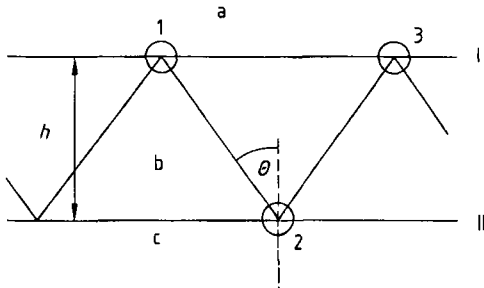
Figure 35. Propagation of a confined light ray within a planar waveguide mounted on a substrate
Most substrate materials are glasses or crystals with a refractive index of 1.4–2.4, and the cover material is air; the thickness of the film ranges from 0.1 to several micrometers.
a) Cover, $n = n_c$; b) Film, $n = n_f$; c) Substrate, $n = n_s$

pseudohalides [175]. A sensor of this type is based on the dynamic quenching of fluorescence emission from glass-immobilized, heterocyclic, acridinium and quinolinium indicators. The fluorometric process involved is known to obey the Stern–Volmer equation [176]. Al(III) ions have been determined with morin (3,5,7,2′,4′-pentahydroxyflavone) [177]. Thus, aluminum displaces the hydrogen ion when it binds to morin, leading to a change in fluorescence properties. Be(II) has also been determined using morin [178]. Furthermore, reports have appeared on sensors for Mg(II), Zn(II), and Cd(II) based on quinoline-8-ol sulfonate immobilized electrostatically on an anion-exchange resin [179].

Gas sensors for carbon dioxide [180] and ammonia [181] in aqueous solution have been constructed on the basis of a gas-permeable membrane and an optical pH subsensor. Gases diffusing through the membrane produce a change in the pH in the sensing region of a fiber-optical pH sensor. Sulfur dioxide [182] and oxygen [183] sensors take advantage of the quenching of the energy transfer (Förster transfer) between excited pyrene and perylene, both of which can be immobilized on a silicone matrix. Other O_2 sensors rely on the effect of fluorescence quenching as described by the Stern–Volmer equation. These generally involve metal complexes like tris(2,2′-bipyridine)ruthenium(II) [184]. In most cases fluorescence quenching also leads to a shortening of the fluorescence lifetime, which can be determined, for example, by phasesensitive detection [185]. The advantage of this method is that fluorescence lifetime is independent of the fluorophore concentration, and there is no signal drift due to bleaching.

28.2.3.2.2. Integrated Optical Chemical and Biochemical Sensors

Generally speaking, integrated optical devices can be constructed using techniques similar to those for manufacturing semiconductor devices. This offers great potential for cost reduction and facilitates the miniaturization often required in sensor applications. Furthermore, such sensors retain all the advantages of the well-known fiber-optical sensors.

The propagation of a light ray confined within a planar dielectric film (or *waveguide*) is depicted in Figure 35. Guidance of the light is possible only if total internal reflection occurs at the interfaces film–cover (I) and film–substrate (II). This condition is met if the refractive index of the film n_f is greater than that of both the cover n_c and the substrate n_s. Moreover, the angle of incidence Θ must be greater than the critical angles for total internal reflection at the interfaces I (Θ_{cI}) and II (Θ_{cII}), given by the expressions $\sin(\Theta_{cI}) = n_c/n_f$ and $\sin(\Theta_{cII}) = n_s/n_f$. To fulfill the self-consistency condition within the waveguide, the phase shift of the light wave as it travels from point 1 to point 3 must be an integral multiple m of 2π, where m is called the *mode number*.

$$\left(\frac{2\pi}{\lambda}\right) \cdot L(n_f, h, \Theta) + \Phi_I(n_f, n_c, \Theta)$$
$$+ \Phi_{II}(n_f, n_s, \Theta) = 2\pi m \qquad (19)$$

Φ_I and Φ_{II} are the phase shifts caused by reflection at the interfaces I and II, and $(2\pi/\lambda) L$ is the shift accompanying travel along the optical path $L = 2h\cos(\Theta)$. The values of n_f, n_c, n_s, and m are usually fixed by the fabrication process for the waveguide, and λ is established by the light source. The value of Θ can therefore be calculated from Equation (19). It is useful to define an *effective refractive index* n_{eff} for the guided mode, defined as $n_{eff} = n_f \sin(\Theta) = n_{eff}(n_f, n_c, n_s, h, m, \lambda)$, which, apart from λ, depends only on the material and geometric parameters of the waveguide. The parameter n_{eff} can be used to treat all types of light-propagation effects within the waveguide.

Even though light is totally reflected internally at the interfaces, its intensity does not fall to zero outside the waveguide. Just as with optical fibers, the small part of the guided light located outside the waveguide is referred to as the evanescent field, with a penetration depth ranging from 100–300 nm depending on the wave-guide performance. A more detailed mathematical treat-

a

d_1 d_2

h   b

c
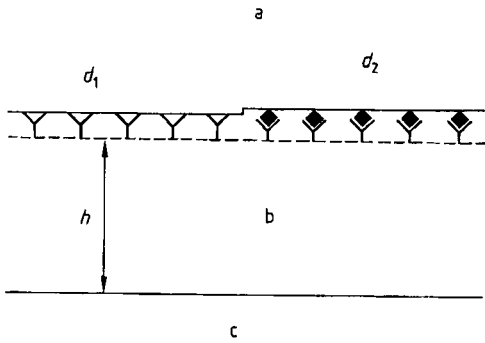
Figure 36. Specific antibody – antigen binding at the surface of a planar waveguide, causing the thickness of the adlayer to change
a) Cover, $n = n_c$; b) Film, $n = n_f$; c) Substrate, $n = n_s d_1$, d_2 = Layer thicknesses before and after binding

Figure 37. General arrangement of a Mach – Zehnder interferometer (A) and an input grating coupler (B)
In A only the signal arm is exposed to an active chemical layer while the reference arm is covered, whereas in B the grating is provided with an active layer and exposed to the sample.
a) Input beam; b) Output beam; c) Incoupled beam

ment of planar and strip waveguides is presented in [186].

Waveguides intended as chemical or biochemical transducers can be operated in either of two modes. A comprehensive physical description of these operating modes and the corresponding efficiencies is provided in [187].

First, it is possible to use a waveguide for acquiring refractive index data. This can be accomplished by coating the waveguide with a reactive but transparent chemical layer of a thickness greater than the penetration depth of the evanescent field. In this case the layer can be treated as if it were a cover material. Any chemical reaction with the layer will result in a change in the refractive index n_c, which in turn leads to a change in the effective refractive index n_{eff}. This approach has been used successfully in the production of sensors for CO_2 and SO_2 [188].

The adsorption and desorption of molecules is depicted in Figure 36 as an example of the second operational mode. The waveguide in this case is coated with a thin chemical layer consisting, for example, of an antibody. When the antibody – antigen binding reaction occurs, the effective thickness [187] of the entire waveguide increases, which again causes a change in the effective refractive index of the guide. Direct immunosensors based on this method have often been reported in the literature [189], [190].

Detecting a change in n_{eff} is a problem that can be solved in various ways. An interferometer approach analogous to the Mach – Zehnder interferometer [191], [192], the Fabry – Perot interferometer [188], or the difference interferometer [193] has often been used. The

general arrangement of a Mach – Zehnder interferometer is illustrated in Figure 37 A. The diameter of the strip waveguide is usually on the order of 5 µm, with a distance between the two arms of < 1 mm, and a total length for the device of < 3 cm. Only the signal arm with its reactive chemical layer is exposed; the reference arm is covered by a dielectric film with a refractive index lower than that of the waveguide. The output power P_{out} of the device depends on the extent of the light phase shift $\Delta\Phi = (2\pi/\lambda) L n_{eff}$ between the signal arm and the reference arm; e.g., $P_{out} = P_{in} \cos^2(\Delta\Phi/2)$.

Grating couplers can also be used to determine the value of n_{eff}, as depicted schematically in Figure 37 B. For efficient light *incoupling* the angle α must be consistent with the expression $n_{eff} = n_0 \sin(\alpha) + l(\lambda/\Lambda)$, where l is the diffraction order and Λ the grating period. According to this relationship, a change in n_{eff} leads to a change in the angle α that will reinforce light guided within the waveguide. A grating coupler can also be used in an *outcoupling* arrangement. In this case it is the angle for the outcoupled light that would conform to the grating equation. A detailed description of grating-coupler methods is presented elsewhere [194].

28.2.3.2.3. Surface Plasmon Resonance

Transducers for optical sensing systems in chemical and biochemical sensors have also been developed based on the sensitivity of surface plasmon resonance (SPR) devices. Since first described in 1983 [195], SPR transducers have been used in many different applications, including immunoassays [195], [196] and sensors for gases [195], [197] and liquids [198].
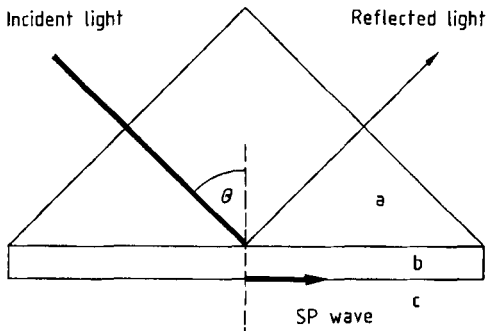
Figure 38. Attenuated total reflection coupler used to excite surface plasmons at the metal – dielectric interface; excitation is observed as a minimum in the intensity of the reflected light by scanning the angle of incidence
a) Prism; b) Metal; c) Dielectric

Electron charges located at a metal – dielectric interface can be caused to undergo coherent fluctuations known as *surface plasma oscillations* or *surface plasmons* (SP). The frequency ω of such longitudinal oscillations is related to the wave vector k_{sp} by *the dispersion relationship.*

$$k_{sp} = \left(\frac{\omega}{c}\right) \cdot \sqrt{\frac{\epsilon_M \cdot \epsilon_D}{\epsilon_M + \epsilon_D}} \qquad (20)$$

where $\epsilon_M(\omega)$ is the permittivity of the metal and $\epsilon_D(\omega)$ that of the surrounding dielectric (cf. Fig. 38); c is the velocity of light in a vacuum. The permittivity of the metal ($\epsilon_M = \epsilon_{M1} + i \cdot \epsilon_{M2}$) is complex and the value of $\epsilon_D = n_D^2$ is real, where n_D is the refractive index of the dielectric. A surface plasmon can be excited only if the conditions $\epsilon_{M1} < 0$ and $|\epsilon_{M1}| > \epsilon_{M2}$ are fulfilled. For example, gold, silver, and palladium are potentially applicable metals with negative values of ϵ_{M1} for frequencies corresponding to visible light [199].

SP excitation can be accomplished in several ways. A very simple and often used approach is illustrated in Figure 38. RATHER [199] has described this as the *attenuated total reflection* (ATR) coupler. Here the base of a prism is coated with a thin metal film approximately 50 nm thick. Incident monochromatic light is reflected at this interface in such a way that a projection of the light vector k_0 on the interface becomes $k_x = k_0 n_p \sin(\Theta)$. If the resonance (Re) condition $k_x = Re\{k_{sp}\}$ is fulfilled, the evanescent field of the incident light excites an SP at the metal –dielectric interface, although this is only possible with light polarized parallel to the plane of incidence (p-polarized). The excitation is observed as a sharp minimum in the intensity of the reflected light,

and it can be detected by scanning the incident angle.

Using such an ATR device, small changes in $\epsilon_D = n_D^2$ and the build-up of a thin layer at the metal – dielectric interface can easily be detected by measuring the shift in the *resonance angle* Θ.

For an immunosensing application the metal layer would be coated with appropriate antibodies, either immobilized covalently or embedded into a matrix layer. Upon exposure to a specific antigen the resonance angle should shift relative to that observed in an analyte-free environment due to the increase in layer thickness that accompanies binding of the analyte. For example, a thickness change of 0.1 nm results in a 0.01° change in the resonance angle, a shift that can be measured quite accurately (see also Section 28.3.3.1).

If an ATR device is to be used for chemical sensing, the metal surface must be coated with a thick dielectric layer of a reactive chemical. Chemical reaction within the layer then produces changes in n_D that can be detected by the corresponding shift in Θ.

28.2.3.2.4. Reflectometric Interference Spectroscopy

Gauglitz introduced reflectometric interference spectroscopy (RIFS) into the field of chemical and biochemical sensing [200], [201]. Figure 39 shows the underlying optical principle. The interference of the schematically shown partial light rays 2 and 3 leads to an intensity modulation of the overall reflected light spectra. This must be happening within the so-called coherence length of light after leaving the substrate. The coherence length of light using, e.g., a tungsten light bulb in the wavelength range of 400–1000 nm is about 20 μm. Therefore, the thickness of the interference causing layer has to be less than 20 μm and has to exceed the wavelength of the used light to give a measurable interference pattern. A layer thickness between of 0.5 and 10 μm with $n_2 \approx 1.4$ fits well this demand.

Light rays which undergo multiple reflections at the different interfaces contribute little because of their lower intensity. The scattered light should be as low as possible as with light from other sources. The equation for the interference pattern is:
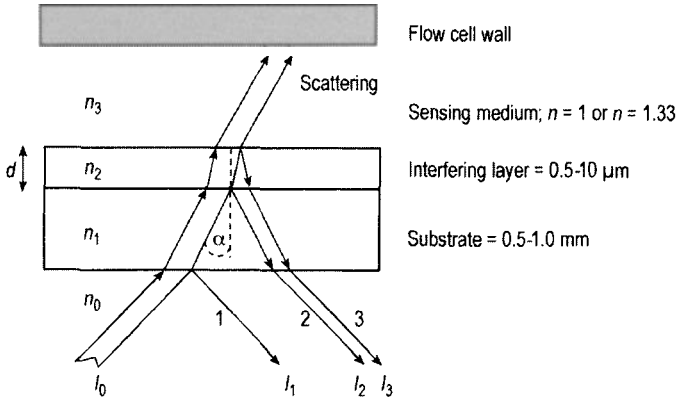
Figure 39. Schematic sketch of the principles of RIFS in a flow through set-up. Schematically shown light rays I_2 and I_3 can be re-collected with a fiber optic and lead to a monochromator with a CCD detector
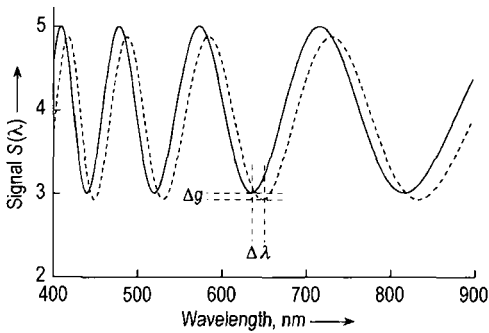
Figure 40. Observed interference pattern of RIFS normally recorded via a CCD array; comprehensive software [202] is needed to gain the needed information from this spectrum

$$I_{all}(\lambda) = I_1 + I_2(n_2) + I_3(n_2) + 2$$
$$\cdot (I_2(n_2) \cdot I_3(n_2))^{1/2} \cdot \cos\{(4\pi/\lambda) \cdot (n_2 \cdot d)\}$$
$$(20a)$$

Figure 40 shows a typical interference pattern after an alteration of the refractive index and the thickness of the interference layer. The shift $\Delta\lambda$ is caused by Δd and Δn_2 whereas the chance in the modulation depth Δg is caused by Δn_2 only.

Thus, the basic response of an RIFS based polymer or immunosensor is the alteration of the optical pathway. For the evaluation of the interference pattern several algorithms have been developed. After the determination of the minima and maxima the interference order k is determined using a series of calculations and curve fitting techniques using newly developed software. Before a measurement is carried out a reference spectrum is obtained from an uncoated sensor includ-

ing all spectral information except for the interference pattern. All further spectra obtained are divided by the reference spectrum to obtain the natural interference pattern.

28.2.3.3. Mass-Sensitive Devices

28.2.3.3.1. Introduction

Mass changes are usually detected with the aid of a gravimetric device, such as a balance. Ideally, balance sensitivity can be extended down into the microgram range, but the desire for further improvements in sensitivity and miniaturization of the experimental setup has led to the development of other mass-sensitive devices: the so-called quartz microbalance (QMB), which is the best-known system, and other, more recent devices such as the surface acoustic wave oscillator (SAW), the Lamb wave oscillator (LW), and the acoustic plate-mode oscillator (APM), which represent the present state of the science (Section 28.2.3.3.2). These devices are all based on the *piezoelectricity* of solids, (see Section 28.2.3.3.3), a phenomenon that permits the generation of vibrations or the propagation of waves. A piezoelectric substrate, usually a quartz plate, is stimulated to vibration by an oscillating electric circuit. This corresponds to an interconversion of electrical and mechanical vibrations (deformations of the crystal), resulting in a device that can be regarded as either an oscillator or a resonator.

Bulk waves, surface waves, and waves of other types (see Section 28.2.3.3.3) are associated with characteristic resonance frequencies correspond-

ing to the most stable frequencies of vibration. Such a resonance frequency is itself highly dependent on the mass of the oscillating plate. A slight change in the mass of the plate will result in a considerable change in the resonance frequency, which can be measured very accurately. An oscillator of this type can therefore be used as a mass-sensitive transducer. Other factors that may influence the observed frequency shift include changes in the viscosity and density of surface layers as well as such physical parameters as forces, temperature, and pressure.

To construct a mass-sensitive sensor the surface of a piezoelectric material must be coated with an appropriate chemically or biochemically reactive layer capable of interacting with the proposed analyte and — ideally — recognizing it as well (see Section 28.2.3.3.4). Such an interaction causes the mass of the crystal to increase, resulting in a measurable shift in the resonance frequency. A relationship thus exists between the amount of analyte adsorbed or absorbed and the frequency shift of the oscillating system. Resonance frequency changes can be established very accurately, so very small changes in mass can be detected. A theoretical approach applicable to simple cases is described in [203].

To date, mass-sensitive devices have been used mainly for sensing analytes in the gas phase; they are also applicable in principle to liquids, but this presupposes a more sophisticated physical arrangement, a much more complicated signal control, and very careful interpretation. Many published papers have dealt with the potential use of QMB, SAW, and, more recently, APM and LW devices as mass-sensitive detectors in chemical and biochemical sensors, but no such device has yet established itself commercially — with the exception of QMBs employed for thin-layer thickness measuring, which cannot be regarded as chemosensors in the strictest sense of the term [204].

Historical Development. The first report concerning the use of piezoelectric resonance from a quartz crystal in a gravimetric microbalance is by J. STRUTT (Lord RAYLEIGH), who in 1885 described a shift in resonance frequency accompanying an infinitesimal mass change in a mechanical oscillator [205]. Fundamental studies into the mass sensitivity of quartz resonators and their potential application for measuring the thickness of thin layers were reported in 1959 by SAUERBREY [203], [206]. KING was the first to use a quartz microbalance "QMB" (in 1964) as a gas-phase

detector in an analytical application [207], [208]. Numerous more recent papers discuss the fundamentals [209] – [216] and applicability [217] – [220] of bulk acoustic wave (BAW) piezoelectric resonators as chemical and biochemical sensors in the gas phase as well as in liquids, but physical devices for the measurement of pressure, temperature, or distance have attracted greater commercial interest [221], [222]. Nevertheless, quartz microbalances based on bulk acoustic waves do find frequent use in analytical research and development.

RAYLEIGH in an early paper [205] also noted the existence of waves that are propagated only in relatively thin surface layers, but it was not until 1979 that WOHLTJEN and DESSY adapted these so-called surface acoustic waves (SAWs) for incorporation into a chemical gas sensor [223] – [225]. The chief advantages of SAWs are increased sensitivity relative to BAWs of lower resonant frequencies and a greater potential for miniaturization. On the other hand, SAWs are associated with a variety of circuitry problems resulting from the relatively high operating frequencies required for adequate sensor sensitivity. A modification of this type of wave was introduced by CHANG and WHITE [226], [227], who adapted a SAW sensor by creating an exceptionally thin membrane-like piezoelectric region in the substrate directly beneath the acoustic path. This in turn led to the production of *Lamb waves*, which proved to be particularly suitable for chemical sensor applications, opening the way to the so-called plate-mode oscillator, a device representing the present state of the art.

28.2.3.3.2. Fundamental Principles and Basic Types of Transducers

Solids and other vibrationally active mechanical structures are capable of being stimulated by electrical, thermal, or optical means either to vibration or to the further propagation of waves. Piezoelectric solids, preferably quartz (SiO_2), are especially susceptible to emission and reception of mechanical vibrations and waves. The anisotropic nature of quartz supports various types of waves, each with its characteristic frequency, amplitude, and propagation properties.

Survey of the Different Types of Piezoelectric Substrates. A summary of the various types of transducers involving piezoelectric solids, includ-
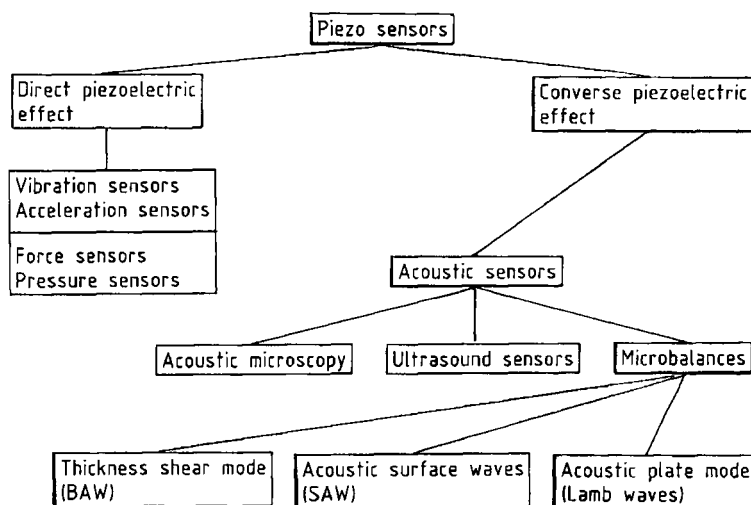
Figure 41. Selected applications of piezoelectric transducers with respect to both the direct piezoelectric effect (active transducer) and the converse piezoelectric effect (passive transducer) [222] (with permission from Springer-Verlag, Heidelberg)

ing their chemical and physical applications, is presented in Figure 41. Acoustic gravimetric sensors can be categorized generally on the basis of their design, their wave mode (BAW, SAW, APM, LW), and their frequency range. Figure 42 offers a schematic overview of the principles and design characteristics of various piezoelectric transducers. Criteria for establishing the usefulness of a particular vibrational system for the determination of mass include [209]:

1) The system should be readily excitable, preferably by electrical means
2) The system must lend itself to coupling with a frequency- (or period-) measuring device without significant disturbance
3) The system must be associated with sharply defined resonant frequencies so that the frequency (or period) of vibration can be determined precisely within a reasonably short period of time
4) In order to provide the required mass sensitivity, the mass-induced change in resonant frequency must be greater than the instability level of the resonant frequency and such uncertainties in the frequency (or period) as may accompany the measurement technique employed
5) Resonant frequency changes due to environmental disturbances (e.g., fluctuations in temperature or pressure, the presence of electrical or magnetic fields, or external mechanical

stress) must be small relative to those caused by mass changes
6) An equation must be available in analytical form relating mass changes to the corresponding shifts in resonant frequency

Bulk Acoustic Wave Transducers. Waves of the bulk acoustic type (BAWs) can be invoked for microgravimetric applications through the use of either AT- or BT-cut plates derived from a single crystal of α-quartz, where "AT" and "BT" refer to cuts with specific orientations relative to the main axis of the crystal (see below and [229]). Figure 43 A represents a schematic diagram of such a quartz crystal with its associated electrodes and sensing layers. This diagram also indicates how a species to be detected might be absorbed into the sensing layer. Both sides of the quartz are in contact with electrodes. The corresponding electrodes are usually 300–1000 nm thick and 3–8 mm in diameter, and are made of gold, nickel, silver, or aluminum [217]. Introduction of oscillating electrical energy through an appropriate electrical circuit causes the mechanical structure to begin vibrating. In the case of AT-cut quartz the crystal vibrates in the thickness shear mode, as shown in Figure 43 B (see also Fig. 58).

Theory. Microgravimetric application of acoustic devices presupposes a quantitative relationship between an observed relative shift in the resonance frequency and an added mass. In the case of a quartz plate vibrating in a liquid an
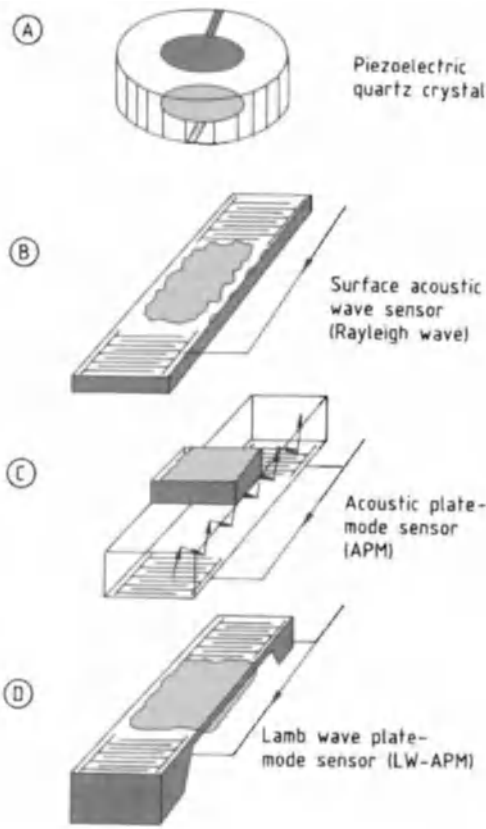
Figure 42. Schematic overview of acoustic gravimetric devices, illustrating the principles of wave propagation and design [228] (with permission from VDE-Verlag, Berlin)
A) BAW; B) SAW; C) APM; D) LW substrates

Figure 43. Schematic diagram of a quartz crystal augmented with electrodes and sensing layers, illustrating the way in which the species to be detected is absorbed into the sensing layer (A) and the ideal thickness shear mode of a quartz plate (B) [230], [203]
a) Quartz crystal; b) Electrodes; c) Sensing layers x, y, z = Coordinates of the quartz plate; X, Y, Z = (IEEE-standard) coordinates for a single crystal of α-quartz, where X is the polar axis and Z is the optical axis; t_q = Thickness of the quartz plate, where the wavelength $\lambda = 2 t_q$; ξ = Shear amplitude; ϑ = Angle of rotation

$$\Delta f_q = -f_q \frac{\Delta m}{m} \tag{21}$$

which finally provides the following expression for an AT-cut crystal:

$$\Delta f_q = -2.3 \times 10^6 f_q^2 \frac{\Delta m}{A} \tag{22}$$

where f_q is expressed in megahertz, Δm in grams, and A in square centimeters. Since f_q for a quartz crystal is typically 10 MHz, and a frequency change of 0.1 Hz can be readily detected by modern electronics, it is possible in this way to detect mass changes of as little as about 10^{-10} g/cm^2 in the gas phase (!) [230].

LU has developed a model involving propagation velocities for sound waves in the substrate (for quartz, v_q) and in the sensing layer (v_f) [216]. Reflection and refraction occur at the interface between the crystal (q) and the film (f) (Fig. 44) analogous to the optical reflection and refraction observed at a boundary between two materials with different optical densities. The shear velocities in the quartz crystal and the film coating, respectively, are:

quartz $v_q = \sqrt{\dfrac{\mu_q}{\varrho_q}}$ $\qquad(23)$

film $v_f = \sqrt{\dfrac{\mu_f}{\varrho_f}}$

extended approach is required with respect to changes in the viscosity and density. The first useful relationship of this type was developed by SAUERBREY, who treated the added mass as an "added thickness" of the oscillator. The relationship itself was based on a quartz crystal vibrating in its thickness shear mode in the gas phase (Fig. 44). The observed oscillation frequency (f_q) is inversely proportional to the thickness t_q of the crystal; $f_q = N/t_q$, where N is a frequency constant ($N = 0.168$ MHz cm for AT-cut quartz at room temperature). The mass of the crystal is $m = \varrho A t_q$, where ϱ is the crystal density and A its cross-sectional area. In this derivation the mass increment due to a foreign mass Δm is treated as an equivalent change in the mass of the crystal. Invoking the approximation $\varrho_q = \varrho_f$ leads to the equation:

Figure 44. Shear wave (with permission from Plenum Publishing, New York)
A) In a quartz crystal; B) After an incremental increase in the crystal thickness; C) With the addition of a filmm_q, t_q = Mass and thickness of the quartz; m_f, t_f = Mass and thickness of the film

Given a knowledge of the shear moduli μ_f and μ_q, Equation (23) can be used to calculate the shear velocity of the wave. Ideal wave-propagation behavior is observed with a crystal upon which such a film has been deposited provided there is a negligible amount of frictional loss. In this case the resonance condition is valid, which leads to Equation (24):

$$\tan\left(\frac{\pi \cdot f_c}{f_q}\right) = \left(\frac{\varrho_f \cdot v_f}{\varrho_q \cdot v_q}\right)\tan\left(\frac{\pi \cdot f_c}{f_f}\right) \qquad (24)$$

where $f_c = \omega/2\,\pi$ is the resonance frequency of the crystal upon which material has been deposited. The terms $\varrho_f \cdot v_f$ and $\varrho_q \cdot v_q$ are the acoustic impedances Z_f and Z_q of the film and crystal, respectively. Their ratio, $Z = Z_f/Z_q$, is an important parameter with respect to acoustic matching of the materials. For optimum resonance conditions Z should be as close as possible to unity, because its value affects the resilience of the whole assembly, and maximum stability corresponds to $Z \sim 1$ [231].

For example, Z has a value of 1.08 for a piezoelectric crystal covered with aluminum electrodes (shear-mode impedance of SiO_2: $Z_{SiO_2} = 8.27 \times 10^6$ kg s m^{-2}; that of Al: $Z_{Al} = 8.22 \times 10^6$ kg s m^{-2}), indicative of the good match associated with this material combination. Deposition of gold electrodes on an SiO_2 surface leads to a Z-value of 0.381, which is barely acceptable (shear-mode impedance of gold: $Z_{Au} = 23.2 \times 10^6$ kg s m^{-2}) [231].

From a practical point of view it is important to develop a relationship involving more experimentally common parameters influencing the frequency response. The frequency response of a piezoelectric sensor in the presence of a gas at a particular pressure can be related generally to three effects: the hydrostatic effect (p), the impedance effect (x), and the sorption effect (m):

$$-\frac{\Delta f}{f_0} = \left(\frac{\Delta f}{f}\right)_p + \left(\frac{\Delta f}{f}\right)_x + \left(\frac{\Delta f}{f}\right)_m \qquad (25)$$

For gravimetric measurements and sensor applications, the latter effect (that is, to added mass) is the most important, and the other two effects can be viewed as nonspecific interferences. Because both of these terms are three orders of magnitude smaller than the mass term, the mass relationship alone is sufficient for characterizing measurements in the gas phase. However, situations do exist in which the other two effects cannot be ignored. For instance, if the pressure is changed from that of the surrounding air to a vacuum the first term is no longer negligible, and in the case of liquids any change in viscosity or density alters the impedance, so this term must be taken into consideration.

Other theoretical approaches relevant to applications in the gas phase and dealing with thickness-shear mode acoustic wave quartz sensors are presented in [213], [232]–[237].

Bulk Acoustic Waves in Liquids. When a quartz crystal is placed in a liquid, there is a significant change in both the density and the viscosity of the surrounding medium. Both become much larger than in the gas phase, and the consequences must be taken into account with respect to the second term in Equation (25). Thus, the Sauerbrey equation is no longer applicable if the viscoelastic properties of the contacting medium (liquid and/or film) change during the course of an experiment. The liquid now represents an additional mass load, which produces coupling between the substrate elastic shear wave and the
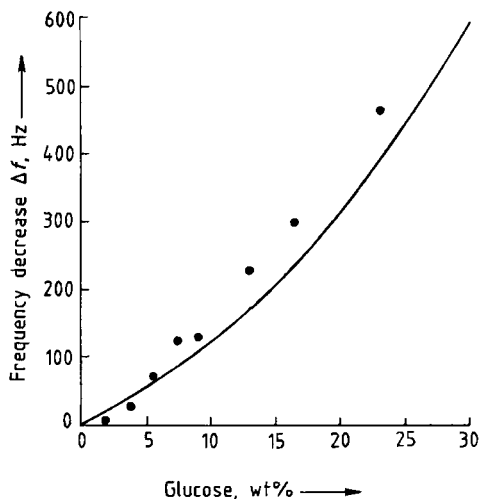
Figure 45. The effect of kinematic viscosity on resonance frequency as a function of glucose concentration (in water) for a 5-MHz quartz crystal
———— = Calculated; ● = Measured (with permission from Am. Chem. Soc., Washington)

liquid. KANAZAWA derived the first theoretical relationship linking the properties of a liquid to the frequency response of a piezoelectric crystal. In a liquid medium, the shear motion of the crystal causes motion to occur in the adjacent layer of solvent molecules (usually water). Up to a layer thickness of approximately 1 µm the solvent molecules move with a certain degree of "slip," which corresponds to a phase shift in the third and subsequent layers. Therefore, vibrational energy from the oscillator is continuously dissipated into the liquid. If only one side of the crystal is coupled to the liquid, the result is a standing wave perpendicular to the substrate surface. For liquids with Newtonian behavior this wave can be described theoretically as a damped wave. As is usual for this type of wave, a decay constant k is used to describe the magnitude of the damping. In this case the constant k can be related to $(\omega/2v)^{1/2}$, where v is the kinematic viscosity. The damping distance (limit of wave propagation) is on the order of micrometers, while the frequency shift can be expressed [238] as:

$$\Delta f = f_0^{3/2} \left(\frac{v}{\pi \mu_q \varrho_q} \right)^{1/2} \tag{26}$$

Figure 45 shows the effect of microscopic kinematic viscosity at the crystal – liquid interface,
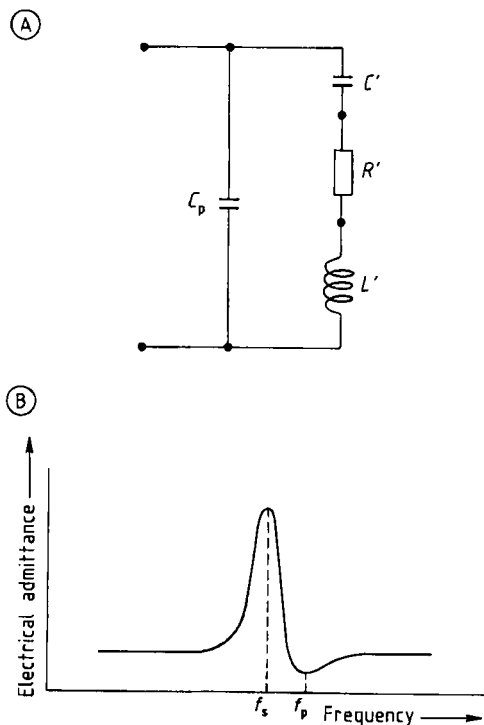
Figure 46. Equivalent electrical circuit (A) and resonance frequency of a mounted piezoelectric crystal oscillator (B) [231], [239]
C_p = Capacitance of the mounted crystal; C', L' = Rigidity and mass of the added material; R' = Equivalent to the mechanical loss, which also represents the acoustic (mechanical) load; f_s, f_p = Series and parallel resonance frequencies
(with permission from Plenum Publishing, New York)

where the magnitude of the observed frequency shift has been plotted against the corresponding increase in density for a glucose solution.

The frequency shift for quartz is highly dependent on changes in the viscosity and density of a surrounding liquid. This situation can be discussed in a straightforward way on the basis of the *quartz equivalent circuit* [231], [239], shown in Figure 46, which consists of the following elements:

1) The $L' - C'$ combination, which determines the resonance frequency f_m of the motional arm
2) A parallel capacitance C_p, caused by the electrodes on the quartz and any stray capacitance
3) A resistive component R'

The resistance R' is a measure of mechanical losses from the vibrating quartz, and it therefore reflects the viscoelastic properties of the contact-
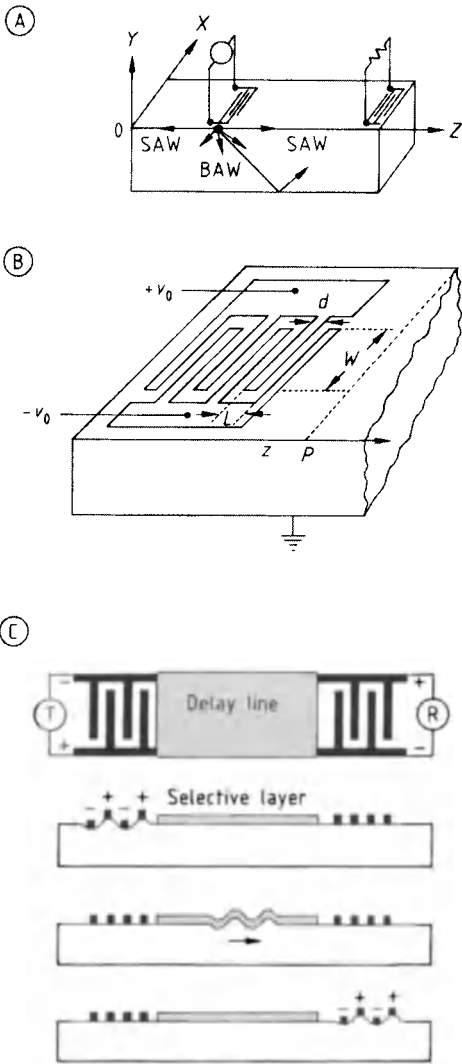
Figure 47. Interdigital metal pattern of a uniform transducer; the IDT behaves like a sequence of ultrasonic sources or receivers [247]

A) An SAW device with input and output transducers; B) A uniform transducer defined by half-spatial period L, electrode width d, and aperture width W; the observation line P used for derivation of the impulse response is at a distance $z+L$ from the first source; C) Schematic diagram of an SAW sensor with a transmitter T, a receiver R, and a chemically selective layer deposited in the form of a delay line [231] (with permission from Plenum Publishing, New York)

ing liquid [240], [241] or a deposited film [242]. Surface roughness also influences the value of R' [243]. This is often not considered in certain applications. It is highly unlikely that in the case of a directly sensed immuno-surface-reaction the bind-

ing events are uniformly smeared out. Likewise, immobilized peptide molecules may alter the tertiary shape by changing the ionic strength, resulting in a different surface roughness. Changes in the vibrating mass correspond to changes in the self-inductance L' [244], [245]. Thus, the resonance frequency of the motional arm (f_m; see Fig. 46) changes in proportion to the mass. In deriving Equation (26) KANAZAWA considered only those L' values for the motional arm that contribute to a frequency shift, but in a real experiment any change in $\Delta R'$ is normally accompanied by a change in L' (typical values for R' in air and in water are $7\,\Omega$ and $350\,\Omega$, respectively, for a 6 MHz quartz crystal). Thus, when a quartz crystal is immersed in liquids of different viscosities, both L' and R' can be expected to vary. Use of an impedance analyzer makes it possible to determine the increase in R' associated with operation of a quartz crystal in air versus a liquid. The frequency shift of the motional arm can be calculated from the values of C' and L' using the formula $f_m = 1/[2\pi(L'\,C')^{1/2}]$ [239].

SAW Transducers. Surface acoustic wave (SAW) sensors have been designed with sensitivity to many different physical quantities, including force, acceleration, hydrostatic pressure, electric field strength, dew point, and gas concentration [246]. The operating principle of an SAW for use with gases is conceptually quite simple. A surface acoustic wave — a periodic deformation perpendicular to the material surface — is transmitted across the surface of some appropriate solid that has been exposed to the atmosphere subject to analysis. With a homogeneous substrate, usually a piezoelectric solid like quartz or $LiNbO_3$, the required SAW (sometimes also called a Rayleigh wave) is generated by means of an interdigital transducer (IDT).

In contrast to a quartz microbalance, the requisite IDT components are deposited only on one side of the crystal. These take the form of planar, interleaved, metal electrode structures, (with permission from Plenum Publishing, New York) where adjacent electrodes are supplied with equal but opposite potentials (Fig. 47). Application of a time-varying r. f. potential causes the crystal to undergo physical deformations, and if these are confined to the surface region of the crystal, the result is a surface acoustic wave (Fig. 47 B) [248]. Such vibrations will interfere constructively only if the distance $L/2$ between two adjacent "fingers" (see Fig. 37 A) is equal to one-half the elastic
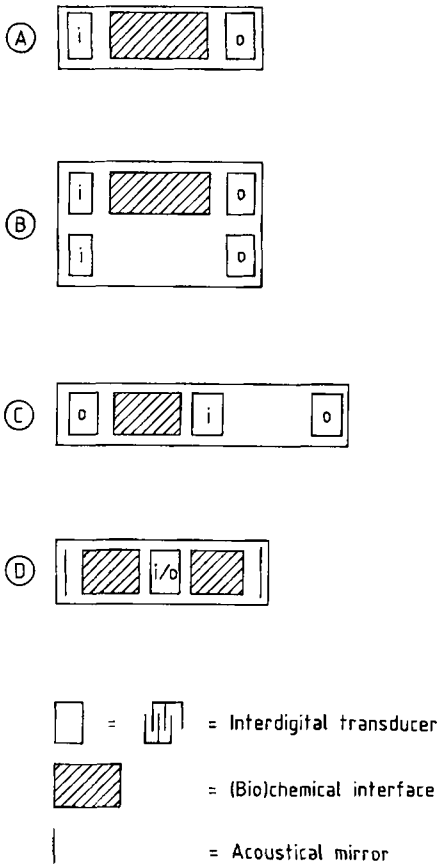
Figure 48. Various possible sensor configurations for SAW chemical sensors [230]
A) Single delay line; B) Dual delay line; C) Three-transducer device; D) Resonator
i = Input transducer; o = Output transducer

wavelength. The frequency $f = v_R/L$ that corresponds to such a cumulative effect is called the *synchronous frequency* or *resonance frequency* (f) [230].

The waves that are generated have their mechanical and electrical components in a single plane, the *sagittal* plane, which is perpendicular to the substrate surface. The phase velocity (v_R) and the amplitude (A) of an SAW are determined by elastic, piezoelectric, dielectric, and conductive properties of the substrate as well as by its mass. If one of these parameters can be modulated in an appropriate way by the quantity of interest, a sensing device can be constructed. Such a modulation could be induced in the transducer itself and/or in the transmission region (waveguide or *delay line*). With layered substrates the physical properties of

each layer as well as the thicknesses of the layers determine the phase velocity and the amplitude of the resulting SAW [230]. Various energy-loss mechanisms also operate on the propagating SAW: scattering loss due to finite grain size, thermoelastic loss due to nonadiabatic behavior of the acoustic conductor, viscous loss caused by the dissipation of energy in a direction perpendicular to the transducer – environment interface, and hysteresis absorption due to irreversible coupling between SAW energy and an adsorbate [231]. Any addition of mass during the sensing step may change the magnitude of one or more of these phenomena, and may also lead to changes in v_R and A, especially a frequency shift (Δf) or a relative frequency shift ($\Delta f/f$). An exact description of the physical processes involved in the generation, propagation, and detection of SAWs is complex, and is available from such sources as [248], [249].

The considerations presented above provide the basis for using SAWs in gas-monitoring devices, because observed frequency shifts are proportional to added mass. However, the characteristic frequency of an SAW is also sensitive to changes in temperature or pressure. The ideal substrate would be one with a zero temperature delay coefficient but a high piezoelectric coupling effect. Despite extensive materials research efforts, it is currently necessary to choose between these parameters; e.g., ST-cut quartz is the preferred substrate if a zero temperature coefficient is most important, and cuts of lithium niobate are selected for high piezoelectric coupling.

Sensor Configurations. A single SAW sensor may sometimes suffice for a gas sensing application (Fig. 48 A), but often the sensors display undesirable sensitivity to such effects as temperature, pressure, and ambient humidity. Overcoming these problems requires the establishment of a relative signal output. For example, VETELINO et al. have described a dual-line SAW gas sensor [230], [250], [251]. One delay line in this case is used for measuring, while the other acts as a reference (see Fig. 48 B). The advantage of a dual delay-line configuration is that it transforms the output signal into a relative change in oscillator frequency, which can be attributed only to the effects of the analyte. D'AMICO et al. have described a system based on a three-transducer delay line (see Fig. 48 C). This device also has two dual-delay lines, but the paths are established in this case by one input IDT and two output IDTs. Again, one path of the delay line is coated and
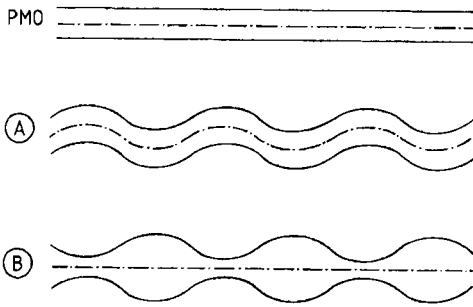
Figure 49. Lamb waves propagating in a thin plate-mode oscillator (PMO) [226], [231] (with permission from Plenum Publishing, New York)
A) Symmetric waves; B) Antisymmetric waves

the other remains uncoated as a way of minimizing the effects of both temperature and pressure [252].

The first use of an *SAW resonator* sensor was reported by MARTIN et al. [253]. The corresponding resonator incorporated a single IDT located between two acoustically reflecting mirrors. The distance between the mirrors was so adjusted as to ensure constructive interference between successive reflections, resulting in a maximum for one particular frequency (Fig. 48 D).

Theory. The SAW velocity v_R can be affected by many factors, each of which is associated with a potential sensor response:

$$\frac{\Delta v}{v_R} = \frac{1}{v_R}\left(\frac{\partial v}{\partial m}\Delta m + \frac{\partial v}{\partial c}\Delta c + \frac{\partial v}{\partial \sigma}\Delta \sigma + \frac{\partial v}{\partial \varepsilon}\Delta \varepsilon + \frac{\partial v}{\partial T}\Delta T + \frac{\partial v}{\partial P}\Delta P + \cdots\right)$$

$$(27)$$

where v_R is the phase velocity unperturbed by such external factors as m, c, σ, and ε (the mass, elastic constant, electrical conductivity, and dielectric constant of the solid medium, respectively), as well as T and p, the environmental temperature and pressure [246]. Three types of films have been considered for use as reactive layers:

1) Nonconductive isotropic overlay film
2) Electrically conducting overlay film
3) Metal-oxide semiconducting film

A theoretical treatment can be developed for each of these film types with respect to its properties, involving, in the nonconductive case, the thickness t_f, the mass density ϱ, λ (the Lamé constant), and μ (the shear modulus).

Regarding the relationship $\Delta v/v = -\Delta f/f$ between SAW velocity and the variation in the

SAW dual-delay line oscillator frequency, perturbation analysis leads to the following result for the relative frequency shift [249]:

$$\Delta f = (k_1 + k_2)f_0^2 t_f \varrho - k_2 f_0^2 t_f \left[\frac{4\mu(\lambda + \mu)}{v_r^2(\lambda + 2\mu)}\right]$$

$$(28)$$

where k_1 and k_2 are material constants for the SAW substrate. Additional information with respect to these equations is available from [224], [246], [249].

Thus, the film mass per unit area is $t_f \varrho$. If chemical interaction does not alter the mechanical properties of the film, the second term in Equation (28) can be neglected, in which case the frequency shift Δf can be attributed exclusively to the added mass:

$$\Delta f = (k_1 + k_2)f_0^2 t_f \varrho$$

$$(29)$$

Coating a polymer film 1 μm thick onto a quartz SAW sensor operating at 31 MHz should therefore cause a frequency shift $\Delta f = -130$ kHz [231].

Plate-Mode Oscillators (PMOs). *Lamb Wave Oscillators.* In all the SAW devices discussed above, the acoustic wave propagates in a slab of material whose thickness is infinitely larger than the wavelength λ of the propagating wave. When the thickness of the plate is reduced to such an extent that it becomes comparable to λ, the entire plate becomes involved in the periodic motion, producing a symmetric and antisymmetric *Lamb wave* (LW) (Fig. 49). This behavior is observed in plate-mode oscillators with thicknesses of a few micrometers. A cross-sectional view through a typical PMO is shown in Figure 50.

Certain performance factors make the LW-PMO potentially attractive as a chemical sensor. Both surfaces contribute to the signal, so the observed sensitivity is greater than for a corresponding SAW device. However, the most important advantage follows from the fact that the velocity of the lowest-order wave in the antisymmetric mode is much lower than in the case of the corresponding SAW oscillator. This may be important for applications involving mass-sensing in liquids, which are problematic with high-frequency SAW devices. As the frequency of the Lamb wave decreases below the velocity of the compressional wave in the liquid, energy loss in the perpendicular direction decreases as well. A
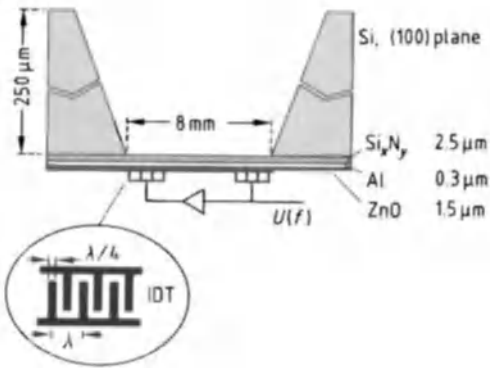
Figure 50. Cross section through a Lamb wave PMO, including the corresponding interdigital transducers [228]

Figure 51. Cross-sectional view of (A) APM and (B) Love plate devices [254]

Lamb wave also has a greater frequency dispersion than the corresponding surface acoustic wave, resulting in increased resolution and sensitivity.

However, there are certain practical problems. The fragility of the thin plate and its consequent sensitivity to external pressure is particularly serious. Performance analyses as well as details regarding fabrication of the device can be found in the original paper [227].

Acoustic Plate-Mode Oscillators and Love Plate Devices [254]. These two types of acoustic PMOs, like BAW sensors, are characterized by surface-guided, shear-horizontally (SH) polarized waves, which have a horizontally polarized component of displacement (see also Fig. 58). In principle the performance of such APMs is similar to that of the Lamb wave sensors described above, but the thickness of the substrate is greater, so waves propagate from the surface into the bulk of the material. Energy is confined mainly in the bulk of the plate as the wave propagates through multiple reflections, generating displacements in both the upper and lower surfaces. The Love plate sensor includes a waveguide structure, whereby the SH wave is confined within an elastic layer deposited on a SAW substrate capable of supporting SH waves. Figure 51 provides cross-sectional views of APM and Love plate devices.

Acoustic sensors based on SH waves have become rather widespread in recent years due to their ability to function in liquid media.

Relative Sensitivities of BAW, SAW, LW, APM, and Love Plate Devices. Because of excitation and propagation differences among the various transducers described, the mass-deposition sensitivity of the several devices would be expected to vary considerably, with some geometries proving more attractive than others for particular sensing applications. Table 10 offers a performance comparison for these transducers. Table 10 also summarizes the calculated frequency-to-mass sensitivities for the various devices. Sensitivity variations can be explained on the basis of energy confinement on the sensing surface and a particular device's operating frequency. A decrease in the thickness of the substrate leads to a sensitivity increase, which can be realized with either an acoustic plate-mode or a Lamb wave device. These are therefore associated with the highest sensitivities, followed by the Love plate, which also has a very thin layer. Less sensitivity is available with BAW devices, for which the sensitivity depends on the thickness of the crystal.

The above statements have to be corrected because of very recent developments. With the commercial availability of a 155 MHz BAW quartz at reasonable prices the QMB should be more favored. According to the Sauerbrey equation the new BAW device shows a much higher sensitivity than the traditional 10 MHz QMB. Table 11 summarized recent data of the different noise levels published in the literature. By this it becomes evident that the new 155 MHz device will be the transducer of choice when an extremely sensitive gas sensor transducer is needed. This sensor also shows almost no drift. Thus drift compensation methods often used with SAW devices are no longer needed. The price of only a few Euro for this mass-produced unit will make it highly competitive in sensor applications.

Table 10. Comparison of sensitivities and other characteristics of acoustic gravimetric sensors [228], [255], [256]; international symbols (short form) have been used to denote specific cuts of the various materials

Acoustic wave type	Sensitivity formula, $cm^2/\mu g$	Calculated sensitivity, $cm^2/\mu g$	Frequency, MHz	Thickness, μm	Material
Bulk wave resonator (BAW)	$S_m = -\dfrac{2}{\varrho\,\lambda} = -\dfrac{1}{\varrho\,t_f}$	-23	10	165	AT-cut quartz
Surface (Rayleigh, SAW)	$S_m = -\dfrac{K(\sigma)}{\varrho\,\lambda}$	-129	100	NA*	ST-cut quartz, LiNbO₃
	$0.8 < K < 2.2$	-516	400	NA*	$K = 1.16$; $\sigma = 0.35$
Lamb plate (LW), A_0 mode	$0 < \sigma < 0.5$ $S_m = \dfrac{1}{\varrho t_f}$	-450	4.7	3	ZnO on SiN
		-951	2.6	(not given)	
Acoustic plate mode (APM, SH)		-3000	NA*	1	ST-cut quartz
		-300	NA*	10	
		-30	NA*	100	
Love plate (transverse surface)		-37	100	NA*	LiNbO₃
		-182	500	NA*	

* not available.

Table 11. Transducer noise of the new polymer-coated 155 MHz BAW in air in comparison with mass-sensitive an SAW-transducer of different operating frequencies from literature values. The integration times of the frequency counters were in the range between 1.0 s and 2.0 s; resonance frequency f_0

Transducer	f_0, MHz	Researchgroup/[257]	Noise, Hz
AT-quartz			
QMW; self-build	155	Cammann, Reinbold [257]	0.2–1.0
SAW LiNbO₃			
Plessey Semiconductors	67	Wohltjen et. al.	9–10
SAW$_{quartz}$ covered with different Me(II)-	39		11
phthalocyanines, leading to	52	Nieuwenhuizen et al.	7
different noise	78		17–29
Xensor Integration, Delft	80	Göpel et al.	1–5
Microsensor Systems, Inc.	158	Zellers, Patrash et al.	11–15
Microsensor Systems, Inc.	158	Grate, Mc Klusty et al.	5–7
Sawtek, Inc. Orlando, FL	200	Grate, Mc Klusty et al.	3–10
Microsensor Systems, Inc.	300	Grate, Mc Klusty et al.	26–55
Microsensor Systems, Inc.	400	Grate, Mc Klusty et al.	15–40
R2632 Siemens; self-build	433	Göpel et al.	1–5

28.2.3.3.3. Theoretical Background

Piezoelectricity. *The Piezoelectric Effect.* Because all acoustic gravimetric sensors are based on the phenomenon of piezoelectricity, it seems appropriate to discuss briefly the effect itself. Piezoelectricity was first observed by the CURIE brothers (JAQUES and PIERRE) in 1880 [258]. It is a reversible phenomenon, consisting of linear electromechanical interactions between mechanical and electrical properties in certain crystals (Fig. 52). The effect is generated, as already mentioned, by application of an AC potential to the piezoelectric material via contact electrodes. One can distinguish between bulk- and surface-generated AW (BAW and SAW, respectively) when the electric field is applied across the substrate or only at its surface, respectively.

A piezoelectric effect occurs only with those ionic crystalline solids whose crystals contain a polar axis along which the physical properties are not constant. The *direct* piezoelectric effect is observed when an applied force (F) produces an electrical polarization (P), and a *converse* (re-
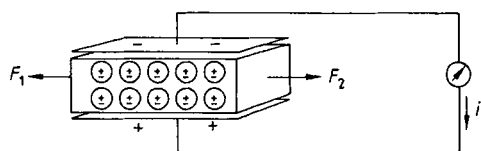
Figure 52. Schematic illustration of the way mechanical and electrical energy can be interconverted with the aid of an appropriate crystal [259]

F_1, F_2 = Mechanical forces; i = Electrical current

Figure 53. Schematic representation of the direct piezoelectric effect (A) and the converse (reciprocal) piezoelectric effect (B) for an X-quartz plate (right-handed quartz) [222]

F = Applied force; P = Polarization; E = Applied electrical field

(with permission from Springer-Verlag, Heidelberg)

ciprocal) piezoelectric effect results when an applied electrical field (E) induces a strain (S, deformation of the crystal) (Fig. 53).

A simplified view of piezoelectricity at the microscopic level is provided by the example of α-quartz (Fig. 54). Here a net dipole moment will arise if the hexagonal structure is stretched or compressed along a direction parallel (Fig. 54 B) or perpendicular (Fig. 54 C) to one of the three in-plane symmetry axes. When symmetry is achieved, the net dipole moment of the molecule vanishes.

Structural Aspects of Piezoelectric Solids. In order for a crystal to be piezoelectric it must be noncentrosymmetric; i.e., it cannot contain a center of symmetry. Solids are usually grouped for structural characterization purposes into seven crystal systems (from lowest to highest symmetry: triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal, and cubic), which can in turn be divided into 32 point groups, depending upon point symmetry. Of these, 11 classes are centrosymmetric and 20 are piezoelectric. An exceptional case is class 432 from the cubic crystal system, because it is neither centrosymmetric nor piezoelectric. In addition to the 21 noncentrosymmetric classes, 11 classes can be distinguished on the basis that they possess no plane of symmetry. This means that they are associated with both right- and left-handed forms, which cannot be interconverted by simple rotation, a phenomenon known as enantiomorphism [261].

Table 12 summarizes on the basis of international point-group symbols the structural distribution with respect to piezoelectric classes, centrosymmetric classes, and enantiomorphism. For example, LiNbO₃, which belongs to the class 3 *m*, is piezoelectric, but not enantiomorphic, whereas α-quartz, in class 32, is both piezoelectric and enantiomorphic. The latter characteristic is illustrated in Figure 55.
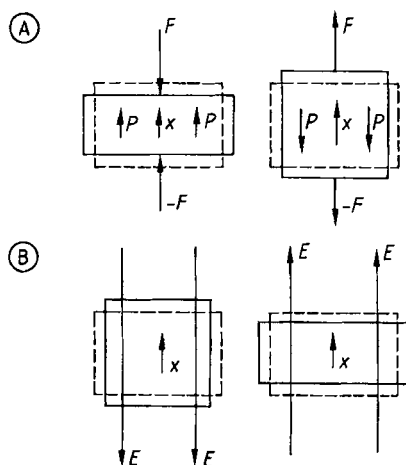
Not every solid material associated with a piezoelectric class is suitable for practical applications, because in some cases the piezoeffect is too weak. Large piezoelectric coefficients are a prerequisite for an electromechanical transducer, like those found, for example, with α-quartz or LiNbO₃. Therefore, these are the solids most often used as mass-sensitive plates or substrates.

Theory and Physics of Piezoelectricity. The discussion that follows constitutes a very brief introduction to the theoretical formulation of the physical properties of crystals. If a solid is piezoelectric (and therefore also anisotropic), acoustic displacement and strain will result in electrical polarization of the solid material along certain of its dimensions. The nature and extent of the changes are related to the relationships between the electric field (E) and electric polarization (P), which are treated as vectors, and such elastic factors as stress T and strain (S), which are treated as tensors. In piezoelectric crystals an applied stress produces an electric polarization. Assuming the dependence is linear, the direct piezoelectric effect can be described by the equation:

$$P_i = d_{ijk} T_{jk} \quad \text{or} \quad P_i = e_{ijk} S_{jk} \tag{30}$$

The quantities d_{ijk} and e_{ijk} are known as the *piezoelectric strain* and *piezoelectric stress* coefficients, respectively.
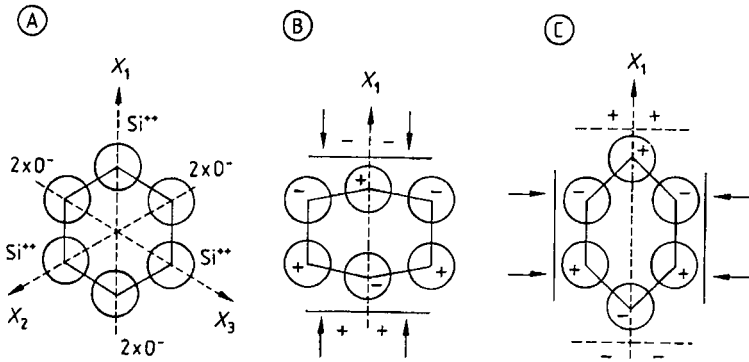
Figure 54. Physical interpretation of the piezoelectric effect [260]
A) Simplified microscopic model of a quartz crystal lattice; B) Longitudinal effect; C) Transverse effect
"Si^{++}" and "2O^{-}" refer to centers of gravity (circles) for charges associated with the two types of atoms, where the tetrahedral "SiO$_4$" structure has been projected onto a plane (as a hexagon)

Table 12. Distribution of crystal structures (point groups) with respect to centrosymmetric, piezoelectric, and enantiomorphic classes [261]

Crystal system	Centrosymmetric classes	Piezoelectric classes	Classes with enantiomorphism
Triclinic	$\bar{1}$	1	1
Monoclinic	2/m	2, m	2
Orthorhombic	mmm	222, mm2	222
Tetragonal	4/m, 4/mmm	4, $\bar{4}$, 422, 4 mm, 42 m	4, 422
Trigonal	$\bar{3}$, $\bar{3}$ m	3, 32, 3 m	3, 32
Hexagonal	6/m, 6/mmm	6, $\bar{6}$, 622, 6 mm, 6 m2	6, 622
Cubic	m3, m3 m	23, $\bar{4}$3 m	23, 432

Conversely, if such a crystal is placed in an electrical field it will become deformed, a phenomenon known as the converse or reciprocal piezoelectric effect (Fig. 53 B). The contributions of stress and strain in this case can be expressed:

$$S_{jk} = d_{ijk} E_i \quad \text{or} \quad T_{ij} = e_{ijk} E_i \tag{31}$$

Since the stress and strain tensors are symmetric, the piezoelectric coefficients can be converted from tensor to matrix notation. Table 13 provides the piezoelectric matrices for α-quartz together with several values d_{ijk} and e_{ijk}, including the corresponding temperature coefficients [259], [261].

For sensor applications, the magnitudes of d_{ijk} and e_{ijk} should be as great as possible, and the temperature dependence should be as small as possible. Although lithium niobate displays larger stress and strain constants than α-quartz, the tem-

perature influence in this case is also great. Therefore, α-quartz is the material most often selected for piezoelectric sensor applications. Because of the anisotropic behavior of a piezoelectric material, properties like resonance frequency and temperature dependence can be optimized by cutting a plate from a single crystal in a particular way. Figure 56 shows two possible cuts with respect to a natural α-quartz crystal. For use as acoustic gravimetric sensors, only AT- and BT-cut quartz plates are useful. These y-rotated cuts provide two different high-frequency plates that vibrate in a shear mode along an axis parallel to the major surface, as indicated previously (see also Fig. 43 B). Minimizing temperature effects requires that the plate be cut at a very precise orientation. The temperature dependence of the relative frequency for various cuts of quartz crystal is shown as a function of the angle of cut in Figure 57 [262].

Table 13. Piezoelectric matrices for α-quartz and LiNbO₃, together with the corresponding piezoelectric stress and strain coefficients [259], [261]

Classification	α-Quartz: Trigonal system, class 32		LiNbO₃: Trigonal system, class 3 m			
Matrix	$\begin{pmatrix} d_{11} & -d_{11} & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & -d_{14} & -2d_{11} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$		$\begin{pmatrix} 0 & 0 & 0 & 0 & d_{15} & -2d_{22} \\ -d_{22} & d_{22} & 0 & d_{15} & 0 & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{pmatrix}$			
Piezoelectric strain constants	d_{11}	d_{14}	d_{22}	d_{31}	d_{33}	d_{15}
	2.3	−0.67	21	−1	6	68
	2.31	−0.670	22.4	−1.2	18.8	78.0
	2.31	−0.727	20.8	−0.85	6.0	69.2
Temperature coefficients of piezoelectric strain constants (×10⁻⁴/°C)	$T_{d_{11}}$	$T_{d_{14}}$	$T_{d_{22}}$	$T_{d_{31}}$	$T_{d_{33}}$	$T_{d_{15}}$
	−2.0	17.7	2.34	19.1	11.3	3.45
	−2.15	12.9				
Piezoelectric stress constants	e_{11}	e_{14}	e_{22}	e_{31}	e_{33}	e_{15}
	0.173	0.04	2.5	0.2	1.3	3.7
	0.171	0.0403	2.43	0.23	1.33	3.76
Temperature coefficients of piezoelectric stress constants (×10⁻⁴/°C)	$T_{e_{11}}$	$T_{e_{14}}$	$T_{e_{22}}$	$T_{e_{31}}$	$T_{e_{33}}$	$T_{e_{15}}$
	−1.6	−14.4	0.79	2.21	8.87	1.47
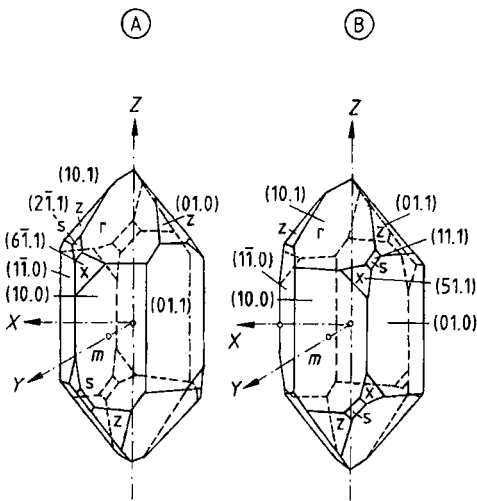
Figure 55. Major crystal surfaces of α-quartz together with their Bravais–Miller indices (hk.l) (with permission from The Institute of Electrical and Electronics Engineers, New York)
A) Left-handed form; B) Right-handed form; Cartesian coordinates are specified in accordance with the IEEE standard of 1987 [229]

Figure 56. Oriented quartz crystal cuts in relation to the natural crystal [209] (with permission from Elsevier, Am-

More detailed theoretical treatments of piezoelectric solids and their properties are provided by [222], [263]–[265]. Another more sophisticated approach based on dynamic quantum mechanics is presented in [266]. Interest has also developed in recent years in piezoelectric polymers, especially as a result of their strongly piezoelectric properties [222].

Vibrations and Waves. The theory of acoustic waves in solids is well understood, and many comprehensive descriptions exist [249], [267]. The description presented here should suffice to explain sensor function despite the fact that it is limited and qualitative.
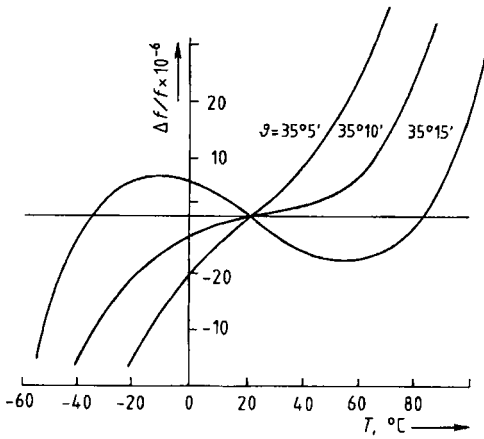
Figure 57. Temperature dependence of the relative vibrational frequency for different AT-cuts of a quartz crystal [262] (with permission from Elsevier, Amsterdam)

Acoustic waves in solids consist of oscillatory motions of the material comprising the solid. The stress – strain relationship treated in the preceding section leads ultimately to two types of acoustic displacements: *longitudinal* or *tensile* motions that cause the dimensions of the solid body to change along a single direction without accompanying changes along the orthogonal directions, and *shear* motion, which produces changes in the shape of the solid body. Figure 58 illustrates these two types of motion [256].

In piezoelectric solids, an applied force generally creates both quasi-longitudinal and quasi-shear waves, where the acoustic displacement in a wave will be largely either longitudinal or shear. Pure longitudinal (L) and shear waves (SH = horizontally polarized shear waves; SV = vertically polarized shear waves) are generated if wave propagation occurs along certain crystallographic axes or with specific orientations. The generation of mixed wave types is also possible, but normally not desirable for sensor applications.

As far as microgravimetric sensors are concerned, two types of wave-propagation devices can be distinguished: those characterized by wave propagation perpendicular to the plate surface (BAW) and those in which the waves propagate along the surface (SAW) or in the bulk of the substrate (LW, APM). Transducers of the former type are normally resonators. A given device might be designed to utilize either longitudinal or shear waves propagated between the faces, as in microbalances based on AT- or BT-cut quartz
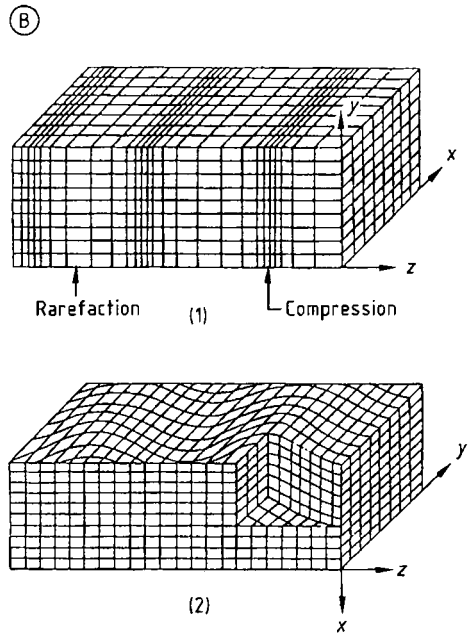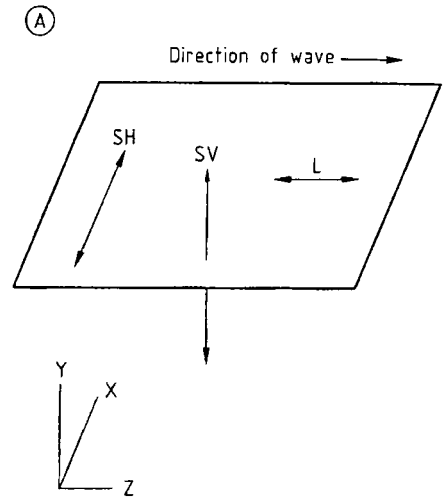
Figure 58. A) Schematic representation of longitudinal and shear displacement [256]; (with permission from Elsevier, Amsterdam) B) Grid diagrams for plane uniform waves propagating along the z axis in a material of infinite extent [248] (with permission from J. Wiley & Sons, New York): (1) Tensile motion (one dimension); (2) Shear motion (two-dimensional; y-polarized in this case)
L = Longitudinal (tensile) motion; SH = x-Polarized shear (horizontal); SV = y-Polarized shear (vertical)

plates operating in their SH high-frequency fundamental mode. Only acoustic waves satisfying boundary conditions for the resonator surfaces will propagate efficiently across such a device. The waves can be visualized as standing sinusoidal waves with displacement nodes at the resonator surfaces (see 28.2.3.3.2).

Transducers in the second category include wave delay-line oscillators (see Figs. 42 and 47), which are based primarily on two types of waves. One is a coupled linear combination of two wave-equation solutions, an L-wave and an SV-wave. In this wave, also called a Rayleigh wave, particle displacement takes the form of an elliptical motion in the plane perpendicular to the surface (sagittal plane) and containing the direction of propagation. Some displacement also occurs outside the sagittal plane, but the deviations are minor (Fig. 59). Two of the most important cuts providing these wave types, and which are used in chemical sensors, have been illustrated in Figure 56. Waveguides that rely on Rayleigh wave devices for their operation are often referred to as surface acoustic wave (SAW) devices, and they have been extensively employed as sensors.

The second type of wave in this context is a horizontal shear (SH) wave, whose acoustic displacement is in the plane of the waveguide. An SH wave generally travels at a different velocity than a Rayleigh wave, and for piezoelectric waveguides the SH wave becomes a true surface wave provided the acoustic impedance at the surface differs from that in the substrate bulk. Such a difference can be caused by piezoelectric stiffening of the SH mode, deposition of a layer of dissimilar material on the surface, the presence of a surface grating, or liquid in contact with the surface. Depending on their origin, SH surface waves are referred to in the literature as Bleustein – Gulayev (BG) waves, Stoneley waves, SH-SAW, or surface transverse waves. Pure surface transverse waves (shear waves) appear to show promise for applications in liquids, especially biosensor applications, because the dissipation of energy across the phase boundary into the liquid is approximately zero, which greatly reduces the generation of longitudinal sound waves in the liquid itself.

Plate-mode waves are analogous in form to Rayleigh and SH waves in a semi-infinite waveguide. Two groups of plate modes can be distinguished: the first consists of Lamb waves or flexural waves, where the acoustic displacement is a combination of longitudinal and vertical shear motion; a second family involves SH acoustic

displacement [256]. Because of their frequency range (100 kHz to the GHz region) these sound waves fall in the ultrasound category, with a propagation velocity of $1-10$ km/s (cf. electrical signals, which travel at $100\,000-300\,000$ km/s).

Interdigitated Transducer (IDT). The IDT was invented by White and Voltmer in 1965. The IDT is a planar, interleaved metal structure at the surface of a highly polished piezoelectric substrate (see Figure 60). The adjacent electrodes are given equal potentials of the opposite sign. The resulting spatially periodic electric field produces a corresponding periodic mechanical strain pattern employing the piezoelectric effect. This gives rise to generation of surface acoustic waves (SAW), provided that the surface is stress free. In general both SAW and BAW may be generated by IDT.

The IDT behaves as a sequence of ultrasonic sources. For an applied sinusoidal voltage, all vibrations interfere constructively only if the distance $a/2$ between two adjacent fingers is equal to half the elastic wavelength. The frequency $F_0 = V_{AW}/\lambda = V_{AW}/a$ that corresponds to this cumulative effect is called the synchronous frequency or the resonance frequency. The bandwidth of an IDT is narrower, when there are more fingers. When IDT is used the AW velocity is determined by the plate material and orientation, while wavelength depends only on the ITD periodicity.

SAW are emitted in both opposite directions, which result in an inherent minimum of 3 dB transducer conversion loss at F_0. A minimum insertion loss of 6 dB is found for a delay line. The acoustic aperture A defines the effective region of transduction between two adjacent electrodes. The IDT is uniform if a constant aperture is obtained.

SAW component design is based on the application of an equivalent circuit model [29] using the values of the piezoelectric coupling coefficient of the material, F_0, and the static capacitance [30].

The frequency at which the AW device operates depends on:

- the acoustic wave velocity in the substrate material
- the IDT finger spacing (for SAW)
- the substrate (plate) thickness

Additionally, the type of AW generated by IDT depends on:

- the crystal cut
- the orientation of the IDT relative to the crystal cut

Cut	Designation	Type of wave	Operating frequency
AT-Quartz	(YXl)-35°	Thickness shear mode	0.5-100 MHz
GT-Quartz	(YXlt)-51°/-45°	Longitudinal extension	50-550 kHz

ST-Quartz	(YXl)-42.5°	Surface wave, vertically polarized	50-1500 MHz
Various Y-rotated quartz cuts		Surface wave, horizontally polarized	e.g., 330 Mz

Figure 59. Wave properties associated with several different types of quartz cuts [260]; see also Figure 48
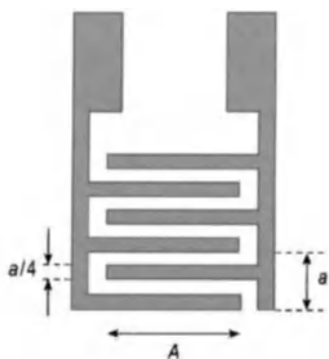
Figure 60. Layout of an Interdigitated Transducer (IDT) with period a and aperture A. The period determines the wavelength of the generated SAW $\lambda = a$. The aperture determines the width of the radiated beam and is typically $10-100$ wavelengths in magnitude. The IDT length is $L = N\lambda$, when IDT contains N electrode pairs

– the substrate (plate) thickness
– the wave guiding mechanism (BAW, SAW, APM)

Acoustic wave devices (AWD). Today the family of AWDs is large. The overview shown in Figure 61 shows the basic types of the acoustic wave devices developed to date, indicating the polarization of the generated acoustic waves, and whether the devices can be operated in liquid.

28.2.3.3.4. Technical Considerations

The (Bio)Chemical Interface. A piezoelectric transducer is transformed into a mass-sensitive sensor by coating it with an appropriate (bio)chemical film or layer, which may be inorganic, organic, or even metallic in nature. Ideally the film should be strongly adherant and deposited in such a way as to obtain a plane surface. Concerning immobilized large biomolecules, certain doubts exist over the demand for a plane surface and the request for a strong adherence. It seems highly unlikely that an immobilized macromolecule will follow any surface movement of the transducer as a whole entity given its flexible structure. The thickness of such a layer is usually on the order of $1-100 \, \mu m$ depending upon the nature of the transducer. The chemically sensitive layer serves as an interface between the environment and a data-processing system, and it is responsible for both the sensitivity and the selectivity of the sensor. This layer should display the following properties:

1) Reversibility, sensitivity, and selectivity
2) Inertness with respect to chemical cross-influences, such as humidity or carbon dioxide
3) Ease of coating, strong adhesion, and long life

Most chemically sensitive layers are insulators, such as organic substances with special functionalities, but conducting materials such as metal films or conducting polymers are also useful. Examples include:

1) Donor – acceptor functionalities; e.g., carbocations [268], M(II)porphyrins [269], or betaines [270]
2) Host – guest binding systems (supramolecular compounds); e.g., molecular cavities [271], clathrate systems [272]
3) Pure metallic layers; e.g., palladium for the detection of hydrogen [273]

Interaction of the Chemical Layer with an Analyte. Interaction with an analyte can take many forms ranging from adsorption – absorption to chemisorption, including the compromise of coordination chemistry. The following types of chemical interactions can be distinguished [274]:

1) *Absorption*
   Here the analyte is distributed between the chemical interface (a liquid film, amorphous solid, or polymer, analogous to the stationary phase in gas chromatography) and the surrounding medium. In the case of a gas interacting with the chemical interfacial layer, once equilibrium has been achieved the amount of analyte present in the layer is a function of the partition coefficient between the gas and the interface (as specified by Raoult's and Henry's laws) as well as the thickness and area of the interface. Because distribution tends to be determined mainly by the polarity of the chemical interface and the analyte, little selectivity is expected.
2) *Adsorption*
   The analyte in this case interacts only at the surface of the interfacial layer, and no chemical bonds are formed or broken. Attraction energies are usually in the range $0-10 \, kJ/mol$ (i.e., attributable to van der Waals forces), extending in the case of hydrogen bonds to as much as $40 \, kJ/mol$. Adsorption may result from physical attraction between a nonreactive gaseous analyte and a metal surface, for example, or formation of a donor – acceptor association (H^+ or e^-). Adsorption is a universal
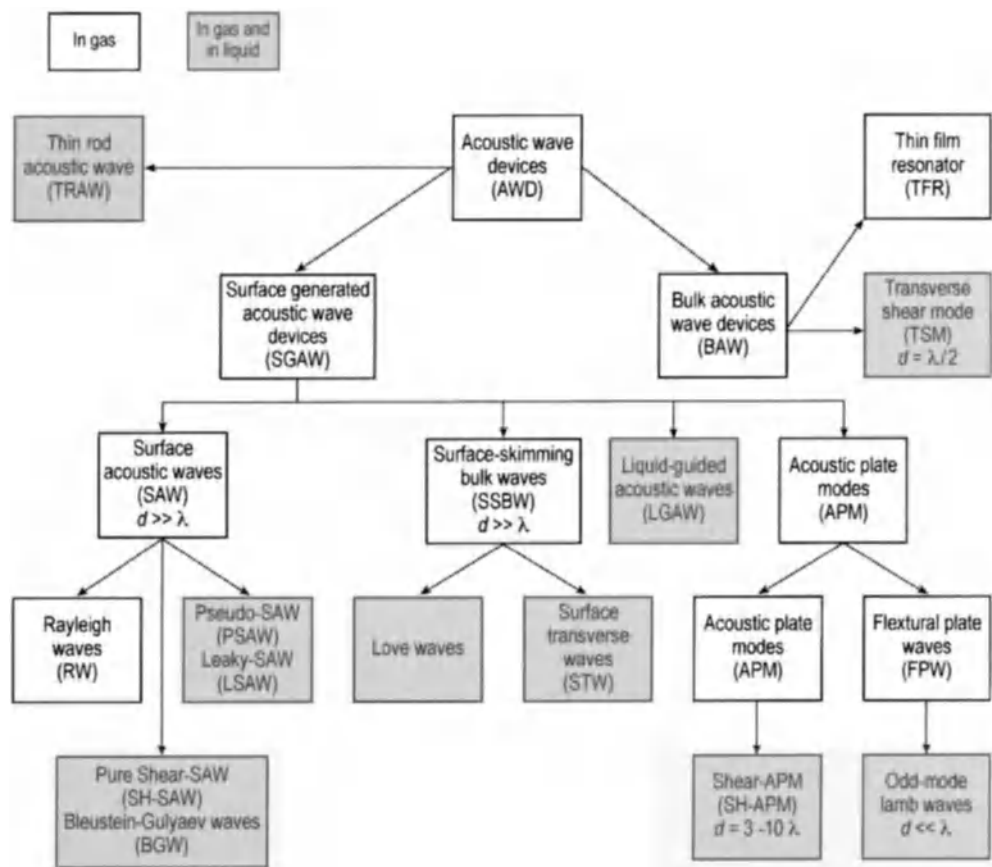
Figure 61. Flow diagram illustrating the basic types of acoustic wave devices. Here, d is the substrate (vibrating plate) thickness and l is the acoustic wave wavelength

phenomenon, and it is quite nonselective. The basic principles of gas adsorption have been described by LANGMUIR [275] and DE BOER [276].

3) *Chemisorption [277]*

Chemisorption involves very strong interaction—up to ca. 300 kJ/mol—developed at the surface of the chemical interface, including the formation and breaking of chemical bonds. High selectivities can therefore be anticipated, but there is a simultaneous loss of reversibility. Because of the partially irreversible nature of bond formation, such a sensor would actually behave more like a dosimeter, although this might be advantageous in some applications.

4) *Coordination chemistry [274]*

A compromise offering both selectivity and reversibility of analyte binding is presented by the area of coordination chemistry. A typical coordination compound consists of a central metal ion M surrounded by a neutral or charged (often organic) ligand. The extent of selectivity can often be influenced by the choice of the metal ion as well as by the choice of the ligand—taking into account both electronic and steric factors. Complexation with an analyte produces changes in the properties of the coordination compound, and these changes can be subject to detection. The selectivity is a function of structural, topological, and polarity parameters.

Other types of chemical interfaces have also been used for microgravimetric gas detection, including polymeric phthalocyanines [278], porphyrins [279], ferrocenes [280], metal
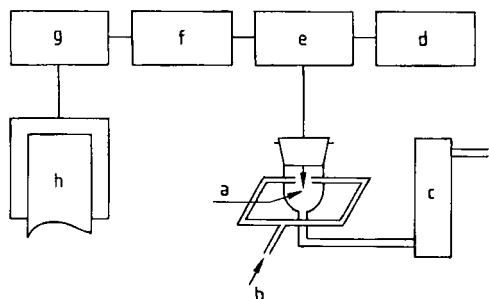
Figure 62. Typical experimental apparatus based on a piezoelectric quartz-crystal detector [217] (with permission from Elsevier, Amsterdam)
a) Piezoelectric quartz crystal; b) Sample; c) Flow meter;
d) Power supply; e) Oscillator; f) Frequency counter;
g) Digital-to-analog converter; h) Recorder

clusters [281], crown ethers [282], and cyclodextrins [283].

In many cases it has proven to be very difficult to find a selective chemical interface for a particular analyte, so the use of sensor arrays has been proposed. Numerous papers have dealt with this approach, invoking both BAW and SAW sensors. Pattern-recognition techniques employing increasingly powerful microprocessors are required to generate the appropriate signals [284].

Coating Methods. The stability of a (bio)chemical interface depends heavily on the attachment of the material to the surface of the sensor device. One or more of the following coating methods is employed according to the nature of the sensitive layer:

Smearing [224]
Spin coating [285]
Drop coating [103], [257]
Solvent evaporation [286]
Spraying [287]
Langmuir – Blodgett film formation [288]
Physical vapor deposition [289]
Sputtering [290]
Chemical immobilization [291]

The first five methods listed involve simple coating procedures in which the interfacial material (usually nonpolar) is dissolved in an organic solvent and then deposited directly onto the substrate surface. Evaporation of the solvent leaves a film that adheres to the surface of the substrate by physical bonds. The Langmuir – Blodgett technique makes it possible to achieve very thin films. This is a dipping procedure in which molecules are

transferred with great accuracy from the surface of a liquid to the surface of the substrate. Even monolayers can be prepared in this way. More complicated methods include physical vapor deposition (PVD) and chemical vapor deposition (CVD). Very planar layers can be achieved in this way, but the reaction conditions are usually quite drastic, so the technique is essentially restricted to materials that are relatively inert, such as inorganic compounds or metals.

28.2.3.3.5. Specific Applications

Having reviewed the fundamental aspects of microgravimetric transducers and piezoelectric solids with respect to the generation and propagation of vibrations and waves, it is now appropriate to turn to the practical considerations of instrumentation and application, addressing first the BAW devices.

Bulk Acoustic Wave Sensors. *Experimental Arrangement.* A typical experimental setup involving a piezoelectric mass-sensitive quartz crystal detector is depicted schematically in Figure 62 [217]. The piezoelectric quartz crystal (a) is shown here inserted directly in a gas stream (b) with a flow velocity of ca. 10 – 100 mL/min, but a measuring cell containing a stationary atmosphere of the analyte gas is also useful. The oscillator (e) is usually powered by a regulated power supply (d; e.g., 5 – 15 V) that drives the quartz crystal. The frequency output from the oscillator is monitored with a frequency counter (f), which should be modified by a digital-to-analog converter (g) to permit the frequency data to be recorded.

Parameters that adversely affect the performance of a piezoelectric sensor include the mass of the chemical layer, built-in stress produced by the chemical layer, stress from the electrodes (clamping), and changes in temperature. A more sophisticated device such as a microprocessor-controlled dual-crystal instrument would be required to eliminate these undesirable effects [292].

The device described is designed for measurements in the gas phase; applications in a liquid medium demand an extended version, including a liquid-tight box, an oscilloscope or network analyzer, and provisions for more detailed signal analysis (impedance analysis). Improvements in the oscillating circuit are particularly important for obtaining satisfactory results with liquids [293] – [295]. A proper interpretation of any ob-
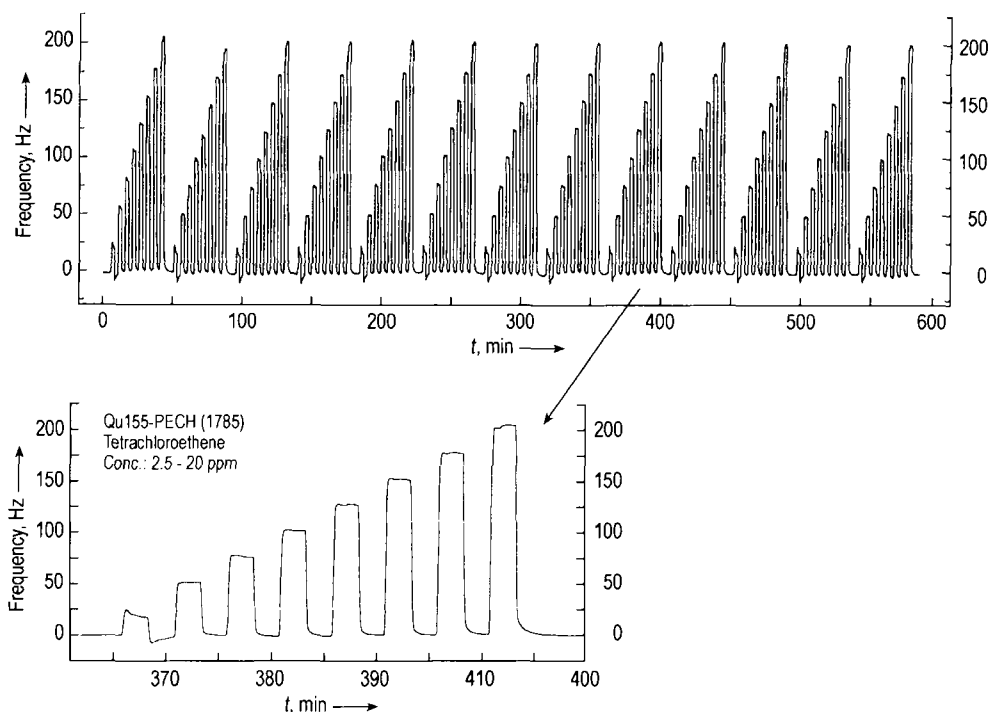
Figure 63. QMB-based gas sensing with polyepichlorohydrin (PECH) as polymer. The upper original recorder traces show 13 complete calibration cycles and demonstrates the very high reproducibility of this gas sensor in this low concentration range. The lower part shows the details of one cycle between 2.5 and 20 ppm tetrachloroethene. With a sensitivity of $\cong$ 10 Hz/ppm tetrachloroethene and an LOD of about 100 ppb this device is more sensitive than any Taguchi gas sensor or the FID considering its low noise, in the range of only about 0.5 Hz! Compared with a 25 MHz QMW the sensitivity could be increased by one order of magnitude [257]

served frequency change in liquid media still remains. Can the densitiy and the viscosity of the sample liquids be controlled so accurately (sometimes within the fifth digit after the decimal point) and does the surface roughness not change during the measurements increasing the viscoelastic effects?

Applications. This section will be limited to a brief consideration of the use of BAW sensors in gas-sensing devices and as sensors for liquid phases. Piezoelectric microbalances can also be used for monitoring such heterogeneous samples as aerosols and suspensions, as discussed in [262].

Gas-Phase Sensing. In principle, both inorganic and organic analytes are subject to detection. Figure 63 provides an example of results obtained from the novel 155 MHz piezoelectric quartz microbalance (QMB) resonator used to detect gaseous tetrachloroethylene (C_2Cl_4). The used absorption layer was polyepichlorhydrin (PECH). Observed response curves are illustrated as a func-

tion of time and for various analyte concentrations. [257].

A wide variety of solvent vapors can be detected with this new QMB with extremely high sensitivity (starting in the ppb range) and showing a dynamic working range of up to 4 decades with response times below 10 s, indicating the potential of this approach to gas-phase monitors and sensors for applications in environmental analysis or process control.

Table 14 provides an overview of the detection of selected gaseous analytes. Information regarding interfering gases can be found in the original literature. The effect of changes in the relative humidity of the environment tends to be a general problem.

Applications in L iquids. Apart from the classical sensors designed for detecting ions, piezoelectric crystal resonators are the preferred devices for conducting biochemical measurements in liquids. Figure 64 illustrates results for the detection

Table 14. Survey of selected reports on BAW sensors for inorganic and organic analytes [230]

Analyte	Range, ppm		Chemical interface	References
Inorganics				
CO	1 – 50		HgO/Hg→Au	[297]
CO₂		≤ 100 000	7,10	[298]
			-	
			dioxa	
			-	
			3,4	
			-diaza-1,5,12,16-hexadecatrol	
COCl₂	8 – 200		methyltrioctylphosphonium dimethylphosphonate	[299]
HCN	13 – 93		bis(pentane-2,4-dionato)nickel(II)	[300]
Organics				
Nitroaromatics	0.001 – 100		carbowax 100	[301], [302]
Dimethylhydrazine	NR*		polybutadienes	[303]
Acetoin	0.008 – 0.120		tetrabutylphosphonium chloride	[304]
Formaldehyde	0.010 – 100		formaldehyde dehydrogenase	[305]
Vinylchloride		≤ 80	amine 220	[306]

* not reported.

of human immunodeficiency virus (HIV) antibodies by means of a synthetic HIV peptide (p 24) immobilized on a piezoelectric quartz sensor. The data refer to the frequency shift observed upon addition of authentic HIV antibody. The quartz device was operated in this case in its fun-damental thickness shear mode at 20 MHz, with an oscillator circuit stability of ca. 0.5 Hz. The quartz plate was coated with HIV peptide (p 24) dissolved in phosphate buffer at a dilution of 1 : 1000. Figure 65 shows the corresponding response as a function of different concentrations of monoclonal HIV antibodies. This device has not yet been brought onto the market despite complete automation and further improvements. Thus, some remarks concerning mass-sensitive transducers employed in liquids must be made, since it is still doubtful if the frequency change observed by passing a sample solution over such a sensor surface really reflects a mass change. Even if the user will manages to control the density and viscosity of the sample solutions within a narrow range determined by the required sensitivity (up to 5 digits after the point) a change in the surface roughness by non-homogeneously attached macromolecules during the immuno reaction may affect the viscoelastic effect. In addition, the complex structure of proteins/antibodies may contain a variable amount of occluded water and ions, which depends on the ionic strength and other factors that have to be carefully controlled to get the true analytical result.

Surface Acoustic Wave Sensors. Many reports have appeared regarding the use of SAWs as gas detectors. These reports also discuss problems associated with the choice of a coating, achieving adequate specificity, selecting a coating thickness, and depositing reproducible levels of polymer coating on the surface of the sensing element. The variety of coatings used is as diverse as the nature of the analytes, covering the range from simple polymers through metallocyanines to metal films [308]. The main drawback after the introduction of the 155 MHz QMB was the lack of availability and the price. Concerning the LOD, their inherent higher noise excludes, in general, the ppb range. The application of SAW sensors in liquid systems is still an active area of research, so this topic will not be treated. Information regarding liquid applications for mass-sensitive devices, especially SAW, APM, and LW devices, is provided in [309], [310].

Experimental Arrangement. A schematic representation of a SAW sensing device is provided in Figure 66. This particular resonator was used by D'Amico et al. [311] for the detection of hydrogen. It consists of a dual delay-line structure in which one of the propagation paths (l_S) is coated with a thin palladium film to act as the sensing channel, while the other (l_R) is uncoated and serves as a reference.

Applications. The type of results to be expected with a variety of analytes in the gas phase is illustrated by the example of an NO₂-SAW sen-
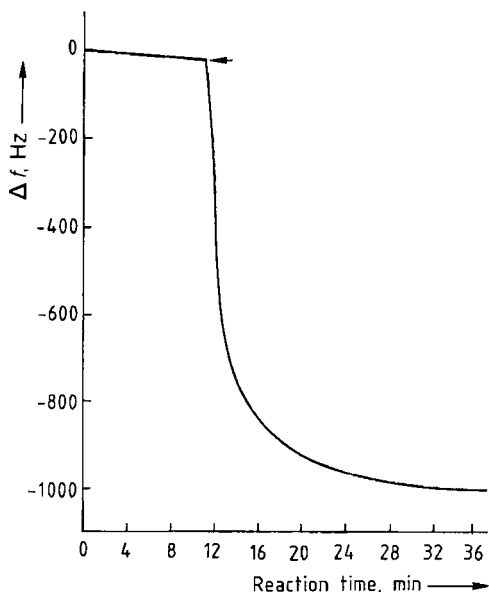
Figure 64. Response for the reaction of anti-HIV antibodies (dilution 1 : 1000) with adsorbed HIV peptide; the saturation value for the frequency shift is reached ca. 20 min after sample introduction (indicated by the arrow) [307] (with permission from Elsevier, Amsterdam)
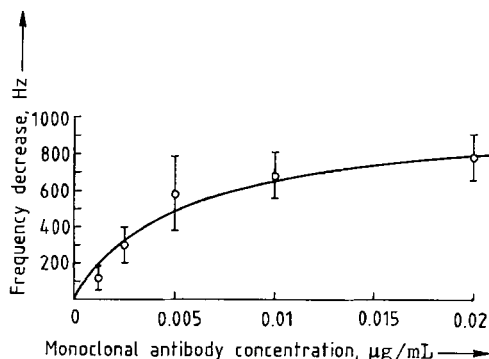
Figure 65. Frequency response for the reaction of HIV antibodies at several concentrations with adsorbed HIV peptide [307] (with permission from Elsevier, Amsterdam)

sor. The device in question was developed by NIEUWENHUIZEN and VENEMA [312], and had an operating frequency of 79 MHz with a metal-free phthalocyanine coating. A frequency change of ca. 670 Hz was observed upon exposure of the sensor to 88 ppm of nitrogen dioxide in air at 150 °C. Figure 67 A demonstrates the sensitivity of the device as a function of nitrogen dioxide concentration, while Figure 67 B shows the re-
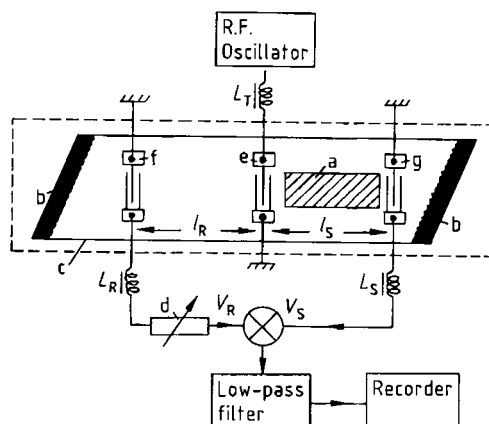
Figure 66. Schematic representation of a SAW hydrogen sensor [311] (with permission from Elsevier, Amsterdam) a) Palladium film; b) Acoustic absorbers; c) YZ-LiNbO$_3$; d) "Stretched line," an adjustable resistance for matching the frequency of V_R to that of V_S; e) Input transducer T; f) Output transducer T_R; g) Output transducer T_S/l_R, l_S = Propagation path (l_R is the distance between T_R and T; l_S between T_S and T); L_T, L_R, L_S = Series inductors (establishing a fixed e.g., 50 Ω electric input – output impedance); V_R, V_S = Output voltage of the dual delay line

sponse of a typical sensor as a function of time with alternating exposure to nitrogen dioxide and air. Possible interferences were also investigated, as described in a similar report published earlier by the same authors [313].

A summary of selected papers describing the use of SAWs for gas sensing is provided in Table 15, subdivided according to the type of analyte.

28.2.3.3.6. Conclusions and Outlook

Mass-sensitive devices offer great promise as chemical sensors, particularly since they rely on the most basic physical effect that accompanies the interaction of one chemical substance with another: the mass effect.

Both BAWs and SAWs have already been successfully applied to the monitoring of many gases, vapors, and more recently, liquids. One of the most critical aspects of chemical sensor research is the search for a suitable chemical interface. The selectivity and sensitivity of mass-sensitive devices can be improved by the use of more carefully designed polymers and other special coating materials. Correction algorithms might also be developed to compensate for variations due to changes in humidity, temperature, pressure, etc. Humidity

Table 15. Survey of selected reports on SAW gas sensors [308] (with permission from Adam Hilger, Bristol)

Analyte	Sensor material	Frequency, MHz	Temperature, °C	Chemical interface	References
Inorganics					
H_2	$LiNbO_3$	75		Pd	[311], [314]
H_2	quartz	23		Pd	[315]
Halogens	quartz	50-80	30-150	Cu-PC[a]	[316]
H_2O	$LiNbO_3$	75	0-100	polymer	[317], [318]
NO_2	$LiNbO_3$	150	70	Cu-PC[a]	[319]
Organics					
DMMP[c]	quartz	31-300	20-80	polymers	[320]
Styrene	quartz	30		Pt complex	[286]
Cyclopentadiene	quartz	31	25	PEM[b]	[321]
Various	ZnO	109	24	ZnO	[322]-[324]
Various	quartz	158	35	polymers	[325]

[a] PC = Phthalocyanine.
[b] PEM = Poly(ethylene maleate).
[c] DMMP = Dimethyl methylphosphonate.
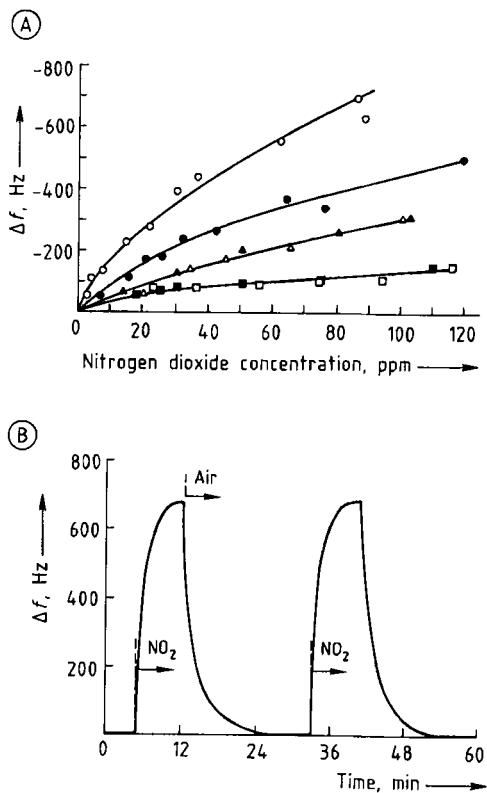
Figure 67. A) Sensitivity of an NO_2-SAW resonator at 120 °C
(●, □, △) and 150 °C (○, □, △) with three different
metal-free pthalocyanine layers: PC 42 (●, ○), PC 46
(□, □) and PC 47 (△, △) [312]; B) Sensor output as a
function of time at 150 °C (PC 42, 88 ppm of NO_2)
(with permission from Elsevier, Amsterdam)

poses one of the greatest problems with respect to most sensing devices. Better design of the electronic oscillator circuits might also increase the stability of the signal in response to a gas or vapor, thereby improving the signal-to-noise ratio and raising the level of sensitivity.

Research on chemical sensors tends to be unusually multidisciplinary. Such new techniques as the use of Love plates and LW-APM make it possible to combine the advantages of microcompatibility, high sensitivity, and suitability for work in both the gas and liquid phases. In particular, the new 155 MHz BAW could open up new horizons in gas sensor developments and electronic noses. The higher sensitivity of this transducer now allows dynamic ranges larger than 4 orders of magnitude with response ranges in the seconds range. Taking the first derivative of the absorption or desorption signal this device can be used as a more or less selective GC detector [257]. A 5 ppm peak of tetrachlorethene can then be sensed with a signal-to-noise ratio of about 3000.

Nevertheless, anyone proposing to conduct investigations in this area should be aware of the complexity of interpreting data accumulated from multicomponent gas mixtures over a prolonged period of time, and of parameter identification and interpretation in the case of liquids [230], [308].

28.2.3.4. Calorimetric Devices

Introduction. Most chemical as well as enzyme-catalyzed reactions are accompanied by changes in enthalpy. For this reason, calorimetric

Table 16. Molar reaction enthalpies for selected processes [328], [329]

Reaction	Examples	Enthalpy $-\Delta H_R$, kJ/mol
Oxidation	methane; O_2	800
Neutralization	NaOH, HCl	55
Protonation	tris(hydroxymethyl) aminomethane; H^+	47
Enzyme catalysis	enzyme: catalase; substrate: H_2O_2	100
	enzyme: glucose oxidase; substrate: glucose	80
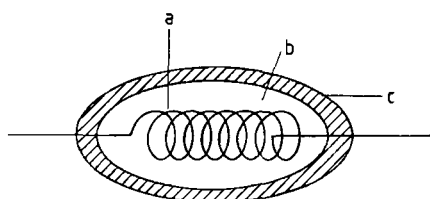
Figure 68. Pellistor device
a) Metal resistor; b) Inactive material; c) Catalyst

transducers represent a universally applicable approach to chemical or biochemical sensors.

Thermal devices for detecting an enthalpimetric effect (either exo- or endothermic) are insensitive to changes in optical properties of the sample, so colored solutions and turbidity do not constitute interferences. There is also no need for a detectable reaction product, whereas amperometric sensors, for example, require the formation of some product subject to transformation at an electrode. Amplification can be achieved by immobilization of a catalyzing reagent that enhances the rate of enthalpy change. Thermal transducers are especially suitable for continuous measurements in flow systems. An important disadvantage, however, is the requirement for thermostatic control of the device. For this reason calorimetric systems tend to be rather large. In addition, calorimetric sensors are relatively complicated to use [326], [327]. Table 16 provides molar enthalpies for several important types of chemical reactions.

The measurement signal in this case, ΔT, is related to the change in enthalpy ΔH, but also to the heat capacity of the system c_s, according to the following equation:

$$\Delta T = n\,\Delta H/c_s \qquad (37)$$

n = Moles of product

This means that the heat capacity of the sensor itself should be minimized to ensure sensitive measurements. A temperature change can easily

be transduced into an electrical signal. The usual measuring device is based on a reference thermal transducer incorporated into a Wheatstone bridge.

Thermal Transducers Commonly Utilized in Chemical and Biochemical Sensors. *Metal-Resistance Thermometers.* The electrical resistance of many metals rises sharply with increasing temperature. Resistors made of platinum (called "Pt 100" because they provide a resistance of 100 Ω at 0 °C) are often used for temperature measurement in the range from -220 °C to $+750$ °C. Other metals like nickel or copper are also applicable for temperatures ≤ 150 °C. Self-heating, caused by a current flow, interferes with the measurement, and the response time of a metal resistor is relatively long (ca. 5 s). The advantages of such resistors are high sensitivity and long-term stability. Platinum resistance devices are especially common in catalytic chemical sensors. A reducing gas may be oxidized in the presence of a heated catalyst (e.g., Rh, Pd, or Pt) in order to increase the reaction rate. In 1962 BAKER [330] described a device called a *pellistor* that effectively separates the catalyst from the platinum wire (see Fig. 68). The detection limit for methane with this device was found to be 20 ppm.

Thermistors. These are semiconductor resistances with temperature coefficients sufficiently high to make them suitable for use in temperature measurement. Many semiconductors have negative temperature coefficients (NTC), which means that their resistance decreases with increasing temperature. Thermistors with reproducible temperature coefficients are difficult to produce, and self-heating within the sensor is always a problem. On the other hand, thermistors are usually both inexpensive and sensitive, and they can be used over a wide temperature range (0 – 1000 °C).

The thermistor most commonly used in chemical or biochemical sensors, known as an *enzyme thermistor* (ET), was designed in 1974 by MOS-BACH et al. [331]; it consists of an enzyme reactor
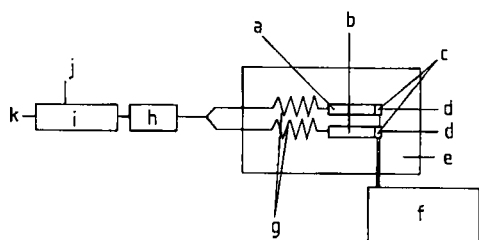
Figure 69. Enzyme thermistor system (with permission from Academic Press, Orlando)
a) Reference reactor; b) Enzyme reactor; c) Thermistors;
d) Waste; e) Insulation; f) Wheatstone bridge; g) Heat exchangers; h) Pump; i) Injector; j) Sample; k) Buffer

specially adapted for use in a flow system. A split-flow arrangement was suggested in 1976 [332], whereby one portion of the analyte solution serves as a reference and flows through a blank column, while another passes through the enzyme reactor itself. The reference signal is subtracted from the signal produced by the enzyme reactor (Fig. 69) [333]. Samples are analyzed at a rate of ca. 20/h. Since the enzyme reaction and the detecting system are separated in space, an enzyme sensor of this type cannot be strictly interpreted as a "biosensor" according to the definition provided in Section 28.3.

Many other devices featuring lower detection limits have been developed in recent years, including a four-channel enzyme-based thermistor that permits the detection of four substrates simultaneously [334]. Another approach involves fabrication of a thermal microbiosensor on a silicon chip. In this case the sensor consists of a micromachined enzyme reactor, inlet and outlet flow channels, and a microthermistor [335].

Thermocouples/Thermopiles. Thermocouples operate on the basis of the Seebeck effect. If wires fabricated from two different metals or semiconductors are soldered together to form a circuit, any temperature difference that exists between the joined points leads to a measurable potential difference, the magnitude of which depends on the extent of the temperature difference and the materials involved. The use of several thermocouples connected in series (*thermopiles*) increases the sensitivity of the sensor, but if a single thermocouple is damaged the entire sensor is affected. It is more difficult to miniaturize thermoelectric sensors than resistance devices, and the long-term stability of thermocouples is often not good. A typical response time is less than one second,

and sensors of this type can be used over a temperature range from $-200\,°C$ to $+1600\,°C$.

Thermal devices based on thermocouples have been extensively investigated in recent years. GUILBEAU et al. described in 1987 a thermoelectric sensor for the measurement of glucose [336]. This device incorporated an antimony–bismuth thin-film thermopile with enzyme immobilized on the side containing the active junctions but not on that of the reference junctions. Exposing the sensor to a solution containing the substrate causes the enzyme-catalyzed reaction to take place, resulting in heat exchange that is detected by the thermoelectric transducer. At the ICB Muenster, a thermopile produced with microsystem technology in the "Institut für physikalische Hochtechnologie" IPHT Jena (Germany) with the extreme sensitivity of about 10^{-4} K has been used for constructing a GOD-based glucose sensor which does not need any calibration. Figure 70 shows the construction of the thermopile with the microsystem technology.

Figure 71 shows the experimental set-up for absolute glucose sensing via the total heat of this specific GOD-based enzymatic reaction, which is strictly proportional to the number of moles reacting [337]. The heat measured in the diagram ΔH versus the glucose concentration follows the theoretical predictions. Thus, no calibration has to be performed, and no standard is needed!

Summary. Most chemical calorimetric sensors in use today are based on the pellistor device, and are designed for monitoring gases; sensor systems of this type have been subjected to intensive investigation [338]. In general, however, sensor determination of the heat of a reaction is rare, primarily because of the availability of other, more sensitive transducers. On the other hand, thermometric biosensors are applicable to a wide range of analytes, including enzyme substrates, enzymes, vitamins, and antigens. Many applications have been reported in clinical analysis, process and fermentation control, and environmental analysis [339]. Thermal biosensors are seldom used in industry because they are difficult to control and relatively expensive. Universal applicability and new techniques for fabricating microthermosensors suggest that these may find wider use in the future.
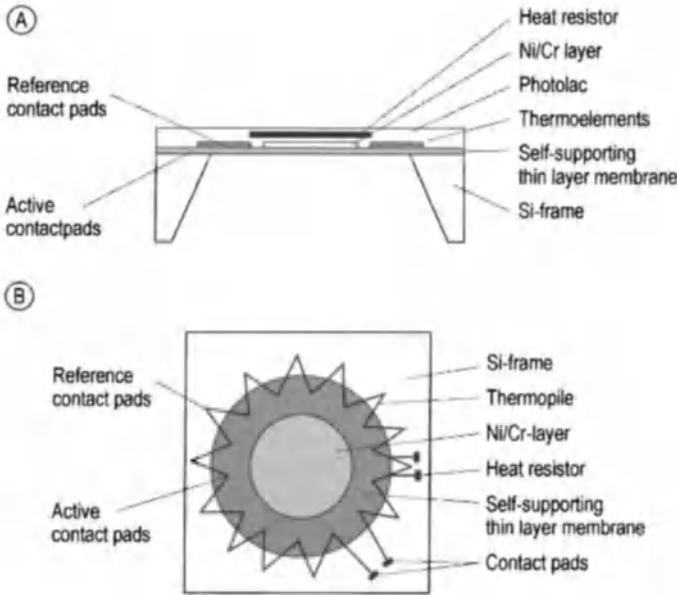
Figure 70. Construction details of a microsystem technology approach towards thermopiles. The extremely sensitive thermopile was produced in the "Institut für physikalische Hochtechnologie" IPHT - Jena
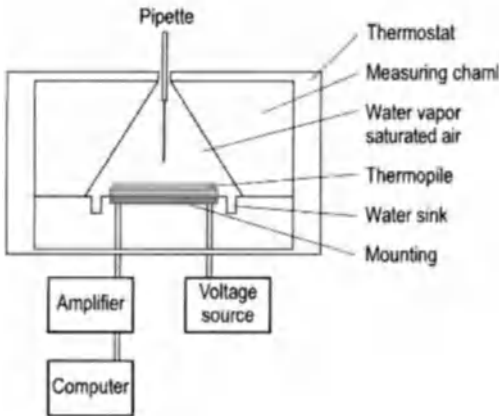
Figure 71. *Measuring cell for absolute calorimetric GOD-based glucose determinations.* The opening must be minimized because otherwise the water evaporates too fast and interferes with the measurement by its heat of evaporation. In order to accelerate the enzymatic reaction peroxidase was added simultaneously. As soon as the drop of blood was added the temperature – time course was recorded and integrated. The glucose concentration can then be directly calculated via the published value for ΔH for this reaction as Figure 72 demonstrates

28.2.4. Problems Associated with Chemical Sensors

All nonbiochemical sensors suffer from a lack of that selectivity which would permit their wide application without detailed knowledge regarding the matrix composition of the samples under investigation. Application is therefore restricted to those areas in which the sample matrix is thoroughly understood and not subject to change. Compared to the selectivity problem, all other problems (e.g., stability, sensitivity, lifetime) are relatively unimportant, since they can be resolved by incorporating the sensor into some type of intelligent instrument capable of carrying out all necessary corrections, recalibrations, and performance tests. In some cases insufficient selectivity of a chemical sensor can be circumvented by integrating a selective chemical reaction into the instrumental design. In this way many interfering compounds are eliminated and/or the analyte of interest is selectively transformed into a species capable of being detected unambiguously. A simple approach to this end takes advantage of the flow-injection analysis (FIA) technique.

Any chemometric method of error correction with respect to lack of selectivity must be checked very carefully, especially if it is to be applied to an open system with varying and/or unknown inter-

fering compounds. In some cases a prior chemical separation and/or selection step is required.

Despite the superb selectivity or even specificity of biochemical sensors for recognizing particular analyte molecules, interfering compounds must be dealt with even here in real sample applications. The compounds in this case do not disturb the actual measurement of the analyte; instead they influence the recognition process. For this reason it is important to ensure the use of a stable and well-defined host molecule with an analyte-fitting geometry that is not subject to change. Anything that might alter the host would influence analyte recognition. Thus, any compound or measuring condition capable of changing the tertiary structure of the sensitive protein skeleton of some biomolecule used for recognition purposes would disturb the measurement: first by modifying the selectivity, but also by altering the sensitivity. It is well established that most selective biomolecules are very stable only in a physiological environment similar to the one in which they were produced; even changes in the ionic strength can influence the tertiary structure of a delicate protein macromolecule. Furthermore, certain metal ions (e.g., Hg^{2+}, Pb^{2+}, Ag^+, Cu^{2+}) are known to be bound to sulfur containing molecules (or destroy a stabilizing $S-S$ bridge in a protein) and thus degrade recognition and thus strongly interfere. A similar effect is produced by surface-active molecules (detergents). Bearing these complications in mind, the superb selectivity of biochemical sensors must be evaluated in a broader perspective.

28.2.5. Multisensor Arrays, Electronic Noses, and Tongues

Theoretically, and under ideal conditions, a sensor array containing several sensors with slightly different selectivities and sensitivities towards the analyte and its main interferences should make it possible to compensate for errors caused by the limited selectivity of each individual sensor. The task resembles the solving of mathematical equations with several unkowns. There must be as many independent equations as unknowns. In sensorics this means: one additional but slightly different sensor is required for each interfering compound. The task then becomes one of solving n equations for n unknowns, which can be accomplished with an algorithm for matrix calculations and an electronic calculator. However,

there are several restrictions that limit the use of this compensating technique.

First, the interfering components must be known in advance so that one can prepare a set of calibration mixtures for obtaining the individual analytical functions for each sensor under variable interference (matrix) concentrations. The approach is therefore limited to a closed sample location, and it presupposes complete information regarding potential interferents within the volume in question. This may prove applicable in certain production control situations, but certainly not in cases where the sample matrix is not sufficiently known a priori, as in the case of most environmental and some clinical analyses. Such precautions cannot be circumvented even with the use of the most sophisticated modern approaches to pattern-recognition analysis and/or so-called adaptive neural-network treatment. All potentially interfering compounds must be known in advance and introduced into the array under study during the so-called learning phase through a permutation of all possible types of mixtures. If for example one analyte is to be determined in a concentration range of 1 to 100 ppm and only five interfering compounds in a similar concentration range are present, and the concentration range should be checked for linearity by five concentrations between 0 and 100 ppm one has a permutation power of about a faculty of 10. Much time may be consumed in the preparation of the necessary mixtures and in performing all the measurements. In the end, the drift associated with most sensors and the limited lifetimes of biochemical sensors could render the whole procedure meaningless!

Second, the linearity and true independence of all effects cannot be guaranteed. In some cases of only one sensor poisoning (e.g. catalyst poisoning) all analytical functions (selectivities and sensitivities) obtained during prior calibration and/or learning are rendered invalid for every individual sensor. How is a sensor array to distinguish between irreversible poisoning effects and the normally reversible disturbances attributable to ordinary interferents? How should it differentiate between the effects of a positive error (where a portion of the signal is due to interferents) and general matrix effects resulting in diminished sensitivity? Especially in environmental trace analysis (<10 ppm) each sample is very likely to contain more than ten interfering compounds. Intimate knowledge of each, as well as the availability of appropriate standards to permit extensive permutation at various concentrations presupposes the
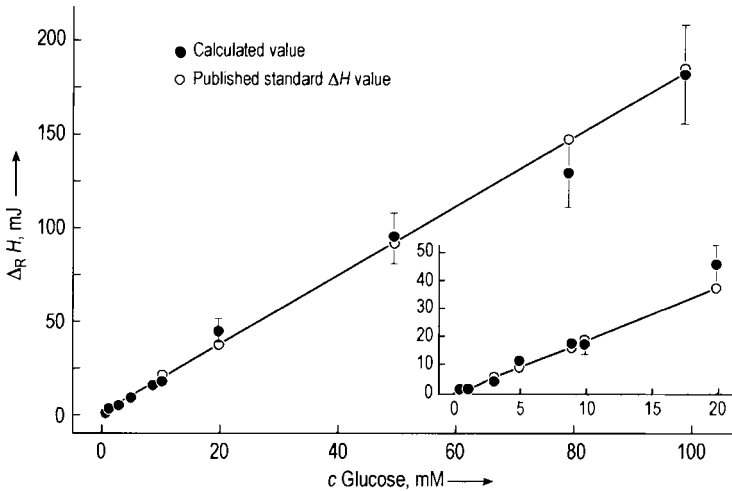
Figure 72. Several measurements with known glucose concentration are plotted on a diagram calculated from the "theoretical" ΔH values for this aliquot. Note the thermopile and the set-up was not optimized further. The only drawback is the fact that both enzymes are lost after each determination as they are part of the receptor buffer solution. This is also the case in traditional enzymatic analysis with photometric evaluation. Here, only a so-called spectral-line photometer can work without further calibration

preparation of countless mixtures to be utilized in the training process for the sensor array. It must also be recognized that even trace amounts of certain compounds may totally disrupt the analyte recognition process.

Taking all this into consideration, as well as the fact that most chemists lack fundamental experience with sensor arrays in real open-environmental trace analysis, considerable caution is warranted with respect to this approach to increasing the selectivity of sensors and correcting for interferants. In general, application of a simple and rapid preliminary separation step will be more productive with respect to the goal of achieving a reliable analytical result. Literature examples of successful sensor-array applications related to the identification of beverages or the characterization of odors [340] – [342] should not be extrapolated to the quantitative analysis of a single analyte. Many of the sensor-array applications (just identification) described in the literature [343] – [351] involve problems that could have been solved by much simpler means, such as exact density measurement, possible now with extreme accuracy. Differentiation among various samples with similar compositions in the context of a closed system (i.e., without the variability associated with an open matrix) can usually be achieved more economically on the basis of simple physical meas-

urements. Nevertheless, electronic noses and tongues on the market have found some application despite the humidity interference of the first category in certain food processing processes, in which they sense any deterioration in the process (e.g., a rotten tomato in ketchup processing or a bad coffee bean in the roasting process. For those applications in the gas phase the ICB Muenster recently introduced a novel electronic nose based on the new sensor class of conductometric gas microsensors (polymer + plasticizer + organic salt). It has already been mentioned that the permutation of polymer, plasticizer, and organic salt together with their concentration allows the construction of an indefinite number of different membranes with different selectivities. This is shown in Figure 73.

The identifying power of an array of only five of such sensors is shown in Figure 74 [144], [352]. The intended application area was the detection of BETX (benzene, ethylbenzene, toluene and the different xylenes) near gasoline stations for reasons of industrial hygiene. The advantages of this electronic nose using only one type of sensor principle is the low energy consumption (conductivity measurements at room temperature) making hand-held devices feasable and the fast exchange of the sensor array in the form of a cheap sensor card not larger than a business card [353].
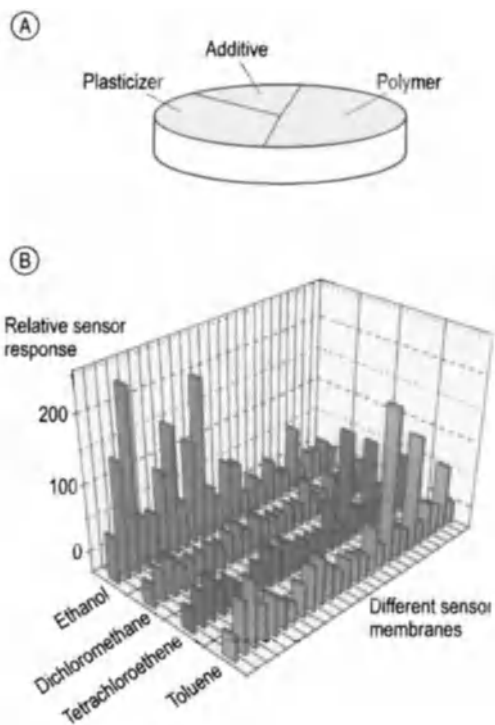
Figure 73. General composition of gas selective conductometric polymer membranes (A) and selectivity pattern against several organic solvents (B)

The above mentioned electronic nose was further improved by incorporating an extremely fast enrichment and water elimination step (in the seconds range) and increasing the resolving power following the time course of a fast thermal desorption step. Since the air sampling time, the desorption time and the number of conductometric gas sensors can be individually set, a very versatile and extremely sensitive electronic nose (called: Air-Check) with sensitivities in the ppb range and almost no interference by humidity was developed in the ICB Muenster. It will be positioned in clean rooms for the purpose of surveillance.

28.3. Biochemical Sensors (Biosensors)

28.3.1. Definitions, General Construction, and Classification

Biosensors are in general small devices based on a direct spatial coupling between a biologically active compound and a signal transducer equipped with an electronic amplifier (Fig. 75). This definition of an I.U.P.A.C. working group is unfortunately very specific. The crucial thing here is the demand for a "direct spatial coupling" of biomolecules and transducers. This might have influenced the development of further commercial available biosensors since much time has been spent on biomolecule immobilization and stabilization oriented towards biosensors that could be transported and stored without a need for a refrigerator. The intimate contact of the less stable biological recognition element with a transducer surface requires either one shot devices and throwing both away after one deteriorates or a reproducible surface rejuvenation with the biological receptor by the user. The latter is the problem since the user is not an expert in controlled coverage of surfaces. Concerning the life and working time and ease of exchanging the more labile biological recognition element, devices which separate the recognition space from the detection space show certain advantages though they cannot be named a biosensor. It is, e.g., quite feasible to employ the labile biomolecules with or without the required co-factors or reagents in a freeze-dried state in a cartridge to be activated only shortly before its use. This cartridge is then to be used upstream of an optical, electrochemical or calorimetric flow through cell. Keeping this cell for longer use allows more sophisticated constructions. With modern microsystem technology the whole device could be rather small and difficult to visualize at a first glance in that the demanded intimate contact for being a biosensor is missing.

Biological systems at various levels of integration are used to recognize a particular substance that is to be determined in a specific way. Different types of biologically sensitive materials can be used for selective recognition, including enzymes, multienzyme systems, organelles, photosensitve membranes from plants and bacteria, membranes generally, protoplasts, whole intact cells, tissue slices, antibodies, lectins, DNA, and transport protein or receptor systems isolated from cell membranes. The macromolecular biological compound is ordinarily immobilized in close proximity to the transducer surface, thereby facilitating direct or mediated signal transfer to the transducer. Besides fixing the receptor on the transducer surface this immobilization also serves the important function of stabilizing the biological material.

The first indicating step is specific complex formation between the immobilized biologically
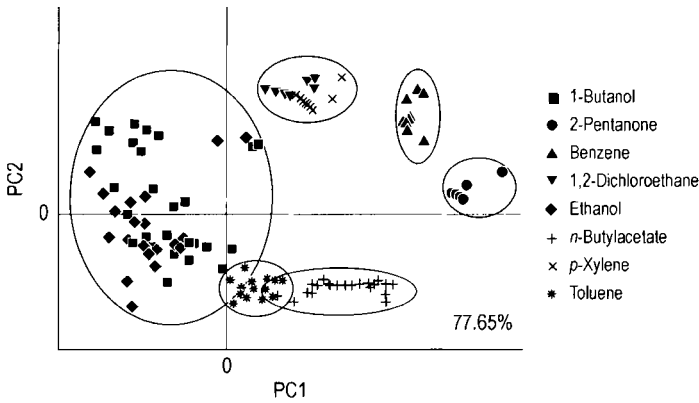
Figure 74. Identifying important BETX compound in the ppm range with only five conductometric gas sensors (ICB-Muenster patent) via principal component analysis (PCA), preliminary results [352]
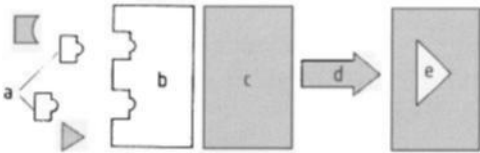
Figure 75. The basic principles of a biosensor
a) Analyte; b) Biological receptor; c) Transducer; d) Electrical signal; e) Signal processing

active compound (in a broader sense, the receptor R) and the analyte (A). This selective and usually reversible interaction produces secondary chemical or physical signals that are recognizable by a suitable electrochemical, optical, enthalpimetric, conductometric, or piezoelectric transducer. The transducer of choice depends on the nature of the molecular-level binding affinity with respect to the analyte, which may cause changes in such parameters as the redox state of the receptor compounds, the concentrations of electrochemically or optically active substances that are consumed or produced, or heat, mass, or electromagnetic radiation (see Table 17) [354] – [357].

Depending upon the mechanism of biochemical interaction between the receptor and the analyte, two basic types of biosensors can be distinguished: *biocatalytic* (metabolic) sensors, and *bioaffinity* sensors (Table 18).

Among the analytes detectable with biosensors are carbohydrates, organic acids, alcohols, phenols, amines, heterocyclic compounds, proteins, enzymes, cofactors, inorganic ions, hormones, vitamins, peptides, drugs, xenobiotics, microorganisms, gases, and pesticides. Analyte concentrations

measurable with various types of biosensors range from 10^{-7} mol L^{-1} (biocatalytic sensors) to as little as 10^{-15} mol L^{-1} (affinity sensors). Based on their level of integration, biosensors can be subdivided into three generations:

1) *First Generation.* The receptor is entrapped between or bound to membranes, and the combination is fixed on the surface of an appropriate transducer or known chemical sensor .
2) *Second Generation.* The receptor is bound covalently to the transducer's surface, thereby eliminating the need for a semipermeable membrane.
3) *Third Generation.* The receptor is bound to an electronic device that transduces and amplifies the signal (e.g., the gate of a field-effect transistor, an array of amperometric ultramicroelectrodes, or interdigitated microelectrodes), utilizing the techniques of microelectronics and micromechanics.

It should be mentioned that this classification is not also meant to be an order of increasing performance. In general rugged systems have more advantages than complicated and expensive constructions.

28.3.2. Biocatalytic (Metabolic) Sensors

Metabolic biosensors are based on the specific recognition of some enzyme substrate and its subsequent chemical conversion to the corresponding product(s). The biocatalytic reaction itself accomplishes specific binding between the active site of

Table 17. Transducer systems used in the design of biosensors

Transducer	Indicating a change in:	Caused by:
Amperometric electrodes	electron transport	enzyme-catalyzed oxidation or reduction of an analyte
Potentiometric electrodes, ISFETs, MOSFETs, conductometric devices	ion concentration	enzyme-catalyzed oxidation, hydrolysis, or cleavage of C–C, C–O, C–N, or other bonds
Calorimetric devices (thermistors, thermopiles)	enthalpy	enzyme-catalyzed reactions
Piezoelectric crystals	mass	bioaffinity reaction (between Ab–Ag, Ab–hapten, an analyte and receptor protein, or analyte and lectin
Fluorometric fiber optics	fluorescing concentration of enzyme substrates/products or immunolabels	enzyme and bioaffinity reactions
Waveguides	optical path (adlayer)	bioaffinity reaction
Devices based on surface plasmon resonance	resonance (adlayer)	bioaffinity reaction

Table 18. Biosensor specification according to the nature of the primary sensing reaction [356]

Biocatalytic (metabolic) sensors		Bioaffinity sensors	
Indicating reaction:*			

$$E \diagdown \quad \diagup A$$
$$\quad R A$$
$$E \diagup \quad \diagdown P$$

$$R \diagdown \quad \diagup A$$
$$\quad RA$$

Analytes	Receptors	Analytes	Receptors
Substrate	enzyme	protein	dye
Cosubstrate	enzyme system	glycoprotein	lectin
Cofactor	organelle	cells	antigen
Prosthetic group	microorganism	antigen/hapten	antibody
	tissue slice	antibody	RNA/DNA
		neurotransmitter	receptor protein
		pesticide	
		enzyme (system)	

*E = Enzyme; A = Analyte; P = Product; R = Receptor; RA = Receptor–analyte complex.

the enzyme and the substrate (recognition process), conversion of the substrate to product, and release of the product from the active center, which in turn leads to regeneration of the binding affinity and catalytic state of the enzyme.

The linear measuring range for metabolic sensors depends on the "apparent" Michaelis–Menten constant (K_{Mapp}) of the immobilized enzyme (not free in solution and approachable from all directions, thus, "apparent"), which describes a particular substrate concentration associated with half the maximum rate of the enzyme reaction. The reaction rate of an immobilized receptor enzyme is controlled by the rate of substrate and product diffusion through both the semipermeable membrane covering the enzyme and the layer of immobilized enzyme itself, resulting in a higher K_M value than that anticipated under conditions of kinetic control. Enzymes, coupled enzyme sequences, organelles, microorganisms, or tissue slices can all be used as receptor materials for metabolic sensors. Under certain conditions, cosubstrates, effectors (inhibitors, activators), and enzyme activities can also be determined by metabolic sensors.

28.3.2.1. Monoenzyme Sensors

Enzyme sensors rely on an immobilized enzyme as the biologically active component. Many biocatalytic reactions, especially those of oxidoreductases, hydrolases, and lyases, are associated with the consumption or formation of electroactive substrates or products in concentrations

Table 19. Common configurations for monoenzyme sensors

Enzyme class	Detectable compound involved in the enzyme reaction	Typical transducer
Oxidoreductases		
Oxidases	O_2, H_2O_2	pO_2 electrode, noble-metal – carbon redox electrode
Oxygenases	O_2, quinones, catechol	pO_2 electrode, (amperometric) carbon electrode
NAD(P)-dependent dehydrogenases	NAD(P)H/NAD(P)	(amperometric) mediator-modified redox electrode, fluorometer, photometer
PQQ-dependent dehydrogenases	$PQQ/PQQH_2$	
Reductases	NAD(P)H/NAD(P) Heme-, Cu-, Mo-, FeS-containing centers	
Hydrogenases	H_2, NAD(P)H, NAD(P), $FADH_2$, $FMNH_2$, FeS-, Ni-containing centers	
Hydrolases	$NH_3/(NH_4)$	(potentiometric) ammonia-sensitive electrode
	CO_2	(potentiometric) pCO_2 electrode
	H^+	pH electrode, ISFETs, coulometer, conductometric and enthalpimetric devices
	redox-active products (phenolic, thio-group-containing compounds)	(amperometric) redox electrodes

proportional to the analyte concentration. For this reason most commercially available biosensors are enzyme sensors based on electrochemical transducers.

Typical electrochemically detectable (co-) substrates and products include oxygen, hydrogen peroxide, hydrogen ion, ammonia, carbon dioxide, reduced or oxidized cofactors, and redox-active (oxidized or reduced) prosthetic groups in oxidoreductases, all of which can easily be converted into electrical signals by suitable transducers (e.g., amperometric or potentiometric electrodes or conductometric sensors, Table 19).

Immobilization of an enzyme in or at an artificial matrix or a membrane usually stabilizes the biocompound, thereby increasing the half-life for activity. Techniques of enzyme immobilization include physical or chemical adsorption, ionic and covalent bonding (perhaps to functionalized transducer surfaces), cross-linking by bifunctional agents, and entrapment in such polymer matrices as natural or artificial gels and conducting polymers. The first biosensor was described by CLARK and LYONS in 1962, who introduced the term "enzyme electrode" [358]. In this first enzyme sensor, an oxidase enzyme in a sandwich membrane was placed next to a platinum redox electrode. A platinum anode polarized at +0.6 V (vs. SCE) responded to peroxide produced by the resulting enzyme reaction with the substrate. Application of a polarization voltage of − 0.6 V per-

mitted the amperometric detection of oxygen consumption. The primary target substrate was β-D-glucose (Fig. 76). Stoichiometric turnover of the glucose oxidase- catalyzed reaction yields a sensor response proportional to the glucose concentration up to the apparent K_M value of the immobilized enzyme. This in turn led to the first glucose analyzer for measuring glucose in dilute blood on the basis of an amperometric enzyme sensor. A commercial analog, the Yellow Springs Instrument Model 23 YSI, appeared on the market in 1974. The same indicating technique has more recently been applied to many other oxygen-mediated oxidase systems, which constitute the most widely exploited enzyme classes for the construction of biosensors.

Amperometric Biosensors Using Artificial Electron Mediators. Amperometric biosensors for monitoring oxidase-catalyzed oxygen consumption or the production of hydrogen peroxide are, under certain circumstances, restricted in their use because of a limited availability of dissolved oxygen. This may necessitate careful pretreatment of the sample with rigorous exclusion of oxygen-consuming compounds, as well as application of a high polarization voltage, which may cause interfering currents in complex matrices.

It should be possible to overcome such problems by direct electron transfer between the redox-active prosthetic group of an enzyme and the elec-

β-ᴏ-Glucose + O₂   Gluconolactone + H₂O₂
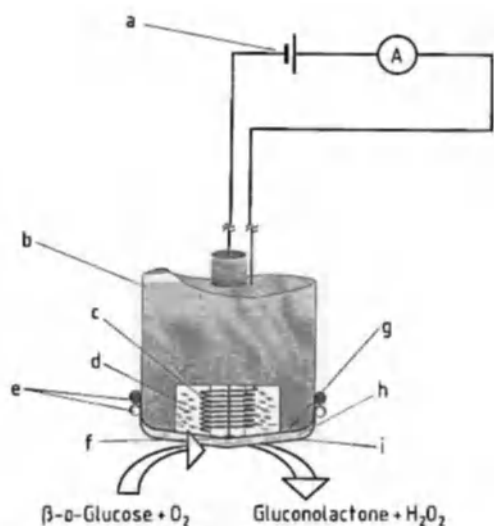
Figure 76. Configuration of an amperometric enzyme elec-
trode for glucose determination
a) Polarization voltage (−600 mV or +600 mV vs. Ag/AgCl);
b) Electrode body; c) Ag/AgCl pseudo-reference electrode;
d) Electrolyte solution; e) O-Rings; f) Platinum working
electrode; g) O₂/H₂O₂-permeable membrane; h) Glucose
oxidase (immobilized layer); i) Semipermeable membrane

Figure 77. Operating principle of a mediator-modified am-
perometric biosensor, involving an oxidase or a dehydro-
genase (in the case of a reductase, electron transfer would
proceed from the electrode to the analyte substrate)
a) Amperometric enzyme sensor; b) Mediator-modified
electrode; c) Redox electrode

trode surface. In a few cases, especially ones in-
volving small redox proteins such as cytochrome
c, efficient electron transfer has been achieved in
this way with promoters that direct the active re-
dox centers to the electrode surface. Macromolec-
ular redox proteins exhibit rather low rates of het-
erogeneous electron transfer because according to
the Marcus theory tunneling is not possible over
the long distance from the active center to the
electrode surface. Thus, the transfer is often ac-
companied by a large overpotential, but the use of
redox mediators may increase the electron-transfer
rate dramatically. These low-molecular-mass
(in)organic redox species undergo facile reversible
redox reactions with both the prosthetic groups of
the enzyme and the electrode surface according to
the general scheme depicted in Figure 77. The
resulting electron shuttle reduces the overvoltage
of the biological redox system to the formal po-
tential of the mediator. Suitable redox mediators
are redox systems which show a high exchange
current density and include quinoid redox dyes,
quinones, organometallic compounds, and fulva-
lenes (see Table 20) [359] – [363].

In constructing such a biosensor the redox me-
diators can be immobilized at the electrode sur-
faces in various ways, including chemisorption or

covalent bonding, chemical or physical deposition
of a polymer film containing the mediator mole-
cules, or physical incorporation of the redox me-
diators (which are only slightly water soluble)
directly into the electrode material. One ingenious
approach involves covalent binding of the media-
tor to the electrode via an appropriate *spacer*, or
electrical "wiring" of the enzyme through an elec-
tron-relaying redox polymer network. In both
cases electron transfer from the redox centers to
the electrode surface is thought to occur by elec-
tron tunneling [364] – [366]. The use of a redox
mediator increases the dynamic range of a sensor,
minimizes interferences, and provides an indicator
reaction that is usually independent of the oxygen
content of the matrix. Furthermore, such mediators
undergo reversible redox reactions with redox-ac-
tive coenzymes, dehydrogenases, reductases, and
hydrogenases, as well as the redox-active products
of enzyme-catalyzed reactions, which opens many
new possibilities for incorporating these classes of
oxidoreductases into new types of biosensors.

28.3.2.2. Multienzyme Sensors

Coupled enzyme reactions mimic certain met-
abolic situations in organelles and cells that have
themselves already been exploited in biosensors
[367], [368]. The use of enzyme sequences facil-
itates:

1) Broadening the scope of detectable analytes
2) Improving selectivity
3) Increasing both sensitivity and dynamic range
4) Eliminating interferences

Table 20. Redox mediators leading to efficient electron transfer with redox proteins

Mediator type	Examples	Structures
Quinones	benzoquinone	
Quinodimethanes	tetracyanoquinodimethane (TCNQ)	
Tetrathiafulvalene (TTF) and its derivatives		
Metallocenes and their derivatives	ferrocene	
Phthalocyanines (Co, Fe, Mn, Cu)	cobalt phthalocyanine (Co-PC)	
Quinoid redox dyes		
Phenazines	N-methylphenazonium (NMP$^+$) salts	
Phenoxazines	meldola blue	
Phenothiazines	methylene blue	
Indamines	toluylene blue	
Viologens	methyl viologen	
Complex compounds		
Ruthenium complexes		$[Ru(NH_3)_6]^{3+/2+}$, $[Ru(en)_3]^{2+/3+}$
Osmium complexes		$[Os(bpy)_2Cl]^{3+/2+}$
Cobalt complexes		$[Co(bpy)_3]^{3+/2+}$, $[Co(en)_3]^{2+/3+}$
Iron complexes		$[Fe(CN)_6]^{3+/4+}$, $[Fe(phen)_3]^{2+/3+}$
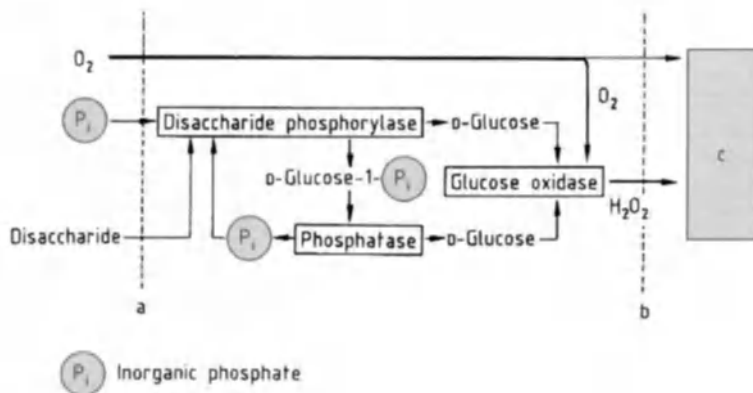
Figure 78. Enzyme sequence for the detection of inorganic phosphate, based on the principle of chemical signal amplification [369]
a) Dialysis membrane; b) H_2O_2/O_2 permeable membrane; c) Working electrode

Depending on the target analyte, sample matrix, required measuring range, etc., enzymes can be coupled to produce reaction sequences operating either in series or in parallel. An alternative approach to realizing multienzyme sensors is the use of cell complexes, whole cells, or tissue slices comprising physiologically integrated enzyme systems (see below).

Enzyme Sequences Operating in Series. The number of substances that can be determined directly by monoenzyme sensors is limited, because many enzyme-catalyzed reactions involve no physicochemically detectable substances. To form readily detectable compounds it is often necessary, therefore, to couple together several different enzyme reactions in a sequential way. The analyte of interest is first converted by some specifically catalyzing enzyme into a product that is itself convertable in the course of one or more subsequent enzyme reactions involving detectable secondary (co)substrates or products. Biosensors based on such enzyme sequences have been developed for the determination of many substrates, including disaccharides (sucrose, maltose, cellulose, lactose), cholesterol, creatinine, and nucleic acids [356].

With certain enzyme sequences it is possible to achieve chemical signal amplification through the accumulation of a particular product by cycling the substrates or cofactors. The sensitivity of some biosensors can be enhanced in this way by as much as three orders of magnitude. Signal-amplifying enzyme systems have been based on oxidase – dehydrogenase, dehydrogenase – transferase, and phosphorylase – phosphatase – oxidase sequences, permitting the determination, for example, of lactate, ATP, and phosphate at the picomole per liter level (Fig. 78) [357] – [369].

Enzyme Sequences Operating in Parallel. Multienzyme sequences operating in parallel are used for the elimination of interfering agents, expansion of the dynamic measuring range, and (in principle) the indirect detection of an electrochemically inactive reaction catalyzed by one of the enzymes in a substrate competition. *Anti-interference* enzyme systems are the most important. Substances interfering with an indicating reaction can be eliminated as one way of improving the analytical quality of the measurement by covering the immobilized enzyme with an anti-interference enzyme layer, in which a second enzyme converts the disturbing substance into an inert product. This strategy has made it possible to remove such interfering agents as ascorbic acid and endogenous glucose [367].

Biosensors Based on Integrated Biological Enzyme Systems. Cellular organelles and membranes, as well as whole cells or tissue slices from animal or plant sources comprising physiologically coupled multienzyme systems are often employed as receptor materials in biosensors based on electrochemical detection. Native physiologically integrated enzyme systems display enhanced stability due to their in vivo environment. Biological material of this type offers the potential for

low cost and simple preparation. Furthermore, this opens the way to analytes that are converted to electrochemically detectable products not by simple enzyme reactions, but rather by complex multienzyme reactions within an intact biological structure. Analytes subject to such receptor systems include chloroaromatics, caprolactam, naphthalene, tryptophan, several hormones, and acetic acid [354], [370].

28.3.2.3. Enzyme Sensors for Inhibitors – Toxic Effect Sensors

Sensors for inhibitors are based on competitive inhibition between a substrate and an inhibitor for an enzyme binding site, inhibition of an enzyme reaction by product accumulation, irreversible binding of an inhibitor to an active binding site, or redox-active heme or SH groups of enzymes that lead to the blocking of enzyme activity.

When thylakoid membranes, protoplasts, or whole cells are used as receptor materials, various inhibitors cause metabolic "short circuits," or inhibition of the (photosynthetic) electron-transport system, which results in a decrease in such metabolic parameters as oxygen consumption or production, photocurrent development, dehydrogenase activity, or fluorescence. In contrast to metabolic sensors, the first sensing step here is a result of binding of the inhibitor to an active binding site or redox center on the enzyme. Sensors of this type usually operate at low enzyme-loading factors and with substrate saturation in order to ensure a highly sensitive signal of limitation in the kinetically controlled reaction due to the presence of an inhibitor. Examples of such inhibitors (analytes) include toxic gases, pesticides, chloroorganic compounds, phenols, and heavy-metal ions [371], [372]. To detect toxic effects on vital enzymes in a reproducible manner, to include also synergistic or antagonistic effects, is only possible with biosensors or bio-tests. This advantage as a fast warning device is still not fully used. Too often biosensors are developed as if they should compete with instrumental analysis. Given the pace of development in this area, especially concerning miniaturization, speed, and versatility, the biosensor as a mono-analytical device will have problems in gaining a larger market share. However, as a toxic warning device triggering some screening and sampling it is without competition.

28.3.2.4. Biosensors Utilizing Intact Biological Receptors

Molecular receptors are cellular proteins (often membrane-bound) that bind specific chemicals in such a way as to produce a conformational change in the protein structure. This conformational change in turn triggers a cellular response: opening of an ion channel, for example, or secretion of an enzyme. In general, molecular receptors can be distinguished from other types of receptors (including larger multicomponent systems) on the basis of their composition, which is a single protein, though this protein may contain more than one subunit. Among the important receptors known to date are receptors for hormones (control agents for a wide range of cellular and body functions), amino acids, insulin, and neurochemical transmitters, in addition to receptors capable of binding synthetic bioactive chemicals, such as drugs [373], [374].

Molecular receptors have two important properties relevant to their possible incorporation into biosensor devices: intrinsic signal amplification and high specificity. The specificity is a consequence of a highly evolved binding region within the receptor that is fully optimized for binding one particular ligand by a variety of electrostatic, hydrogen bonding, van der Waals, and hydrophobic forces.

Limitations affecting the study and use of molecular receptors include:

1) Relatively difficult experimental procedures for the isolation of the receptor proteins (even prior to purification), which are both time-consuming and expensive, and often require animal sources
2) An absence of techniques for obtaining more than very small amounts of pure receptors
3) The rapid loss of biological function that usually follows isolation

Molecular biology has brought considerable progress in the isolation and purification of receptors, while advances in artificial membrane technology and reconstitution have recently reduced the problem of instability.

Comparatively few biosensors of this type have been reported [373], [374], primarily because of problems associated with the successful isolation of molecular receptors (in sufficient quantity and purity) and their subsequent stabilization in

artificial environments in such a way that appreciable biological activity is retained.

A clear exception, however, is a fiber-optic sensor based on evanescent fluorescence developed by ELDEFRAWI et al. [374], [375] that contains nicotinic acetylcholine receptor (nAChR). This was the first neurotransmitter receptor to be purified and characterized in vitro, and it is involved in chemical signalling. The sensor in question consisted of nAChR immobilized on the surface of a quartz fiber. Determination of the concentration of acetylcholine and other natural neurotransmitters (e.g., nicotine) was accomplished by measuring the extent to which they inhibited the binding of a standard solution of a fluorescence-labeled toxin (fluorescein isothiocyanate-α-bungarotoxin, FITC-α-BGT). Binding determinations were based on measurements of fluorescence intensity in the region of the evanescent wave (that is, at the fiber–solution interface, see Section 28.2.3.2.1), where the nAChR was immobilized. The sensor had an acceptable signal-to-noise ratio of 10^2 and an analysis time < 10 min (based on the initial rate of signal increase rather than a steady-state value), and it was capable of operating under conditions of high ionic strength. However, it suffered from the problem of non-regeneration of the sensing element because of the irreversibility of FITC-α-BGT binding to the nAChR.

A similar receptor-based biosensor using electrochemical detection was reported for the detection and measurement of riboflavin (vitamin B_2) [373], [374]. This device is based on the competition for aporiboflavin-binding protein (apoRBP) between riboflavin present in the sample subject to analysis and a riboflavin analogue bound in a membrane. In the absence of riboflavin all the apoRBP is bound to the analogue on the membrane. When riboflavin solution is added, the five-fold greater affinity of apoRBP for its natural cofactor (riboflavin) results in displacement of apoRBP from the membrane into the solution, causing a change in the electrical potential of the membrane. A riboflavin measurement range of $0.1 - 2 \times 10^{-6}$ mol/L was reported.

Further progress in molecular receptor-based sensors will be a function of advances in molecular biology with respect to enhancing the quantity, purity, and stability of receptors that can be used in vitro.

28.3.3. Affinity Sensors – Immuno-Probes

Affinity sensors take advantage of physicochemical changes that accompany complex formation (e.g., changes in layer thickness, refractive index, light absorption, or electrical charge). Once a measurement has been made the initial state must be regenerated by splitting off the complex. Because of this they cannot be used for continuous monitoring nor to follow declining concentrations of the analyte. Therefore, they cannot be named biosensors into the strict sense of the I.U.P.A.C. definition to be "reversible". Thus, they are named immuno-probes. Since the splitting of the complex can be very fast (e.g., with pH < 2.0 or chaotropic reagents) and with the process being automated (e.g., in FIA-systems) the response time of an immuno system can reach less than 1 min and, thus, becomes comparable to enzymatic biosensors working under diffusion control.

Examples include immunosensors utilizing antibodies or antigens for recognition purposes, lectins, and "true" biological receptors in matrix-bound form (see Section 28.3.2.4).

Immunosensors Based on Antibody Molecules. Biosensors based on the antigen–antibody interaction are called *immuno-probes*. The high specificity and high sensitivity that typifies an antigen–antibody (Ag – Ab) reaction has been used in a vast range of laboratory-based tests incorporating antibody components (immunoassays). The analytes detected and measured have included many medical diagnostic molecules, such as hormones (e.g., pregnancy-related steroids), clinical disease markers, drugs (therapeutic and abused), bacteria, and such environmental pollutants as pesticides. A crucial distinction must be made between an immunoassay and an immuno-probe, however, with respect to the technology relevant to the analysis. The ultimate aim is a low-cost, disposable immuno-probe, with the entire device being used only once. For example, it has now proven possible in practice to develop an instrument containing a disposable "sensor chip" that can be used for Ab or Ag tests.

Two different types of immuno-probes have been described:

1) *Direct immunosensors* register immunochemical complex formation at the transducer surface via electrochemical, mass or optical changes. The advantage of a direct immunosensor is that measurement of the antigen – an-
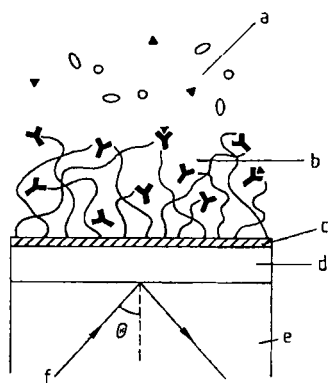
Figure 79. Operating principle of an SPR immunosensor
a) Bulk solution: b) Carboxymethylated dextran matrix;
c) Linker layer, $HO-(CH_2)_n-SH$; d) Gold film; e) Glass
substrate; f) Light
(with permission from GEC Journal of Research, Essex)

tibody interaction can be accomplished immediately without any need for additional antibodies or markers (enzymes, fluorescent labels, etc.).

2) *Indirect immunosensors* are based (like homogeneous and heterogeneous immunoassays) on labeling of one of the partners in an immune interaction. One class of indirect immunosensors derived mainly from enzyme immunoassays (EIA) encompasses the enzyme immunosensors. These combine the high selectivity of an immunoreaction with the high sensitivity characteristic of the enzymatic amplification effect. However, the enzyme can affect the molecular recognition step because of its large size and needs in general special attention (cooling) during storage. In the case of the widely used assay technique of Enzyme Linked Immuno Sorbent Assay (ELISA) the use of enzymes generates two further manipulations: a) the addition of a suitable sustrate and b) the addition of a reagent which stops the enzymatic reation. This is not necessary when using modern fluorescence markers.

28.3.3.1. Direct-Sensing Immuno-Probes without Marker Molecules

Among direct immunosensors the most favored are optical sensors, because they incorporate the advantages of fiber optics or optoelectronic transducers. The nonelectrical properties of optical systems offer important safety advantages for medical, industrial, and military applications. Remote sensing is clearly advantageous, for example, for explosives and toxins. Optical-fiber probes are also mechanically flexible, small, inexpensive, and disposable.

The detection of an evanescent field in an optical waveguide (see Section 28.2.3.2.1) is a highly sensitive technique especially suitable for monitoring the antibody–antigen interaction [373], [374]. One partner of the immuno pair (e.g., the antibody) is immobilized at the surface of the waveguide, and its reaction with the other partner in the test sample is monitored. An evanescent wave generally penetrates a distance of several hundred nanometers extending the full length of an antibody–antigen complex, so in contrast to conventional techniques there is no need to carry out a prior separation of the non-specific biological components. There is no need for washing steps.

The technique of surface plasmon resonance (SPR) represents a very sensitive way to follow an immuno-reaction on an interface , and its feasibility has been adequately demonstrated [376]. As noted previously (Section 28.2.3.2.3), SPR is a phenomenon produced by a beam of light directed onto a glass–metal interface, usually a glass prism in contact with a layer of gold or silver (Fig. 79). At some specific angle (the resonance angle) a component of the electromagnetic light wave propagates in the metal along the plane of the interface, taking the form of surface plasmons. The resonance angle is quite sensitive to changes in refractive index and dielectric constant at the interface up to a distance of several hundred nanometers from the actual metal surface, with the sensitivity decreasing exponentially as a function of distance from the surface. Immobilization of an antibody on the surface causes a measurable shift in the resonance angle, and binding of a high-molecular antigen to the immobilized antibody leads to a further change. For typical biological systems this binding-induced shift in resonance angle (expressed in resonance units) is approximately linearly proportional to the concentration of bound antigen (or antibody if antigen has been pre-immobilized).

The SPR-based biosensor systems BIAcore and BIAlite, developed by Pharmacia Biosensor AB and now available commercially, represent a breakthrough in immunosensor technology, particularly in the use of SPR techniques. Apart from high cost, the only limitation associated with this technique is that the sensitivity depends on the

molecular mass of the adsorbed layer, which in turn controls the optical thickness, so low concentrations of small molecules (molecular mass < 250, including most haptens) are unlikely to be measurable directly without additional labeling. In addition to SPR, Mach–Zehnder interferometers [377], RIFS [201], [202], and so-called grating couplers [378] are also used for optical immunosensing. An interesting step towards miniaturization was introduced by the development of practicable fiber optical SPR systems by JORGENSON AND YEE [379] and KATERKAMP et al. in the ICB Muenster [380], [381]. The latter device was so small and low-cost that application as a hand-held refractometer with a resolution in the fifth digit after der decimal point was feasible. The fiber optical SPR tips were rapidly exchangeable and bending of the fiber had no negative influence on the result.

All the above mentioned transducers measure changes in the refractive index and layer thickness together! With respect to quantitative assays one should bear in mind that those techniques only sense a mean layer thickness. Thickness extensions that occur because of large macromolecules are leveled out. Futhermore, the change of the refractive index inside the evalescent field (within the surface layer) may also depend on ions or water molecules entrapped in the proteins. The latter is dependent on the ionic strength. Thus, changes in the ionic strength of the sample solution may also produce a signal. In addition, if antibody molecules are immobilized on the transducer surface, how can anyone differentiate between an immuno-reaction with the analyte or a change in the tertiary and quarternary structure of the antibody molecule caused by other effects giving rise to a change in the thickness of the layer?

Alternative transducer technologies that have been applied in immunosensor research include piezoelectric and electrochemical systems. *Piezoelectric* immunosensors [383], [384] (see Section 28.2.3.3) tend to suffer from significant levels of nonspecific binding to the piezoelectric substrate, which makes accurate analyte quantification difficult. However, this type of immunosensor may have an important role to play in the search for traces of volatile drugs in security monitoring, and it offers the advantages of relatively small size and low manufacturing cost. The main disadvantage from the standpoint of immunosensing is that reproducible measurements have so far been achieved only in the gas phase, whereas the Ab–Ag interaction is restricted to aqueous solution. As has already been mentioned above, working with mass-sensitive transducers in liquids gives rise to doubts about the correct interpretation of frequency shifts. Here, the same as in case of the optical transducers, they need to be differentiated from real mass changes. Any structural changes of large protein molecules which induce associated changes of entrapped water and ions will result in a measurable signal. Despite the fact that these frequency changes have nothing to do with specifically bound analyte molecules these signals can be very reproducible. Without a thorough and convincing validation with real samples, the analytical accuracies are questionable. This is not to say that qualitative binding studies or determination of association constants could not be measured. All transducers needing no label compound are ideal for screening purposes. They are mostly used for the detection of specific receptor functions.

Potentiometric immunosensors [356] (see Section 28.2.3.1.1) are unfortunately associated with a low signal-to-noise ratio as a result of the low charge density on most biomolecules relative to such background interferences as ions. These sensors also show a marked dependence of signal response on sample conditions, including pH and ionic strength. Similarly, *immunoFET* devices (see Section 28.2.3.1.4), which measure very small changes in an electric field when an Ab–Ag binding reaction occurs, tend to suffer from practical problems associated with membrane performance and various artifacts. The theoretical basis of the sometimes observed voltage change remains unclear. The only plausible explanation for the fact is that clearly the binding of a neutral antigen molecule to a large antibody molecule with multiple charges (depending on the pH and isoelectric point) may change the charge distribution at the interface or may lead to a different mixed potential. If the whole immuno-reaction producing an analyte proportional signal is not watched carefully and controlled by convincing validation studies, a certain bias can easily be overlooked by all so called direct sensing techniques.

Main Drawback of all Direct Sensing Techniques. It has become a well known fact that the sensitivity (or LOD) of all direct sensing techniques is not any signal-to-noise ratio nor any theoretical mass-loading sensitivity! It is the lack of control over non-specific binding processes, which are not caused by the very specific molecular interaction under study. All direct sensing immuno-techniques cannot differentiate between

competing surface processes leading to adsorbed particles with various heats of adsorption. Even if working with very hydrophilic surfaces to prevent this, the ultimate LOD of such an immuno-probe depends, according to the laws of Analytical Chemistry, on the standard deviation of such non-specific binding (LDO = 3 x σ_{blank})! This renders many extreme sensitive transducers useless for practical applications in real world samples.

Amperometric or fluorescence-based immuno-technology (immuno-electrodes or capillary fill devices with a fluorescence label, for example) has shown more promising results at the research prototype stage than other types of transducers. One preferred technique continues to be the indirect amperometric immunosensor approach described below, which uses enzymes as markers. The second emerging technique is based on fluorescence dyes which are now commercially available together with a simple covalently binding protocol (one to two steps only). Fluorescence can be as sensitive as enzyme-based amperometry especially if modern fluorophores are used with excitation wavelengths in the near IR and the emission shifted about 100 nm to higher wavelengths. Around 600 – 800 nm is an observation window for biological systems since no natural fluorescence (e.g., tryptophane) occurs in that region. Furthermore, cheap laser diodes of around 680 nm with high energy are available.

28.3.3.2. Indirect-Sensing Immuno-Probes using Marker Molecules

The most satisfactory marker molecules for use in this context have been fluorophores and enzymes. As regards the sensors themselves, several fundamental types exist.

Reactor immuno-systems are based on a cartridge or microcolumn packed with appropriate immobilized antibodies (which serve as so-called catching antibodies). As in the case of enzyme immunoassays, the competition, sandwich, and replacement principles are all subject to exploitation. The amount of bound (or replaced) antigen is determined via the extent of enzyme activity remaining after substrate addition, which can be established with an electrochemical or optical sensor. For many applications (on-line measurements, coupling with flow-injection analysis), repeated and automated measurements are required. Because of the extremely high loading of catching antibodies, reactor immunosensors have the ad-

vantages of high sensitivity, regenerability, and reproducibility. On average they can be reused 50 – 100 times without loss of activity. The US Naval Research Laboratories were very successful with respect to sensitivity, selectivity, reliability, and simplicity with several prototypes of reactor-based systems they developed under the guidance of Frances Ligler [382]. The developed immuno-systems allow field determinations and have been extensively tested and validated in the field of environmental analysis. In this respect, it is worthwhile to mention that the US EPA has adopted immuno-kits for certain important analytes in the environment, e.g. gasoline spills. If the fluorophore is carefully selected sensitivities in the ppb range with a FIA-like set-up and response times under 1 min are possible!

Membrane immuno-probes employ antibodies immobilized directly at the surface of a sensor, or at a membrane covering the sensor, as shown in Figure 80. An example of the use of this type of immunosensor is illustrated schematically in Figure 81. The analyte to be detected is an early myocardial infarction marker, the fatty acid binding protein (FABP). A modified Clark electrode is covered by a membrane bearing immobilized monoclonal catching antibodies directed against FABP. After binding of the analyte and washing, a secondary polyclonal antibody is added to the solution. This secondary antibody is conjugated with glucose oxidase (GOD). The next step is incubation followed by further washing, after which glucose is added to the measuring solution. Glucose is oxidized by GOD in an amount directly proportional to the amount of FABP, yielding H_2O_2, the concentration of which is monitored with an amperometric sensor, as shown in Figure 82. This system was developed at the ICB Muenster with the aim of developing a rapid test for a myocardial infarct to be performed actually in the ambulance with only one drop of blood. It works fully automatically. One drawback was, however, the use of a labile enzyme. Despite the fact that GOD is the most stable enzyme the experts know today, some precautions had to be taken and also the need for substrate addition and stopping of the reaction needed computer-controlled timing.

One example of a nice commercial immunosensor is produced by Serono Diagnostics. This fluorescence-based evanescent-wave immunosensor [374], [375], [376] incorporates a novel capillary-fill design, and is shown in Figure 83.

The system consists of two glass plates separated by a narrow capillary gap of 100 mm. The
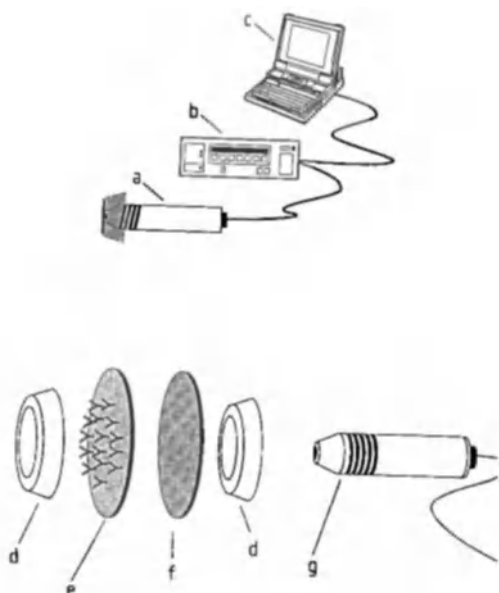
Figure 80. Design principle of an immunosensor system
a) Immunosensor; b) Electronic unit; c) Computing unit;
d) O-Rings; e) Catching antibody membrane; f) Dialysis
membrane; g) Pt vs. Ag/AgCl electrode

lower plate acts as an optical waveguide and contains on its surface an immobilized layer of antibodies. The system benefits from its capillary-fill system, by means of which a fixed volume of sample is drawn into the space between the plates regardless of the volume of the bulk sample (a blood droplet for example). Delivery of a highly reproducible volume of sample eliminates a significant source of error in analyte measurement. The Serono immunosensing format constitutes a competitive assay, though a sandwich-type assay could also be used. The upper plate contains on its surface a layer of fluorescence-labeled analyte trapped in a water-soluble matrix. On addition of sample the labeled analyte is released into solution, where it competes for antibody sites on the lower plate with any unlabeled analyte present in the sample. After a fixed incubation period the fluorescence signal of the evanescent waves is measured and related to analyte concentration.

This technology does have shortcomings, however, some of which are related to the practical problems associated with carrying out measurements on whole blood. These include poor capillary flow because of the "stickiness" of blood compared with water, increased incubation time resulting from slow dissolution of the labeled an-

alyte matrix, and the presence of fluorescent and highly colored molecules in blood, such as hemoglobin. Drawbacks include the necessity for an incubation period of several minutes (due to Ag – Ab interaction) and increased manufacturing costs (each analyte to be tested requires a dedicated sensor, which involves labeling and immobilization prior to physical construction).

While immunological analysis with one of the different ELISA methods works very well and reliably in the laboratory, the fabrication of pre-calibrated small stick-like probes has usually failed until recently. The main problem was the lack of information one has about the surface density of active immunopartners and a way of reproducibly manufacturing those single-use devices with antibodies or analyte molecules bound to the surface. Thus, one or more calibration solutions to be run simultaneously on all surfaces with immobilized biomolecules were always necessary. Only if full control of the surface density and state (active or not, freely accessible or not, etc.) were be given pre-calibrated immuno-probes for such fields other than clinical diagnostic and environmental field analysis with quantitative results would be feasible.

The Integrated Optical System — IOS. The solution to overcoming the main drawbacks of all hitherto developed biosensors, with the exception of the glucose sensor (poor storage capability), is a single-use sensor chip on which the labile biomolecules could be kept active forever by freeze-drying them. The solution to overcoming the pre-calibration problems is to employ a kinetic data evaluation method. The solution to overcoming the mostly diffusion limited immuno-reaction at an interface is to reduce the Nernstian diffusion layer thickness significantly so that differences in the association kinetics become apparent. To overcome poor sensitivity the solution is to use stable fluorophores with high quantum yield and emission in the near IR. In order to overcome washing steps, the evanescent field method for excitation and back-coupling has to be employed used. The IOS-chip (developed at the ICB Muenster) consists of a plastic baseplate with a refracting index of about 1.50. The mostly aqueous samples lie in the range of about 1.33. Thus, if the interface plastic base sample solution is illuminated, total reflection takes place and the light is reflected back into the plastic chip material. However, the reflected light does not fall abruptly to zero in the sample phase but fades exponentially (evanescent field). The evanescent field reaches out into the
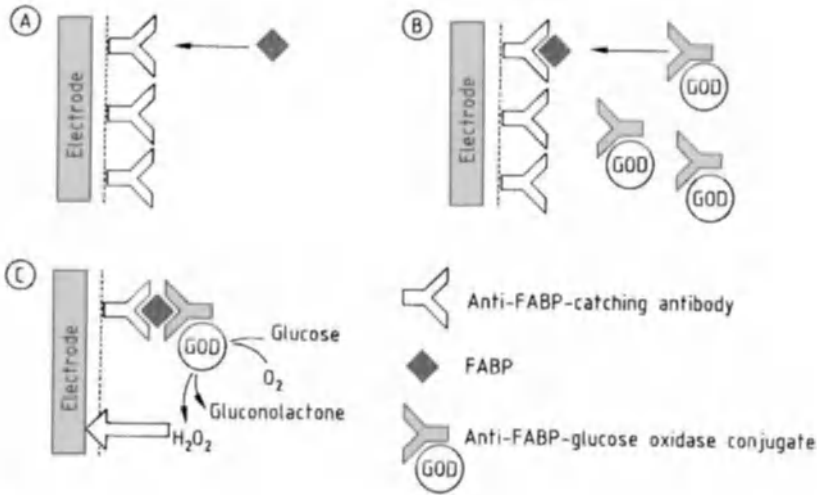
Figure 81. Schematic representation of the operation of an immunosensor for detecting the fatty acid binding protein (FABP)
A) Analyte binding step; B) Introduction of a secondary polyclonal antibody conjugated with glucose oxidase (GOD);
C) Determination of GOD activity after addition of glucose
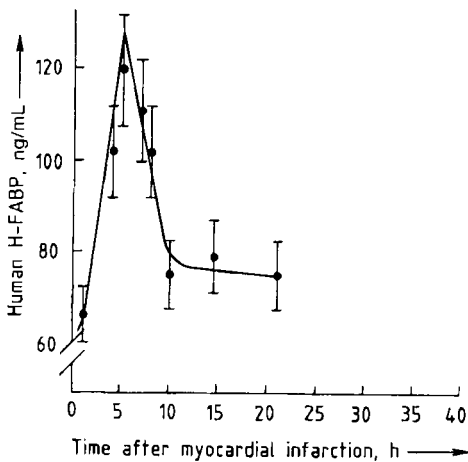
Figure 82. Analytical results obtained with the human heart FABP sensor described in Figure 81 as a function of time after onset of a myocardial infarction

sample medium only about 100 nm. The first 20 nm contain about 20 % of the optical power radiated onto the interface. The dimensions of most antibodies lie in a range between 5 and 7 nm. This means that only those fluorophore-labeled antibodies that are located directly on the surface are excited to fluorescence. Labeled but not specifically bound antibodies in the sample volume above the interface are not excited. Together with further improvements the IOS technology was pat-

ented wordwide by the ICB Muenster (prior art: DE 19628002 A2/DE 19711281 A2 inventors: Katerkamp, Meusel, Trau).

The plastic base chip, onto which a diode laser beam (λ near the absorption wavelength of the fluorophore) is located, is directed from underneath. The corresponding antibodies or antigens are immobilized on the chip (about 30×10 mm) surface in a simple and straightforward batch process of several thousands or more. Then, in order to reduce the Nernstian diffusion layer thickness a very thin double adhesive spacer (about 50 μm or lower) and a cut-out sample flowing channel is put onto the side with the immobilized molecules. A plastic cover with two openings for the sample in- and outlet is put on forming the necessary narrow detection channel for the hydrodynamics. The sample is brought into the measuring channel by suction forces applied at the outlet within the apparatus. However, on the sample inlet side of the chip, an easy to change container (dwell) is located. On its wall all necessary reagents, buffer, etc. are freeze-dried. The type of recognition molecule is determined by the method. In general all known ELISA methods are feasible with higher speed and sensitivity. Of particular interest are generic chip designs in which, e.g., a labelled antibody (e.g., with a Cy5 kit) reacts with the antigen of the sample in the small dwell at the inlet and is then filtered into the latter via a membrane filter with immobilized antigen. Thus, the
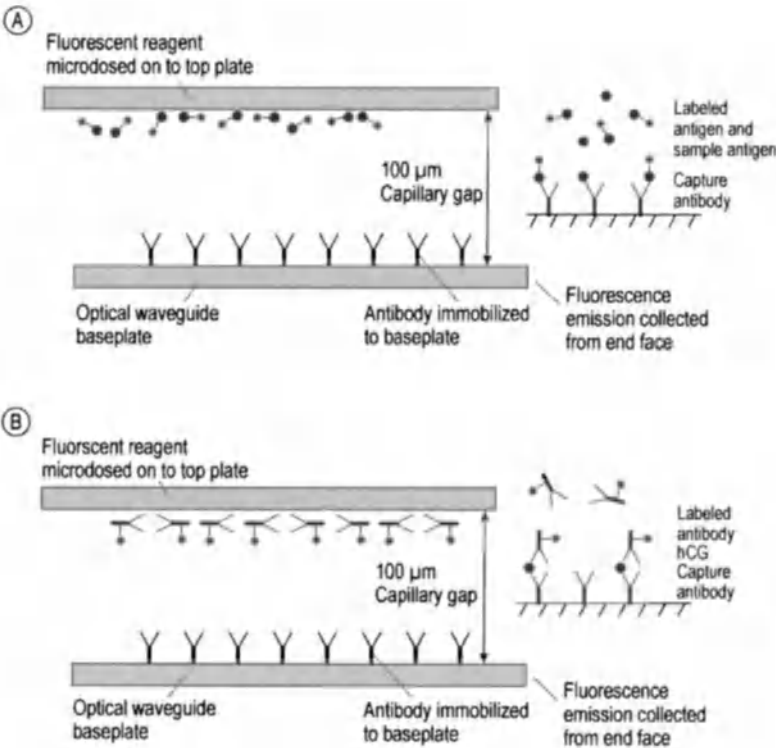
Figure 83. Principle of the patented Capillary-Fill Device. The upper part shows schematically a competitive immuno-assay; the lower part a sandwich assay. The optical waveguide is always the lower baseplate. Here a diode laser beam is used to produced an evanescent field in the sample inside the capillary

excess antibodies are not passing through since they are captured on the filter surface. However, the antigen – antibody complex reaches the optical channel after application of a slight vacuum on the outlet side. Here a catching layer for general antibody recognition makes the whole device generic! The dynamic measurement of the slope of the increasing fluorescence signal falling onto a simple optical detector below the chip is important. In the described operational mode this slope strictly depends on the concentration of the antigen. This slope can be measured in a few seconds! In cases of extremely low levels, the measuring time can be increased. By this kinetic evalution and viewing of the whole active area the inactivation of receptor molecules on the surface is more or less evened-out. The slope depends much more on the concentration of the partners and the constant hydrodynamics.

This IOS technology has recently been licensed to several companies in the field of diagnostics and will change the future of biosensing because of its inherent advantages:

- no washing and separation steps are needed for performing solid phase immunoassays in this arrangement
- reduction in measuring time from several hours to a few seconds
- dramatic increase in sensitivity ($<$ ppb)
- automatic computer control of the whole process:
  1) insert the analyte specific sensor chip in the read-out unit
  2) fill the sample container (sample volume needed for slope determination $<$ 200 µL)
  3) start the measurement process (chip moves into read-out unit and the outlet is punctured (closed previously to prevent uncontrolled intake by capillary forces) and the constant metering pump forces the sample through the measuring channel.

The low-cost chip manufactured in mass-produced plastic injection molding can also contain several functional layers for optimum sample treatment:
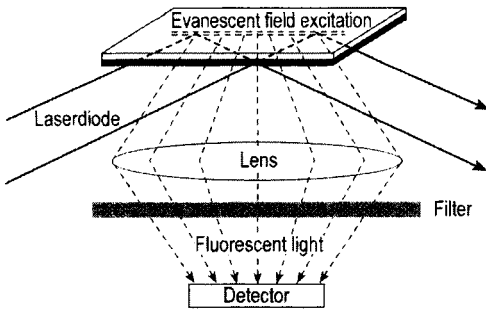
Figure 84. Optical set-up and construction of the IOS-chip. The laser light from a low-cost laser diode is directed towards the lower plastic baseplate of the immuno-probe chip. The evanescent field is produced within the flow channel produced by the spacer and the cover with sample entrance and outlet holes. The fluorescence of labeled molecules bound to the surface of the baseplate is collected, after passing an interference filter corresponding to the fluorescence wavelength, by a low-cost detector
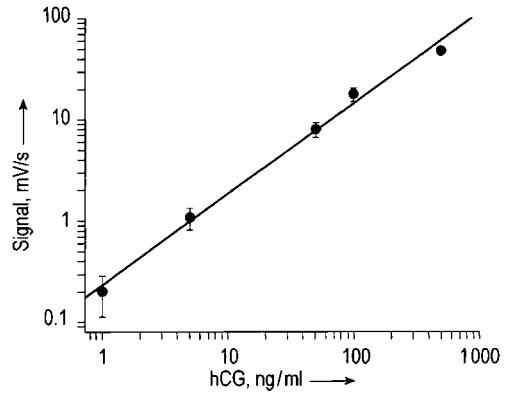
Figure 86. Calibration curve for the pregnancy hormone hCG. Each measurement was made with a new IOS sensor. The change of the fluorescence signal is plotted vs. the concentration
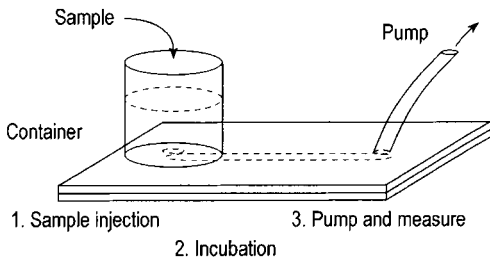
Figure 85. Simplicity of the whole measuring process. In the sample container (small dwell) all necessary reagents are present. After sample injection the reagents (with appropriate antibodies) dissolve in the sample and after a variable incubation for the immunological reaction the IOS-chip is automatically transported into the readout instrument where the outlet is punctured and a pump starts to suck the sample through the narrow measuring channel

- adjustment of pH and/or ionic strength
- filtration of sample (with or without analyte coating)
- separation of cellular blood
- components
- release of further immuno reagents or hydrodynamic puffers.

The Figure 84 shows the IOS chip design and Figure 85 the simple optical set-up. In Figure 86 a typical calibration plot for the quantification of the important pregnancy hormone hCG is presented. Note: This calibration curve was obtained by using single-use sensors at the corresponding concentration. The error brackets show the standard devia-

tion obtained in each case with five different IOS sensors from one production batch!

What is evident from that calibration curve is its extremely high dynamic range, not typical in the ELISA techniques! The reproducibility can be brought well below 5 % relative, which is outstanding in the trace analytical range. The matrix effect between an aqueous buffer solution and urine can be about 10 % (less) and between buffer and serum about 20 % (less) relative. Both can be compensated for with sufficient accuracy.

Outlook for the IOS technology-DNA-Chip. The IOS technology developed at the ICB Muenster is a base technology with high market orientation, with only as much sophisticated high-tech as needed, optimized for ease of production and quality control, increased production units by generic approach (one base chip for antibody capturing can be used for a wide number of different analytes depending only on the sample container at the inlet). All these advantages can even be topped by the fact that with another optical system delivering a focussed image of sensing spots onto a CCD camera, multi-analyte detection is feasible! Even further, fluorescence-based hybridization events are also sensed with the multi-analyte IOS technology. The upper limit is only given by patent violation reasons and lies at 400 detection spots. This application can further trigger mass-production and has the great advantage compared with other DNA-arrays that it quantifies automatically in real time!
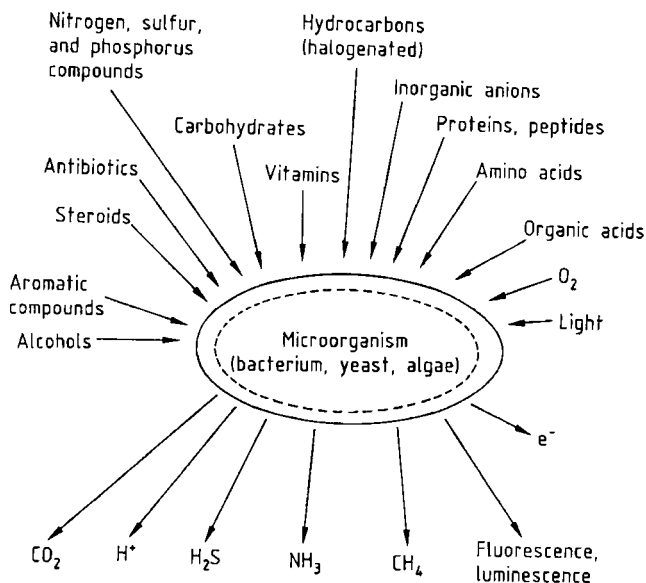
Figure 87. Microorganisms as receptor systems in biosensors useful for detecting specific groups of compounds

28.3.4. Whole-Cell Biosensors

Biological cells can be regarded as small membrane-bound structures containing a high concentration of chemicals, including enzymes, nucleic acids, ions, many types of protein, small organic molecules, and a host of others. A major challenge in exploiting whole (intact) cells in biosensor research is the problem of achieving selective transduction of a specific biochemical event or process within the context of this complex cellular chemistry [357], [370], [373], [374], [386] (Fig. 87).

The sensitivity and specificity of a particular cell with respect to a particular chemical varies enormously depending on the type, source, and environment of the cell and the precise nature of the chemical itself. It is not possible to provide any universal guidelines beyond such generalizations as the observation that most mammalian cells respond to a range of highly toxic chemicals, including cyanides. Even in this case it is known that hydrogen cyanide can interact with a number of different cellular components (such as electron transport proteins and enzymes) because of its small molecular size (which facilitates diffusion through biological membranes) and the ability of the cyanide anion to bind strongly to a wide variety of biomolecules.

The stability of biological cells is of critical importance in biosensor applications. In general, cells derived from higher organisms are rather difficult to isolate and maintain in a viable form for the extended periods of time required in most sensor formulations. The most biologically robust cells are derived from less complex organisms, particularly yeast, bacteria, and algae. Cells from these organisms can generally be stabilized in an in vitro environment more easily than complex cells; indeed, many of the organisms themselves are unicellular (yeast, bacteria, algae), leading to biosensors incorporating truly "living" biological components.

Natural cell differentiation and the growth characteristics of such whole-cell biosensors sometimes cause problems associated with signal drift over time. Nevertheless, there are potential advantages to be gained by exploiting whole cells in a sensor format, including:

1) Versatility (cells respond to a wide range of chemicals)
2) The presence in situ of biological structures supporting multicomponent or multistep biological reaction sequences useful for detection purposes and ensuring optimum enzyme activity (many enzymes lose activity when isolated in vitro)

3) Low preparation cost; i.e., growth in culture

For the reasons outlined above, nearly all whole-cell biosensors reported to date have employed relatively simple bacterial or algal cells [372], [387] – [390]. Existing whole-cell biosensor technology has also been based exclusively on electrochemical transduction of cellular responses.

The increasing interest in whole-cell biosensors since the first reports by DIVIES [387] and by RECHNNITZ [388] has led to the development of a large number of analyte-specific amperometric and potentiometric sensors. In these devices an immoblized layer of whole cells (entrapped in a gel, for example, or membrane-supported) is placed at the surface of an electrode. The cells then function as an enzyme source, where the enzyme of interest is present at a high relative concentration and is effectively able to function in its natural optimized environment. Bacterial cells in particular may contain very large amounts of particular enzymes relative to all others, especially if they are grown on a single source of carbon or nitrogen.

KARUBE [389] has pioneered in the development of many bacterial electrodes based on this principle using various bacteria depending on the target analyte. Ion-selective electrodes for NH_3, O_2, CO_2, H_2S and H^+ have all been used in conjunction with immobilized whole cells. RIEDEL et al. [390] have recently demonstrated that preincubation of certain bacterial electrodes with the desired analyte (substrate) can enhance the sensitivity of a sensor toward that chemical by a factor of as much as 25. This induction approach may prove to be widely applicable. The shelf-life of a whole-cell electrochemical sensor can extend up to several weeks with fully optimized storage conditions (low temperature, for example). Microbe thermistors (sensors that respond to the heat evolved during bacterial metabolism of a substrate) have also been developed, but these present problems once again with respect to analyte specificity.

A significant advance in whole-cell biosensor technology accompanied the development by RAWSON et al. [372] of a mediated electrochemical sensor employing a layer of blue-green algae (cyanobacteria). These authors demonstrated that a low molecular mass redox-active mediator, such as $Fe(CN)_6^{3-}$ or benzoquinone, could accept electrons from the photosynthetic energy chain of a cyanobacterium and relay them, via diffusion, to an electrode. The sensor proved capable of detecting aqueous herbicides on the basis of their inhibition of electron transport in the cells. This resulted in a dramatic fall in current, since fewer electrons reached the mediator and hence the electrode. The advantage of this approach to the use of whole cells is that it permits reasonably specific detection of a group of chemicals (herbicides, in this case) rather than limited-range detection of a single substance.

The principal application at the present time for cell sensors is in the measurement of biochemical oxygen demand (BOD) in wastewater. For this purpose yeast and/or bacterial cells (mainly the yeasts *Trichosporon* sp. or *Issatchenkia* and the bacterium *Rhodococcus erythropolis* [370], [386]) are immobilized on top of a Clark oxygen electrode and the oxygen consumption is measured. The great advantage of BOD sensors is a short measurement time (20 s to several minutes) compared with the classical BOD test in wastewater treatment, which requires five days. The main disadvantage is that a BOD sensor reflects only low molecular mass and easily assimilable substrates. High molecular mass substrates such as proteins and starch cannot penetrate the protective dialysis membrane of the sensor. Pretreatment with hydrolytic enzymes could solve this problem. In general, however, BOD sensors, represent a promising approach to the control of wastewater facilities.

28.3.5. Problems and Future Prospects

Compared with traditional sensors and classical methods of analysis, biosensors offer an attractive alternative for the rapid, selective, and highly sensitive assay of a wide variety of analytes. Further advantages of this cheap and economical tool emerge with the possibility of direct measurements within complex matrices, demanding only a rudimentary sample pretreatment, and the prospect of an uncomplicated measuring procedure that avoids the use of toxic reagents and organic solvents. Potential fields of application for biosensors are suggested in Table 21.

Current biosensor research is concerned with a variety of unsolved problems posing barriers to further expansion within the analytical market. Until recently, biosensors were available for only a few analytes (roughly two dozen). This is mainly a consequence of insufficient reliability associated with poor stability of the biomaterial, including a dependence on physical and chemical parameters in the environment and interferences within the transducers. In the case of implantable biosensors

Table 21. Potential fields of application for biosensors

Field of application	Biosensor opportunities
Medicine	patient self-control. home health care (glucose, lactate, creatinine, phenylalanine, histamine)
	in vivo analysis
	long-term in vivo control of metabolites and drugs
	control element (biotic sensor) for artifical prosthesis and organs
	rapid analysis at intensive-care units
	surface imaging of organs during implantation
	bedside monitoring
Clinical chemistry	diagnostics for metabolites. drugs, enzymes, vitamins, hormones, allergies, infectious diseases, cancer markers, pregnancy, etc.
	laboratory safety
Environmental protection	pollution control
	monitoring/screening of toxic compounds in water supplies, solid and liquid wastes, soil and air (e.g., pesticides, inorganic ions, explosives, oils, PAHs, PCBs, microorganisms, volatile vapors and gases)
	self control of industrial companies and farms
	alarm systems for signaling hazardous conditions
	determination of organic load (BOD)
Chemical, pharmaceutical, food industry	monitoring and control of fermentation processes and cell cultivation (substrates, metabolites, products)
	food quality control (screening/detection of microbial contaminations; estimation of freshness, shelf life; olfactory qualities and flavor; rancidity; analysis of fats, proteins, carbohydrates in food)
	study of efficiency of drugs
	detection of leakage and hazardous concentrations of liquids and gases in buildings and mine shafts
	indoor air quality checks
	location of oil deposits
Agriculture	quality control of soils
	estimation of degradation/rottage (e.g., of biodegradable waste, or in wood or plant storage)
	rapid determination of quality parameters of milk
Military	detection of chemical and biological warfare agents (e.g., nerve gases, pathogenic bacteria, viruses)

for in vivo monitoring of metabolites or feedback control of artificial organs (e.g., the pancreas), a lack of biocompatibility with the body environment, organic interferences, and inadequate stability of the biological receptor have so far prevented long-term implantation. Nevertheless, with respect to rapid, sensitive, and selective low-cost determination of a great variety of relevant compounds no suitable alternative to the principle of biosensors is apparent at this time.

Constructing biosensors with greater reliability requires a better understanding of the basic principles of enzyme catalysis and immunochemical reactions, protein structures, pathways for receptor-based signal amplification, and the interfacial behavior of biocompounds at the artificial transducer surface. Biosensor research must therefore be directed toward the following points:

1) Protein engineering, with the aim of enhancing the stability of enzymes, catalytic antibodies, and bifunctional enzymes
2) Screening stable enzymes from microorganisms (extremophiles)
3) Utilization of antibody fragments (Fab, Fab'), bispecific antibodies, and single-chain and single-domain antibodies (tailored through genetic engineering and its expression in bacteria)
4) Recombinant antibodies
5) Modification of the efficiency, specificity, and modulation of antibodies by site-directed mutagenesis
6) Use of DNA libraries for the production of monoclonal antibodies to meet specific demands of the desired transducer

7) Selection and specification of biological receptor systems applicable to the quantification of olfaction and taste

8) Manipulating enzymes and cofactors by electrically wiring redox-active centers to transducer surfaces, utilizing the principles of both molecular electronics and biological electron transfer

9) Investigation of interfacial effects and interactions between biocompounds and transducer surfaces at the molecular and atomic levels by exploiting new methods of surface analysis (e.g., scanning microscopy techniques, fast-atomic-bombardment mass spectrometry, laser-assisted mass spectrometry, time-of-flight secondary ion mass spectrometry, Fourier-transform infrared spectrometry, ellipsometry, X-ray photoelectron spectroscopy, and electron microscopy)

10) Materials research, including biocompatible membranes (based on compounds found within the body) and redox-active electrically conducting (metal)organics and conjugated polymers

11) Development of (multifunctional) silicon-based ultramicroelectrodes covalently linked to biomaterials

12) Preventing the fouling of the biological compounds through microbiological degradation and attacks from proteases

Concerning methodological applications of biosensors, most of the recent developments relate to enzyme electrodes integrated into laboratory analyzers for the determination of low molecular mass compounds in the medical and biotechnological fields. A further important line of development is directed toward disposable biosensors for screening and semiquantitative analysis in environmental control, patient home diagnostics, and food-quality control. Direct testing would avoid the time-consuming procedures of sample storage and transfer as well as expensive laboratory analyses. A medium-term application is expected to be the use of biosensors in (quasi)continuously operating random-process analyzers based on micromachined fluidics elements, together with low-maintenance systems for the control and monitoring of water supplies and clinical bedside monitoring. Finally, a long-term target is the development of highly reliable implantable biosensors for long-term implantation. Biosensor applications in the field of instrumentation should also contribute to the solution of many important problems in conventional measuring techniques, opening major new areas to modern analysis.

28.4. Actuators and Instrumentation

The instrumentation associated with chemical sensors is rather simple and inexpensive, especially in the case of electrochemical transducers and the new IOS technology developed at the ICB Muenster . All that is normally required is a high-ohmic voltmeter (pH meter), a conductometer, and/or polarographic equipment. Fiber-optical devices, on the other hand, can be quite expensive depending on quality requirements with respect to the monochromatic light source. Spectrophotometers with fiber-optical inputs and outputs are already commercially available. A laser source is required in certain cases of sensitive fluorescence measurements with small-volume samples. Exploiting the modern techniques of acoustic wave modulation also requires more than a simple frequency counter. For example, a network analyzer can be of great help in following the shift in a resonance curve caused by interaction of an analyte with the transducer surface.

The highest costs are usually related to development of a selective analyte recognition layer, perhaps the most important part of a complete chemical sensor. Transducer technology is much more highly developed; indeed, there seems to be little need here for further improvement without equivalent improvements in the selectivity of sensor elements.

One of the most important areas of current sensor research is microsystem technology (mst). A microsystem is defined as one too small to be fabricated by traditional miniaturized mechanical techniques, especially one that is dependent on modern mass-production techniques involving silicon or LIGA technology. A *system* must, by definition, consist of at least two interactive processes. A typical example is the antiblocking brake system (ABS) in an automobile, in which acceleration sensors are capable of detecting a blocked wheel and selectively controlling the brake power to that wheel via actuators.

Examples of analogous complete systems in chemistry are still rare. The best example is perhaps the "artificial pancreas," in which a glucose sensor measures the actual blood-glucose concentrations in order to control an appropriate actuator, the insulin pump. Bedside devices for this purpose

have operated successfully for many years, but all attempts at implantation have so far failed, mainly because of the limited lifetime of the glucose biosensor and biofouling of the biosensor membrane by macrophages and other body constituents.

A complete system providing both a sensor and an actuator would be ideal in the field of process control, but because of a lack of truly reliable chemical sensors on the market the concept has not been widely implemented. One exception relates to the analytical method of coulometry, a technique that offers great potential for delivering chemical compounds to a controlled reaction. Especially attractive in this context is the method of constant-current coulometry, which can be carried out with an end-point sensor and a coulometric actuator for maintaining a generator current until the end-point has been reached. In this case both of the required devices can be miniaturized and constructed with the same technology.

28.5. Future Trends and Outlook

Surveys that have been conducted on the subject of chemical sensors all predict a growing market during the late 1990s and the early years of the 21st century [391] – [393]. Anticipated developments in automation, environmental protection, health care, and high technology ensure a strong demand for sensors. However, devices of this type must become increasingly reliable, must survive a thorough validation, and become even simpler to use. Japanese researchers see the first area for mass application of sensors to be the health care industry, especially with respect to the elderly. The demographic prospect of an increasing number of retirees and a lack of sufficient care centers may trigger the use of many sensors for monitoring the health of people living for as long as possible in their own homes. It is now clear that the application area of health care will dominate the application of biosensors. The "point of care" idea will replace large central laboratories to which the labile biological sample has to be transported. The reliabilty, built-in validation checks, and ease of use will allow their use by the normal physician who will have then the result within a few minutes. While the numbers of DNA arrays used in the genom research area will become saturated shortly, the use of simple arrays for looking for specific protein expression patterns will increase. Combining bio-informatics with the latter

in finding super-marker molecules for extremely early tumor diagnostics will be possible. This leads to the corresponding mass-production of these diagnostic chips.

With respect to environmental chemistry, controls over the emission of organic compounds will undoubtedly be strengthened in the future. For example, municipal incinerators without provisions for the effective measurement of all hazardous organic compounds will soon be strictly forbidden everywhere. It seems likely that a large fraction of the moneys previously spent for defense purposes will be redirected toward improvements in continuous measurement technology. Sensors will be required to compete in this respect with such fast separation methods as capillary electrophoresis or flash chromatography, which permit the separation of most compounds within only a few minutes.

One critical remark is in order regarding papers published in the field of chemical sensors. Many such reports reflect a lack of input from qualified analytical chemists during the development and testing stages. Insufficient understanding is often apparent with respect to the analytical process itself and the difficulties involved in obtaining a truly reliable result; indeed, many sensor researchers seem to be unaware of such basic problems as sample – matrix effects. Anyone who claims to have developed a new sensor should be required to demonstrate its effectiveness with a real sample under real conditions. The latter is required and called validation by international quality standards (ISO 25). Publications that fail to reveal this most important feature of a sensor are also essentially worthless. Chemical sensors are analytical devices, so they should be subject to the same standards of characterization as other instrumental tools. Quality-control considerations mandate critical evaluation of any new molecular sensor presented to the scientific community. Every report concerning a new sensor development should clearly delineate the crucial features, with none omitted.

The most important parameters for characterizing or validating a chemical sensor or a biosensor include:

1) *Selectivity.*
   The selectivity with respect to important known interferents for the particular application in question should always be expressed numerically; expressions like "relative selectivity" toward the analyte, or "not affected by

typical concentrations of interfering compounds" are inappropriate.

2) *Sensitivity and Limit of Detection (LOD).*
   The slope of a calibration curve expressed in the appropriate units (e.g. microamps, millivolts, or hertz) per millimole or micromole should always be provided both for a matrix-free solution and for a sample in its matrix. The corresponding LOD should also be reported in both cases. In the absence of a blank the signal 3 × larger than the noise determines the LOD. The limit of quantification is typical 10 × LOD.

3) *Stability of the Analytical Function (Drift).*
   A zero-point drift value should be provided for a sensor in a matrix-free solution and also with respect to real samples involving a matrix. The same applies to the stability of the slope (varying sensor sensitivity).

4) *Response Time.*
   Since response time is also a function of the sample matrix it should be measured both in pure calibration solutions and with real samples. Generally two different response times are observed, one for a concentration rise and one for decreasing analyte concentration, with the latter usually slightly greater. Since IUPAC has not yet recommended a specific procedure for measuring this parameter, one should express the response time in the form of a time constant ($=1/e$ of the final value), since this is the approach taken with electronic circuits. Only in the case of a sensor that never reaches its final value should other approaches be considered for characterizing the rate of response to a changing analyte concentration.

5) *Sensor Lifetime.*
   Especially in the field of biosensors the lifetime of a sensor in contact with a calibration solution or a real sample, or during storage in a refrigerator with an optimal conditioning solution, is of considerable interest, and it should always be mentioned. Conditions leading to rapid sensor fouling should also be identified in order to permit appropriate precautions to be taken. Examples include the fouling of certain enzymes by inhibitors or heavy metals, or microbiological attack. Any use of conditioning solutions containing metal complexation reagents, inhibitor scavengers, sodium azide, etc., should be clearly described.

Other important information might include total power consumption, auxiliary equipment costs, total elapsed time between failures, temperature compensation measures, or maintenance requirements.

Finally, sensor developers must remain aware of progress within instrumental analytical chemistry generally, especially in the area of fast high-resolution chromatography. Indeed, progress has already been so great in the latter field that more than 20 anions (including phosphate) can now be separated completely in less time than would be required for a response from a multienzyme phosphate biosensor! The serious challenge this poses to sensor advocates will be obvious.

28.6. References

[1] W. Göpel, J. Hesse, J. N. Zemel (eds.): *Sensors – A Comprehensive Survey,* vols. 1 – 8 + **updates,** VCH Verlagsgesellschaft, Weinheim 1989.
[2] T. Seiyama, S. Yamauchi (eds.): *Chemical Sensor Technology,* vols. 1 + 4, Elsevier Science Publisher B, Amsterdam 1988.
[3] R. W. Murray et al.: "Chemical Sensors and Microinstrumentation," *ACS Symp. Ser.* **403** (1989).
[4] J. Janata: *Principles of Chemical Sensors.* Plenum Press, New York 1989.
[5] K. Cammann, H. Galster: *Das Arbeiten mit ionenselektiven Elektroden,* 3rd ed., Springer Verlag, Berlin 1996.
[6] J. Janata, R. Huber: *Solid State Chemical Sensors.* Academic Press, New York 1985.
[7] K. Cammann et al.: *Analytiker Taschenbuch,* **vol. 15,** Springer-Verlag, Berlin – Heidelberg 1997, pp. 3 – 40.
[8] J. Reinbold, K. Cammann: *Analytiker Taschenbuch,* **vol. 16,** Springer-Verlag, Berlin – Heidelberg 1997, pp. 3 – 42.
[9] K. Cammann (ed.): *Instrumentelle Analytische Chemie,* Spektrum-Verlag, Heidelberg 2000.
[10] T. Takeuchi, *Sens. Actuators* **14** (1988) 109.
[11] G. Hötzel, H. M. Wiedemann, *Sens. Rep.* **4** (1989) 32.
[12] J. P. Pohl, *GIT Fachz. Lab.* **5** (1987) 379.
[13] K. Cammann et al.: *Angew. Chem.* **103** (1991) 519; *Angew. Chem. Int. Ed. Engl.* **30** (1991) 516 – 539.
[14] W. Masing: *Handbuch der Qualitätssicherung,* Hanser Verlag, München 1988.
[15] K. Cammann: *Working with Ion-Selective Electrodes,* Springer Verlag, Berlin 1979.
[16] A. K. Covington: *Ion Selective Electrode Methodology,* vols. I and II, CRC Press, Boca Raton, Fla., 1979.
[17] H. Freiser (ed.): *Ion Selective Electrodes in Analytical Chemistry,* Plenum Press, New York 1979.
[18] K. Schierbaum, *Sens. Actuators B* **18 – 19** (1994) 71 – 76.
[19] J. Reinbold et al., *Sens. Actuators B* **18 – 19** (1994) 77 – 81.
[20] N. Taguchi, JP 4 5 38 200, 1962.
[21] T. Seiyama, A. Kato, K. Fujiishi, M. Nagatani, *Anal. Chem.* **34** (1962) 1502.
[22] Y. Nakatani, M. Sakai, M. Matsuoka in T. Seyama et al. (eds.): *Proc. Int. Meet. Chem. Sensors (Fukuoka, Japan),* Elsevier, Amsterdam, Sept. 1983, pp. 147 – 152.

[23] S. R. Morrison: "Semiconductor Gas Sensors," *Sens. Actuators* **2** (1982) 329.

[24] P. J. Shaver, *Appl. Phys. Lett.* **11** (1967) 255.

[25] J. C. Loh, JP 4 3 28 560; FR 1 545 292, 1967.

[26] N. Yamazoe, Y. Krakowa, T. Seiyama: "Effects of Additives on Semiconductor Gas Sensors," *Sens. Actuators* **4** (1983) 283.

[27] S. Matsushima, Y. Teraoka, N. Miura, N. Yamazoe, *Jpn. J. Appl. Phys.* **27** (1988) 1798.

[28] M. Nitta, S. Kanefusa, S. Ohtani, M. Haradome, *J. Electr. Mater.* **13** (1984) 15.

[29] A. Chiba: "Development of the TGS Gas Sensor," in S. Yamauchi (ed.): *Chemical Sensor Technology,* vol. 4, Elsevier, Amsterdam 1992, pp. 1–18.

[30] T. A. Jones: "Characterisation of Semiconductor Gas Sensors," in P. T. Moseley, B. C. Tofield (eds.): *Solid State Gas Sensors,* Adam Hilger, Bristol 1987.

[31] G. Heiland, D. Kohl, *Sens. Actuators* **8** (1985) 227.

[32] W. Weppner: "Halbleiter-Sensoren" in R. Grabowski (ed.): *Sensoren und Aktoren,* VDE-Verlag, Berlin 1991, pp. 167–183.

[33] D. E. Williams: "Characterization of Semiconductor Gas Sensors," in P. T. Moseley, B. C. Tofield (eds.): *Solid State Gas Sensors,* Adam Hilger, Bristol 1987.

[34] N. Yamazoe, N. Miura: "Some Basic Aspects of Semiconductor Gas Sensors," in S. Yamauchi (ed.): *Chemical Sensor Technology,* vol. 4, Elsevier, Amsterdam 1992, pp. 19–41.

[35] G. Heiland: "Homogeneous Semiconducting Gas Sensors," *Sens. Actuators* **2** (1982) 343.

[36] H. Windischmann, P. Mark, *J. Electrochem. Soc.* **126** (1979) 627.

[37] S. Chang, *J. Vac. Sci. Technol.* **17** (1980) 366.

[38] J. Watson: "The Tin Dioxide Gas Sensor and Its Applications," *Sens. Actuators* **5** (1984) 29.

[39] N. Komori, S. Sakai, K. Konatsu in T. Seyama et al. (eds.): *Proc. Int. Meet. Chem. Sensors (Fukuoka, Japan),* Elsevier, Amsterdam, Sept. 1983, p. 57.

[40] B. Ahlers: *Dissertation,* University of Muenster 1996.

[41] M. L. Hitchmann, H. A. O. Hill, *Chem. Br.* **22** (1986) 1177.

[42] D. Ammann: *Ion Selective Microelectrodes – Principles, Design and Application,* Springer Verlag, Berlin 1986.

[43] P. L. Bailey: *Analysis with Ion-Selective Electrodes,* Heyden & Sons, London 1976.

[44] R. Bock: "Nachweis- und Bestimmungsmethoden," in *Methoden der Analytischen Chemie,* vol. 2, part 2, Verlag Chemie, Weinheim 1984, pp. 33–83.

[45] K. Cammann: "Fehlerquellen bei Messungen mit ionenselektiven Elektroden," in *Analytiker-Taschenbuch,* vol. 1, Springer Verlag, Berlin 1980.

[46] R. A. Durst: "Ion-Selective Electrodes," *NBS Special Publication No. 314.* 1969.

[47] S. Ebel, W. Parzefall: *Experimentelle Einführung in die Potentiometrie,* Verlag Chemie, Weinheim 1975.

[48] E. Bakker, *Trends Anal. Chem.* **16** (1997) 252–260.

[49] G. Horvai, *Trends Anal. Chem.* **15** (1997) 260–266.

[50] G. Eisenman: *Glass Electrodes for Hydrogen and Other Cations, Principles and Practice,* Marcel Dekker, New York 1967.

[51] D. J. G. Ives, J. Janz: *Reference Electrodes,* Academic Press, New York 1961.

[52] R. Koryta: *Ion Selective Electrodes,* Cambridge University Press, Cambridge 1975.

[53] T. S. Ma, S. S. M. Hassan: *Organic Analysis Using Ion-Selective Electrodes,* 2 vols., Academic Press, New York 1982.

[54] G. J. Moody, J. D. R. Thomas: *Selective Ion Electrodes,* Merrow Publishing, Watford 1977.

[55] F. Oehme: *Ionenselektive Elektroden: CHEMFETs – ISFETs – pH-FETs, Grundlagen, Bauformen und Anwendungen,* Hüthig Verlag, Heidelberg 1991.

[56] E. Pungor: *Ion-Selective Electrodes.* Akadémiai Kiadó, Budapest 1973.

[57] F. Honold, B. Honold: *Ionenselective Elektroden.* Birkhäuser Verlag, Basel 1991.

[58] K. Umezawa, Y. Umezawa: *Selectivity Coefficients for Ion Selective Electrodes,* University of Tokyo Press, Tokyo 1983.

[59] E. Bakker, P. Bühlmann, E. Pretsch: "Carrier-Based Ion-Selective Electrodes and Bulk Optodes. 1. General Characteristics," *Chem. Rev.* **97** (1997) 3083–3132.

[60] P. Bühlmann, E. Pretsch, E. Bakker: "Carrier-Based Ion-Selective Electrodes and Bulk Optodes. 2. Ionophores for Potentiometric and Optical Sensors," *Chem. Rev.* **98** (1998) 1593–1687.

[61] Y. Mi, S. Mathison, R. Goines, A. Logue, A. Bakker, *Anal. Chim. Acta* **397** (1999) 103–111.

[62] A. Schwake: *Dissertation,* University of Muenster 1999.

[63] F. Zuther: *Dissertation,* University of Muenster 1997.

[64] A. J. Bard, L. R. Faulkner: *Electrochemical Methods,* J. Wiley, New York 1980.

[65] A. M. Bond, H. B. Greenhill, I. D. Heritage, J. B. Reust, *Anal. Chim. Acta* **165** (1984) 209.

[66] A. J. Bard: *Electroanalytical Chemistry,* vol. 9, Marcel Dekker, New York 1976.

[67] G. Henze, R. Neeb: *Elektrochemische Analytik,* Springer Verlag, Berlin 1985.

[68] K. Cammann: "Elektrochemische Verfahren," in H. Naumer, W. Heller (eds.): *Untersuchungsmethoden in der Chemie,* Thieme Verlag, Stuttgart 1986.

[69] P. T. Kissinger, W. R. Heinemann, *J. Chem. Educ.* **60** (1983) 702.

[70] J. Osteryoung, J. J. O'Dea in A. J. Bard (ed.): *Electroanalytical Chemistry,* **vol. 14,** Marcel Dekker, New York 1986, pp. 209–308.

[71] P. Vadgama, G. Davis, *Med. Lab. Sci.* **42** (1985) 333.

[72] L. C. Clark, US 2 913 386, 1958.

[73] H. Suzuki et al., *Sens. Actuators B* **2** (1990) 297.

[74] S. C. Cha, M. J. Shao, C. C. Liu, *Sens. Actuators B* **2** (1990) 239.

[75] W. Sansen et al., *Sens. Actuators B* **1** (1990) 298.

[76] H. Suzuki et al., *Sens. Actuators B* **1** (1990) 528.

[77] H. Suzuki et al., *Sens. Actuators B* **1** (1990) 275.

[78] D. S. Austin et al. *J. Electroanal. Chem. Interfacial Electrochem.* **168** (1984) 227.

[79] R. M. Wightman, *Anal. Chem.* **53** (1981) 1125.

[80] T. Hepel, J. Osteryoung, *J. Electrochem. Soc.* **133** (1986) 752.

[81] L. Zhaohui et al., *J. Electroanal. Chem. Interfacial Electrochem.* **259** (1989) 39.

[82] C. L. Colyer, J. C. Myland, K. B. Oldham, *J. Electroanal. Chem. Interfacial Electrochem.* **263** (1989) 1.

[83] J. Osteryoung, S. T. Singleton, *Anal. Chem.* **61** (1989) 1211.

[84] R. E. Howard, E. Hu: *Science and Technology of Microfabrication,* Materials Research Society, Pittsburgh 1987.

[85] C. D. Baer, N. J. Stone, D. A. Sweigart, *Anal. Chem.* **58** (1988) 78.

[86] J. Ghoroghchiana et al., *Anal. Chem.* **58** (1988) 2278.

[87] H. Meyer et al., *Anal. Chem.* **67** (1995) 1164–1170.

[88] M. Wittkamp, G. Chemnitius, K. Cammann, M. Rospert, M. Mokwa, *Sens. Actuators B* **43** (1997) 87–93.

[89] W. Göpel, J. Hesse, J. N. Zemel: *Sensors*, vol. 2, VCH Verlagsgesellschaft, Weinheim 1991.

[90] W. Göpel, J. Hesse, J. N. Zemel: *Sensors*, vol. 3, VCH Verlagsgesellschaft, Weinheim 1991.

[91] P. T. Moseley, B. C. Tofield: *Solid State Gas Sensors*, Adam Hilger, Bristol 1987.

[92] P. T. Moseley, J. O. W. Norris, D. E. Williams: *Techniques and Mechanisms in Gas Sensing*, Adam Hilger, Bristol 1991.

[93] H. Emons, G. Jokuszies, *Z. Chem.* **28** (1988) 197.

[94] R. Kalvoda, *Electroanalysis* **2** (1990) 341.

[95] F. Opekar, A. Trojanek, *Anal. Chim. Acta* **203** (1987) 1.

[96] ISO/TC 147 (draft):*Electrolytic Conductivity*, Beuth Verlag, Berlin.

[97] DIN 38 404:*Deutsche Verfahren zur Wasser-, Abwasser- und Schlammuntersuchung*, VCH Verlagsgesellschaft, Weinheim 1993.

[98] F. Oehme in W. Göpel, J. Hesse, J. N. Zemel (eds.): *Sensors*, vol. 2, VCH Verlagsgesellschaft, Weinheim 1991.

[99] F. Oehme, R. Bänninger: *ABC der Konduktometrie*, Polymetron AG, Mönchaltorf (Switzerland).

[100] K. Rommel, *VGB Kraftwerkstech.* **65** (1985) 417.

[101] K. Rommel: *Kleine Leitfähigkeitsfibel*, Wissenschaftlich-Technische Werke, Weilheim 1988.

[102] F. Oehme, M. Jola, *Betriebsmeßtechnik*, Hüthig Verlag, Heidelberg 1982.

[103] M. Niggemann: *Dissertation*, University of Muenster 1999.

[104] W. H. Brattain, J. Bardeen, *Bell Syst. Tech. J.* **32** (1952) 1.

[105] G. Heiland, *Z. Phys.* **138** (1954) 459.

[106] J. Watson, *Sens. Actuators B* **8** (1992) 173.

[107] N. Taguchi, US 3 644 795, 1972.

[108] D.-D. Lee, D.-H. Choi, *Sens. Actuators B* **1** (1990) 231.

[109] B. C. Tofield in P. T. Moseley, B. C. Tofield (eds.): *Solid State Gas Sensors*, Adam Hilger, Bristol 1987.

[110] R. Lalauze, C. Pijolat, S. Vincent, L. Bruno, *Sens. Actuators B* **8** (1992) 237.

[111] J. Watson, *Sens. Actuators* **5** (1984) 29.

[112] R. M. Geatches, A. V. Chadwick, J. D. Wright, *Sens. Actuators B* **4** (1991) 467.

[113] S. R. Morrison, *Sens. Actuators* **12** (1987) 425.

[114] S. Matsushima et al., *Sens. Actuators B* **9** (1992) 71.

[115] T. Maekawa et al., *Sens. Actuators B* **9** (1992) 63.

[116] S. Nakata, H. Nakamura, K. Yoshikawa, *Sens. Actuators B* **8** (1992) 187.

[117] U. Kirner et al., *Sens. Actuators B* **1** (1990) 103.

[118] N. Li, T.-C. Tan, *Sens. Actuators B* **9** (1992) 91.

[119] U. Lampe, M. Fleischer, H. Meixner, *Sens. Actuators B* **17** (1994) 187–196.

[120] N. Koshizaki, K. Yasumoto, K. Suga, *Sens. Actuators B* **9** (1992) 17.

[121] M. Egashira, Y. Shimizu, Y. Takao, *Sens. Actuators B* **1** (1990) 108.

[122] P. T. Moseley, D. E. Williams, *Sens. Actuators B* **1** (1990) 113.

[123] E. Ye. Gutman, I. A. Myasnikov, *Sens. Actuators B* **1** (1990) 210.

[124] R. S. Tieman, W. R. Heineman, J. Johnson, R. Seguin, *Sens. Actuators B* **8** (1992) 199.

[125] Y. Sadaoka, T. A. Jones, W. Göpel, *Sens. Actuators B* **1** (1990) 148.

[126] S. Dogo, J. P. Germain, C. Maleysson, A. Pauly, *Sens. Actuators B* **8** (1992) 257.

[127] J. W. Gardner, M. Z. Iskandarani, B. Bott, *Sens. Actuators B* **9** (1992) 133.

[128] M. Trometer et al., *Sens. Actuators B* **8** (1992) 129.

[129] J. D. Wright, *Prog. Surf. Sci.* **31** (1989) 1.

[130] T. A. Temofonte, K. F. Schoch, *J. Appl. Phys.* **65** (1989) 1350.

[131] S. Kanefusa, M. Nitta, *Sens. Actuators B* **9** (1992) 85.

[132] W. Göpel, K. D. Schierbaum, D. Schmeisser, D. Wiemhöfer, *Sens. Actuators* **17** (1989) 377.

[133] B. Bott, S. C. Thorpe in P. T. Moseley, J. O. W. Norris, D. E. Williams (eds.): *Techniques and Mechanisms in Gas Sensing*, Adam Hilger, Bristol 1991.

[134] T. Hanawa, S. Kuwabata, H. Yoneyama, *J. Chem. Soc. Faraday Trans. 1* **84** (1988) 1587.

[135] C. Nylander, M. Armgarth, I. Lundström, *Anal. Chem. Symp. Ser.* **17** (1983) 203.

[136] P. N. Bartlett, P. B. M. Archer, S. K. Ling-Chung, *Sens. Actuators* **19** (1989) 125.

[137] P. N. Bartlett, P. B. M. Archer, S. K. Ling-Chung, *Sens. Actuators* **19** (1989) 141.

[138] H. Arai, T. Seiyama: "Humidity Control," in W. Göpel, J. Hesse, J. N. Zemel (eds.): *Sensors*, vol. 2, VCH Verlagsgesellschaft, Weinheim 1991.

[139] A. K. Michell in P. T. Moseley, J. O. W. Norris, D. E. Williams (eds.): *Techniques and Mechanisms in Gas Sensing*, Adam Hilger, Bristol 1991.

[140] T. Nenov, S. Yordanov, *Sens. Actuators B* **8** (1992) 117.

[141] Y. Sakai, Y. Sadaoka, H. Fukumoto, *Sens. Actuators* **13** (1988) 243.

[142] P. Bergveld, *IEEE Trans. Biomed. Eng.* **BME-17** (1970) 70.

[143] K. Buhlmann: *Dissertation*, University of Muenster 1997.

[144] K. Buhlmann, B. Schlatt, K. Cammann, A. Shulga, *Sens. Actuators B* **49** (1998) 156–165.

[145] T. Matsuo, M. Esashi, K. Inuma, *Proc. Dig. Joint Meeting*, Tohoku Sect. IEEEJ, Oct. 1971.

[146] J. E. Moneyron, A. De Roy, J. P. Besse, *Solid State Ionics* **46** (1991) 175.

[147] S. M. Sze: *Physics of Semiconductor Devices*, Wiley, New York 1981.

[148] P. Bergveld, A. Sibbald: *Analytical and Biomedical Applications of Ion Sensitive Field Effect Transistors*, Elsevier, Amsterdam 1988.

[149] Y. G. Vlasov, D. E. Hackleman, R. P. Buck, *Anal. Chem.* **51** (1979) 1579.

[150] W. Moritz, I. Meierhöffer, L. Müller, *Sens. Actuators* **15** (1988) 211.

[151] S. D. Moss, J. Janata, C. C. Johnson, *Anal. Chem.* **47** (1975) 2238.

[152] A. Sibbald, P. D. Whalley, A. K. Covington, *Anal. Chim. Acta* **159** (1984) 47.

[153] D. J. Harrison et al., *J. Electrochem. Soc.* **135** (1988) 2473.

[154] P. D. van der Wal et al., *Anal. Chim. Acta* **231** (1990) 41.

[155] A. van den Berg, A. Griesel, E. Verney-Norberg, *Sens. Actuators B* **4** (1991) 235.

[156] R. W. Catrall, P. I. Iles, J. C. Hamilton, *Anal. Chim. Acta* **169** (1985) 403.

[157] R. Smith, D. C. Scott, *IEEE Trans. Biomed. Eng.* **BME 33** (1986) 83.

[158] A. van den Berg, P. Bergveld, D. N. Reinhoudt, E. J. R. Sudhölter, *Sens. Actuators* **8** (1985) 129.

[159] N. J. Ho, J. Kratochvil, G. F. Blackburn, J. Janata, *Sens. Actuators* **4** (1983) 413.

[160] C. Dumschat et al., *Sens. Actuators B* **2** (1990) 271.

[161] K. Domansky, J. Janata, M. Josowicz, D. Petelenz, *Analyst (London)* **118** (1993) 335.

[162] H. H. van den Vlekkert et al., *Sens. Actuators* **14** (1988) 165.

[163] C. Dumschat et al., *Anal. Chim. Acta* **243** (1991) 179.

[164] P. Bergveld, *Sens. Actuators B* **4** (1991) 125.

[165] W. F. Love, L. J. Button, R. E. Slovacek: "Optical Characteristics of Fiberoptic Evanescent Wave Sensors," in D. L. Wise, L. B. Wingard (eds.): *Biosensors with Fiberoptics*, Humana Press, Clifton, N.J., 1991.

[166] S. J. Lackie, T. R. Glass, M. J. Block: "Instrumentation for Cylindrical Waveguide Evanescent Fluorosensors," in D. L. Wise, L. B. Wingard (eds.): *Biosensors with Fiberoptics*, Humana Press, Clifton, N.J., 1991.

[167] I. M. Walczak, W. F. Love, T. A. Cook, R. E. Slovacek: "The Application of Evanescent Wave Sensing to a High-sensitivity Fluoroimmunoassay," *Biosensors Bioelectron.* **7** (1992) 39–47.

[168] B. J. Tromberg, M. J. Sepaniak, T. Vo-Dinh, G. D. Griffin: "Fiberoptic Probe for Competitive Binding Fluoroimmunoassay," *Anal. Chem.* **59** (1987) 1226–1230.

[169] R. B. Thompson, F. S. Ligler: "Chemistry and Technology of Evanescent Wave Biosensors," in D. L. Wise, L. B. Wingard (eds.): *Biosensors with Fiberoptics*, Humana Press, Clifton, N.J., 1991.

[170] W. Trettnak, P. J. M. Leiner, O. S. Wolfbeis: "Fiberoptic Glucose Biosensor with an Oxygen Optrode as the Transducer," *Analyst (London)* **113** (1988) 1519–1523.

[171] I. J. Peterson, S. R. Goldstein, R. V. Fitzgerald, D. K. Buckhold: "Fiberoptic pH Probe for Physiological Use," *Anal. Chem.* **52** (1980) 864–869.

[172] G. F. Kirkbright, R. Narayanaswamy, N. A. Welti: "Studies with Immobilized Chemical Reagents Using a Flow-cell for the Development of Chemically Sensitive Fiberoptic Devices," *Analyst (London)* **109** (1984) 12–19.

[173] L. A. Saari, W. R. Seitz: "pH Sensor Based on Immobilized Fluoresceinamine," *Anal. Chem.* **54** (1982) 821–823.

[174] Z. Zhujun, W. R. Seitz: "A Fluorescence Sensor for Quantifying pH in the Range from 6.5 to 8.5," *Anal. Chim. Acta* **160** (1984) 47–55.

[175] E. Urbano, M. Offenbacher, O. S. Wolfbeis: "Optical Sensor for Continuous Determination of Halides," *Anal. Chem.* **56** (1984) 427–430.

[176] O. S. Wolfbeis, E. Urbano: "Fluorescence Quenching Method for Determination of Two or Three Components in Solution," *Anal. Chem.* **55** (1983) 1904–1907.

[177] L. A. Saari, W. R. Seitz: "Immobilized Morin as Fluorescense Sensor for Determination of Aluminum(III)," *Anal. Chem.* **55** (1983) 667–670.

[178] L. A. Saari, W. R. Seitz: "Optical Sensor for Beryllium Based on Immobilized Morin Fluorescence," *Analyst* (London) 109 (1984) 655–657.

[179] Z. Zhujun, W. R. Seitz: "A Fluorescent Sensor for Aluminum(III), Magnesium(II), Zinc(II) and Cadmium(II) Based on Electrostatically Immobilized Qui-

nolin-8-ol-sulfonate," *Anal. Chim. Acta* **171** (1985) 251–258.

[180] Z. Zhujun, W. R. Seitz: "A Carbon Dioxide Sensor Based on Fluorescence," *Anal. Chim. Acta* **170** (1984) 209–216.

[181] M. A. Arnold, T. J. Ostler: "Fiberoptic Ammonia Gas Sensing Probe," *Anal. Chem.* **58** (1986) 1137–1140.

[182] A. Sharma, O. S. Wolfbeis: "Fiberoptic Fluorosensor for Sulfur Dioxide Based on Energy Transfer and Exciplex Quenching," *Proc. SPIE Int. Soc. Opt. Eng.* **990** (1989) 116–120.

[183] A. Sharma, O. S. Wolfbeis: "Fiberoptic Oxygen Sensor Based on Fluorescence Quenching and Energy Transfer," *Appl. Spectrosc.* **42** (1988) 1009–1011.

[184] O. S. Wolfbeis, M. J. P. Leiner, H. E. Posch: "A New Sensing Material for Optical Oxygen Measurement, with Indicator Embedded in an Aqueous Phase," *Mikrochim. Acta* 1986, no. 3, 359–366.

[185] K. W. Berndt, J. R. Lakowicz: "Electroluminescent Lamp-Based Phase Fluorometer and Oxygen Sensor," *Anal. Biochem.* **201** (1992) 319–325.

[186] "Integrated Optics" in T. Tamir (ed.): *Topics in Applied Physics*, Springer Verlag, Berlin 1975.

[187] K. Tiefenthaler, W. Lukosz: "Sensitivity of Grating Couplers as Integrated-optical Chemical Sensors," *J. Opt. Soc. Am. B Opt. Phys.* **B 6** (1975) 209–220.

[188] A. Brandenburg, R. Edelhauser, F. Hutter: "Integrated Optical Gas Sensors Using Organically Modified Silicates as Sensitive Films," *Sens. Actuators B* **11** (1993) 361–374.

[189] D. Schlatter et al.: "The Difference Interferometer: Application as a Direct Immunosensor," in: *The Second World Congress on Biosensors*. Oxford, Elsevier Advanced Technology, 1992, pp. 347–355.

[190] Ph. Nellen, K. Tiefenthaler, W. Lukosz: "Integrated Optical Input Grating Couplers As Biochemical Sensors," *Sens. Actuators B* **15** (1988) 285–295.

[191] R. G. Heidemann, R. P. H. Kooyman, J. Greve: "Performance of a Highly Sensitive Optical Waveguide Mach–Zehnder Interferometer Immunosensor," *Sens. Actuators B* **10** (1993) 209–217.

[192] N. Fabricius, G. Gaulitz, J. Ingenhoff: "A Gas Sensor Based on an Integrated Optical Mach–Zehnder Interferometer," *Sens. Actuators B* **7** (1992) 672–676.

[193] Ch. Fattinger, H. Koller, P. Wehrli, W. Lukosz: "The Difference Interferometer: A Highly Sensitive Optical Probe for Molecular Surface-Coverage Detection," in: *The Second World Congress on Biosensors*. Oxford, Elsevier Advanced Technology, 1992, pp. 339–346.

[194] K. Tiefenthaler: "Integrated Optical Couplers As Chemical Waveguide Sensors," *Adv. Biosensors* **2** (1992) 261–289.

[195] B. Liedberg, C. Nylander, I. Lundström: "Surface Plasmon Resonance for Gas Detection and Biosensing," *Sens. Actuators* **4** (1983) 299–304.

[196] P. B. Daniels, J. K. Deacon, M. J. Eddowes, D. Pedley: "Surface Plasmon Resonance Applied to Immunosensing," *Sens. Actuators* **14** (1988) 11–17.

[197] J. Gent et al.: "Optimization of a Chemooptical Surface Plasmon Resonance Based Sensor," *Appl. Opt.* **29** (1990) 2343–2849.

[198] K. Matsubara, S. Kawata, S. Minami: "Optical Chemical Sensor Based on Surface Plasmon Measurement," *Appl. Opt.* **27** (1988) 1160–1163.

[199] H. Rather: *Surface Plasmons.* Springer Verlag, Berlin 1988.

[200] G. Gauglitz et al.: DE 4 200 088 A1, 1993.

[201] A. Brecht, J. Ingenhoff, G. Gauglitz, *Sens. Actuators B* **6** (1992) 96 – 100.

[202] G. Kraus, G. Gauglitz, *Fresenius' J. Anal. Chem.* **349** (1994) no. 5, 399 – 402.

[203] G. Sauerbrey, *Z. Phys.* **155** (1959) 206 – 222.

[204] H. K. Pulker, J. P. Decostered in C. Lu, A. W. Czanderna (eds.): "Applications of Piezoelectric Quartz Crystal Microbalances," *Methods and Phenomena*, vol. 7, Elsevier, Amsterdam 1984, Chap. 3.

[205] Lord Rayleigh, *Proc. London Math. Soc.* **17** (1885) 4 – 11.

[206] G. Sauerbrey, *Z. Phys.* **178** (1964) 457.

[207] W. H. King, Jr., *Anal. Chem.* **36** (1964) 1735 – 1739.

[208] W. H. King, Jr., US 3 164 004, 1965.

[209] A. W. Czanderna, C. Lu: "Applications of Piezoelectric Quartz Crystal Microbalances," in C. Lu, A. W. Czanderna (eds.): *Methods and Phenomena*, vol. 7, Elsevier, Amsterdam 1984. chap. I.

[210] M. Thompson et al. *Analyst (London)* **116** (1991) 881 – 890.

[211] L. Wimmer, S. Hertl, J. Hemetsberger, E. Benes, *Rev. Sci. Instrum.* **55** (1984) 605 – 609.

[212] C.-S. Lu, O. Lewis, *J. Appl. Phys.* **43** (1972) 4385 –4390.

[213] J. G. Miller, D. I. Bolef, *J. Appl. Phys.* **39** (1968) 5815 – 5816.

[214] E. Benes, *J. Appl. Phys.* **56** (1984) 608 – 626.

[215] D. R. Denison, *J. Vac. Sci. Technol.* **10** (1973) 126 – 129.

[216] C.-S. Lu, *J. Vac. Sci. Technol.* **12** (1975) 578 – 583.

[217] G. G. Guilbault: "Applications of Piezoelectric Quartz Crystal Microbalances," in C. Lu, A. W. Czanderna (eds.): *Methods and Phenomena*. vol. 7, Elsevier, Amsterdam 1984. chap. 8.

[218] H. Beitnes, K. Schrøder, *Anal. Chim. Acta* **158** (1984) 57 – 65.

[219] A. Kindlund, H. Sundgren, I. Lundström, *Sens. Actuators* **6** (1984) 1 – 17.

[220] G. G. Guilbault: *Ion-Selective Electrode Review*, vol. **2**, Pergamon Press, Oxford 1980, pp. 3 – 16.

[221] R. M. Langdon: "Resonator Sensors—a Review," *J. Phys. E* **18** (1985) 103 – 115.

[222] J. Tichy, G. Gautschi: *Piezoelektrische Meßtechnik*, Springer Verlag, Berlin 1980.

[223] H. Wohltjen, R. Dessy, *Anal. Chem.* **51** (1979) 1458 – 1464.

[224] H. Wohltjen, R. Dessy, *Anal. Chem.* **51** (1979) 1465 – 1470.

[225] H. Wohltjen, R. Dessy, *Anal. Chem.* **51** (1979) 1470 – 1478.

[226] C. T. Chang, R. M. White: "Excitation and Propagation of Plate Mode in an Acoustically Thin Membrane," *Proc. IEEE Ultrasonics Symp.* 1982, 295 – 298.

[227] R. M. White, P. W. Wicher, S. W. Wenzel, E. T. Zellers: "Plate-Mode Ultrasonic Oscillator Sensors," *IEEE Trans. Ultrason. Dev. Ferroelectr. Freq. Contr.* **UFFC-34** (1987) 162 – 171.

[228] T. G. Giesler, J.-U. Meyer: *Fachbeilage Mikroperipherik*, vol. 6 (1992) XLVI.

[229] IEEE Standard on Piezoelectricity (ANSI/IEEE Std. 176-1987), The Institute of Electrical and Electronics Engineers, New York 1987.

[230] M. S. Nieuwenhuizen, A. Venema: *Sensors* – A Comprehensive Survey. vol. II, VCH Verlagsgesellschaft, Weinheim 1991, p. 652.

[231] J. Janata: *Principles of Chemical Sensors*, Plenum Press, New York 1989. pp. 55 – 80.

[232] J. G. Miller, D. I. Bolef, *J. Appl. Phys.* **39** (1968) 45.

[233] K. H. Behrndt, *J. Vac. Sci. Technol.* **8** (1971) 622.

[234] A. P. M. Glassford in A. M. Smith (ed.): *Progress in Astronautics and Aeronautics*, vol. **56**, American Institute of Aeronautics and Astronautics, New York 1977, p. 175.

[235] A. P. M. Glassford, *J. Vac. Sci. Technol.* **15** (1978) 1836.

[236] V. Mecea, R. V. Bucur, *Thin Solid Films* **60** (1979) 73.

[237] R. A. Crane, G. Fischer, *J. Phys. D* **12** (1979) 2019.

[238] K. K. Kanazawa, J. G. Gordon, *Anal. Chem.* **57** (1985) 1770.

[239] C. Fruböse, K. Doblhofer, C. Soares, *Ber. Bunsen-Ges. Phys. Chem.* **97** (1993) 475 – 478.

[240] R. Beck, U. Pittermann, K. G. Weil, *Ber. Bunsen-Ges. Phys. Chem.* **92** (1988) 1363 – 1368.

[241] F. Eggers, Th. Funck, *J. Phys. E* **20** (1987) 523 – 530.

[242] A. Glidle, R. Hillman, S. Bruckenstein, *J. Electroanal. Chem. Interfacial Electrochem.* **318** (1991) 411 – 420.

[243] R. Beck, U. Pittermann, K. G. Weil, *J. Electrochem. Soc.* **139** (1992) no. 2, 453 – 461.

[244] R. Beck, U. Pittermann, K. G. Weil, *Ber. Bunsen-Ges. Phys. Chem.* **92** (1988) 1363 – 1368.

[245] D. Salt: *Handbook of Quartz Crystal Devices*, van Nostrand Reinhold, New York 1987.

[246] Z. P. Khlebarov, A. I. Stoyanova, D. I. Topalova, *Sens. Actuators B* **8** (1992) 33 – 40.

[247] V. M. Ristic: *Principles of Acoustic Devices*, J. Wiley and Sons, New York 1983.

[248] P. Hauptmann, *Sens. Actuators A* **25** (1991) 371 – 377.

[249] B. A. Auld: *Acoustic Fields and Waves in Solids*, vol. II, J. Wiley and Sons, New York 1973.

[250] J. F. Vetelino, D. L. Lee, WO 83 1511, 1983.

[251] A. Bryant, D. L. Lee, J. F. Vetelino, *Ultrason. Symp. Proc.* 1981, 171 – 174.

[252] A. D'Amico, *Sens. Actuators* **3** (1982/83) 31 – 39.

[253] S. J. Martin, S. S. Schwartz, R. L. Gunshor, R. F. Pierret, *J. Appl. Phys.* **54** (1983) 561 – 569.

[254] E. Gizeli, A. C. Stevenson, N. J. Goddard, C. R. Lowe: *Chemical Sensors*, Waseda University Int. Conference Center, Tokyo, Sept. 13 – 17, 1992.

[255] P. Hauptmann: *Sensoren, Prinzipien und Anwendungen*, Hanser Verlag, München 1990.

[256] G. J. Bastiaans in S. Yamauchi (ed.): *Chemical Sensor Technology*, Elsevier, Amsterdam 1992, pp. 181 – 204.

[257] J. Reinbold: *Dissertation*, University of Muenster 2000.

[258] J. Curie, P. Curie, *Bull. Soc. Min. Paris* **3** (1880) 90.

[259] R. A. Heising: *Quartz Crystal for Electrical Circuits – Their Design and Manufacture*, van Nostrand, New York 1947, p. 16.

[260] S. Büttgenbach: *Mikromechanik*, Teubner, Stuttgart 1991, p. 39.

[261] L. E. Halliburton, J. J. Martin in E. A. Gerber, A. Ballato (eds.): *Precision Frequency Control*, vol. 1, Acoustic Resonators and Filters, Academic Press, New York 1985.

[262] M. H. Ho: "Applications of Quartz Crystal Microbalances in Aerosol Measurements," in C. Lu, A. W. Czanderna (eds.): "Applications of Piezoelectric Quartz Crystal Microbalances," *Methods and Phenomena*, vol. 7, Elsevier, Amsterdam 1984.

[263] W. G. Cady: *Piezoelectricity*, McGraw Hill, New York 1946.

[264] J. F. Nye: *Physical Properties of Crystals,* Oxford University Press, London 1957.

[265] J. C. Brice: "Crystals for Quartz Resonators," *Rev. Mod. Phys.* **57** (1985) 105.

[266] M. Born, K. Huang: *Dynamical Theory of Crystal Lattices,* Clarendon Press, Oxford 1954.

[267] E. K. Sittig: "Acoustic Wave Devices," in *Encyclopedia of Physical Science and Technology,* vol. 1, Academic Press, New York 1987, pp. 83–109.

[268] F. L. Dickert, M. Vonend, H. Kimmel, G. Mages, *Fresenius Z. Anal. Chem.* **333** (1989) 615–618.

[269] A. W. Snow, J. R. Griffith, N. P. Marullo, *Macromolecules* **17** (1984) 1614.

[270] F. L. Dickert, *Chem. Unserer Zeit* **26** (1992) 138–143.

[271] B. A. Cavic-Vlasak, L. J. Rajakovic, *Fresenius J. Anal. Chem.* **343** (1992) 339–347.

[272] E. Weber, A. Ehlen, C. Wimmer, J. Bargon, *Angew. Chem.* **105** (1993) 116–117.

[273] I. Lundström, M. S. Shivaraman, C. M. Stevenson, *J. Appl. Phys.* **46** (1975) 3876–3881.

[274] M. S. Nieuwenhuizen, W. Barendsz, *Sens. Actuators* **11** (1987) 45–62.

[275] I. Langmuir: "The Adsorption of Gases on Plane Surfaces of Glass, Mica and Pt," *J. Am. Chem. Soc.* **40** (1918) 1361.

[276] J. H. de Boer: *The Dynamical Aspects of Adsorption,* Academic Press, Oxford 1953.

[277] A. Clark: *The Chemisorptive Bond.* Academic Press, New York 1974.

[278] F. M. Moser, A. L. Thomas: *Phthalocyanine Compounds,* Reinhold Publ. Co., New York 1963, chap. 8.

[279] L. F. Barringer, D. P. Rillema, J. H. Ham IV, *J. Inorg. Biochem.* **21** (1984) 195.

[280] J. R. Lenhard, R. W. Murvay: "Chemically Modified Electrodes," *J. Am. Chem. Soc.* **100** (1978) 7880.

[281] J. S. Miller (ed.), *ACS Symp. Ser.* **182** (1981) part 15.

[282] M. T. Reed et al.: "Thermodynamic and Kinetic Data for Cation Macrocyclic Interactions," *Chem. Rev.* **85** (1985) 271.

[283] D. W. Armstrong, *J. Liq. Chromatogr.* **7, Suppl. 2** (1984) 353–376.

[284] M. S. Nieuwenhuizen, A. Venema: "Mass-Sensitive Devices," in W. Göpel et al. (eds.): *Sensors,* vol. II, chap. 13, VCH Verlagsgesellschaft, Weinheim 1991, pp. 657–658.

[285] J. G. Brace, T. S. Sanfelippo, S. Joshi, *Sens. Actuators* **14** (1988) 47–68.

[286] E. T. Zellers, R. M. White, S. M. Rappaport, S. W. Wenzel, *Proc. Int. Conf. Sens. Actuators 4th* 1987, Tokyo: IEE Japan, pp. 459–461.

[287] M. S. Nieuwenhuizen, A. J. Nederlof, *Sens. Actuators B 2* (1990) 97–101.

[288] A. W. Snow et al., *Langmuir* **2** (1986) 513–519.

[289] A. J. Ricco, S. J. Martin, T. E. Zipperian, *Sens. Actuators* **8** (1985) 105–111.

[290] A. Byrant et al., *Sens. Actuators* **4** (1983) 105–111.

[291] M. Janghorbani, H. Freund, *Anal. Chem.* **45** (1973) 325–332.

[292] A. Kindlund, H. Sundgren, I. Lundström, *Sens. Actuators* **6** (1984) 1.

[293] D. Soares, *J. Phys. E,* in press.

[294] C. Barnes, *Sens. Actuators A* **29** (1991) 59–69.

[295] S. Bruckenstein, M. Shay, *Electrochim. Acta* **30** (1985) 1295–1300.

[296] F. L. Dickert, A. Haunschild, V. Maune, *Sens. Actuators B* **12** (1993) 169–173.

[297] M. Ho, G. G. Guilbault, E. P. Scheide, *Anal. Chem.* **52** (1982) 1998–2002.

[298] J. M. Jordan, Dissertation, University of New Orleans 1985.

[299] A. Suleiman, G. G. Guilbault, *Anal. Chim. Acta* **162** (1984) 97–102.

[300] J. F. Alder, A. E. Bentley, P. K. P. Drew, *Anal. Chim. Acta* **182** (1986) 123–131.

[301] Y. Tomita, M. H. Ho, G. G. Guilbault, *Anal. Chem.* **51** (1979) 1475–1478.

[302] J. A. O. Sanchez-Pedreno, P. K. P. Drew, J. F. Alder, *Anal. Chim. Acta* **182** (1986) 285–291.

[303] G. M. Varga, Jr., US NTIS AD Report 780171/5 GA, 1974.

[304] E. C. Hahn, A. Suleiman, G. G. Guilbault, J. R. Cananaugh, *Anal. Chim. Acta* **197** (1987) 195–202.

[305] G. G. Guilbault, *Anal. Chem.* **55** (1983) 1682–1684.

[306] M. J. van Sant, Dissertation, University of New Orleans 1985.

[307] C. Kößlinger et al., *Biosens. Bioelectron.* **7** (1992) 397–404.

[308] C. G. Fox, J. F. Alder: "Surface Acoustic Wave Sensors for Atmospheric Gas Monitoring," in P. T. Moseley, J. Norris, D. E. Williams (eds.): *Techniques and Mechanism in Gas Sensing,* Adam Hilger, Bristol 1991, chap. 13.

[309] J. Auge, P. Hauptmann, F. Eichelbaum, S. Rösler, *Sens. Actuators,* in press.

[310] J. W. Grate, S. J. Martin, R. M. White, *Anal. Chem.* **65** (1993) 987–996.

[311] A. D'Amico, A. Palma, E. Verona, *Sens. Actuators* **3** (1982/83) 31–39.

[312] A. Venema et al., *Sens. Actuators* **10** (1986) 47.

[313] A. Venema et al., *Electron. Lett.* **22** (1986) 184.

[314] A. D'Amico, A. Palma, E. Verona, *Appl. Phys. Lett.* **41** (1982) 300–301.

[315] A. D'Amico, M. Gentilli, P. Veradi, E. Verona, *Proc. Int. Meet. Chem. Sens. 2nd* 1986, 743–746.

[316] A. W. Snow et al., *Langmuir* **2** (1986) 513–519.

[317] J. G. Brace, T. S. Sanfelippo, S. Joshi, *Sens. Actuators* **14** (1988) 47–68.

[318] J. G. Brace, T. S. Sanfelippo, *Proc. 4th Int. Conf. Sens. Actuators* 1987, 467–470.

[319] M. S. Nieuwenhuizen, A. J. Nederlof, M. J. Vellekoop, A. Venema, *Sens. Actuators* **19** (1989) 385–392.

[320] H. Wohltjen, *Sens. Actuators* **5** (1984) 307–325.

[321] A. W. Snow, H. Wohltjen, *Anal. Chem.* **56** (1984) 1411–1416.

[322] S. J. Martin, S. S. Schwartz, R. L. Gunshor, R. F. Pierret, *J. Appl. Phys.* **54** (1983) 561–569.

[323] S. J. Martin, K. S. Schweizer, S. S. Schwartz, R. L. Gunshor, *Ultrason. Symp. Proc.* 1984, 207–212.

[324] A. J. Ricco, S. J. Martin, T. E. Zipperian, *Sens. Actuators* **8** (1985) 319–333.

[325] J. W. Grate et al., *Anal. Chem.* **60** (1988) 869–875.

[326] K. Mosbach, *Biosens. Bioelectron.* **6** (1991) 179.

[327] P. T. Walsh, T. A. Jones in W. Göpel et al. (eds.): *Sensors: A Comprehensive Survey.* **vol. 2,** VCH Verlagsgesellschaft, Weinheim 1991, p. 529.

[328] A. F. Hollemann, E. Wiberg: *Lehrbuch der anorganischen Chemie,* De Gruyter, Berlin 1985.

[329] B. Danielsson, F. Winquist in A. E. Cass (ed.): *Biosensors: A Practical Approach,* Oxford University Press, Oxford 1989, p. 191.

[330] A. R. Baker, GB 892530, 1962.

[331] K. Mosbach et al., *Biochim. Biophys. Acta* **364** (1974) 140.

[332] B. Mattiasson et al., *Anal. Lett.* **9** (1976) 217.

[333] B. Danielsson et al., *Appl. Biochem. Bioeng.* **3** (1981) 103.

[334] H. G. Hundeck et al., *GBF Monogr. Ser.* **17** (1992) 321.

[335] B. Xie et al., *Sens. Actuators B* **6** (1992) 127.

[336] E. J. Guilbeau et al., *Trans. Am. Soc. Artif. Intern. Organs* **33** (1987) 329.

[337] G. Steinhage: *Dissertation*, University of Muenster 1996.

[338] M. G. Jones, T. G. Nevell, *Sens. Actuators* **16** (1989) 215.

[339] B. Danielsson et al., *Methods Enzymol.* **137** (1988) 181.

[340] S. Wold et al.: *Food Research and Data Analysis*, Applied Sciences, Barking 1983, pp. 147–188.

[341] T. Nakamoto, K. Fukunishi, T. Moriizumi, *Sens. Actuators B* **1** (1989) 473.

[342] K. Ema et al., *Sens. Actuators B* **1** (1989) 291.

[343] J. R. Stetter, P. C. Jurs, S. L. Rose, *Anal. Chem.* **58** (1986) 860.

[344] U. Weimar et al., *Sens. Actuators B* **1** (1990) 93.

[345] W. P. Carey et al., *Sens. Actuators* **9** (1986) 223.

[346] D. Wienke, K. Danzer, *Anal. Chim. Acta* **184** (1986) 107.

[347] B. R. Kowalski, *Anal. Chem.* **47** (1975) 1152.

[348] B. R. Kowalski, *Anal. Chem.* **52** (1980) 112 R.

[349] E. Frank, B. R. Kowalski, *Anal. Chem.* **54** (1982) 232 R.

[350] K. R. Beebe, B. R. Kowalski, *Anal. Chem.* **59** (1987) 1007 A.

[351] D. W. Osten, B. R. Kowalski, *Anal. Chem.* **57** (1985) 908.

[352] B. Schlatt: *Diplomarbeit*, University of Muenster 1996.

[353] A. Schulga, K. Cammann in G. Henze, M. Köhler, J. P. Lay (eds.): *Umweltdiagnostik mit Mikrosystemen*, Wiley-VCH, Weinheim 1999, pp. 14–32.

[354] A. F. P. Turner, I. Karube, G. S. Wilson (eds.): *Biosensors – Fundamentals and Applications*, Oxford University Press, Oxford 1987.

[355] H.-L. Schmidt, R. Kittsteiner-Eberle, *Naturwissenschaften* **73** (1986) 314–321.

[356] F. Scheller, F. Schubert: *Biosensors, Techniques and Instrumentation in Analytical Chemistry*, vol. 11, Elsevier, Amsterdam 1992.

[357] E. A. H. Hall: *Biosensors*, Open University Press, Buckingham 1990.

[358] L. C. Clark, Jr., C. Lyons: *Ann. N. Y. Acad. Sci.* **102** (1962) 29–45.

[359] J. J. Kulys, A. S. Samalius, G. J. S. Sviermikas, *FEBS Lett.* **114** (1980) 7–10.

[360] M. J. Eddows, H. A. O. Hill, *Biosci. Rep.* **1** (1981) 521–532.

[361] M. L. Fultz, R. A. Durst, *Anal. Chim. Acta* **140** (1982) 1–18.

[362] L. Gorton, *J. Chem. Soc. Faraday Trans. I* **82** (1986) 1245–1258.

[363] B. Gründig et al., in F. Scheller, R. D. Schmid (eds.): "Biosensors: Fundamentals, Technologies and Applications," *GBF Monogr.* **17** (1992) 275–285.

[364] Y. Degani, A. Heller, *J. Phys. Chem.* **91** (1987) no. 6, 1285–1289.

[365] Y. Degani, A. Heller, *J. Am. Chem. Soc.* **111** (1989) 2357–2358.

[366] W. Schuhmann, T. J. Ohara, H.-L. Schmidt, A. Heller, *J. Am. Chem. Soc.* **113** (1991) no. 4, 1394–1397.

[367] F. W. Scheller, R. Renneberg, F. Schubert, in S. Colowick, N. O. Kaplan, K. Mosbach (eds.): *Methods in Enzymology*, **vol. 137**, Academic Press, San Diego 1988, pp. 29–44.

[368] F. Schubert, D. Kirstein, K. L. Schröder, F. W. Scheller, *Anal. Chim. Acta* **169** (1985) 391–406.

[369] A. Warsinke, B. Gründig, DE 4 227 569, 1993.

[370] K. Riedel, B. Neumann, F. Scheller, *Chem. Ing. Tech.* **64** (1992) no. 6, 518–528.

[371] M. H. Smit, A. E. G. Cass, *Anal. Chem.* **62** (1990) 2429–2436.

[372] D. M. Rawson, A. J. Willmer, *Biosensors* **4** (1989) 299–311.

[373] R. S. Sethi: *Biosensors & Bioelectronics* **9** (1994) 243–264.

[374] M. P. Byfield, R. A. Abuknesha: "Biochemical Aspects of Biosensors," *GEC J. Res.* **9** (1991) 97–117.

[375] M. E. Eldefrawi et al., *Anal. Lett.* **21** (1988) 1665–1680.

[376] S. Lofas, B. Johnsson, *J. Chem. Soc. Chem. Commun.* 1990, 1526–1528.

[377] F. Brosinger et al., *Sens. Actuators B* **44** (1997) 350–355.

[378] D. Kröger, A. Katerkamp, R. Renneberg, K. Cammann, *Biosensors & Bioelektronics* **13** (1998) 1141–1147.

[379] WO 9416312 A1, 1994 (R. C. Jorgenson, S. S. Yee).

[380] A. Katerkamp et al., *Mikrochim. Acta* **119** (1995) (1–2) 63–72.

[381] M. Niggemann et al., *Proc. SPIE-Int. Soc. Opt. Eng.* (1995) 2508. Chemical, Biochemical, and Environmental Fiber Sensors VII, 303–311.

[382] U. Narang, P. R. Gauger, F. S. Ligler, *Anal. Chem.* **69** (1997) no. 14, 2779–2785.

[383] J. E. Roederer, G. J. Bastiaans, *Anal. Chem.* **55** (1983) 2333–2336.

[384] G. G. Guilbault, J. H. Luong, *J. Biotechnol.* **9** (1988) 1–10.

[385] M. P. Byfield, R. A. Abuknesha, *GEC J. of Res.* **9** (1991) 108.

[386] M. Hikuma et al. *Biotechnol. Bioeng.* **21** (1979) 1845–1850.

[387] C. Diviès, *Ann. Microbiol. (Paris)* **126A** (1975) 175–186.

[388] G. A. Rechnitz, *Science* **214** (1981) 287–291.

[389] I. Karube, S. Suzuki, *Ion Sel. Electrode Rev.* **6** (1984) 15–58.

[390] K. Riedel, R. Renneberg, F. Scheller, *Anal. Lett.* **53** (1990) 757–770.

[391] Frost & Sullivan: *The European Market for Industrial Gas Sensors* and The European Market for Portable and Transportable Analytical Instruments, Frost & Sullivan Eigenverlag, Frankfurt 1990.

[392] INFRATEST: *Chemische und biochemische Sensoren – Marktübersicht*, Infratest Industria, 1992.

[393] MIRC: *The European Market for Process Control Equipment and Instrumentation*, Frost & Sullivan Eigenverlag, Frankfurt 1992.

29. Microscopy

ANDRES KRIETE, Klinikum der Justus-Liebig-Universität Gießen, Institut für Anatomie und Zellbiologie, Gießen, Federal Republic of Germany (Chap. 29)

HEINZ GUNDLACH, Carl Zeiss, Jena, Federal Republic of Germany (Chap. 29)

SEVERIN AMELINCKX, Universiteit Antwerpen (RUCA), Antwerpen, Belgium (Sections 29.2.1, 29.2.2)

LUDWIG REIMER, Physikalisches Insitut Universität Münster, Münster, Federal Republic of Germany (Section 29.2.3)

29.1. Modern Optical Microscopy

29.1.1. Introduction

Optical microscopy goes back as far as the 16th century, when magnifiying glasses and optical lenses became available. ANTONI VAN LEEUWENHOEK (1623 – 1723) observed structures lying beyond the resolution limit of the eye, such as bacteria, with a 270 fold magnifying, single lens. The exact design of the microscope was established from a theory of image formation developed by ABBE in 1873 [1]. With the introduction of apochromatic lenses and oil immersion, a theoret-
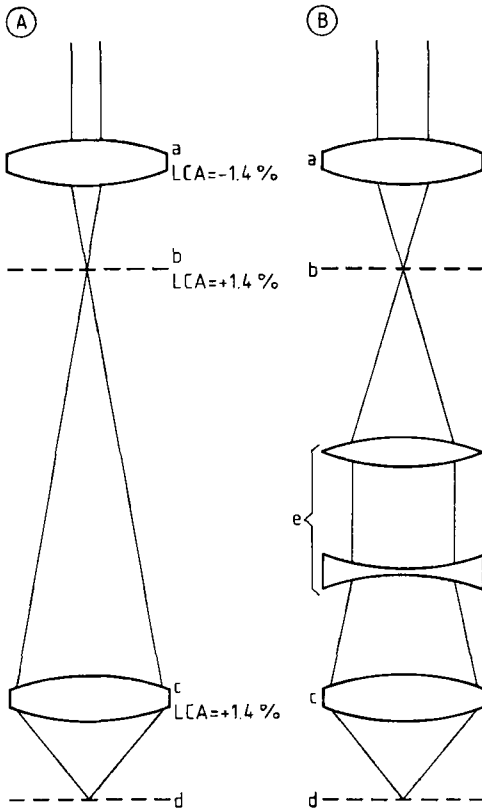
Figure 1. Schematic optical path for objectives of finite length; compensating eyepieces correct for chromatic aberrations

A) Conventional microscope with finite optics; B) Microscope with finite optics and telan optics for insertion of optical components

a) Eyepiece; b) Primary intermediate image; c) Objective; d) Object plane; e) Telan system (incident light fluorescence, Optovar, Bertrand lens)

ical resolution limit of 0.2 μm became available at this early stage. Light microscopy advanced further by new lens designs, optical contrast enhancement methods, and application of fluorescent and immunofluorescent dyes, also in combination with inverted microscope designs. Optoelectronic methods, such as video techniques, have further broadened the spectrum of light microscopic applications. Finally, laser scanning techniques and confocal imaging allowed the investigation of microscopic structures three-dimensionally. Compared with a standard microscope, the confocal microscope has enhanced lateral and axial resolution, as well as it has improved contrast, and

in particular it can remove out-of-focus blur. Thus, an optical, noninvasive sectioning capability is achieved in thick specimens. The final breakthrough of confocal microscopy in biomedicine and material science boosted the development of a variety of commercial confocal laser scanning microscopes. Highly developed optoelectronic components such as laser light sources and sensitive detectors, as well as the application of specific fluorescent dyes, enabled wide acceptance of this new imaging technique together with effective use of powerful computers to digitally store, visualize, and analyze three-dimensional data.

29.1.2. Basic Principles of Light Microscopy

29.1.2.1. Optical Ray Path

The rigid *stand* of the microscope is the "carrying" element, with mechanical and optical parts attached. The *objective* is the central optical unit, generating an intermediate image that is magnified by the *eyepiece* for visual observation [2] – [4]. *Condenser* and *collector* optimize specimen illumination. Lens designs corrected for a *finite length* form a mirrored, magnified, aerial (intermediate) image, which is projected infinity by the eyepiece (Fig. 1). The distance between the back focal plane of the objective and the primary image plane (which is the front focal plane of the eyepiece) is defined as the *optical tube length*. The *mechanical tube length*, measured from the mounting plane of the objective to the mounting plane of the eyepiece, is 160 mm for currently used microscopes. Any aberrations present in the intermediate image, such as the lateral chromatic aberration (LCA), are compensated by the eyepiece. Optical – mechanical components such as the Bertrand lens, or epi-illuminating reflectors mounted within the optical path require a so-called telan system. Objectives of *infinity-corrected designs* generate an infinite image projected into a real image by the tube lens (Fig. 2). The infinity color-corrected system (ICS) optic is an example of how the tube lens together with the objective can correct aberrations, particularly chromatic ones [5]. The intermediate image is monochromatic; consequently, photographic and video recording do not require compensating lenses. Moreover, the path between lens and tube lens may be used by all optical – mechanical components for various illuminating and contrast generating methods without telan optics (Fig. 2).
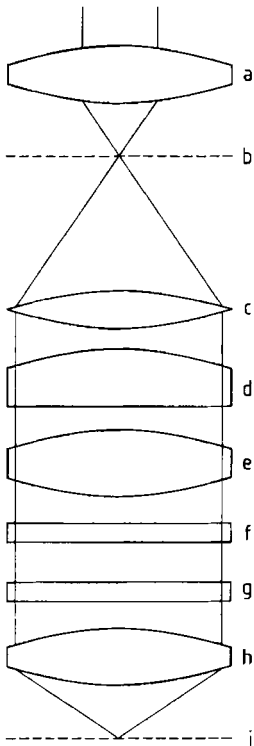
Figure 2. Schematic optical path for objectives corrected for infinity with color-corrected intermediate image
a) Eyepiece; b) Plane of intermediate image; c) Tube lens; d) Bertrand lens slider; e) Optovar 1 X, 1.25 X, 1.6 X, reflector for incident light; f) $\lambda/4$ plate, plate, compensators; g) Differential interference contrast (DIC) slider; h) Objective; i) Specimen plane

29.1.2.2. Imaging Performance and Resolution

The image generation in microscopy cannot be explained sufficiently by means of geometric optics alone; wave optics must also be considered [1], [6]. Imaging a point source generates a diffraction pattern in the focal plane. The central parts of this Airy disk contain most of the intensity; therefore, the diameter of this kernel can be defined as the smallest resolvable image element (Fig. 3 A – D). The diameter of the kernel d_k is given by

$$d_k = 1.22\lambda/n \cdot \sin\alpha \tag{1.1}$$

with λ being the wavelength, n the refractive index of the material, α the angle between the optical axis and the marginal beam of the light cone that enters the objective, and $n \cdot \sin\alpha$ the numerical aperture (NA).

The resolution limit for two points is reached when the maximum of the first diffraction spot coincides with the first minimum of the second one. ABBE developed this theory and proved it experimentally in his historical papers [1]. The resolution is calculated from the available numerical aperture and the wavelength λ by

$$d_r = 1.22\lambda/NA_{obj} + NA_{cond} \tag{1.2}$$

As an example, a lens and condenser of $NA = 1.4$ and $\lambda = 530$ nm give a smallest separation of $d = 200$ nm. Resolution can be improved either by a higher NA, which is difficult to obtain in lens design, or by using shorter wavelength, as in ultraviolet microscopy [7].

Both the size of the diffraction spot and the resolution of the eye define the utilizable magnification. Since the resolution of the eye is ca. $2-4$ arcminutes, total magnification should not exceed $500-1000$ times the numerical aperture. Magnification above this limit does not deliver any additional information, but may be used for automatic measuring procedures or counting.

29.1.2.3. Characteristics and Classification of Lenses

The imperfections of optical lenses give rise to various distortions, either monochromatic or chromatic, that are visible in the diffraction pattern. Monochromatic distortions include aberration, curvature, astigmatism, and coma. Chromatic aberrations are caused by differences in the refractive index of glasses. These types of errors are minimized by combining several pieces of glass with different form and refractive index, or by using diaphragms. The three main categories of lenses are

1) Achromatic objectives
2) Fluorite or semiapochromatic objectives
3) Apochromatic objectives

Fluorite and apochromatic lenses are mostly designed as plane (i.e., corrected for a flat field of view). Computers allow calculation of the diffraction pattern under various imaging conditions to optimize imaging performance for a wide range of applications. The present advantage in modern lens design and production is indicated by the image quality of the ICS Plan Neofluar class. At a field-of-view number of 25, color photography, as well as all optical contrast enhancement methods, are supported with achromatic correction
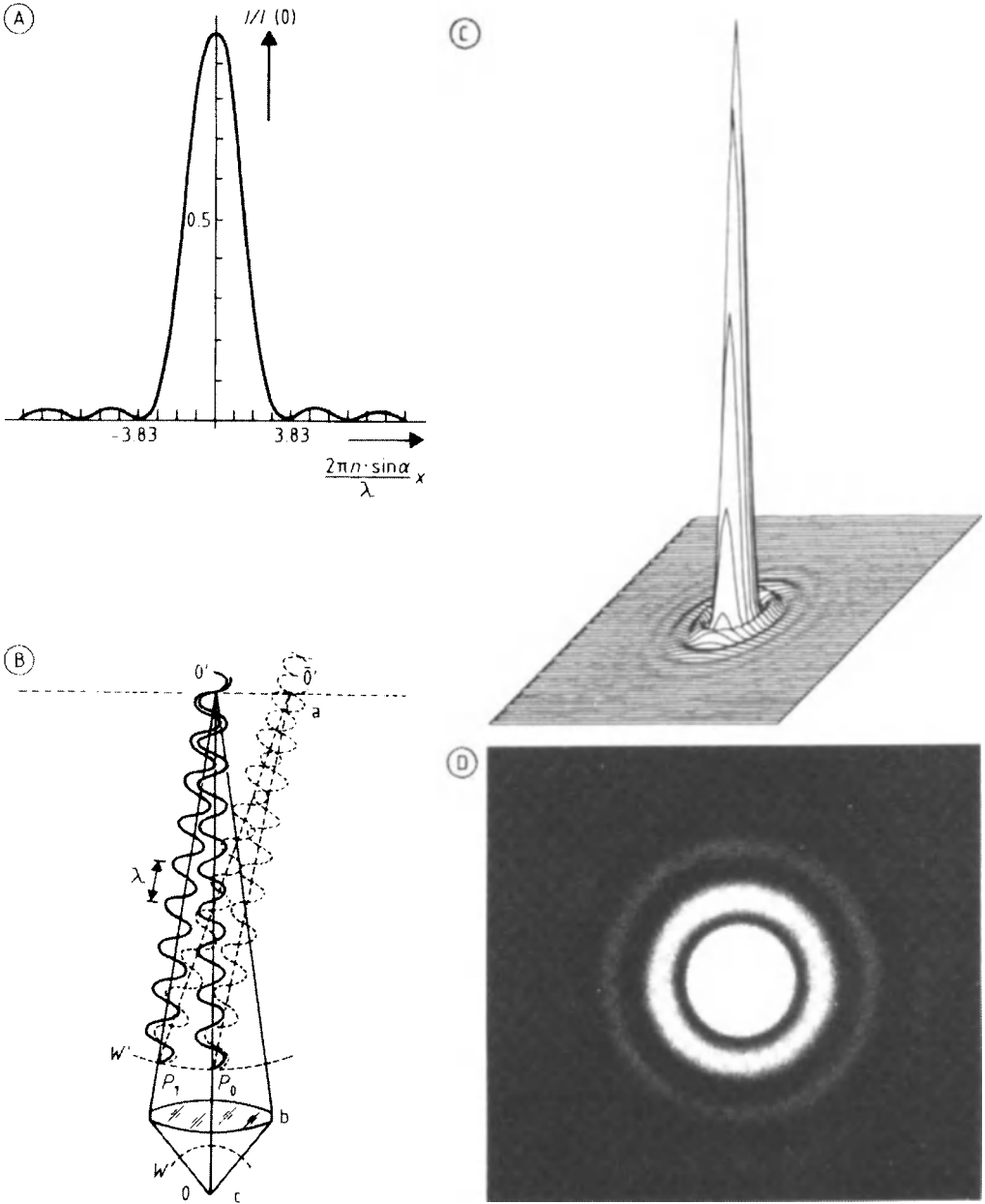
Figure 3. A) and B) Intensity distributions in the image of a luminous point; a) Image plane; b) Lens; c) Objects plane; C) Computed intensity distribution; D) Airy disk

comparable to apochromatic lenses with a good transmission down to UV at 340 nm. These objectives are generally applicable for all fluorescent methods (see Sections 29.1.3.2, 29.1.6.4). Plan apochromates feature a still higher aperture: chro-

matic rendition is improved, as well as correction at the field-of-view boundary; therefore these objectives are very suitable for color microphotography. In addition, a complete correction not only at the focus level but also for a limited range above
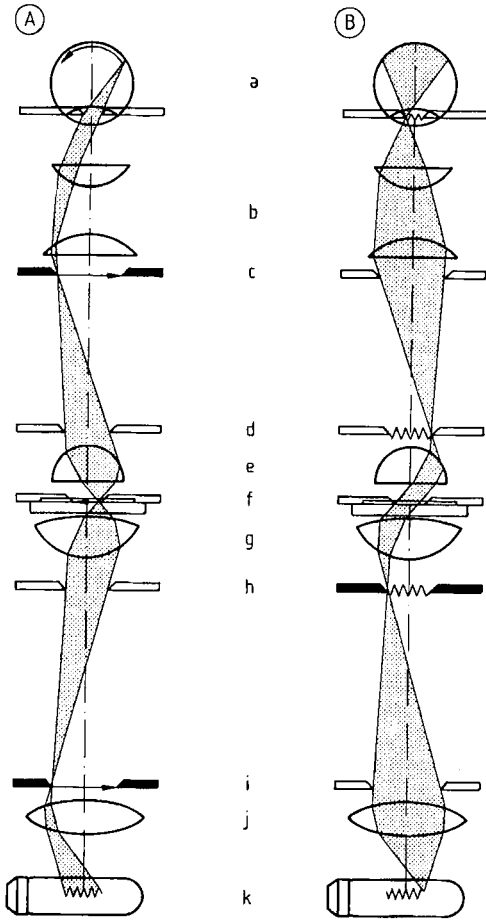
Figure 4. Schematic of Köhler illumination
A) Image-forming ray path; B) Illuminating ray path
a) Eye; b) Eyepiece; c) Eyepiece field stop; d) Exit pupil of objective; e) Objective; f) Specimen; g) Condenser;
h) Aperture diaphragm; i) Lamp field stop; j) Light collector; k) Light source

and below the focal plane is achieved. For specific applications, the following lenses are also available:

1) Multi-immersion objectives of type Plan Neofluar, corrected for use with oil, glycerol, or water, with or without a cover glass
2) Objectives for specimen without cover slide
3) Objectives with a long working distance
4) Objectives with iris diaphragm
5) Objectives for phase contrast investigation
6) Objectives, strain free for polarized light microscopy
7) Ultrafluars, corrected between 230 and 700 nm

29.1.2.4. Eyepieces and Condensers

The main characteristics of *eyepieces* are magnification and type of correction, such as color compensation and correction of viewing angle. The correction of the *condensers* must match the properties of the lenses. Special or combined condensers are required for certain methods of contrast generation or for longer working distance with inverted microscopes.

29.1.3. Illumination and Contrast Generation

Illumination and contrast are central points in the discussion of imaging performance. Without sufficient contrast, neither the required magnification nor a maximal resolution can be obtained. A. KÖHLER developed optical illumination for microscopy in 1893 [8]. The main advantage of Köhler illumination is that in spite of the small filament of the light bulb, the entire area of the lamp field stop, as well as the illuminated part of the specimen, have a uniform luminance (Fig. 4). The twofold ray path of diaphragms and lenses suggests that adjustment of the condenser iris, being the illumination aperture, controls resolution, contrast, and resolution depth. This principle is also important for correct setup of the confocal microscope as well as electron microscopy.

29.1.3.1. Optical Contrast Generation

For the following considerations the principles of wave optics are used, which describe image formation as a result of diffraction and interference.

Dark-Field Illumination. In dark-field illumination the illuminating aperture is always set higher than the lens aperture. Only light waves being diffracted by the microstructures are captured by the lens. Objects appear as bright diffraction patterns on a dark background.

Phase Contrast. Based on ABBE's theory the Dutch physicist FRITS ZERNIKE (Nobel prize in 1953) developed phase contrast imaging in 1935, to convert the invisible phase shift within the specimen into recognizable contrast [9]. A phase plate located in the lens modifies the phase and amplitude of the direct (iluminating) wave in such a way that a phase shift of $\lambda/2$ is obtained (Fig. 5). The
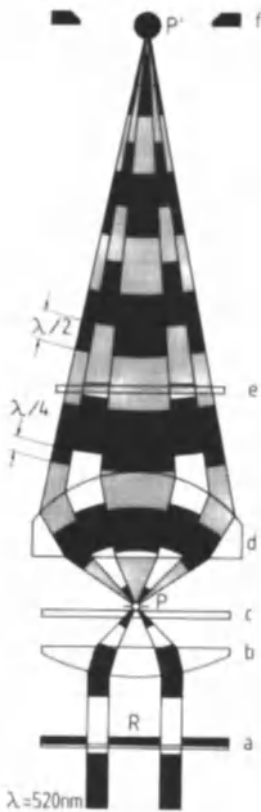
Figure 5. Image path of the phase contrast microscope, converting phase shifts into a noticeable contrast
a) Illuminating diaphragm ring; b) Condenser; c) Specimen generating a phase shift; d) Objective plate; e) Phase plate; f) Intermediate image

Figure 6. A) Image path of a two-beam interference microscope; B) Nomarski prism with direction of optical axes and position of interference plane indicated
a) Polarizer (45°); b) First Wollaston prism; c) Condenser; d) Specimen; e) Objective; f) Second Wollaston prism; g) Analyzer (135°); h) Intermediate image
Contrast is generated by a phase shift of the polarized and prism-split illuminating beams, which are combined by the second prism

amplitudes of the direct diffracted waves are equalized, and by destructive interference, dark-phase objects appear on a bright background. The application of this technique ranges from routine diagnosis to cell biology research [6], [10]. Today this method is also used advantageously in combination with epifluorescence. The specimen is adjusted under phase contrast before switching to fluorescence light in order to avoid bleaching, or both imaging modes are combined to match the fluorescent signal with the morphological image.

Nomarski Differential Interference Contrast (DIC) . In 1955 the physicist GEORGE NOMARSKI simplified the two-beam interference microscope in a way that it became available for routine microscopy [11] (Fig. 6). DIC uses modified wollaston prisms (Fig. 6 B), lying outside the focal

areas of condenser and objective [11]–[13]. Normal lenses are used; the contrast can be adjusted by shifting the lens-sided prisms. As in phase contrast, differences in the optical path length are visualized. Objects appear as a relief, and thick specimens lack the halo-typical pattern of phase contrast imaging. Full aperture of both the objective and the condenser allow optical sectioning (Fig. 7). Combined with inverted microscopes (see Section 29.1.4) and contrast-enhanced video microscopy (see Section 29.1.5), ultrastructures in cells and tissues can be imaged. Certain applica-
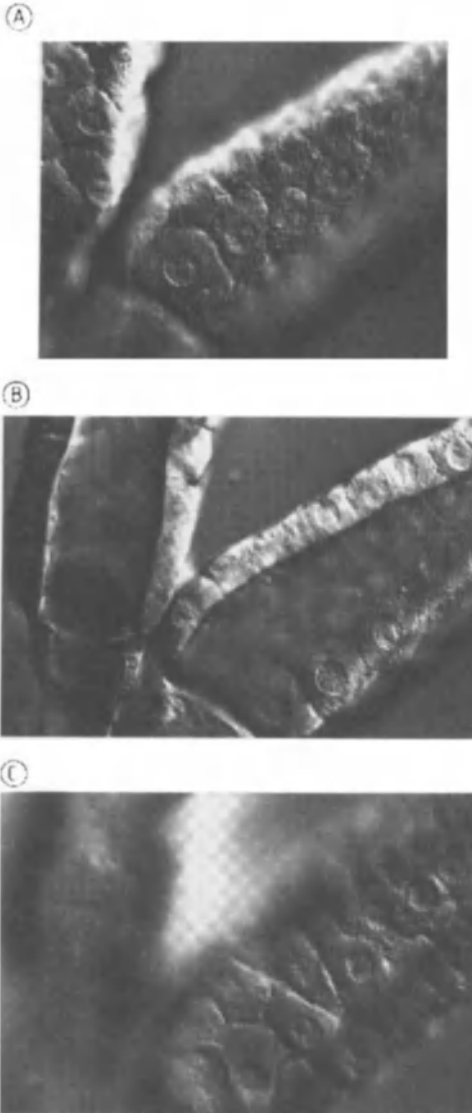
Figure 7. A)–C) Optical sectioning of salivary gland of drosophila by DIC Nomarski (Specimen courtesy of L. Rensing, Universität Bremen, Germany)

tions combine DIC with phase contrast. For routine applications, phase contrast lenses may give better results at lower magnification. The DIC image is well suited to be combined with a fluorescent image as well.

Polarization and Reflection Contrast Microscopy. Polarization microscopy is used preferentially for certain applications in cell research,

to enhance double diffraction (anisotropic) structures such as muscle fibers or the mitotic apparatus. Quantitative studies require strain-free polarization lenses for the polarized light used. With the help of reflection contrast microscopy [14], [15], visible interference is generated at epi-illumination, for example, to study adhesive properties of living cells on glass surfaces [16]. Moreover, this method may be combined with phase contrast or Nomarski DIC contrast. For the evaluation of immunogoldmarked specimens, immunogold staining (IGS), reflection as well as transmission contrast is feasible.

29.1.3.2. Fluorescence Microscopy

Basic Principles. The development of immunofluorescence greatly influenced medical practice and basic biomedical research [17]–[20]. Many biomedical specimens emit *autofluorescence* if illuminated properly. Today a wide range of fluorescent microstructures can be excited by application of fluorochromes (*secondary fluorescence*). Immunofluorescence is based on antigen–antibody connectivity [21]. This couples the high specificity of the immune reaction with the extreme sensitivity of fluorescence microscopy. Various markers for antibodies exist and derivates of fluorescein, mainly fluorescein–isothiocyanate (FITC), are the preferred fluorochromes. For this kind of microscopy, special lamps and filters are required. Light sources used include high-pressure bulbs as well as lasers. Fluorescence microscopy is used either in transmission or in epi-illumination; the latter combines phase contrast and Nomarski DIC [22], [23]. Fluorescence microscopic objectives with high aperture and a good transmission between 340 and 700 nm are preferred. Multi-immersion lenses of plan-type Neofluar have been designed specifically for such applications, as well as low-magnification lenses featuring high apertures (primary magnification 40, *NA* 1.3 oil). Plan apochromates are more suitable for illumination in visible light. The brightness *b* of fluorescent images increases proportionally to the square of the numerical aperture of lens L and condenser C, and decreases inversely with the square of the total magnification *M*

$$b \sim \frac{NA_L^2 + NA_C^2}{M^2} \tag{1.3}$$

This makes ocular lenses with low primary magnification an ideal choice. At epi-illumination the
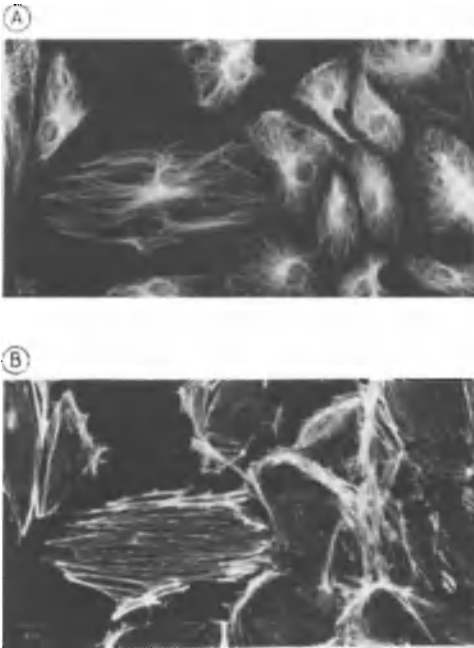
Figure 8. Double fluorescence micrograph of epithelial kidney cells
A) Microtubules, fluorescent dye FITC; B) Actin filament, fluorescent dye tetramethylrhodamine-B-isothiocyanate (TRITC)

brightness depends on the power of four of the aperture, since the lens acts as the condenser as well:

$$b \sim \frac{NA_L^4}{M^2} \qquad (1.4)$$

In the 1980s, fluorescence methods became a major tool for sensitive and highly specific detection in biomedical research and medical diagnosis.

Markers. The development of reagents based on fluorescence includes the use of biological molecules: antibodies, RNA, DNA, proteins, and other macromolecules and physiologic indicators [17]–[19]. These methods have enabled new findings in cell biology, molecular biology, developmental biology, and cancer research. The cytoskeleton (Fig. 8) is responsible for the cell shape, cell division, capacity of the cell to move, and the intracellular transport of cell organelles [24], [25]. Multiparameter fluorescence microscopy uses a broad selection of fluorescent dye molecules to examine several properties or components of a single cell simultaneously (Fig. 9).

A series of new probes suitable for this purpose has been developed by extending the usable spectrum to the near infrared (Fig. 10). *Fluorescence-based in situ hybridization* (FISH) is rapidly being recognized as a powerful tool for detecting and quantifying genetic sequences in cells and isolated chromosomes [26]. FISH is based on the highly specific binding of pieces of DNA, labeled with a fluorescent probe, to unique sequences of DNA or RNA in interphase nucleus or metaphase chromosomes. Recent developments enable the simultaneous visualization of three or more DNA regions (chromosome painting) [27].

Dyes and Applications. The green fluorescence protein (GFP) is a strongly fluorescent reporter molecule that stems from the jellyfish *Aquorea victoria* [28]. It has a fluorescein-like characteristic, it requires no cofactors or substrates, and is species-independent. GFP engineering has produced different color mutants which have been used for gene expression, tracing of cell lineage, and as fusion tags to monitor protein localization [29].

Time-Resolved Methods. Special methods and dyes have been developed to observe functional phenomena in living cells. Fluorescence lifetime imaging microscopy (FLIM) monitors fluorecence lifetime as a function of environmental parameters of cellular components. These include Ca^{2+}/pH-sensitive indicators, also called biosensors [19]. Typically a fluorescent dye is chemically linked to a macromolecule, creating an analogue of the natural molecule. The fluorescent analogue is specifically designed to measure chemical properties of the macromolecule, instead of tracking its location within the cell. Time-gated multichannel plates (MCPs) are used for time- and spatially resolved measurements.

Fluorescence recovery after photobleaching (FRAP) and the measurement of the fluorescence resonance energy (FRET), which allow the diffusional mobility of cellular components and their interaction at the molecular level to be monitored are other varieties of time-resolved methods [30]. A novel method is fluorescence correlation spectroscopy (FCS), which measures the statistical fluctuations of fluorescence intensity within a confocally illuminated volume. Correlation analysis allows the concentration of particles and their diffusion to be determined [31], [32]. FCS has proved
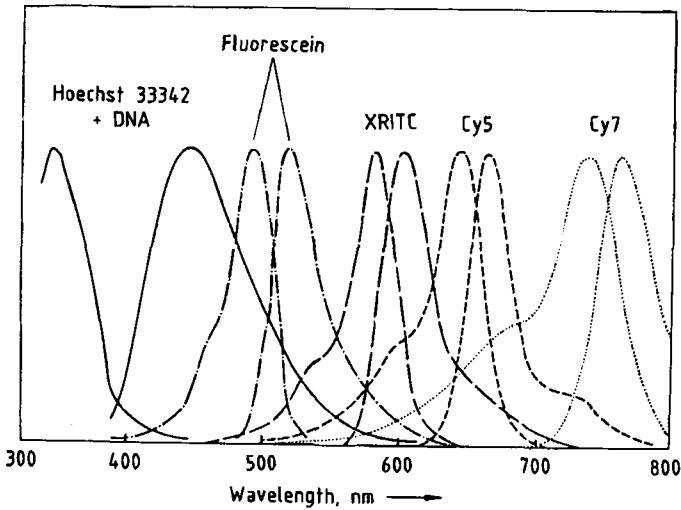
Figure 9. Excitation and emission curves of various fluorochromes
Filter curves depict dyes Hoechst 33342 (suitable for DNA), FITC (for actin), rhodamine (endosomes), and cyanine 5 (mitochondria)
(Courtesy of L. TAYLOR, University of Pittsburgh, United States)

be useful in the study of enzyme kinetics, DNA/protein interactions, receptor/ligand interactions, and cellular measurements [33]. Molecular interactions can be studied without interfering with the natural binding conditions, since only the reference substances in competition experiments need to be fluorescently labeled.

29.1.4. Inverted Microscopy

Inverted microscopes were originally designed for the evaluation and counting of plankton and sediments [34]. The first inverted research microscope which included photomicrography, microphotometry, cinemicrography, and video imaging was the Axiomat [4], [35]–[37]. Inverted microscopes in U-design (Fig. 11) with a rigid stage are of great advantage for investigations in cell and developmental biology, as well as in cell physiology experiments that are performed in optical microscopes [37]–[40]. Identification and selective micropreparation on live nuclear components in oocytes of *Xenopus* in the DIC Nomarski technique facilitate highresolution optical microscopy in different planes of the specimen (optical sectioning), which can be isolated by selective micropreparation for subsequent analysis in the electron microscope [41].

In the life cycle of the green alga *Acetabularia*, high-resolution phase and DIC microscopy have made part of the endoplasmic reticulum and lampbrush chromosomes visible [42], [43]. The distinct advantage of inverted compared to upright microscopes is the unhindered application of cell and tissue culture chambers in conjunction with micromanipulators, micropipettes, microelectrodes (patch clamp method), and injection systems for experimental cell research. Computer-controlled microinjection can be performed automatically [44]. Manipulation of cells, organelles, and genomes is also possible by means of laser microbeam and optical trapping [25].

29.1.5. Optoelectronic Imaging

Electronic imaging methods such as video microscopy and laser scanning microscopy can extend resolution slightly, but the ability to extend the detectibility of subresolved structures and subresolution movements is more important. A video camera can generally be mounted to the same port employed for normal photomicrography. In *video-enhanced contrast microscopy* (VEC) (Fig. 12), high-resolution video cameras based on a vidicon tube are used. This principle of adjusting gain, and offset and digital image processing is also used in

A)

B)
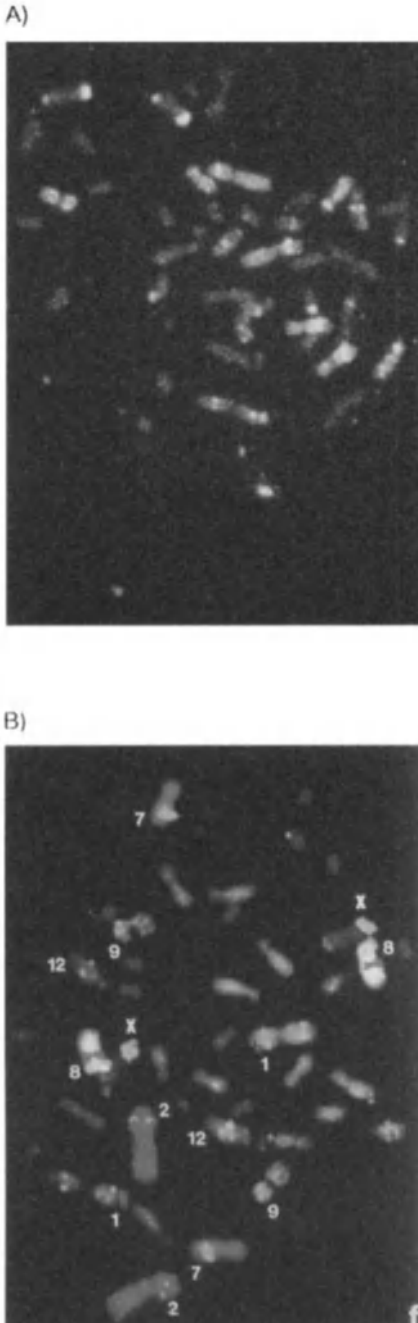
Figure 10. A) Micrograph by triple exposure of color slide film using conventional fluorescence microscopy; B) Pseudocolored digital image recorded with the CCD camera for chromosome 1, 2, 7, 8, 9, 12 and the human X chromosome Green = FITC; Red = TRITC; Blue = DAPI
(Courtesy of T. Cremer, Heidelberg, Germany and Oxford University Press)
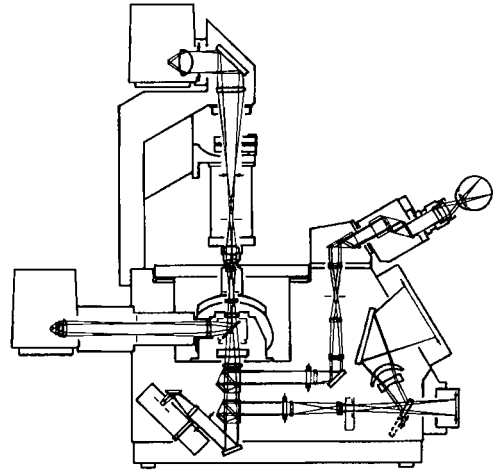
Figure 11. Light path of an inverted microscope

laser scanning microscopes (see Section 29.1.6). Digital imaging and image processing make it possible to see structures that are smaller than the resolution limit [12], [45]–[48]. By means of this method, single microtubules can be observed [49], [50]. The size of each microtubule is only one-tenth of the visible-light resolution limit. In the 1980s, solid-state cameras were introduced into microscopy [51], [52]. Most of them are based on a *charge coupled device* (CCD), which produces digital data directly for computer processing and image analysis. For low light level (LLL) or video intensification microscopy (VIM), i.e., weak fluorescence or small FISH signals (see Fig. 10 B), cooled, sensitive CCD cameras for black and white or color are available now and have replaced tube cameras [53]. Digital imaging techniques have gained increasing attention as supplements to conventional epifluorescence and photomicrography [54]. *Microphotometry* permits spectrophotometric analysis of substances, as well as statistical and kinetic measurement and compilation of two-dimensional intensity profiles. [40].

29.1.6. Confocal Laser Scanning Microscopy

29.1.6.1. Basic Principles

The fundamentals of confocal microscopy go back to 1959, when MINSKY described a new kind of microscopic apparatus [55]. The historical de-

Figure 12. Principle of video-enhanced contrast microscopy
A) Original image; B) Offset (brightness); C) Contrast gain

Figure 13. Ray path in the confocal microscope
a) Dichroic beam splitter; b) Scanner; c) Tube lens; d) Objective; e) Specimen; f) Focal plane; g) Filter; h) Pinhole; i) Detection system (photomultiplier)

Figure 13. The laser light is focused via the scanner (b) through the tube lens (c) and the objective (d), and illuminates a small spot in the specimen (e). Emitted light emanating from the focal plane and the planes above and below (dotted and dashed lines) is directed via the scanner to the dichroic beam splitter (a) where it is decoupled and directed onto a photomultiplier (i). A pinhole (h) in front of the photomultiplier is positioned at the crossover of the light beams emerging from the focal point. This plane corresponds to the intermediate image of the Köhler illumination described in Section 29.1.3. Light emanating from above and below the focal point has its crossover behind and before the pinhole plane so that the pinhole acts as a spatial filter. Numerous papers elucidate the basic aspects of confocal image formation [59] – [64].

Since only a small specimen point is recorded, the entire image of a specimen is generated by either moving the specimen (*stage scanning*) or moving the beam of light across the stationary specimen (*beam scanning*). The latter is utilized in all commercial instruments (Fig. 14). Theoretical considerations have shown that both scanning forms give identical resolution. The optical sectioning capability at lateral scanning opens a way to image a three-dimensional structure with a slight axial stepping of the stage (see Fig. 15).

29.1.6.2. Imaging Performance

The imaging characteristic in confocal microscopy is given by the properties of the specimen weighted by the applicable spatial confocal response function. An explanation based on wave optics suggests that the point spread functions of the illuminating path and the detection path (resembling the photon detection possibility) are con-

velopment of confocal microscopy is reviewed elsewhere [56], [57]. Two different kinds of confocal instruments are available: the tandem scanning microscope and the laser scanning microscope. The *tandem scanning microscope* uses normal light chopped by a rotating Nipkow disk [58]. The light beams reflected by the specimen cross the rotating disk on its opposite side (thus tandem), and due to the symmetric pattern of holes, out-of-focus blur is removed. This system allows real-time and real-color imaging, but is limited in sensitivity because of the fixed size of the pinholes.

The more comonly applied *confocal laser scanning microscope* (CLSM) features a variable pinhole and allows adjustment of magnification by modifying scan angles. A schematic example of the principle of confocal laser scanning is given in

Figure 14. Setup of the confocal microscope, based on a research light microscope (Zeiss)
a), b) Photomultiplier tubes; c) Confocal variable pinhole; d) Scanners; e) Barrier filter; f) He – Ne laser (543 nm, 633 nm);
g) Stage; h) Ar laser (488 nm, 514 nm); i) UV laser (350 – 360 nm); j) Antivibration table
Light of internal (He – Ne) and external (Ar ion or UV) lasers passes the scanner unit; switchable mirrors and barrier filters permit
multiwavelength operation. Additional barrier filters are located between the confocal variable pinhole and the beam splitter
adjacent to the photomultipliers.

volved, consequently resulting in a narrow point resolution [65].

Describing image formation of an reflecting object and the corresponding coherent imaging properties must be distinguished from a fluorescent object viewed in incoherent imaging because in the latter the illuminating wavelength and the detected wavelength are different. In terms of resolution, the Rayleigh criterion determines the minimum resolvable distance between two points of equal brightness (see Eq. 2). For *incoherent imaging* using a lens with a certain numerical aperture NA at a wavelength λ, the *lateral resolution* d_r obtainable in confocal imaging is [62]:

$$d_r \text{ (conf.)} = 0.46\lambda/NA \qquad (1.5)$$

With a wavelength of 514 nm (green argon laser line) and a numerical aperture of 1.4, this equation gives a resolution of $d_r = 157$ nm. Compared to conventional imaging (see Eq. 1.2), the resolution is ca. 32 % higher [66].

An approximation of the *axial resolution* d_z is given by

$$d_z \text{ (conf.)} = 1.4\lambda/NA^2 \qquad (1.6)$$

The values (514 nm and $NA = 1.4$) give a d_z of 367 nm.

One way to measure the axial response of a confocal microscope is to record light intensity from a thin layer with infinite lateral extension. In a conventional microscope, no difference — and therefore no depth discrimination — are obtained. However, in confocal microscopy, a strong falloff occurs if the object is out of focus. The free width at half maximum (FWHM) of such a function is proportional to the optical section thickness [64], [67]. The optical section thickness also depends on the diameter of the pinhole. The optimum pinhole size D with a maximum deterioration of 10 % in reflection is given by [63]

$$D = 0.95\lambda M/NA \qquad (1.7)$$

with magnification M, λ, and D given in micrometers. In fluorescence microscopy the size has to be reduced 10 – 30 %, depending on the ratio of excitation to emission wavelengths. Another way to describe the imaging performance is by the modulation transfer function (MTF); such measured and theoretically derived three-dimensional functions have been published [68], [69].

29.1.6.3. Instrumentation

Confocal microscopic equipment is available as an add-on to a standard microscope or in the

Figure 16. Multichannel confocal image of macrophages (J774, mouse cell line)
Extrathyroidal release of thyroid hormones from thyroglobulin is marked by fluorescence dye (first channel, green); actin cytoskeleton is labeled with rhodamine-phallodin (second channel, red). Morphological information (third channel), providing a dark background in the cells, is detected in transmission.
(Permission of K. Brix, V. Herzog, Institut für Zellbiologie, Bonn, Germany)

Figure 15. A) Conventional versus confocal imaging (inserted region of interest) of a lung structure in fluorescence; B) Multislice image sequence obtained by optical sectioning

form of more fully integrated and stable devices, including computer control equipment. Both upright and inverted types of microscopes are available. The inverted confocal microscope allows bigger specimens to be studied and offers easy handling of micromanipulating or injecting devices (see also Section 29.1.4). Beam scanning is obtained by rotating galvanometric mirrors. Axial control is realized by a stepping motor drive or a piezoelectric translator. Most instruments are equipped with several detectors to record multi-

fluorescence events simultaneously. Some also offer transmission detectors. Signals from different channels can then be mixed electronically (see Fig. 16). Recently, confocal transmission microscopy has also been studied. Improvement of contrast and resolution depending on the pinhole size of a thin specimen has been documented in a confocal transmission arrangement using differential interference contrast [70].

Variations of the confocal design have been developed, such as bilateral scanning in real time with a scanning slit [71]. Because of the fast scanning, TV detectors such as a sensitive CCD camera are best suited. Slit-scan confocal microscopes offer a better signal-to-noise ratio and real-time imaging capabilities but a slightly reduced resolution compared to a conventional CLSM.

Various types of lasers can be linked to the microscope [72], but up to now only continous-wave gas lasers have been used routinely. This includes argon-ion lasers having excitation lines at 488 and 514 nm and helium – neon lasers with a major line at 633 nm and a weaker line at 543 nm. The use of UV lasers requires achromatic lenses and specifically designed scanning mirrors. Lasers
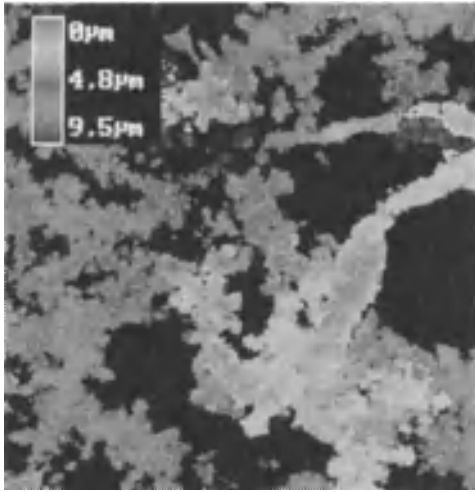
Figure 17. Golgi-stained neuron: three-dimensional display of a confocal data set acquired in reflection
Color depth coding visualizes the three-dimensional neuron; various spines of different size give a brushlike impression. (Specimen courtesy of B. FORTMANN, Institut für Ernährungswissenschaften, Gießen, Germany)

used in the UV range are argon-ion UV or argon-ion tunable lasers, as well as krypton-ion and helium – cadmium lasers. Depending on the emission lines of the fluorescent dyes, various filters are used in front of the photomultiplier, characterized by quantum efficiency and spectral response. The total sensitivity of a system is given by the transfer efficiency of the various optical components [73] and, in addition, the quantum efficiency of the detector and its associated electronics [74].

Simplified types of laser scanning microscopes, such as the *fiber scanning optical microscope* have also been proposed [68]. Here the fiber can act both as a light delivering source and as a pinhole. A totally different physical approach to reduce out-of-focus blur is based on a physical phenomenon called *double-photon (2 p) excitation*, whereby fluorescence is generated by simultaneous absorption of two photons of long wavelength at a single molecule. The necessary energy emitted by pulse lasers is present only at the focal plane; thus a "confocal" effect is produced. At turbid tissue media, the signal level under 2 p-excitation drops much faster than that under single-photon excitation, although image resolution is higher in the former case [75], [76].

The diffraction pattern of the illuminating beam is elongated axially compared to the lateral direction (see Eq. 1.5 versus 1.6). This anisotropy

is caused basically by the limited numerical aperture of the microscopic lens. An optimal diffraction pattern would be achieved by an unlimited, 360° aperture. This concept is known as the *4-pi microscope*. Prototypes use two facing lenses or two lenses oriented at an angle (theta microscopy). The object in focus is illuminated from two sides, and fluorescence emanating from the object is detected through both lenses [77].

29.1.6.4. Imaging Modalities and Biomedical Applications

Imaging Modalities. Confocal contrast generation include reflectance, rescattering, and fluorescence. The first results in confocal imaging were obtained on unstained tissue in *reflectance* [78]. Reflectance is also the preferred imaging modality used in industrial inspection of semiconductors, often combined with methods such as optical beam induced conductivity (OBIC).

In biomedical science, the study of unstained tissue is limited due to reflectance. The reflectance must be intensified by metal impregnation (e.g., Golgi stain); one such example is given in Figure 17. Other examples include peroxidase – labeling with nickel intensification in neurobiology [79] or silver-stained nucleolar organizer region-associated proteins (AgNORs). Gold immunolabeling has also been used [80]. In *rescattering*, variations in the refractive index can give rise to reasonable contrast. One particular example is in ophthalmology [81] where structures in the transparent cornea such as cell nuclei, keratocytes, neurons, or the fibers of the lens can be visualized (see Fig. 18).

Fluorescence is the most commonly used imaging mode in confocal microscopy, such as, applications in cell biology [82], [83], which include the use of autofluorescence, specific dyes in combination with antibodies, and in situ hybridization. Comprehensive reviews of available fluorochromes are published elsewhere [80], [84] – [87]. UV lasers broaden the spectrum of available fluorescence dyes; even the autofluorescence present in reduced pyridine nucleotides [NAD(P)H] can be monitored [88].

Biomedical Applications. Optical sectioning devices, being noninvasive and having low radiation damage, are ideal for studying living specimen, organs, and tissues. With confocal microscopes, structures can be observed in vivo at full three-dimensional microscopic resolution [87].
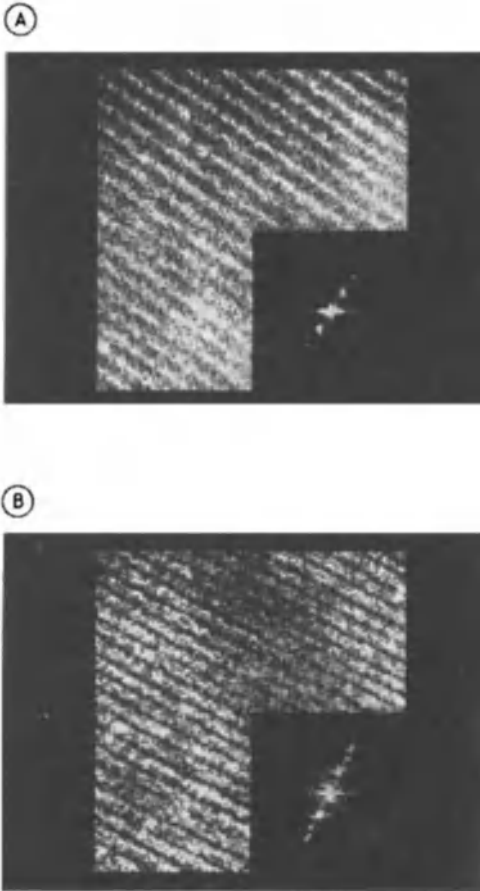
Figure 18. Confocal images of the eye's lens fibers, resulting from rescattering at microstructures with different indices of refraction
A) Imaging through the 400 μm transparent cornea;
B) Imaging with cornea removedThe inserted Fourier power spectrum documents the loss in spatial frequency content.
(Specimen courtesy of B. R. MASTERS, USUHS, Bethesda, United States)

Figure 19. Application of confocal microscopy to plant cells (*Egeria densa Planchon*), in vivo
Cell boundaries and the movement of chloroplasts are visualized by autofluorescence (green) and FITC labeling (red). (Permission of I. DAHSE, Institut für Biophysik, Jena, Germany)

Small organisms or organisms in an early stage of development can be observed, as in embryogenesis [89], [90]. Concerning individual structures, morphological changes, such as the movement of cell compartments (see Fig. 19), growth of neurons [91], or synaptic plasticity in the retina, have also been studied [92]. Moreover, confocal arrangements allow optical trapping (i.e., user-controlled movement of small spheres filled with biochemical solution into and within cells). A new application of confocal microscopy is in clinical imaging, which includes the observation of wound healing, flow processes in veins, and opthalmology [81], [86] (see Fig. 18).

29.1.7. Computer Applications in Digital Microscopy

Computer applications in digital microscopy include image storage in databases, compression, enhancement, analysis, and visualization. Digital analysis applied to microscopical data sets has the aim of automatically extracting structures of interest out of digitized images and then quantifying them. This includes field-specific or stereological, geometrical or object-specific, and intensity-related parameters. In the case of 3D images, typically generated by a confocal imaging technique, a computer graphical 3D visulization of the data stack is performed, which allows inspection of the original data and subsequent image-processing operations.

Different digital image formats are used in microscopy, for example, the TIFF format. In 1999, a microscopic imaging standard was established by the ARC-NEMA Institute as part of the DICOM standard (Digital Imaging and Communication in Medicine). This format, DICOM-VL (Visible Light), regulates storage of instrumental
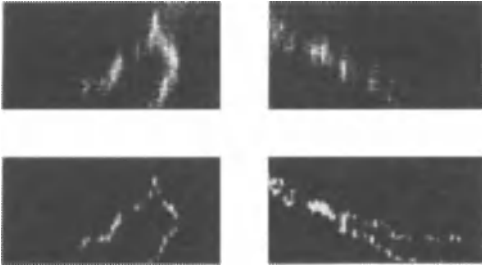
Figure 20. Projectional side views of rat hippocampal dendrites stained with Lucifer yellow
Top: Original confocal data set: Bottom: Restored data following 100 iterations of the maximum likelihood blind deconvolution algorithm
(Permission of T. HOLMES, Rensselaer Insitute, Troy, United States)

and specimen parameters besides the image information. This fosters the use of microscopic images in telemicroscopic applications and in medical networks.

29.1.7.1. Image Analysis

Image quality assurance is an important control mechanism for image data to be analyzed quantitatively. Appropiate techniques include sampling [93], focus setting techniques [94], image quality measurement [95], and calibration routines [96]. Some of the undesired phenomena which influence or prevent successful application of image analysis require preprocessing. This includes attenuation of the laser beam in thick specimens, photobleaching or saturation of the fluorescence, autofluorescence, and background noise. and the distortions caused by the embedding media [80], [86], [97]. The loss of resolution in thick specimens is shown in Figure 18. The correction of such phenomena is often limited by the underlying theoretical models selected, since the effects are highly object dependent.

Undesired effects in fluorescence confocal microscopy can be corrected by ratio imaging. Ratioing (i.e., the compensation of two channels by digital division) can substantially reduce cross-talk effects in double detectors [63]. Environmental markers indicating the pH value or the level of free calcium ions (Ca^{2+}) also require ratio imaging to compensate for structural densities or section thickness [19].

Due to the anisotropic resolution of the data sets, various ways of improving axial resolution

are available; *deconvolving methods* use the point spread function of the optics, which has to be measured first [98]. The algorithms can be differentiated further by their ability to incorporate noise. The classical filter is the Wiener filter; new developments include the iterative *maximum likelihood estimation* [99]. This method was extended with a blind deconvolution algorithm, using certain constraints to model the point spread function, which is a promising approach for biomedical users (see Fig. 20). However, calculation times are still in the range of 1 h for some hundred interations. A different approach is the *nearest-neighbor deconvolution* developed for bright-field imaging [100], by taking into account the blur present in the planes above and below the focal plane. More recently, this method has been contrasted with a no-neighbor deconvolution method [101]. Besides directly improving data, isotropy can also be obtained by interpolating the lateral dimension [80]. Transformation of data from a cubic format into other lattices has also been proposed, based on mathematical morphology [102].

Sets in digital image processing to enhance, filter, segment, and identify structures are frequently applied to three-dimensional microscopic data sets. This can be done either section by section or by using the corresponding three-dimensional extension of digital filters, such as the Gauss, Laplace, Sobel, and Median filter or noise removal algorithms [103] – [106]. Segmentation often requires interaction of the user. since automatic segmentation algorithms generally rely on the intensities of the voxels only. Once the binary images have been obtained, binary morphological operations using structuring elements for erosion and dilatation can be performed [107].

Digital measurements address the size and volume of structures, quantities, or forms and relation between volumetric subtleties. To measure volumes, all voxels that identify a certain structure are summed up. It is particularly difficult to measure the surface of structures, and most methods offer an estimation based only on stereological methods. Object counting in volumes usually requires the definition of guard boxes, which are moved throughout the volume. Objects falling into such boxes are evaluated by following certain stereological rules to avoid any bias. Stereological considerations are also used to estimate spatial statistics, such as numerical densities or nearest-neighbor distances; such methods have been used extensively in three-dimensional cytometry [80].
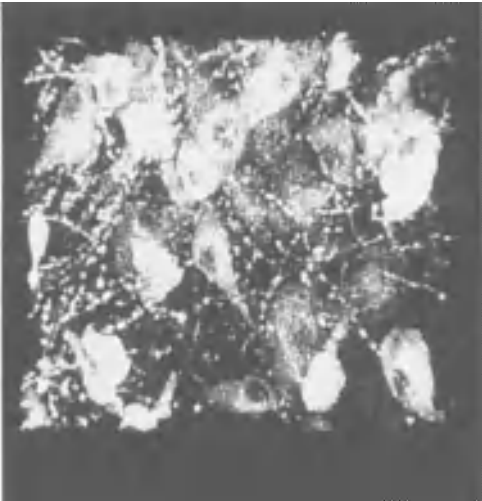
Figure 21. Volume rendering of a confocal sequence
The neural network of immunohistochemically stained cells, together with connecting fibers in the neural optical tract of hamsters is visualized by this projection
(Specimen courtesy of H. KORF, Zentrum für Morphologie, Frankfurt, Germany)

29.1.7.2. Visualization

Visualization of serial sections is the most frequently used computer technique in confocal microscopy, independent of specific applications [83], [108], [109]. Most commercial instruments have three-dimensional software implementations ready for the user. Three-dimensional objects are displayed on the computer screen by staking up the individual sections. Such techniques have been reviewed in several articles [110]–[114]. Besides surface rendering based on contours, volumetric representations that rely on the intensities of the voxels are the preferred techniques in confocal microscopy. A comparative assessment of both techniques is given in [115]. Reconstructions are presented three-dimensionally on a computer graphics screen by using either visual cues, animation, of stereoscopic displays.

One method used in *volume rendering* is to define imaginary rays through the volume (*ray-casting techniques*) [116], [117]. The way of casting is accomplished by the definition of certain projection geometry. Given a central projection, the angles of the rays are defined by the matrix of the screeen and the center of projection. A common implementation traverses the data in a front-to-back order, which has the advantage that

the algorithm can stop when a predefined property of the accumulation process has been reached. This class of methods is also referred to as *image order rendering*, since the matrix of the screen is used as the starting point for the accumulation process. Because of the projection geometry, such a ray-firing technique does not necessarily pierce the center of a voxel, and an interpolation has to be taken into account. To improve the output of ray-casting renderings, preprocessing of the voxel scenes may be performed as, for example, in the renditions of complex chromatin arrangements [106]. Modalities available in image order rendering include maximum projection (Fig. 21), integration, and surface mode (Fig. 18).

A much simpler geometry is realized in a parallel projection, such as in *object order rendering*. Here the volume is rotated according to the current viewing angle and then traversed in a back-to-front order [118]. During traversal, each voxel is projected to the appropiate screen pixel where the contributions are summed or blended, a situation that is referred to as compositing. During the accumulation process other properties of voxels besides the intensities may be considered, such as opacity, color, gradient values [119], [120], or the simulation of fluorescence processes (SFP) [121]. However, since volume rendering relies entirely on voxel attributes, it has some disadvantages if analytical representations are required. Therefore, elements of computer graphics are combined with volumetric renditions. Examples can be found in vector descriptions of volumetric data sets [122] or in geometric attributes expressing local properties and topologies that can be embedded in the three-dimensional reconstruction [113], [120], [123]. The vast amount of data available in four-dimensional studies initiated new concepts in computing and visualization that utilize special data compression techniques [124].

29.2. Electron Microscopy

29.2.1. Introductory Considerations [138]

There are two main classes of electron microscopy (EM) techniques. In the first class, the electron probe is a stationary beam incident along a fixed direction. This incident beam can be parallel [conventional transmission electron microscopy (CTEM), high-resolution transmission electron microscopy (HRTEM), high-voltage transmission

electron microscopy (HVTEM), selected-area electron diffraction (SAED)] or convergent [convergent-beam electron diffraction (CBED), convergent-beam electron microscopy (CBEM)]. The resolution is determined by the quality of the imaging optics behind the specimen. Instruments implementing these techniques are conceptually related to classical light microscopes.

In the second class of methods, a fine electron probe is scanned across the specimen, and transmitted electron (TE) [scanning transmission electron microscopy (STEM)] or the desired excited signal such as secondary electrons (SE) or backscattered electrons (BSE) [scanning electron microscopy (SEM)], and/or AE [Auger Electron spectroscopy (AES)/scanning Auger microscopy (SAM)] is selected, detected, and the signal amplified and used to modulate the intensity of another electron beam which is scanned synchronously eith the first over the screen of a TV monitor. The stationary beam methods are based on image formation processes, whereas the scanning methods are essentially "mapping" techniques. Their resolution is mainly determined by the probe size, i.e., by its electron probe formation optics. The magnification is geometric; it is determined by the ratio of the areas scanned by the electron beam on the screen and synchronously by the electron probe on the specimen.

Analytical microscopes (AEM) have been developed which make it possible to implement a number of imaging modes, diffraction modes, and analytical modes in a single instrument and often also to apply them to a single specimen. The synergism of these tools has the potential to provide detailed structural and chemical information on the nanoscale. There is a trend of extending stationary beam probe analytical tools such as X-ray microanalysis, Auger electron spectroscopy (AES), and electron energy loss spectroscopy (EELS) into mapping methods which make it possible to image spatial distributions of chemical elements.

29.2.2. Conventional Transmission Electron Microscopy (CTEM)

29.2.2.1. Introduction

The images produced in transmission electron microscopy are essentially due to local diffraction phenomena; absorption contrast plays only a minor role. Not only is electron diffraction responsi-

ble for the image formation; it also allows establishment of the orientation relationship between direct space, as observed in the image, and reciprocal space as imaged in the diffraction pattern (see → Structural Analysis by Diffraction). Images and the corresponding diffraction patterns are equally important for a detailed interpretation; both should always be produced from the same selected area, with the specimen orientation unchanged. For this reason, most commercial microscopes have the ability to switch easily from the diffraction mode to the imaging mode and vice versa (see Section 29.2.2.5).

29.2.2.2. Scattering by Atoms: Atomic Scattering Factor

Electrons are scattered by atoms as a result of Coulomb interaction with the nucleus and the electron cloud. The atomic scattering factor $f_e(\theta)$ thus contains two terms of opposite sign

$$f_e(\theta) = \frac{m e^2 \lambda}{2h^2} [z - f_x(\theta)]/\sin^2\theta \qquad (2.1)$$

where z is the atomic number; $f_x(\theta)$ the scattering factor for X rays; m the electron mass; e the electron charge; λ the wavelength of the electrons; and h Planck's constant [139].

The first term clearly relates to the nucleus, whereas the second term is due to the electron cloud. The interaction with matter is stronger ($\times$ 10^4) for electrons than for X rays or neutrons, which interact only with the electron cloud or with the nucleus, respectively. As a result multiple scattering will not be negligible in electron diffraction experiments. Moreover, electron scattering is oriented mainly in the forward direction. Tables of $f_e(\theta)$ for different atoms are given in [140].

29.2.2.3. Kinematic Diffraction by Crystals

29.2.2.3.1. Lattice, Reciprocal Lattice

The amplitude diffracted by an assembly of atoms results from interference of the waves scattered by the atoms (i.e., the scattered amplitude in any given direction is obtained by summing the amplitudes scattered in that direction by the individual atoms and taking into account the phase differences due to path differences resulting from the geometry of the assembly). For a physical and mathematical description of the important terms
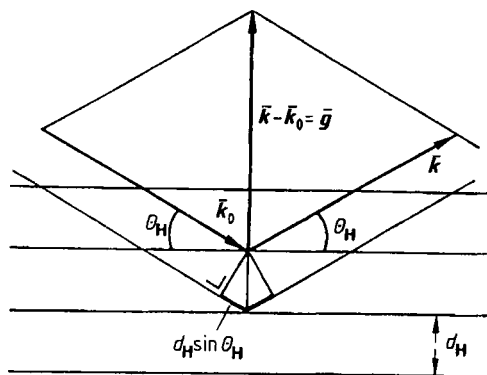
Figure 22. The derivation of Bragg's diffraction law in direct space

lattice and reciprocal lattice, see [141] and → Structural Analysis by Diffraction.

29.2.2.3.2. Geometry of Diffraction

Bragg's Law [142]. Like other diffraction phenomena in periodic structures, electron diffraction can be described in either direct or reciprocal space. Bragg's diffraction condition describes diffraction in direct space. A diffracted beam will be formed whenever the incident beam encloses with the set of lattice planes an angle θ_H, such that the path difference $2 d_H \sin \theta_H$ between waves diffracted by successive lattice planes is an integer n of wavelengths λ of the radiation used:

$$2 d_H \sin \theta_H = n\lambda \qquad (2.2)$$

where **H** denotes the Miller indices of the plane. In Figure 22 this path difference is indicated by a thicker line. For 100-kV electrons, $\lambda = 4$ pm, and as a result, the Bragg angles θ_H are very small — a few degrees at most.

Ewald Construction [143]. Diffraction conditions can also be formulated in terms of reciprocal space as

$$\bar{k} = \bar{k}_0 + \bar{g} \qquad (2.3)$$

where $\bar{k}_0$ is the wave vector of the incident electron beam and $\bar{k}$ the wave vector of the diffracted beam, where $|\bar{k}| = |\bar{k}_0| = 1/\lambda$; $\bar{g}$ is a reciprocal lattice vector. Equations (2.2) and (2.3) have the same physical content. They can be obtained by expressing the conservation of energy and of mo-

mentum of the incident electrons on scattering. More details on Ewald's construction are given in [144] and in → Structural Analysis by Diffraction.

Diffraction by a Thin Foil. Since electrons are strongly "absorbed" in solids the specimen must be a thin foil (< 300 nm); otherwise no diffracted beams will be transmitted. The number of unit cells along the normal to the foil is thus finite, whereas along directions parallel to the foil plane, the number of unit cells can be considered infinite. In such specimens the diffraction conditions are relaxed, and diffraction also occurs for angles of incidence deviating somewhat from the Bragg angle. In reciprocal space this relaxation results in a transformation of the sharp reciprocal lattice nodes into thin rods (so-called relrods) perpendicular to the foil plane. The Ewald sphere intersects such a rod to produce a diffracted beam. The direction of this beam is obtained by joining the center of Ewald's sphere with this intersection point. The resulting diffraction pattern can to a good approximation be considered to be a planar section of the reciprocal lattice. Indexing of the diffraction spots is simple and unambiguous.

The distance $\bar{s}_{\bar{g}}$ by which the lattice node G is missed by Ewald's sphere is called the *excitation error*. It is a vector along the direction of the foil normal joining the reciprocal lattice node G to the intersection point with Ewald's sphere. It is positive when pointing in the sense of the incident beam and negative in the opposite case.

Column Approximation [145]. The small magnitude of the Bragg angles causes the electrons to propagate along narrow columns parallel to the beam direction. The intensity observed in a point at the exit face of the specimen is thus determined by the amplitude scattered by the material present in the column located at that point. In a perfect foil the intensity does not depend on the considered column, but when defects are present, it does.

Kinematic Rocking Curve. The amplitude of the beam scattered by a foil of thickness z_0 is obtained by summing the amplitudes scattered by the volume elements along a column perpendicular to the foil surface and taking into account the phase differences due to their depth position in the column. For a perfect foil the intensity of the scattered beam, I_s for reflection **H** is given by
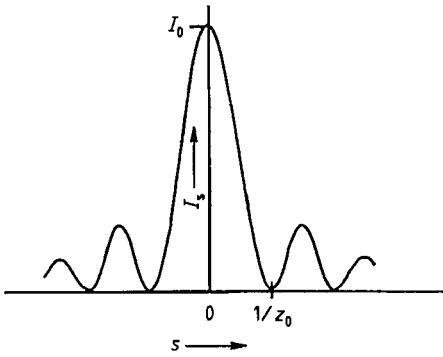
Figure 23. Plot of scattered intensity I_s versus excitation error, i.e., the "rocking curve" according to the kinematic approximation
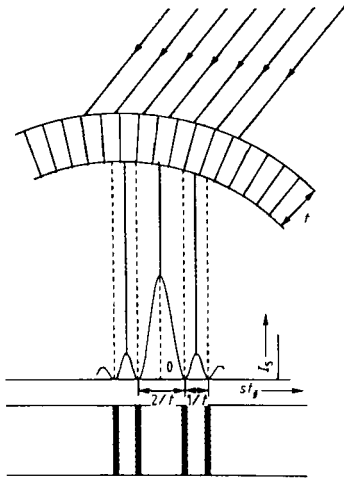
Figure 24. Formation of bent extinction contours by a cylindrically bent foil

$$I_s(s, z_0) = F_H^2 \sin^2 \pi s z_0 / (\pi s)^2 \qquad (2.4)$$

where F_H is the structure factor [127] with indices **H**.

The dependence of I_s on s (i.e., on the direction of the incident beam and on the foil thickness z_0) is called the *rocking curve* (see Fig. 23). The separation of the zeros is $1/z_0$, and the full width of the central peak is $2/z_0$.

Bent Contours [146]. For a curved foil, s varies along the specimen; the loci of constant s, imaged as bright and dark fringes, are called bent extinction contours. They form as shown schematically in Figure 24 for a cylindrically bent foil.
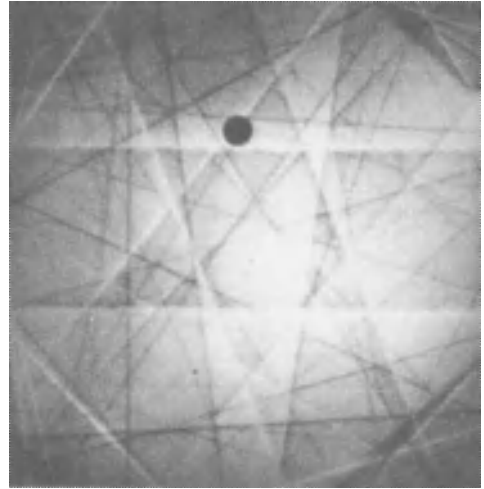
Figure 25. Kikuchi line pattern produced by a silicon foil. Note the parallel pairs of bright (excess) and dark (deficiency) lines. The deficiency lines are close to the center. The excess lines are separated in angle by $2\theta_H$ from the corresponding deficiency lines

Thickness Contours. The thickness dependence of I_s is periodic with period $1/s_g$: the period thus diverges for $s_g = 0$ (i.e., for the exact Bragg condition). The loci of equal thickness are also imaged as lines of equal intensity; they are called thickness extinction contours or wedge fringes.

Kikuchi Lines [147]. In rather thick specimens a second type of diffraction pattern becomes prominent, which consists mainly of lines rather than spots. These lines occur in parallel pairs; the line closest to the center of the diffraction pattern is darker, whereas the one further from the center is brighter than the background (Fig. 25). These lines are the intersections of Ewald's sphere with rather flat double cones having semiapex angles equal to $90° - \theta_H$, where θ_H is the Bragg angle associated with the pair of lines ($\theta_H \approx 1 - 2°$). The angular separation of the bright – dark line pair is $2\theta_H$.

If the Bragg condition for a set of lattice planes is satisfied exactly, the bright Kikuchi line associated with that set of planes will pass through the corresponding Bragg spot, whereas the dark one will pass through the origin. The separation of a Bragg spot from its corresponding Kikuchi line is a measure of the excitation error s of that reflection. As a result, the Kikuchi line pattern provides a means to measure s [148].

29.2.2.4. Dynamic Diffraction by Crystals

29.2.2.4.1. General Considerations

According to kinematic theory the intensity of diffraction spots is proportional to the square of the structure factor (Eq. 4). This simple proportionality is lost in reality because of multiple diffraction effects, which are taken into account adequately in dynamic theory [149].

Although kinematic theory does not predict the intensities of diffraction spots correctly, it is useful because the geometric features of diffraction are well described. For large s or very small foil thicknesses the relative spot intensities are qualitatively reproduced. For $s = 0$, kinematic theory breaks down since a diffracted beam with diffraction vector $\bar{g}_1$ may then acquire an amplitude comparable to that of the incident beam, and thus act as an incident beam and give rise to diffraction in the reverse sense (i.e., with diffraction vector $-\bar{g}_1$).

For most orientations of the incident beam, many diffracted beams are excited simultaneously and a section of reciprocal space is imaged. However by carefully orienting a sufficiently thick specimen the intensity of a single diffracted beam can be maximized: this is called a *two-beam case*. Such specimen orientations should be used, whenever possible, in the quantitative study of defects.

29.2.2.4.2. Basic Equations

The two-beam dynamic theory describes the interplay between transmitted and diffracted beams by means of a set of coupled differential equations of the form [128], [150].

$$\mathrm{d}T/\mathrm{d}z = \left(\pi i/t_{-g}\right)S\exp 2\pi i s z \tag{2.5a}$$

$$\mathrm{d}S/\mathrm{d}z = \left(\pi i/t_g\right)T\exp -2\pi i s z \tag{2.5b}$$

where T and S are the complex amplitudes of the transmitted and scattered beams, respectively; z measures the depth in the foil. The exponential factors take into account the growing phase shift due to the excitation error, with increasing distance z behind the entrance face. The factor i takes into account the phase jump on scattering; t_g and t_{-g} *are called the extinction distances.* The extinction distance, which has the dimension of length, is a measure of the strength of the reflection $\bar{g}$. For low-order reflections of elemental metals it is of the order of a few tens of nanometers, but for weak superstructure reflections it may be several hun-

dred nanometers. Whether a foil is "thick" or "thin" depends on the number of extinction distances in the foil thickness. Kinematic theory is valid only for foils with a thickness that is a fraction of an extinction distance.

Absorption can phenomenologically be accounted for by assuming the extinction distances to be complex [151]. Formally $1/t_{\bar{g}}$ is replaced by $1/t_{\bar{g}} + i/\tau_g$, where τ_g is called the absorption length. Empirically, it is found that τ_g ranges from $5\,t_{\bar{g}}$ to $15\,t_{\bar{g}}$. The substitution can be made, either in the Equation set (2.5) or directly in the solution of this set.

29.2.2.4.3. Dynamic Rocking Curve

By ignoring absorption and taking into account the fact that $T = 1$ and $S = 0$ at $z = 0$ (entrance face), integration of Equation set (2.5) gives

$$T(s,z) = \left[\cos \pi \sigma z - i(s/\sigma)\sin \pi \sigma z\right]\exp \pi i s z \tag{2.6a}$$

$$S(s,z) = \left(i/\sigma t_g\right)\sin \pi \sigma z \exp \pi i s z \tag{2.6b}$$

and thus

$$I_s = S S^* = \left(1/\sigma t_g\right)^2 \sin^2 \pi \sigma z \tag{2.7}$$

$$I_T = 1 - I_s \tag{2.8}$$

where

$$1/\sigma = t_g/\left(1 + s^2 t_g^2\right)^{1/2} \tag{2.9}$$

For $s = 0$ this reduces to $\sigma = 1/t_g$ *(i.e., the depth period is now tg and for large s: $\sigma = s$).* The divergence of kinematic theory for $s = 0$ is now removed. The depth variation of the transmitted and scattered beams, leading to thickness extinction contours, is represented schematically in Figure 26 for $s = 0$ as well as for $s \neq 0$.

29.2.2.4.4. Anomalous Absorption, Bormann Effect [151], [152]

Taking absorption into account gives more complex rocking curves of the type shown in Figure 27 for a foil of thickness $t = 3\,t\bar{g}$ and $\tau_g = 10\,t\bar{g}$. The curve for the diffracted beam (b) is symmetric, whereas the curve representing the transmitted beam (a) is asymmetric in s. For the same absolute value of s, the transmitted intensity is larger for $s > 0$ than for $s < 0$. This asymmetry is known as
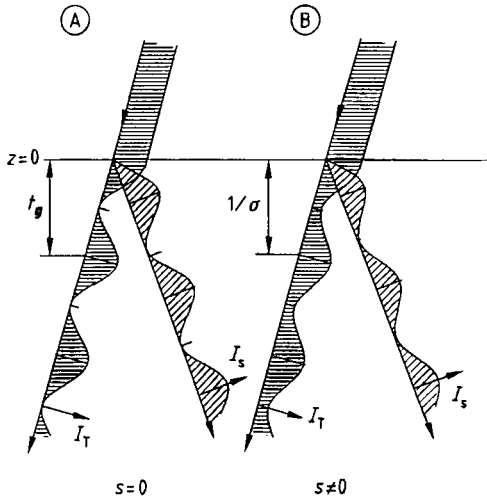
Figure 26. Interplay between incident and diffracted beam according to the dynamic theory
A) Bragg's condition is satisfied exactly, the depth period is t_g: I_S and I_T pass periodically through zero; B) $s \neq 0$ the depth period is now $1/\sigma_g$; I_T varies periodically but does not pass through zero; I_S and I_T are complementary since absorption was neglected

the Bormann effect; it was demonstrated initially for X-ray diffraction under dynamic conditions [152]. The steep change of I_T with s in the vicinity of $s = 0$ causes a large sensitivity of I_T to small changes in orientation.

29.2.2.4.5. Lattice Fringes [153]

The Electron Wave Function at the Exit Face. The amplitude of the transmitted wave for a unit incident wave can be represented by $T \exp 2\pi i \bar{K} \cdot \bar{r}$, where $\bar{K}$ is the wave vector of the incident wave inside the crystal (i.e., corrected for refraction at the vacuum – crystal interface). Similarly the diffracted wave can be represented by $S \exp 2\pi i (\bar{K} + \bar{g}) \cdot \bar{r}$, where $\bar{g}$ is the diffraction vector (i.e., the reciprocal lattice vector corresponding to the excited reflection). The wave function of the electrons emerging at the exit face of the foil is then

$$\psi = \left(T + S \exp 2\pi i \bar{g} \cdot \bar{r}\right) \exp 2\pi i \bar{K} \cdot \bar{r} \qquad (2.10)$$

The observed intensity

$$I_\psi \equiv \psi \psi^* = I_T$$
$$+ I_S + 2 \sqrt{I_T I_S} \sin (2\pi g x + \varphi) \qquad (2.11)$$
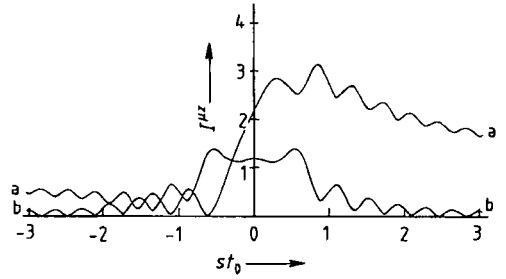
Figure 27. Rocking curves for transmitted (a) and scattered (b) beams according to the two-beam dynamic theory, with anomalous absorption taken into account

with

$$\tan \varphi = (s/\sigma) \tan \pi \sigma t; \qquad (2.12)$$

x is measured along the exit surface in the direction of $\bar{g}$; further, the following abbreviations are used: $I_T \equiv TT^*$; $I_S \equiv SS^*$. This expression represents sinusoidal fringes with a period $1/|\bar{g}|$; they can be considered as forming an image of the lattice planes $\bar{g}$.

29.2.2.4.6. Faulted Crystals

Planar Interfaces. Planar interfaces are imaged as fringes parallel to the foil surface; their characteristics depend on their geometric features. *Translation interfaces* (stacking faults, out-of-phase boundaries, etc.) separate two crystal parts related by a parallel translation $\bar{R}$ (Fig. 28 A), and *domain boundaries* separate domains that differ slightly in orientation (i.e., for which the excitation errors are different $\Delta s = s_1 - s_2$; (Fig. 28 B) [154].

Diffraction Equations for Faulted Crystals. The displacement over R_0 (R_0 = constant) of the exit part of a column with respect to the entrance part (Fig. 28 A), the two parts being separated by a fault plane, can be accounted for by noting that a supplementary phase change $\alpha = 2\pi \bar{g} \cdot \bar{R}$ occurs on diffraction but not on transmission (see Fig. 29). The path difference Δ between diffracted waves 1 and 2 is $\Delta = 2 R_0 \sin \theta_H$, where θ_H is the Bragg angle and hence $2 d_H \sin \theta_H = \lambda$ with $d_H = 1/|\bar{g}|$. This gives for the phase change $\alpha = (2\pi/\lambda) \Delta = 2\pi g R_0$, where R_0 is in general the component of $\bar{R}$ along $\bar{g}$ (i.e., $g R_0 = \bar{g} \cdot \bar{R}$).

The equations describing diffraction by the displaced exit part are obtained by the substitution
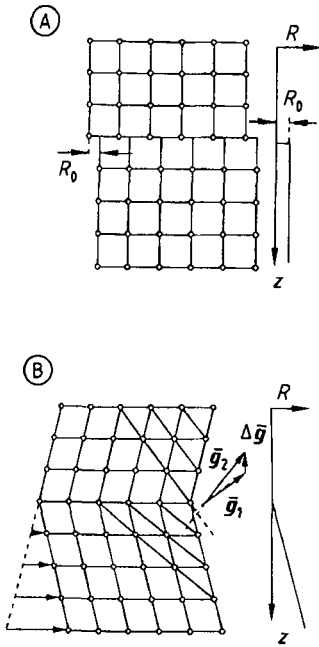
Figure 28. Schematic of two different types of interfaces
A) Translation interfaces with constant displacement vector
$\bar{R}$; B) Domain boundary between two regions for which the
excitation error is slightly different; the displacement R in-
creases with increasing distance from the interface (the ho-
mologous diffraction vectors $\bar{g}_1$ and $\bar{g}_2$ are slightly different:
$\Delta\bar{g} = \bar{g}_2 - \bar{g}_1$)

Figure 29. Diffraction by a faulted crystal
A) Effect of a translation interface with vector $\bar{R}_0$; B) Effect
of a displacement field $\bar{R}(\bar{r})$; C) Schematic of the expressions
for transmitted and scattered beam amplitude for a foil con-
taining a general interface parallel to the foil surfaces

$S \to S \exp -i\,\alpha$ in Equations (2.5). The amplitudes
T and S emerging from a column intersecting such
an interface are then obtained by integrating the
Equation (2.5) set along a column down to the
level of the interface and subsequently further
down to the exit face, using the adapted form of
Equation (2.5) (Fig. 29).

The effect on T and S of the presence of a
defect described by a displacement field $\bar{R}(\bar{r})$ (for
instance due to a dislocation) (Fig. 29) can be
taken into account in Equations (2.5) by replacing
the constant s by an effective local value s_{eff} which
now depends on z, i.e., on the depth along the
columns, and which is given by $s_{eff} = s + \bar{g}\,(d\bar{R}/dz)$.
This can be shown rigorously but it can also be
understood intuitively by the following geomet-
rical considerations (Fig. 29 B).

Let the diffraction vector of the considered set
of lattice planes be $\bar{g}$ at the level z below the
entrance face. At level $z + dz$ the displacement $\bar{R}$
has increased by $d\bar{R}$ and as a result the local vector
$\bar{g}$ is rotated over a small angle $\delta\theta \simeq \tan\theta = dR/dz$.
The change in s corresponding to this rotation of $\bar{g}$

over $\delta\theta$ is given by $\delta s = g\,\delta\theta$, i.e., by $g\,(dR/dz)$ or
somewhat more general by $d\,(\bar{g}\cdot\bar{R})/dz$, expressing
that only displacements which change the orienta-
tion of the vector $\bar{g}$ are operative.

In the particularly simple case of a domain
boundary as modeled in Figure 28 B, $R = kz$
($k =$ constant) in the exit part. Hence, at the level
of the domain boundary, s changes abruptly into
$s + k$ along the integration column.

**Diffraction by a Crystal Containing a Planar
Interface** [155]. The amplitudes of the transmitted
and scattered beams for a foil containing a planar
interface, with displacement vector $\bar{R}$ (Fig. 29 B),
can be formulated as

$$T(s_1,s_2,z_1,z_2,\alpha) = T_1(s_1,z_1)T_2(s_2,z_2)$$
$$+ S_1(s_1,z_1)S_2(-s_2,z_2)\exp i\,\alpha \qquad (2.13)$$
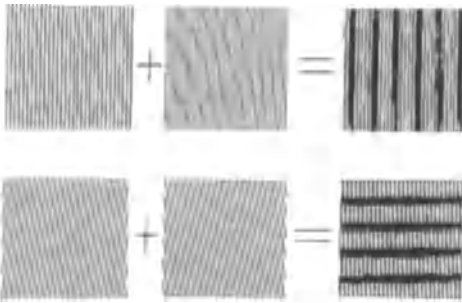
Figure 30. Optical analogue for the formation of moiré fringes in a sandwich crystal
(Upper) The two crystals are parallel, but the lattice parameters are slightly different; (Lower) The two crystal parts exhibit a slight orientation difference

$$S(s_1,s_2,z_1,z_2,\alpha) = T_1(s_1,z_1)S_2(s_2,z_2)$$
$$\cdot \exp(-i\alpha) + S_1(s_1,z_1)T_2(-s_2,z_2) \qquad (2.14)$$

with $\alpha = 2\pi g \cdot \bar{R}$. These relations can also be derived from the adapted set of Equation (2.5).

The first equation states that the amplitude of the transmitted beam T results from the interference between the doubly transmitted beam (i.e., the beam transmitted by the first part T_1 and subsequently transmitted by the second part: $T_1 \times T_2$) and the doubly scattered beam $S_1 \times S_2^{(-)} \exp i\alpha$. The factor $\exp i\alpha$ takes into account the phase shift on the amplitude S_2 scattered by the second part, resulting from the translation of the second part of the foil.

A similar interpretation can be given to the second expression. The amplitudes $T_1(s_1,z_1)$, $T_2^{(-)}$, $S_1^{(-)}$, $S_2^{(-)}$ are given by Equations (2.6 a) and (2.6 b). The indices 1 and 2 refer to the entrance and exit parts, respectively: the minus superscript means that in the corresponding Equations (2.6 a) and (2.6 b), s must be replaced by $-s$, because diffraction occurs from the negative side of the set of lattice planes (i.e., the diffraction vector is $-\bar{g}$).

If $s_1 = s_2$ but $\alpha \neq 0$, Equations (2.13) and (2.14) refer to a pure translation interface. If $s_1 \neq s_2$ and $\alpha = 0$, the expressions are applicable to a domain boundary. For mixed interfaces, $s_1 \neq s_2$ as well as $\alpha \neq 0$. If the interfaces are parallel to the foil surfaces, transmitted and scattered intensities are constant over the specimen area, but they vary in a pseudoperiodic fashion with the thicknesses z_1 and z_2 of entrance and exit parts, with $z_1 + z_2 = z_0$ (total thickness). The expressions describing this vari-

ation explicitly are obtained by substituting Equations (2.6 a) and (2.6 b) into (2.13) and (2.14).

If the interfaces are inclined with respect to the foil surfaces this depth variation is displayed as a pattern of fringes parallel to the foil surfaces in the area where the two wedge-shaped crystal parts overlap.

The fringe profiles of α-fringes ($s_1 = s_2$; $\alpha \neq 0$) and δ-fringes ($\Delta s = s_1 - s_2 \neq 0$; $\alpha = 0$) in bright field (BF) and dark field (DF), exhibit different properties. The symmetry properties allow the fringe patterns due to the two kinds of interfaces to be distinguished. The properties of α-fringes with $\alpha = \pi$ are singular [156].

29.2.2.4.7. Moiré Patterns [157] – [159].

The superposition of two identical crystal foils having a small orientation difference θ (rotation moiré) or of two parallel crystal foils with a small difference in lattice parameters (parallel moiré) gives rise to a fringe pattern, which can to a good approximation be considered as the coincidence pattern of the lattice fringes corresponding to the diffraction vectors $\bar{g}_1$ and $\bar{g}_2$ active in the two crystal foils.

The fringes are parallel to the average direction $\bar{g}$ of the two diffraction vectors $\bar{g}_1$ and $\bar{g}_2$ in the case of rotation moiré patterns and perpendicular to the common direction of the two diffraction vectors in the case of a parallel moiré. From the moiré spacing Δ, small differences can be deduced in the lattice parameters d_1 and d_2 of the two components of the sandwich since $\Delta = d_1 d_2/(d_1 - d_2)$, or the small orientation difference $\theta = |\Delta \bar{g}|/|\bar{g}|$ in the case of a rotation moiré. Figure 30 shows an optical analogue for the two types of moiré fringes produced by the superposition of two line patterns.

29.2.2.5. Operating Modes of the Electron Microscope [128], [130], [131]

29.2.2.5.1. Microscope Optics

The ray paths in an electron microscope are shown in Figure 31, according to the geometrical optics approximation, for the two main operating modes: high-magnification, high-resolution imaging, and selected area diffraction.

The microscope is essentially a three-lens system: an objective lens (c), an intermediate lens (f), and a projector lens (g). Each lens may be a composite lens. A movable selector aperture (b) is present in the image plane of the objective lens; a second aperture (d) is placed at the objective lens, close to the back focal plane. With the first
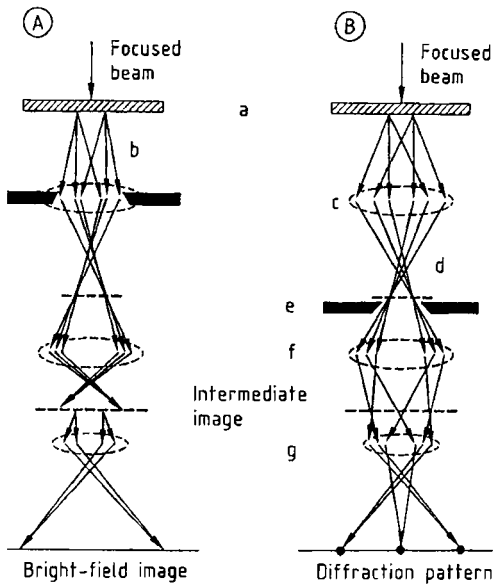
Figure 31. Schematic of the beam path in an electron microscope according to the geometrical optics approximation A) High-magnification, high-resolution mode; B) Selected area diffraction mode
a) Specimen; b) Objective aperture; c) Objective lens; d) Field-limiting aperture; e) Gaussian image plane; f) Intermediate lens; g) Projector lens

aperture a small area (≤ 1 μm) of the image (i.e., of the specimen) is selected, with the second aperture a single beam is selected or a number of the image-forming diffracted beams.

The characteristics of the objective lens are crucial because they determine to a large extent the image resolution and the contrast. The other two lenses mainly provide the desired magnification.

Although magnetic lenses normally rotate the image about the optical axis, in microscopes designed at the end of the 1980s, these rotations are compensated by a suitable device, and the image and the diffraction pattern have the same orientation.

29.2.2.5.2. High-Resolution, High-Magnification Mode

In the first (Fig. 31) imaging ray path the beam produced by a source (tungsten filament, LaB_6, field emission gun; see also Section 29.2.3.2.1) and collimated by a condenser lens system, is scattered by the object, and an image is formed in the image plane of the objective lens. By means

of the selector aperture, an area of the specimen is selected and magnified by the intermediate lens, which is focused on the image plane of the objective lens and provides a magnified image in the image plane of the intermediate lens. This image serves as the object for the projector lens, which forms the final image either on a fluorescent screen, on a photographic plate, or on an image intensifier followed by a TV camera. In the latter case—which is virtually a requirement for successful high-resolution work—the image can be viewed on a TV monitor and the final adjustments performed before the image is recorded photographically. Electronic recording on tape is of course possible directly via the TV signal; this mode is used mainly for so-called in situ dynamic studies.

29.2.2.5.3. Diffraction Mode (Figure 31 B)

When operating in the diffraction mode the intermediate lens (f) is weakened (i.e., its focal length is enlarged) so as to cause the back focal plane of the objective lens (c) to coincide with the object plane of the projector lens (g). This produces a magnified image of the diffraction pattern. In doing so, the selected area is not changed since only the intermediate lens current is changed; the diffraction pattern is thus representative of the selected area. The area selected in the diffraction mode is much larger than the field of view under high-resolution conditions.

As the electron beam passes through the specimen, various interactions occur next to diffraction: characteristic X rays are produced, the electrons suffer energy losses, etc. These effects can be used for analytical applications (see Section 29.2.2.12).

29.2.2.6. Selected-Area Electron Diffraction (SAED)

ED methods (SAED, CBED) are a useful complement to X-ray diffraction, especially when the material is available only as small microcrystals. Then only powder diffractometry can be applied when using X rays. However, even fine powders usually contain single crystals of a sufficient size to produce a single-crystal ED pattern that allows the approximate (but unambiguous) determination of the lattice parameters. These approximate lattice parameters in turn allow unambiguous indexing of the powder diffraction pattern and sub-
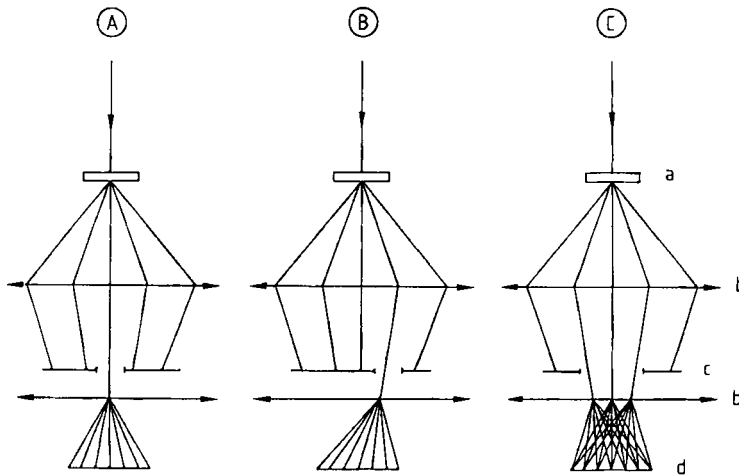
Figure 32. Schematic representation of different imaging modes
A) Bright-field diffraction contrast mode; B) Dark-field diffraction contrast mode; C) High-resolution imaging mode
a) Specimen; b) Lens; c) Aperture; d) Screen

sequently make possible a precise lattice parameter measurement. In addition, ED patterns often exhibit weak reflections due to superstructures, which are not visible in X-ray powder diffraction.

The spatial geometry of diffuse scattering can be reconstructed more easily from ED patterns than from X-ray diffraction patterns, since ED patterns are planar sections of reciprocal space, which is not the case for X rays.

29.2.2.7. Diffraction Contrast Images [128], [148]

29.2.2.7.1. Imaging Modes

Diffraction contrast images are usually obtained under two-beam conditions. By means of the aperture placed close to the back focal plane of the objective lens, either the transmitted or the diffracted beam is selected and the corresponding diffraction spot is highly magnified (Fig. 32). The image obtained in this way is a map of the intensity distribution in the selected diffraction spot. The image made in the transmitted beam is a *bright-field image*, whereas the image made in the diffracted beam is a *dark-field image*. Since the electrons are confined to narrow columns, the latter can be considered the picture elements of this image. The amplitude of the transmitted (or scattered beam) associated with a point at the back surface of the specimen is obtained by summing the contributions of the volume elements along a

column parallel to the incident beam centered on that point. The amplitude depends on how the excitation error *s* changes along the column, which in turn varies with changes in the orientation of the diffracting planes along the column. Strain fields cause such orientation changes and can thus be imaged in diffraction contrast. Diffraction contrast images do not reveal the crystal structure but are very sensitive to the presence of strain fields caused by defects such as dislocations.

29.2.2.7.2. Dislocation Contrast [125], [128]

Consider a foil (Fig. 33) containing an edge dislocation in E, with an orientation such that in the perfect parts of the foil the amplitude of the incident beam is approximately equally divided between the transmitted and the diffracted beams. To the left of the dislocation the considered lattice planes are locally inclined in such a way as to better satisfy Bragg's condition (i.e., *s* is smaller) than in the perfect parts; more electrons are diffracted than in the perfect part. A lack of intensity will then be noted left of the dislocation in the transmitted beam, leading to a dark line in the bright-field image. At the right of the dislocation, Bragg's law is less well satisfied than in the perfect part. This leads to a lack of intensity in the scattered beam and thus to a dark line in the dark-field image. The image is on one side of the dislocation: the side on which the dark line occurs in the bright-field image is called the *image side*. Ac-
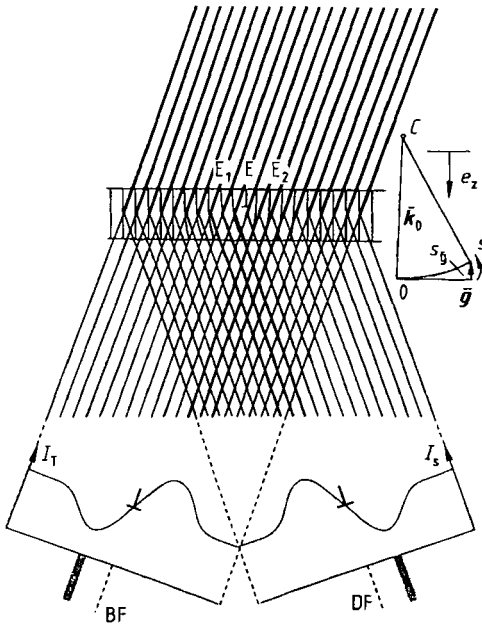
Figure 33. Schematic representation of the image formation at an edge dislocation in a thin foil
The line thickness is a measure of the beam intensity. The image profiles for bright-field (BF) and dark-field (DF) images are represented schematically along with diffraction conditions in the perfect part of the foil

Figure 34. Examples of computed images (right) and the corresponding observed images (left) of the same dislocation by using different diffraction vectors [162]

cording to this model the image side is where the lattice planes are tilted into the Bragg orientation.

The conditions for kinematic diffraction [160] are best approximated in the *weak-beam* method, which consists of making a dark-field image in a weakly excited diffraction spot. The dislocation image then consists of a narrow bright line on a darker background.

29.2.2.7.3. Extinction Conditions for Defects

Diffraction by a family of lattice planes that remain undeformed by the presence of the dislocation will not produce an image of the dislocation [128], [161].

Quantitative images based on dynamic theory can be computer simulated by using the methods developed in [162]. A comparison of simulated and observed images allows determination of all the relevant parameters of various dislocation types (Fig. 34). More details are given in [163].

29.2.2.7.4. Domain Textures

Phase transitions in crystals are usually accompanied by changes in symmetry, the space group of the low-temperature phase being a subgroup of the space group of the high-temperature phase. On cooling, the crystal breaks up into *orientation variants* (or twins) related by the lost symmetry elements of the high-temperature phase and *translation variants* (out-of-phase boundaries) related by the lattice translations lost on forming the superlattice [164]. The composite diffraction pattern of a domain texture is the superposition of the diffraction patterns due to individual domains. The diffraction pattern taken across a *reflection twin interface* contains one row of unsplit spots, which is perpendicular to the mirror plane. All other diffraction spots are split along a direction parallel to the unsplit row of spots [127].

In diffraction contrast images made in a cluster of *split diffraction spots*, orientation domains exhibit differences in brightness (i.e., domain contrast), especially for conditions close to $s = 0$ where the intensity variation with s is steep (Fig. 35). The presence of translation variants is not reflected in the diffraction pattern, except by some diffuse scattering.

Images made in *unsplit reflections* do not reveal brightness differences due to orientation differences of the lattice, but they may exhibit domain contrast due to differences in the structure factor. This is the case for Dauphiné twins in quartz, which are related by a 180° rotation about
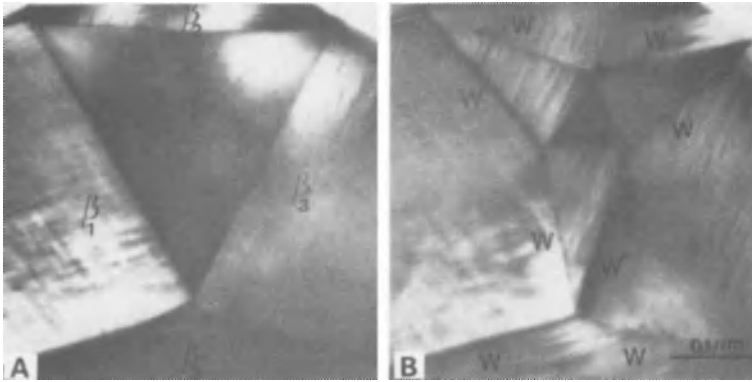
Figure 35. Domain texture in β-lead orthovanadate resulting from the $\gamma \rightarrow \beta$ transformation
A) The central triangular area still consists of the high-temperature γ-phase, whereas the other domains are in the low temperature β-phase; B) The same area as A) after additional cooling. The unstable triangular γ-area has been reduced further (Courtesy of C. Manolikas)

the threefold axis (Fig. 36). The two domains produce the same diffraction pattern since their lattices are common, but certain reflections have structure factors that differ in magnitude [165]. A dark-field image made in such a reflection will exhibit a different brightness in the two domains, which is called the structure factor contrast [128].

29.2.2.7.5. Interface Contrast and Domain Contrast

Interface contrast (i.e., an α-fringe pattern) is produced in the projected area of the interface even if the structure factors in the two domains have the same magnitude, but differ in phase by α; no domain contrast is produced under these conditions. Interface contrast is also produced at *orientation domain boundaries*; δ-fringes are produced in the projected area, since the s-values on both sides of the boundary are different in general [166]. Extinction of the δ-fringe pattern occurs if the s-values are the same for the two domains.

The presence or absence of domain contrast across a fringe pattern helps to distinguish the two types of interface. If the two domains on either side of the interface exhibit the same brightness for all reflections, the fringes must be due to a translation interface; they must exhibit α-character. Domains exhibiting a difference in brightness are separated by a domain boundary and the fringes are of the δ-type.

29.2.2.7.6. Strain Field Contrast

The strain field associated with precipitate particles in a matrix whose lattice parameters differ slightly from those of the precipitate can also be revealed as regions of different brightness [167]. For example, a spherical particle in an elastically isotropic matrix produces a spherically symmetric strain field. In certain parts of this strain field the vector $\bar{R}$ is locally perpendicular to $\bar{g}$; such areas will show up with the background brightness. In other areas, $\bar{g}$ is parallel to $\bar{R}$; such areas will show up with a brightness that is different from the background, either lighter or darker.

29.2.2.8. Convergent Beam Diffraction [132], [168]

29.2.2.8.1. Geometry of Convergent Beam Patterns

In the early 1980s, a class of applications based on the use of a convergent beam of electrons (convergent beam diffraction, CBD) was developed. Whereas parallel beams allow high spatial resolution, *high angular resolution* is more easily achievable with convergent beams. For these applications the electrons are incident along directions within a cone of revolution or along the surface of a cone (hollow cone method) having its apex in (or close to) the sample plane. Under these conditions, electrons are incident on the specimen under all possible directions within a
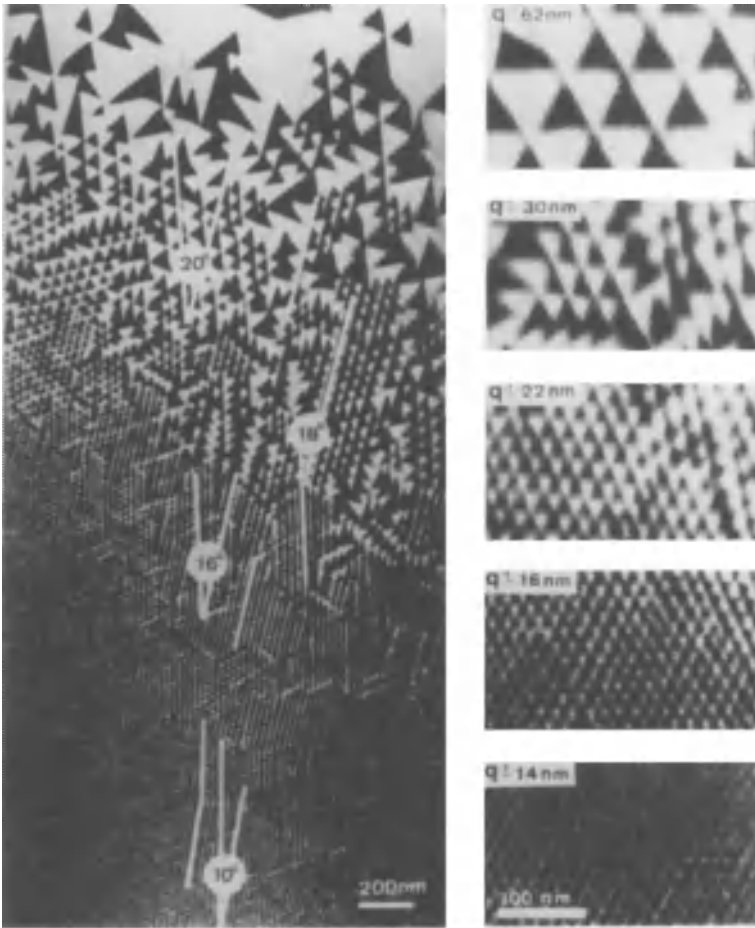
Figure 36. Dauphiné twin lattice in α-quartz close to the phase transition into the high-temperature β-phase viewed along the threefold axis

A temperature gradient is established across the specimen area. The α-phase is broken up in regular arrays of columnar domains whose size decreases with increasing temperature [165]

certain angular range, which should be of the order of the Bragg angles of the lowest-order reflections of the material under study. In practice, the semi-apex angle α varies from 2 to 10 mrad. Also, the corresponding diffracted beams for each recip-rocal lattice node form cones. The intersections of these cones with the photographic plate are circular disks whose radii are proportional to the semiapex angle of the cone of incident directions (Fig. 37). If the specimen is perfectly flat and de-fect free, these disks are images of the angular intensity distribution in each particular reflection.

29.2.2.8.2. Point Group Determination [169], [170]

When the incident beam is parallel to a zone axis, the symmetry of the intensity distribution in the complete disk pattern, including the fine lines in the 000 disk, reflects the point symmetry of the crystal along this zone axis (Fig. 38).

Several techniques can be used to determine the point group [169]. The method developed by BUXTON et al. [170] is based on the use of zone axis patterns and of dark-field diffraction patterns obtained under exact two-beam conditions. By
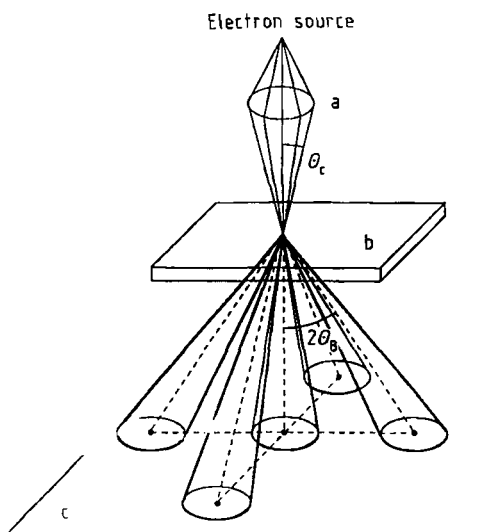
Figure 37. Production of a convergent beam electron diffraction pattern
a) Probe forming optics; b) Specimen; c) Film
The incident beam is convergent. Diffraction disks are centered on each diffraction spot.

Figure 39. Deficiency HOLZ (high-order Laue zone) lines in the central disk of a convergent beam diffraction pattern of silicon [168]

Figure 38. Zone axis convergent beam electron diffraction pattern revealing the presence of a sixfold rotation axis [168]

choosing the appropriate optical conditions, these disks do not overlap and the exact Bragg spots remain in the centers of the corresponding disks. The optics required for the application of these methods are described in [171]. With increasing cone angle the disks start to overlap and the nature of the pattern changes, but the symmetry elements are conserved.

A disk in the CBED pattern usually consists of broad fringes due to the dynamic interaction between beams belonging to the zero-order Laue zone (ZOLZ). Finer lines are formed by the interaction of high-order Laue zone reflections (HOLZ) with the zero-order beams (Fig. 39). The geometry of the HOLZ lines is determined by the accelerating potential and the lattice parameters; comparison of computer-simulated and observed patterns allows determination of one parameter if the other is known [171], [172].

29.2.2.8.3. Space Group Determination

Space group determination is based on the observation of the Gjønnes–Moodie lines [173] in certain "forbidden" diffraction disks. Reflections that are kinematically forbidden may appear as a result of double diffraction under multibeam dynamic diffraction conditions. If such forbidden spots are produced along pairs of different symmetry-related diffraction paths that are equally excited, the interfering beams may be exactly in antiphase for certain angles of incidence if the structure factors have opposite signs. Since the convergent beam disks are formed by beams with con-
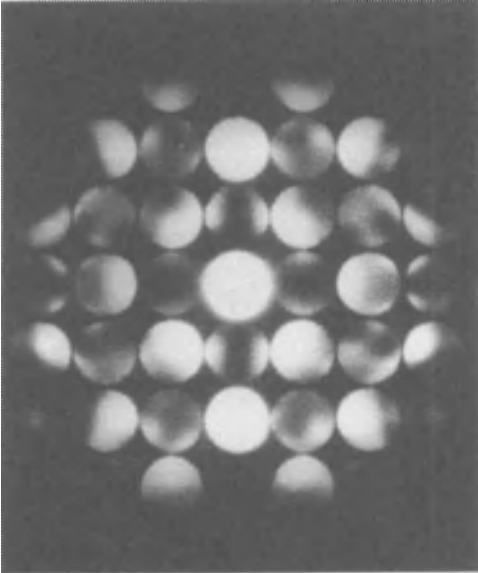
Figure 40. Convergent beam electron diffraction pattern exhibiting Gjønnes – Moodie lines due to the presence of glide mirror planes and twofold screw axis [168]

Figure 41. Image formation in an ideal microscope
a) Specimen; b) Objective lens; c) Back-focal plane, objective aperture; d) Image plane
The diffraction pattern $Q(u, v)$ is the Fourier transform of the object $q(x, y)$; the image is the inverse Fourier transform $\psi(x, y)$ of the diffraction pattern

tinuously varying directions of incidence, this condition will always be satisfied somewhere in the disk. The locus of points for which the phase difference is exactly π is a dark fringe along a diameter of the disk for which the Bragg condition is satisfied exactly (Gjønnes–Moodie lines, Fig. 40).

29.2.2.8.4. Foil Thickness Determination

From the dark-field two-beam rocking curve (Section 29.2.2.8.1) observed in the disk corresponding to the diffracted beam, the specimen thickness can be deduced with high accuracy provided the accelerating voltage and the interplanar spacing of the excited reflection are known [174].

29.2.2.9. High-Resolution Electron Microscopy

In recent years, high-resolution electron microscopy has become an important tool in the study of complicated crystal structures and their defects. Whereas detailed considerations of electron optics are of only marginal importance in the case of diffraction contrast images, they become essential for a discussion of the atomic resolution structure images produced by crystals.

29.2.2.9.1. Image Formation in an Ideal Microscope (Fig. 41)

Let the crystalline object be a thin foil characterized by a two-dimensional transmission function $q(x, y)$ that describes at each point of the exit surface of the specimen the amplitude and phase of the electron beams emerging from the column situated at (x, y) after dynamic diffraction in the foil. The diffraction pattern can be described to a good approximation as the Fourier transform $Q(u, v)$ of the object function $q(x, y)$. This diffraction pattern acts in turn as a source of Huyghens wavelets, which interfere to form the image, after linear magnification by the optical lens systems: the image is, in turn, the Fourier transform $\psi(x, y)$ of the diffraction pattern.

29.2.2.9.2. Image Formation in a Real Microscope

In real microscopes the situation is complicated by the presence of finite apertures and lens aberrations. The apertures truncate the Fourier
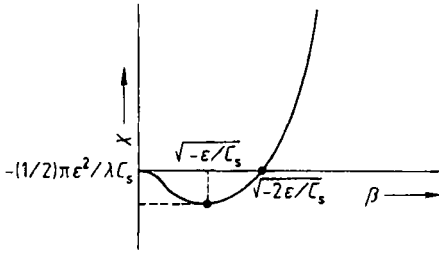
Figure 42. Dependence on β of the phase shift $\chi(\beta)$ of a beam enclosing an angle β with the optical axis of the microscope

transform by admitting only a finite number of beams to the image formation process. Lens aberrations depend on the angle β that a diffracted beam encloses with the optical axis of the microscope, hence they introduce angle-dependent phase shifts between diffracted beams. Moreover, they cause "blurring" of the image, a point of the object plane being represented as a disk in the image plane and vice versa.

Spherical Aberration. Spherical aberration causes electron beams enclosing an angle β with the optical axis to suffer a phase shift χ_S with respect to the central beam:

$$\chi_S = \frac{1}{2} \pi C_S \beta^4 / \lambda \qquad (2.15)$$

where C_S is the spherical aberration constant.

Defocus. Under exact Gaussian focusing conditions the contrast produced by an object that changes only the phase of the incident electron wave (i.e., a so-called phase grating) is minimal. This is the case for thin crystalline foils. Visual contrast improves when working in underfocused conditions. The phenomenon is in a sense similar to phase contrast in optical microscopy of phase objects. Defocusing causes a relative phase shift of the diffracted beams enclosing different angles with the central beam. The phase shift caused by a defocus distance ε with respect to the Gaussian focus is given by

$$\chi_D = \pi \varepsilon \beta^2 / \lambda \qquad (2.16)$$

The total phase shift caused by spherical aberration and defocusing is thus:

$$\chi(\beta) = \left(\frac{1}{2} \pi C_S \beta^4 + \pi \varepsilon \beta^2 \right) / \lambda \qquad (2.17)$$

$\chi(\beta)$ depends on β in the way shown in Figure 42, i.e., there is a flat minimum at $\beta_{min} = (-\varepsilon/C_S)^{1/2}$ and a zero at $\beta = (-2 \varepsilon/C_S)^{1/2}$.

Phase Grating Approximation [175]. The effect of this phase shift on the image parameter can be understood most easily by discussing the case of a pure phase grating, which is a reasonable model for a very thin crystalline foil. Such a specimen causes a phase shift χ of the electron wave on passing through the foil because the wavelength of the electron is different in vacuum and in the crystal. This phase shift in the point (x, y) can be written as $\sigma\varphi(x, y)$ where φ is the projected potential integrated over the sample thickness along the propagation direction of the electrons, and $\sigma = \pi/\lambda E$ (λ = wavelength of electron in vacuum; E = accelerating voltage). Ignoring absorption, the object function for a thin foil can be written as

$$q(x, y) = \exp i \sigma \varphi(x, y) \simeq 1 + i \sigma \varphi(x, y) \qquad (2.18)$$

Taking also into account the phase shift (Eq. 2.17) caused by the lens system on the diffracted beams, and neglecting absorption it can be shown by Fourier transformation [176] that the image intensity is finally given by

$$I(x, y) \simeq 1 - 2 \sigma \varphi(x, y) \qquad (2.19)$$

whereby it was assumed that $\sin \chi = -1$ ($\cos \chi = 0$). The image contrast, defined as $(I - I_0)/I_0 = -2 \sigma \varphi(x, y)$, where I_0 is the incident intensity, is thus directly proportional to the projected potential $\varphi(x, y)$ provided $\sin \chi = -1$.

Provided $\sin \chi = -1$, columns of large projected potential are imaged as dark dots. If $\sin \chi = +1$, the same columns would be imaged as bright dots.

Optimum Imaging Conditions. The most stringent requirement is $\sin \chi = -1$; this condition can only be met approximately and only in a limited β interval. The condition $\sin \chi = +1$ can be satisfied only for a limited number of discrete β-values.

The situation is somewhat comparable to positive and negative phase contrast in optical microscopy. The lens imperfections have been used to introduce a phase shift of $\pi/2$ in the same way as the quarter wavelength ring in the optical microscope. Only beams passing through the "window" in which the condition $\sin \chi = -1$ is met, interfere with the required phase relationship that causes the image to represent maxima in projected potential as dark areas.

From Equation (2.17) it is clear that $\chi(\beta)$ depends on the parameters C_S, ε, and λ. The condition $\sin \chi = -1$ can thus be met by adjusting one or several of these parameters. For a given instrument C_S and λ are fixed and the observer can meet this condition by optimizing the defocus ε. From Equation (2.17) it follows that χ adopts the essentially negative value $\chi = -\pi \varepsilon^2/(2 \lambda C_S)$ for $\beta = (-2 \varepsilon/C_S)^{1/2}$. $\sin \chi$ will be -1 if $\pi \varepsilon^2/(2 \lambda C_S) = \pi/2$, i.e., for

$$\varepsilon_S = - (\lambda C_S)^{1/2} . \qquad (2.20)$$

For the defocus $\varepsilon = \varepsilon_S$ the curve $\sin \chi$ versus β will exhibit a rather flat part in the region around

$$\beta = (- \varepsilon/C_S)^{1/2} \qquad (2.21)$$

where $\sin \chi = -1$. The underfocus value ε_S corresponding to this optimum is called Scherzer defocus [177]. Most high resolution images are made under this defocus condition and in the thinnest part of the sample.

29.2.2.9.3. Resolution Limiting Factors

Aperture. The presence of an aperture admitting only beams that enclose an angle with the optical axis not exceeding β_A imposes a resolution limit, called the Abbe limit [178]. A point in object space is imaged as a circle with a radius

$$\varrho_A = 0.61 \lambda / \beta_A$$

Chromatic Aberration. Instabilities ΔE in the high voltage E of the microscope cause a spread in the wavelength of the incident electron, which blurs the image. Variation ΔI in the lens current I similarly causes a spread in the focal length of the lenses, which also contributes to such blurring. Furthermore, the inelastic scattering of electrons in the specimen causes slight changes in the wavelength of electrons emerging from the specimen. The net result of these different phenomena can be described as an effective spread in the focal length of the objective lens:

$$\Delta f = C_C \left[(\Delta E/E)^2 + 4 (\Delta I/I)^2 \right]^{1/2} \tag{2.22}$$

where C_C is the chromatic aberration constant; the corresponding disk of least confusion in object space is $\varrho_C = \beta \Delta f$. In concrete cases, $\Delta E/E \leq 10^{-6}$; $\Delta F \approx 10$ mm, $C_C \approx 10$ mm.

Astigmatism. Ideal lenses should have cylindrical symmetry. However, in reality, slight deviations may occur, leading to a dependence of the focal length on the azimuth of the ray path considered. Astigmatism can occur for various reasons, including inhomogeneities in the polepiece material, asymmetry in the windings, or a dirty aperture. As a result of astigmatism a point source is imaged as two different line foci at right angles to each other. At exact focus the image deformation disappears but the image is still blurred. Astigmatism can be corrected by a stigmator—a device (usually an octopole) that makes the lens appear perfectly cylindrical—by applying weak additional magnetic fields. Careful correction of astigmatism is essential for the production of high-quality, high-resolution images.

Beam Divergence. The intense illumination required for high-resolution imaging is obtained by a condenser lens system that produces a slightly conical beam with a typical apex angle of the order of 10^{-3} rad. This also leads to some image blurring since the final image is the superposition of images corresponding to the different directions of incidence.

Mechanical Instability. The obtainable resolution depends not only on the lens characteristics but also to a large extent on the mechanical stability of the microscope, which should be installed on a vibration-free heavy concrete block. In many cases the mechanical stability is actually the resolution-determining factor. The specimen holder should moreover be creep free, which is a problem for hot and cold stages. High-resolution work is therefore performed almost exclusively at room temperature.

Ultimate Resolution. The final disk of confusion due to the different nonmechanical blurring effects is given approximately by

$$\varrho = \left(\varrho_A^2 + \varrho_S^2 + \varrho_C^2 \right)^{1/2} \tag{2.23}$$

which depends on β. At low angles the aperture effect is usually dominant, whereas at high angles the spherical aberration is the limiting factor (for 100-kV microscopes, $C_S \approx 8.2$ mm and $C_C \approx 3.9$ mm). The optimum radius of least confusion is achieved for a certain β-value and the radius of confusion corresponding to that optimum is

$$\varrho_0 = 0.9 \lambda^{3/4} C_S^{1/2} \tag{2.24}$$

Therefore, a gain in resolution can be achieved by reducing C_S and by reducing λ (i.e., using a higher voltage). At present, many microscopes used in material science and in solid-state chemistry operate at 200–400 kV.

At higher voltages, radiation damage to the specimen severely limits observation time. Advanced designs aim at reducing C_S to allow a decrease of the high voltage for a given resolution. Practical resolution limits are at present ca. 0.16 nm, but new developments tend to peak this value to below 0.10 nm.

Directly interpretable high-resolution images (i.e., having a direct relationship to either the projected lattice potential or the electron density) are formed by the interference between beams passing through the "window" in the image transfer function (ITF). As a result, only low-order reflections (i.e., low-order Fourier components of the lattice potential) generally contribute to the image, which limits the detail that can be resolved (irrespective of the point resolution) since fine details are carried by the high-order Fourier components. However, information transmitted beyond the window (i.e., by the rapidly oscillating part of the image transfer function) can also be used. This information can be extracted by computational techniques, for instance, by using as input a set of images made at a series of closely spaced defocus values [179].

29.2.2.9.4. Image Formation Models [180]

Fourier Model. High-resolution image formation is considered most conveniently as occurring in two steps. In the first step, electrons propagate through the crystal and produce a two-dimensional periodic electron distribution at the exit face, which images the projected lattice potential or the projected electron density. In a second step, this exit face acts as a planar distribution of point sources of spherical electron waves, which interfere behind the foil and form the diffracted beams that move in the lens system of the
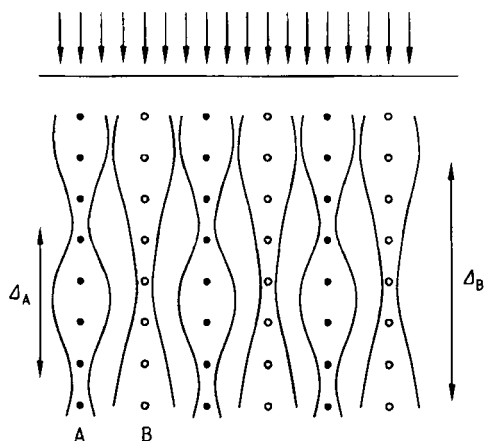
Figure 43. Channeling of electrons along columns of atoms parallel to the incident beam
Alternating focusing and defocusing occurs. The focusing distances Δ_A and Δ_B are different for A and B columns

Figure 44. High-resolution image of the high-temperature superconductor $Y_2Ba_4Cu_7O_{15-\delta}$. The bright dots reveal the heavy atom columns as projected along two different zones. Simulated images are reproduced as insets. The projected unit cell is outlined (Courtesy G. Van Tendeloo). The images were made at optimum defocus. The two images refer to different specimens, with comparable thickness.

microscope. Interference of these beams produces the final image, which will be directly interpretable only if the beams interfere with the correct phases.

The resolution of the image depends on the spatial frequencies of the highest-order beams admitted by the selector aperture that still contribute to the image with the correct phase. Only beams passing through the window in the ITF contribute to a directly interpretable image.

The image will exhibit more detail as the order of included reflections increases, but it will give a faithful representation of the structure only if the Fourier components have correct relative phases; this puts a practical limit on the number of useful reflections.

Complete software packages for the simulation of high-resolution images are available commercially (see, e.g., [181]).

Channeling Model. An alternative model that emphasizes the particle nature of electrons provides a simple intuitive picture [182]. An atomic column parallel to the incident beam is a cylindrically symmetrical potential well, which acts on the moving incident electrons as a succession of alternating convergent and divergent lenses. At the entrance face of the foil the incident electron distribution is uniform, but in the first part of the foil the electrons are attracted toward the core of the atom columns and focusing occurs, transforming part of the potential energy of the electrons into kinetic energy. Subsequently the electrons repel one another and defocusing occurs. The focusing – defocusing motion occurs periodically as a function of the depth in the foil, the depth period being a function of the atomic number of the atoms in the column. The heavier the atoms are, the smaller is the depth period. As a result, for a foil of given thickness, one type of atom column will give rise to a relative maximum in the electron distribution, whereas another type may give rise to a relative minimum. These extremes are not necessarily imaged as extremes of the same nature. A maximum may be imaged either as a bright dot or as a dark dot, depending on the defocus and on the foil thickness (see Fig. 43). This model explains why different atom columns produce dots of differ-

ent brightness. The focusing and defocusing behavior was confirmed by computer simulations, which also showed that the depth periods are in the 4 – 10-nm range depending on the atomic number.

29.2.2.9.5. Image Interpretation [180]

Trial and Error. Digital simulation programs for HRTEM are usually based on the Fourier approach. They allow one to compute the image of a given structure projected along a given zone axis as a function of the foil thickness and of the imaging conditions (defocus) for a microscope with known characteristics (spherical aberration coefficient, accelerating voltage, beam convergence) (Fig. 44).

Identification of a structure or of a defect in a structure proceeds by "trial and error". A model is proposed, and a matrix of images that varies the two main independent variables (foil thickness and defocus) is computed.

These theoretical images are then compared with the observed images at the different thicknesses along a wedge-shaped part of the foil. A solution is considered acceptable when the corre-
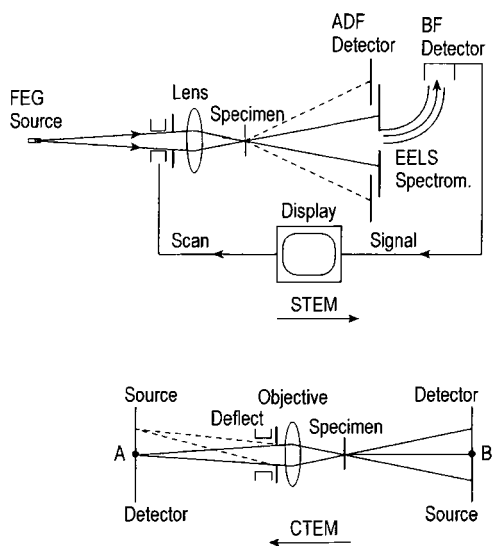
Figure 45. Scanning transmission electron microscopy
Top: Schematic representation of the essential components of a scanning transmission electron microscope (ADF: annular dark field; BF: bright field; FEG: field image gun)
Bottom: Illustrating the reciprocity relationship between the beampaths in STEM and CTEM microscopes; the electrons move in opposite senses (courtesy of J. Cowley)

spondence between observed and computed images is judged visually to be "good enough" for all available thicknesses at a constant defocus. This comparison allows one at the same time to estimate the foil thickness and the defocus; these quantities are usually not known a priori. Numerical criteria to quantify the goodness of fit have been proposed but have seldom been applied as yet. In recent developments, high-resolution images have been used to determine the chemical composition along individual columns in rather special circumstances, such as along the interfaces in synthetic layer structures of semiconductors grown by molecular beam epitaxy. The composition profile across an interface can be obtained at an atomic scale. The application of these methods requires a certain amount of a priori knowledge concerning the structure.

Direct Retrieval. Recently, "direct retrieval" methods have been developed which use as the input a series of images taken at closely spaced defocus values (focus variation method). If the microscope parameters describing the transfer function and its inverse are known, the projected

wavefunction at the exit plane of the foil can be reconstructed. From this, the projected structure (the object) can be retrieved by using an analytical formulation of the channeling model. Knowing the projected structure along more than one zone axis allows one to reconstruct the three-dimensional structure. Less a priori knowledge is required than for the methods based on "trial an error".

29.2.2.10. Scanning Transmission Microscopy [183]

Transmission electron microscopy can also be performed with a scanning incident beam (STEM). In a STEM instrument a fine convergent electron probe, formed by demagnifying a small, but brilliant, electron source, is scanned over the specimen area of interest. The incident beam is focused on the specimen plane and a convergent electron diffraction pattern is formed in a plane behind the specimen. Parts of this diffraction pattern (CBED pattern) can be selected by an appropriate aperture, and the signal detected hereby gives rise to an electronic signal which is displayed on a TV monitor, the scan of which is operated synchronously with the probe scan. A bright field (BF) image is obtained when the directly transmitted beam is selected. If one or several beams outside the central beam are selected, a dark field (DF) image is produced.

High resolution is achieved by making the effective probe size as small as possible. For this purpose use is made of a field emission gun with an effective source diameter on the order of 5 nm; demagnification then allows a probe size on the order of 1 nm in diameter to be obtained. The beam current is of the order of 0.5 nA.

The essential electron optics of STEM and CTEM instruments are in principle not very different, even though the electrons travel in opposite senses in the two instruments. A comparison of the principles of STEM and CTEM is represented schematically in Figure 45. In CTEM the image signal is produced in parallel (i.e., simultaneously) in all parts of the image plane (e.g., a photographic plate), whereas in STEM the signal is generated in serial form, as a time-dependent electric signal, which makes it easy to apply on-line image processing. Like in CTEM various signals, excited by the passage of the electron beam through the specimen, can be displayed by using the appropriate detectors.

Electron
beam

X–
Rays
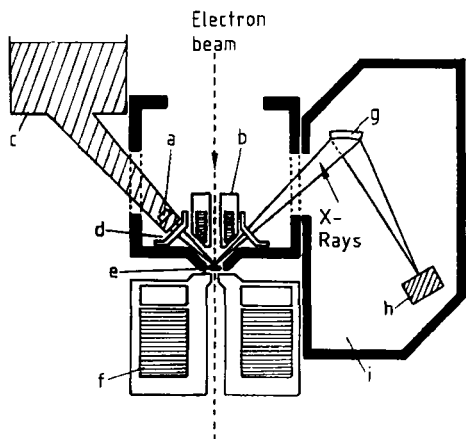
Figure 46. Specimen – detector configuration used in X-ray microanalysis
Left: Energy-dispersive X-ray system (EDX) with Si(Li) detector; Right: Wavelength-dispersive spectrometer (WDS) with bent crystal
a) Si(Li) detector; b) Mini lens; c) Liquid-nitrogen reservoir; d) X-ray window; e) Specimen; f) Objective lens; g) Crystal; h) Detector; i) Crystal spectrometer

29.2.2.11. Z-Contrast Images [184]

The commonly used imaging modes (CTEM, HRTEM) are based on interference and diffraction and rely strongly on coherently scattered electrons. However, simultaneously with the coherent scattering, incoherently scattered electrons also are produced by thermal diffuse scattering and in particular by Rutherford scattering. In the STEM mode these incoherently scattered electrons can be used to image atom columns, provided that the proper electron optics are available. A fine electron probe is obtained by focusing the incident convergent electron beam in the sample and scanning over the foil. Thereby, a significant fraction of the incoherently scattered electrons, emerging from the sample at relatively large scattering angles, is captured in an annular aperture and detected.

in the case of incoherent imaging conditions, the contrast transfer function is a monotonously decreasing function of the spatial frequency, whereas under coherent imaging conditions it is a rapidly oscillating function. This has important consequences. In the coherent case the brightness of a dot imaging a given atom column may change from bright to dark and vice versa as a function of defocus. In contrast, in the incoherent case, the relative brightness of a dot remains consistently

the same, independent of defocus, i.e., there is no contrast reversal. Moreover, the dot brightness increases monotonously with increasing average Z value of the atoms in the column as a consequence of the contribution of the Rutherford scattering to the incoherently scattered electrons. Incoherent images can therefore be interpreted on an intuitive basis even for relatively large thickness. It can be shown that structure retrieval is in principle possible and simpler than in the case of coherent imaging. The method is therefore well suited to the study of geometric defects in crystals.

29.2.2.12. Analytical Methods

29.2.2.12.1. X-Ray Microanalysis [134]

Principle. An important feature of electron microscopy is the possibility of performing in the same instrument both a qualitative and a quantitative elemental analysis of the same small crystal fragment that produced the electron diffraction pattern and the image. The most commonly used method consists in principle of detecting the characteristic X rays excited by the high-energy electrons passing through the specimen.

For this purpose, an X-ray spectrometer, mounted sideways on the microscope column at the level of the specimen holder, captures the X rays emitted in all directions by the specimen within a certain solid angle centered around the chosen takeoff direction and transmitted through an X-ray transparent window (Fig. 46). The spectrum consists of the characteristic X-ray lines superposed on a continuous background of bremsstrahlung (see also → Surface Analysis, → Surface Analysis).

The physical basis of X-ray microanalysis is Moseley's law [185], which establishes a direct and systematic relation between the atomic number Z and the characteristic X-ray wavelength λ in a family of X-ray lines (e.g., associated with similar transitions in different elements):

$$\lambda = A/(Z - C)^2 \qquad (2.25)$$

A and C depend on the considered family; K_α, K_β, ..., L_α, ..., X-ray lines are produced when an inner shell (e.g., K-shell) electron has been ejected from the atom by absorbing an amount of energy from an incident electron, exceeding the critical ionization energy associated with that shell. On deexcitation, this vacant level is filled by an electron originating from a higher-lying shell or

subshell of the same atom. The transition is accompanied either by the emission of X rays, which is the phenomenon of interest here, or by the ejection of an Auger electron ($\rightarrow$ Surface Analysis). Equation (2.25) shows that the atomic number can in principle be determined by measuring the characteristic X-ray wavelength.

Wavelength-Dispersive Spectrometry (WDS). Two types of X-ray spectrometers are in use: wavelength-dispersive spectrometers and energy-dispersive spectrometers.

In wavelength-dispersive spectrometers, dispersion is achieved by an analyzing crystal. Since the Bragg angle for a given set of lattice planes depends on the wavelength of the incident radiation, X rays of different wavelengths can be separated spatially by the crystal, and a spectrum can thus be analyzed by changing the orientation of the analyzing crystal by θ and detecting the amount of radiation passing through a slit whose angular position is rotated simultaneously over $2\,\theta$. Often the analyzing crystal, which acts as a grating, is cylindrically curved so that focusing along a line occurs. The detector, usually a gas proportional counter, is placed in the focus and connected to a single channel analyzer, followed by standard counting electronics.

Depending on the wavelength range to be detected, different analyzing crystals are necessary; provision is usually made on the equipment to allow switching in several crystals from a built-in carousel. Extension of this method to long-wavelength X rays is hampered by the absence of suitable crystals with large interplanar spacings. Synthetic layered structures, prepared by molecular beam epitaxy, and Langmuir–Blodgett layers have been used for this purpose since the end of the 1980s.

Energy-Dispersive Spectrometer (EDS). In energy-dispersive spectrometers, spectral analysis is based on measuring the X-ray energy rather than the wavelength. As a detector, a solid-state ionization chamber is used (e.g., a $p-i-n$ junction in silicon, an intrinsic germanium crystal, or a mercuric iodide crystal). The energy resolution is of the order of 150 eV or less for X-ray energies in the range $1-12$ keV (Fig. 47). The silicon $p-i-n$ junction is produced by diffusing lithium (at elevated temperature) into a p-type silicon slab under a reverse bias, creating an intrinsic region via the exact compensation of fixed acceptors by mobile lithium donors, as a result of electromigra-

tion of the lithium ions. Under reverse bias an electric field is created in the broad intrinsic region, which acts as the sensitive volume. Such detectors must be kept permanently at liquid-nitrogen temperature since lithium ions migrate at room temperature. When an X-ray photon passes through the intrinsic region, its energy is dissipated almost entirely by the formation of electron–hole pairs, which are swept away by the reverse bias field and produce a voltage pulse across the detector crystal that is proportional to the energy of the incoming X-ray photon. The X-ray spectrum is then established by analysis of the pulse heights with a multichannel analyzer. For qualitative analysis, only the line positions are important, but for quantitative analysis the line shape and the peak heights are also used. Fundamental for the interpretation of EDS spectra is the assumption that all the energy of the X-ray photon is deposited in the detector (silicon escape peaks are ignored). The number of charges n created by a photon with energy E is then given by $n = E/E_i$, where $E_i = 3.6$ eV for silicon. Since the charge pulses to be detected are small, reducing the noise of the detector is essential; this is a second reason for cooling the detector. EDS has many advantages over WDS, the main ones being the much shorter acquisition time and the smaller minimum useful probe size. On the other hand, complications also occur that do not arise in WDS instruments; the major ones are escape peaks, pulse pileups, electron beam scattering, peak overlap, and window absorption effects. In practice, the latter prevents detection of light elements with $Z \le 10$ (beryllium window) unless a windowless detector is used.

Quantitative Analysis. Quantitative analysis is based on the assumption that the intensity of the Lorentzian X-ray peak is proportional to the number of atoms of the species responsible for the peak. Since the theoretical relations are complicated and involve a large number of parameters, some of which are difficult to evaluate, well-characterized homogeneous standards are generally used for calibration, the composition of which has been determined by classical chemical macroscopic methods. These standards, which should also have similar shapes to the sample, are then measured in EDS equipment under the same well-defined and reproducible conditions as the unknown sample, and the spectra are compared. This calibration leads to the so-called k_i factors (coef-
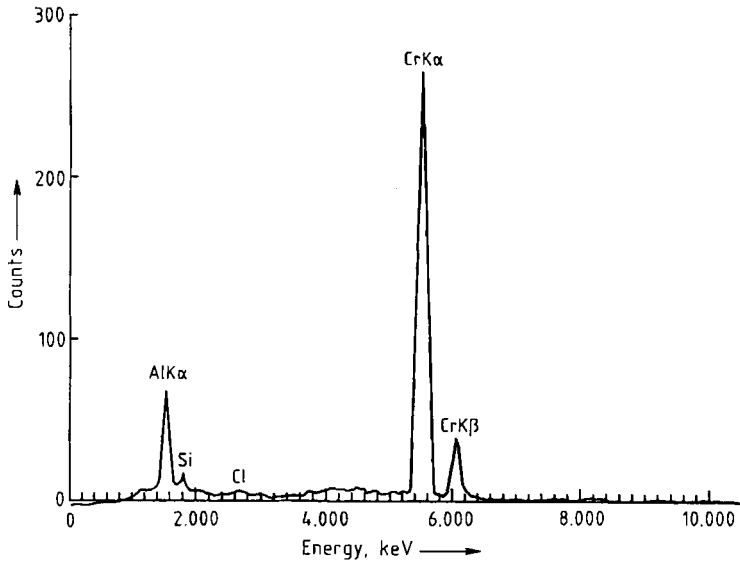
Figure 47. Example of X-ray spectrum obtained by means of an energy dispersive spectrometer. The material is a chromium aluminum alloy.

ficients of proportionality) for the different elements [186]:

$$C_{i_{unknown}}/C_{i_{standard}} = I_{i_{unknown}}/I_{i_{standard}} = k_i \qquad (2.26)$$

where C_i are weight fractions.

Since k_i may depend on the particular matrices, Equation (2.26) must be corrected as follows [134], [187]:

$$C_{i_{unknown}}/C_{i_{standard}} = Z_i \times A_i \times F_i \qquad (2.27)$$

where the matrix correction factors Z_i, A_i, and F_i still depend on the element under consideration; Z_i is the factor due to atomic number; A_i is due to X-ray absorption; and F_i to X-ray fluorescence. Other correction factors are due to the specimen geometry because X-ray production is depth dependent since it is proportional to the local electron energy. In the thin foils used in TEM the ZAF correction factors can often be neglected. X-ray microanalysis allows quantities of material of the order of 10^{-11} g to be analyzed; the concentration can be as low as 0.01 %, and the precision is of the order of 1 %.

For a more detailed discussion of the method and its technicalities, see [186]–[188].

Special Effects [189]. For foil orientations close to an exact Bragg orientation, the transmitted and scattered wave fields both propagate along the lattice planes. In a model

crystal with a primitive structure, one wavefield has its maximum amplitude along planes coinciding with the atomic planes; the other has its maximum along the planes midway between atomic planes. For $s > 0$, the latter wavefield is excited, and therefore easy transmission occurs since the electrons in this wavefield avoid the atom cores (Bormann effect). For $s < 0$, the former wavefield is excited, but strongly absorbed, since the electrons pass close to the atom cores and can excite X-ray emission. X-ray emission is thus expected to be enhanced on the $s < 0$ side of a bent contour ($s = 0$).

In a crystal whose superstructure consists of alternating A and B atomic layers parallel to the considered lattice planes, which are close to the exact Bragg position, one type of wavefield is peaked along the planes of A atoms, whereas the other is peaked along the B-atom planes. Depending on the sign of s, one of the wavefields is enhanced relative to the other; hence the A and B atomic layers will be excited unequally. Under such conditions, X-ray microanalysis may produce erroneous results for the A/B ratio. Therefore, foil orientations should be avoided that cause an extinction contour to pass through the selected area due to anomalous transmission. On the other hand, this effect has been exploited to locate the site occupied by a third atomic species, in an AB superstructure, using a technique called ALCHEMI (atom location by channeling-enhanced microanalysis) [189].

29.2.2.12.2. Electron Energy Loss Spectrometry (EELS) (→ Surface Analysis) [190]

Principles [134], [191]. Apart from elastic scattering, which is responsible for electron diffraction, inelastic scattering events also occur as electrons pass through the foil. Inelastic processes can be caused by (1) single electron excitations,
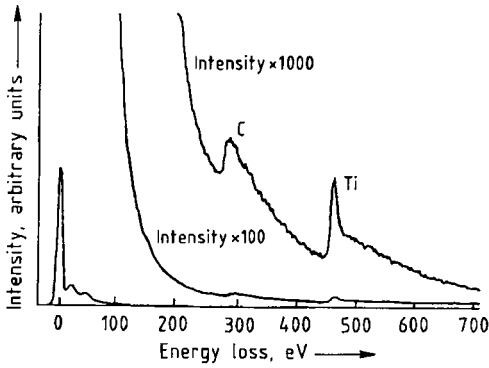
Figure 48. Example of EELS (electron energy loss spectrometer) spectrum of TiC
On the low-energy side the plasmon peak is visible. The fine structure on the high-energy end is caused by interference of waves reflected by the neighboring atoms

such as X-ray and Auger electron production; and (2) collective excitations, such as volume and surface plasma oscillations (plasmons) and phonons. Collective excitation can be revealed indirectly as characteristic energy losses of the incident electrons. The energy of elementary excitations (plasmons and phonons) is low compared to the incident electron energy (100–140 kV).

Plasmons have energies in the range of 10–20 eV, whereas phonon energies are of the order of 10 meV. Individual quantized plasmons are thus much easier to detect than phonons. Since the positions of the plasmon loss peaks are characteristic of the materials, they can be used as analytical tools for aluminum and magnesium, for example.

However, for chemical analysis, the absorption edges in EELS curves are more important (Fig. 48); they reflect the absorption phenomena leading to X-ray production and exhibit a fine structure on the high-energy side. This fine structure is referred to as EXELFS (extended energy loss fine structure) [192], which is the analogue of EXAFS (extended X-ray absorption fine structure). The steep rise on the low-energy side of the absorption edge is due to the excitation of inner-shell electrons and characterizes the element.

The fine structure is produced by the electron wave originating from an inner shell that is partially back-reflected by the surrounding atoms, which leads to a modulation of the excitation probability of inner-shell electrons. This fine structure can therefore provide information not only on the chemical nature of the absorbing atom, but also on the number of nearest neighbors and their distance.

As an analytical tool, EELS is in a sense complementary to X-ray microanalysis and makes use of the same type of phenomena. EELS works well for light elements, whereas X-ray microanalysis is much more difficult to apply to elements with $Z \leq 11$.

Spectrometers. The simplest system uses a magnetic prism to produce spatial separation of electrons with different energies. The spectrum is scanned by means of a slit, and a phosphor coupled to a photomultiplier is used as the detection system.

One of the most advanced types of equipment is the ω electron spectrometer [193]. It consists of four magnetic prisms that bend the incident beam back to its original direction; it can be mounted on an electron microscope column behind the specimen where inelastic scattering is produced. Slits allow a certain energy range to be selected from the spectrum. The spectrometer can thus also act as a filter and can produce, e.g., zero-loss images in which the blurring due to inelastic scattering is eliminated.

Since the 1980s *X-ray microanalysis* based on energy-dispersive spectrometers has become a routine technique. Equipment as well as data processing hardware and software are available commercially as optional accessories for most microscopes. In the late 1990s the range of applications of EELS spectroscopy was considerably expanded and extended to nanoscale samples. The commercial availability of user-friendly equipment compatible with most transmission electron microscopes has greatly stimulated its use. Also the improvements in the interpretation methods and the development of the corresponding computer software greatly contributed to the expansion of its use. It is now possible to map the spatial distribution of selected features of the information contained in the EELS spectrum. In this way spatial distribution of different elements (also light elements) can be mapped with nanometer resolution, but for instance also the spatial distribution of different ionization states of the same element can be visualized. These advances were made possible by the use of arrays of diodes that allow parallel detection instead of the initially used serial detection.

29.2.2.13. Specimen Preparation

The objective of all specimen preparation methods is to obtain a crystal fragment that is sufficiently thin to transmit electrons with energies of 100 keV or more, and has a sufficiently large lateral size. The numerous methods available depend strongly on the type of material and observation method.

29.2.2.13.1. Diffraction Contrast Specimens

Specimens should have a thickness of a few hundred nanometers. If too thick, the foil is not transparent; if too thin, achieving a two-beam orientation is difficult; moreover, the sample is not representative of bulk material for a number of applications.

Electrically conducting bulk samples (metals and alloys) can generally be thinned by electropolishing; *bulk ceramic samples* are often chemically thinned and polished. The resulting sample is usually perforated in the center; the edges of the hole are then wedge shaped and have transparent edges.

Insulating thin foils may become charged and deflect the electron beam; this can be avoided by coating the specimen with a thin layer of amorphous carbon, which is electrically conducting.

Thin films of uniform thickness can be obtained by *vapor deposition* under vacuum or by *chemical deposition* techniques; such samples usually have microstructures that are different from those of samples prepared from bulk material.

Layered crystals exhibiting a pronounced cleavage plane, such as graphite or MoS_2, can be thinned by repeated cleavage to produce rather uniformly thin specimens.

A truly universal thinning method does not yet exist. However, *ion beam bombardment under grazing incidence* is applicable to many materials, in particular to brittle ones such as most semiconductors. In the first step, these materials are mechanically polished to produce slabs with dimples; subsequent ion beam thinning under grazing incidence then produces a hole whose edges are wedge shaped. Thinning can be achieved either from one side only or from both sides of the slabs; this is important, for instance, for quantum wells grown by molecular beam epitaxy. Surface damage left after ion bombardment can often be removed by a final chemical polish. Figure 49 sche-matically represents thinning methods used for semiconductor specimens.

29.2.2.13.2. High-Resolution Specimens

The various methods described above can also of course be used to prepare specimens for use in high-resolution imaging. However much thinner specimens (1 – 10 nm) are required than for diffraction contrast. Nevertheless, more rudimentary methods can often be used successfully, since much smaller lateral dimensions of the specimen can be tolerated. Specimens of brittle materials (e.g., oxide superconductors) are in many cases obtained simply by crushing the material in a quartz mortar under a protecting organic liquid. In such a way the crystal fragments are dispersed in a suitable liquid, and a drop of the suspension is deposited on a grid covered with a holey carbon film and allowed to dry. The edges of fragments protruding over the holes are potentially suitable specimens, but actually only a small fraction of the crystal fragments are sufficiently thin to produce high-quality images. The thickness of the specimen is often the resolution-limiting parameter. Materials that deform plastically when crushed at room temperature often behave as brittle when crushed while immersed in liquid nitrogen.

29.2.2.14. Applications to Specific Materials and Problems

The number and diversity of applications of different forms of electron microscopy have become so large that an exhaustive survey, even when restricted to chemical aspects, is almost impossible. In particular, its application to structural problems in solid-state chemistry has evolved explosively in recent years. The examples given in what follows attempt to show as large a diversity as possible.

29.2.2.14.1. Crystal Structures

Only in simple cases can an "ab initio" structural determination be made by electron microscopy and diffraction, essentially because in electron diffraction the intensity of the reflections is not simply proportional to the square of the structure amplitude, but also depends in a complicated manner on the orientation and the thickness of the sample.
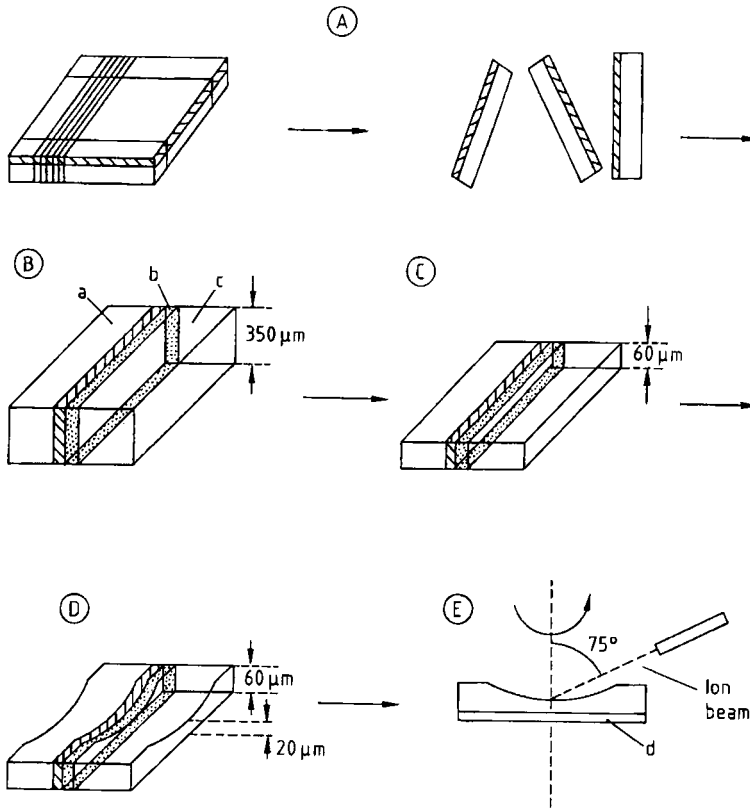
Figure 49. Preparation of a cross section of integrated circuits
A) Specimen is cut into thin slabs; B) Slabs are glued on a glass slide; C) Composite specimen is thinned mechanically; D) Thin specimen is dimpled; E) Final thinning is done by ion bombardment from one side while the specimen is rotated
a) Integrated circuit; b) Glue line; c) and d) Glass slide
(Courtesy of IMEC)

However, whereas X-ray and neutron diffraction provide only average structures (→ Structural Analysis by Diffraction), electron diffraction and microscopy can provide detailed information on local structures, especially when the average structure is known from other techniques. Also planar defects (i.e., local deviations from the average structure) can be studied to suggest possible new structures in which such defects occur periodically.

Since electron microscopy requires minute samples, the crystallography of materials that are available only as small particles (e.g., catalysts) can be studied in direct and reciprocal space. The early crystallographic studies on C_{60} and C_{70} crystals were for instance, performed by electron diffraction [194]. Even fine powders or fine-grained ceramics contain single crystal particles large enough to produce a single crystal electron diffraction pattern; therefore, the lattice parameters and the unit cell can be determined easily, albeit with moderate precision. Subsequently, X-ray powder diffraction patterns can be indexed unambiguously and the lattice parameters determined with high precision. Electron diffraction and especially high-resolution electron microscopy are complementary techniques to X-ray and neutron diffraction.

Imaging Modes for Various Materials. *Polytypes, Mixed-Layer Polytypes.* The structures of polytypes are usually based on the various stacking modes of closely packed layers of spheres. The stacking sequence can best be revealed by imaging the structure along the zone parallel to the direction of the close-packed rows of atoms. Figure 50 shows a foil of SiC – 15 R imaged along the close-packed rows by using three different modes; it
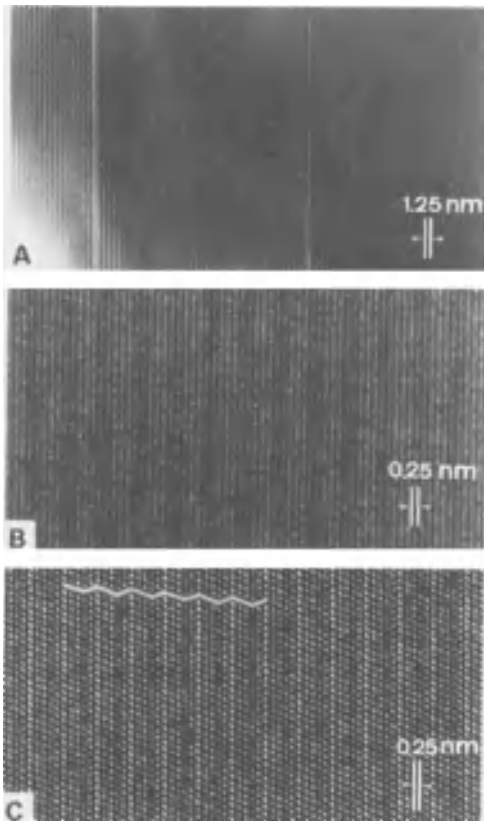
Figure 50. Imaging of the 15 R polytype of SiC along the closely packed rows of atoms
A) Two successive spots of the 0001 row contribute to the image; only the five-layer lamellae are imaged as lines; singular line spacings represent stacking faults; B) Cluster of five successive spots of the 0001 row contributes to the image; single SiC layers are revealed; each lamellae of A) is seen to contain five layers; C) Spots from parallel rows are allowed to contribute to the image; the stacking sequence ABCAC can be deduced from the two-dimensional image since close-packed atom columns are imaged as dots

illustrates clearly the effect on the image detail of increasing the number of beams participating in the image-forming process. In Figure 46 A, only two successive spots out of the densely populated row of 0001 spots are included; only the super-spacing consisting of five-layered SiC lamellae is revealed as a line image. The isolated stacking faults are imaged as anomalously wide line spacings.

Including in the aperture, a cluster of at least five successive spots of the central row between the 0000 and 00015 reflections, but excluding parallel rows, reveals the individual SiC layers and the superperiod due to the five-layered lamellae as lines; it shows in particular that the superperiod contains five layers (Fig. 46).

Finally, when spots from neighboring dense rows are also collected, a two-dimensional dot pattern is obtained and the close-packed rows are imaged as bright dots; the stacking sequence ABCAC... is directly revealed (Fig. 46).

The same imaging mode is used for the study of mixed-layer compounds based on the epitaxy of close-packed layers [195].

Ordered Alloys. The diffraction patterns of ordered binary alloys, derived from the face-centered cubic (fcc) structure, consist of strong basic spots that would also be produced by the disordered fcc lattice, and of weaker superlattice spots characteristic of the superstructure. The first-order superlattice spots correspond to larger interplanar spacings and are thus closer to the origin than the basic spots. Therefore images can be produced by selecting superstructure spots next to the direct beam only and excluding basic spots. The image so obtained (*bright-field superlattice image*) reveals only the configuration of the minority atoms as bright dots, since these atoms determine the superlattice. This information is sufficient to determine completely the projection of the superstructure, provided the basic lattice is known.

An image can also be formed exclusively by means of the superstructure spots situated within a reciprocal unit cell of the basic lattice (*dark-field superlattice image*) [196]. This last mode produces the best-contrasted images of the minority atom sublattice (Fig. 51).

If basic and superlattice spots are included in the image-forming process the basic lattice is also revealed; minority and majority atoms are imaged as dots with a different brightness. Such images contain information that is redundant if only the superstructure is sought. Examples of superlattice images of ordered alloys obtained by using the bright-field superlattice mode are reproduced in Figure 52.

Long-Period Interface Modulated Structures. Many long-period structures are derived from simpler structures by the periodic introduction of planar translation interfaces. Typical examples are the long-period antiphase boundary structures in many ordering alloy systems such as $Au-Mn$, $Cu-Au$, and the shear structures derived from a simple transition-metal oxide structure such as WO_{3-x}, TiO_{2-x}. Such compounds produce charac-
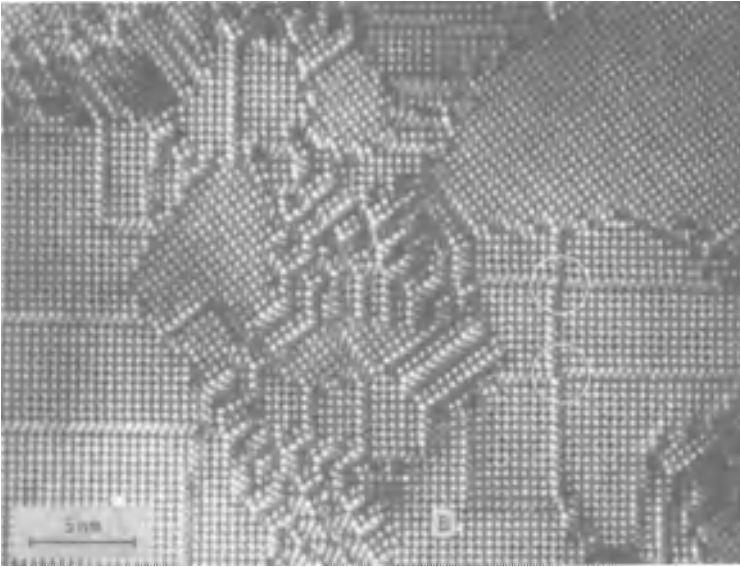
Figure 51. Dark-field superlattice image of Au_4Mn
The manganese atoms are represented as bright dots. Note the presence of two orientation variants and several translation variants [196], [198]

teristic diffraction patterns. In combination with their high-resolution images, their structures can usually be elucidated completely, including local deviations from the average structure, which often occur in such phases.

In a model [197], the superstructure is assumed to consist of identical slabs of basic structure with thickness D limited by planes normal to $\bar{e}_n$. Usually the slab thickness is equal to an integer number n of unit cell parameters a of the basic structure i.e., $D = na$, but this need not be the case. Successive slabs are separated by planar interfaces (stacking faults, anti-phase boundaries, discommensuration walls, etc.) with a displacement vector $\bar{R}$ and a unit normal $\bar{e}_n$.

The diffraction pattern of such a structure consists of clusters of equally spaced superstructure spots (spacing: $1/D$), called satellites, located around the positions of the basic spots. The intensity of the satellites decreases rapidly with separation of the satellite from the basic spot to which it belongs, which also determines its position. The positions $\bar{H}$ of satellites are given by the relation [197]

$$\bar{H} = \bar{h} + (m + \bar{h} \cdot \bar{R})\bar{q} \qquad (2.28)$$

(m = integer; order of the satellite; $\bar{q} = (1/D)\bar{e}_n$. The diffraction pattern thus exhibits main (or basic) spots at $\bar{h}$. With each basic spot $\bar{h}$ a linear sequence of equidistant satellites $m\bar{q}$ is associated. These sequences are perpendicular to the interfaces and they are shifted with respect to the positions of the basic spot over a fraction $\bar{h} \cdot \bar{R}$ of the intersatellite spacing; $\bar{q}$ is the wave vector of the modulation. Provided the basic structure is known, the long-period structure can be determined from the geometry of the diffraction pattern.

If the displacement vector $\bar{R}$ is parallel to the interfaces, the latter are termed conservative since the long-period structure has the same chemical composition as the basic structure.

If $\bar{R}$ is not parallel to the interfaces the superstructure may have a composition different from that of the basic structure; this is the case for shear structures, in which the interfaces perpendicular to $\bar{e}_n$ are generated by removing slabs of material with thickness $\bar{R} \cdot \bar{e}_n$, followed by closing the gaps so created by the displacement $\bar{R}$. If the removed slab has a composition different from that of the basic structure, the overall composition changes in the process.

The alloy, based on a face-centered cubic lattice, with ideal composition $Au_{22}Mn_6$, produces the diffraction pattern shown in Figure 53 A and
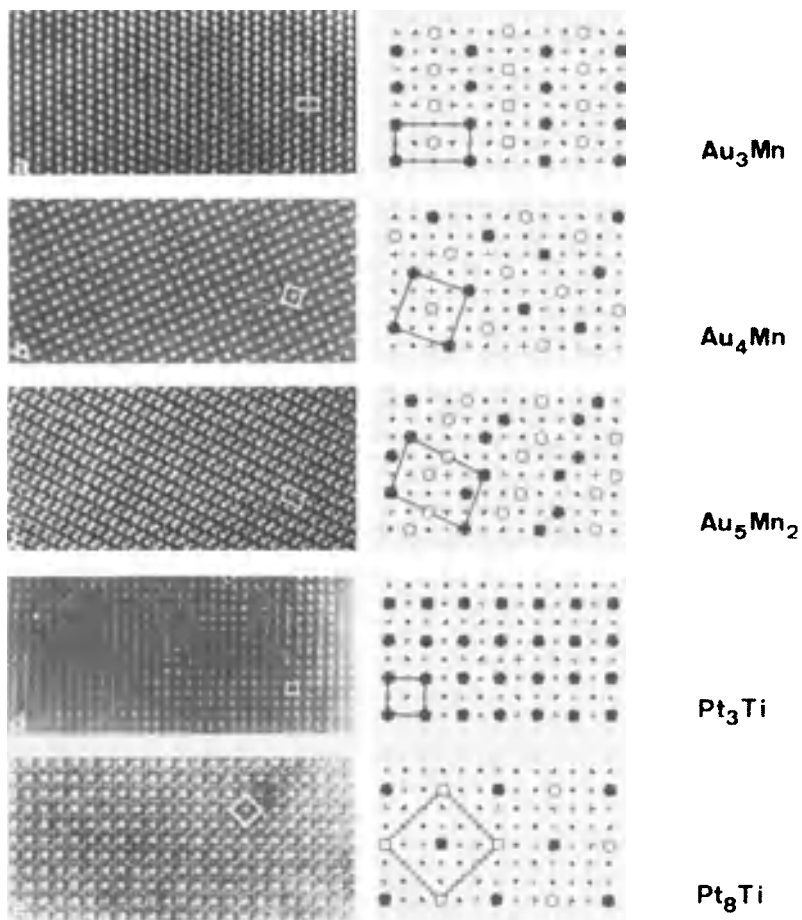
Figure 52. Various superstructures in ordered binary alloys imaged using the bright-field superlattice mode and compared to structural models

represented schematically in Figure 53 B. It consists of sequences of superstructure spots, marked as full dots in Figure 53 B, associated with the basic spots marked as small open disks in Figure 53 B, which are the spots that would be produced by the Au$_4$Mn structure. Both types of spots are present simultaneously in Figure 53 A because the selected sample contained a small area of Au$_4$Mn as well as an area exhibiting the Au$_{22}$Mn$_6$ structure. From the direction of the rows of superstructure spots, their spacing, and the fractional shifts, the model of Figure 53 D can be derived [198]. In this model, only the manganese columns are represented. The high-resolution image (Fig. 53 C) made in the bright-field superlattice mode can be compared directly with the model of Figure 53 D;

only the minority atom columns are revealed as bright dots.

Molecular Crystals: Fullerites. Organic crystals are usually prone to ionization damage and decompose very rapidly under electron irradiation; they can thus be studied for only a short time (a few seconds) and only with a very low electron beam intensity. Transmission electron microscopy has, therefore, seldom been applied to organic crystals. However, the all-carbon molecules C$_{60}$, C$_{70}$, etc., (fullerenes) discovered at the end of the 1980s resist electron radiation fairly well. Early structural studies on the crystalline phases of fullerenes (fullerites) were performed mainly by electron microscopy because only small quantities of sufficiently pure material were available. At room
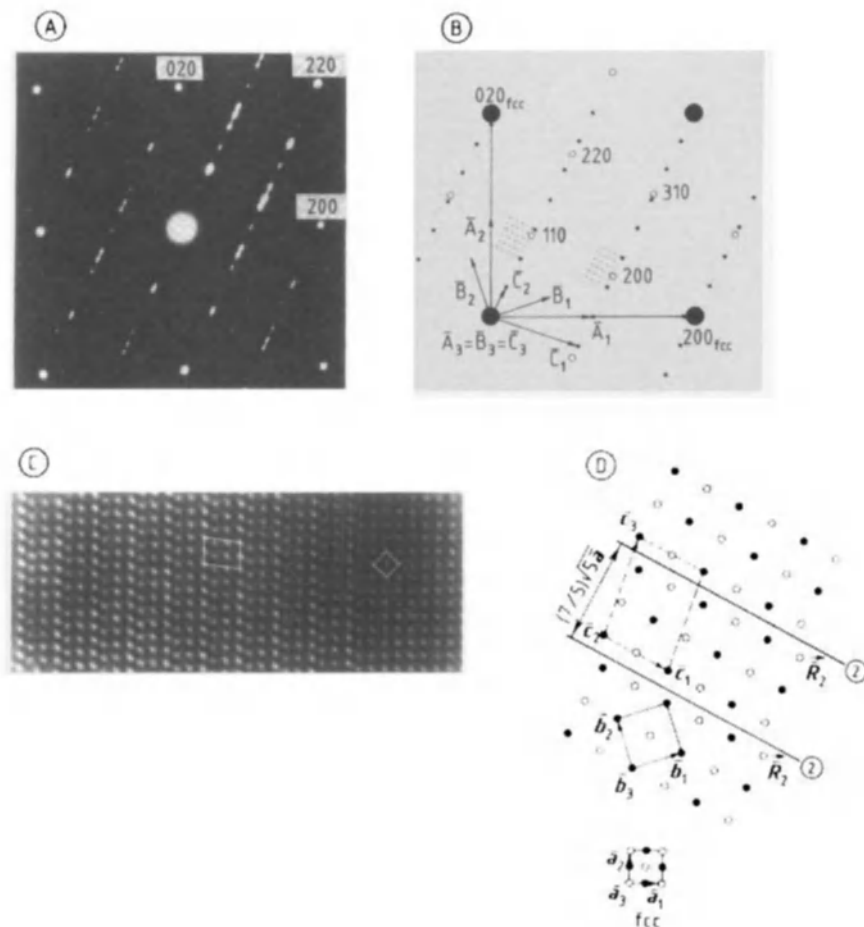
Figure 53. One-dimensional nonconservative long-period superstructure of the Au$_4$Mn structure, with theoretical composition Au$_{22}$Mn$_6$ [198]
A) Composite diffraction pattern along the [001] zone of Au$_4$Mn; B) Schematic representation of composite diffraction pattern (open dots indicated locations of basic spots, full dots are superstructure spots); C) High-resolution image of the superstructure (on the right, a small area of basic Au$_4$Mn structure is also visible); D) Model of the one-dimensional superstructure; only manganese columns are represented

temperature, C$_{60}$ crystals were found to have the fcc structure, often containing intrinsic stacking faults, twins, and other defects characteristic of low stacking fault energy fcc alloys [199]. Figure 54 shows, for instance, various faults in a crystal of C$_{60}$: an intrinsic fault containing a dipole of stair rod dislocations in S, and a Frank partial dislocation and its associated stacking fault in A. The diffraction effects and the microstructure caused by the orientational phase transition in C$_{60}$ at 255 K, from the room temperature fcc phase to the simple cubic low-temperature phase, were studied by means of electron microscopy [200] (Fig. 55).

Vapor-grown C$_{70}$ crystals were found to belong to two different structural types. About 1 % of the vapor-grown crystals are hexagonally close packed; the rest are face-centered cubic. Many crystal fragments exhibit both crystal structures. The hexagonally close-packed crystals undergo several orientational phase transitions on cooling, leading to a monoclinic orientationally ordered superstructure at low temperature [201]. The fcc
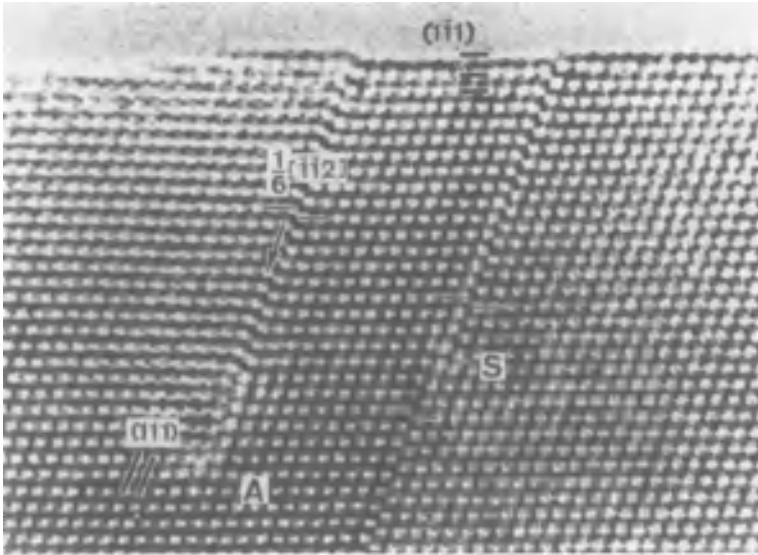
Figure 54. High-resolution image of a fragment of a C_{60} molecular crystal: bright dots represent rows of C_{60} molecules Note the presence of a stepped intrinsic stacking fault (in S). Also a Frank-type partial dislocation and its associated stacking fault are present (in A) [199]
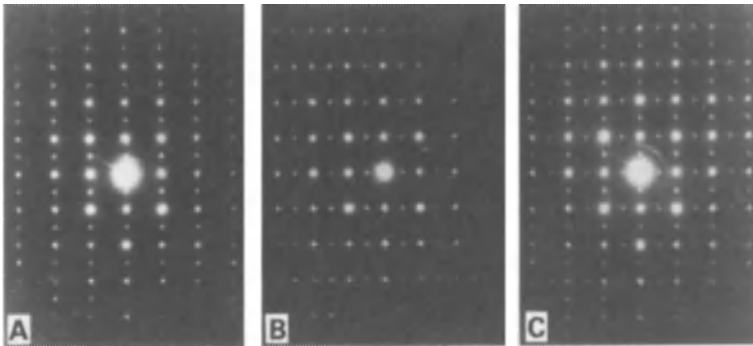
Figure 55. Diffraction patterns of the simple cubic phase of C_{60} along a cube zone [200]
A) One variant; B) Perpendicular variant to A); C) Both variants present

crystals undergo a single sluggish phase transition into a rhombohedral phase.

Quasi Crystals. Until 1984, one of the most firmly established fundaments of crystallography was the postulate that crystals have three-dimensional translation symmetry, and that only symmetry rotation axes with multiplicities 1, 2, 3, 4, and 6 can occur because they are the only ones that are consistent with homogeneous space filling. However electron microscopy and electron diffraction have shown that 5-, 8-, 10-, and 12-fold symmetry axes occur in the structure of certain, usually rapidly cooled, alloy phases with complicated compositions [202], [203]. Careful tilting experiments have demonstrated the presence of icosahedral and dodecahedral symmetry groups in many of these phases. Their electron diffraction patterns consist of sharp spots, whose geometry reflects the above-mentioned symmetry elements, but they do not exhibit translation symmetry. The same conclusions follow from high-resolution images [204].

At present, numerous alloy systems are known to exhibit noncrystallographic symmetry elements and to lack three-dimensional periodicity; they are
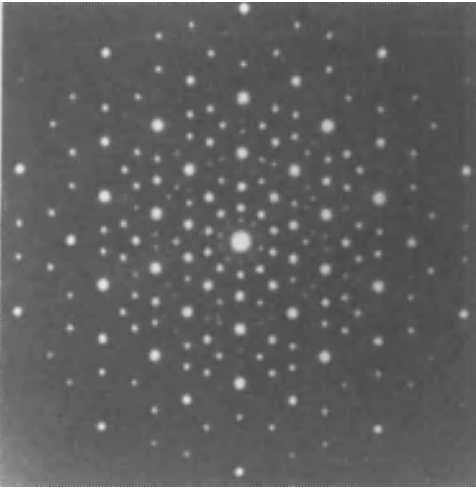
Figure 56. Diffraction pattern of the icosahedral phase of quenched Al – Mn alloy exhibiting fivefold symmetry (Courtesy of G. Van Tendeloo)
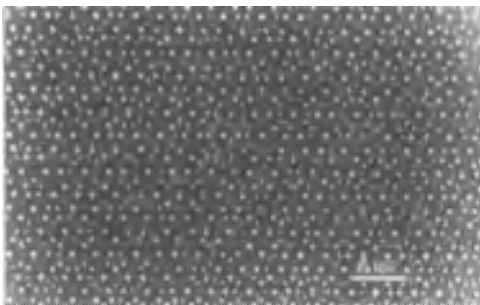
Figure 57. High-resolution electron micrograph of rapidly cooled Al – Mn alloy showing the absence of a periodic structure; pentagonal arrangements of dots can be observed

Figure 58. High-resolution image of quantum well consisting of alternation of pure GaAs and (GaAl)As

called quasi crystals. Figure 56 shows the diffraction pattern of a quasi crystal with composition Al – Mn exhibiting a fivefold axis, whereas Figure 57 shows the corresponding high-resolution image; pentagonal arrangements of bright dots are visible in many places.

Artificial Layer Structures and Semiconductor Devices. In recent years the growth of artificial layer structures, which can be used as quantum wells, has become an important research activity. The properties of such layered structures are determined to a large extent by the perfection of the interfaces, as well as the regularity of their spacing. Both parameters can be studied by means of high-resolution, cross-sectional samples. The steps

in the preparation of cross-sectional samples are discussed in Section 29.2.2.13. The method is clearly destructive and cannot be used for fabrication control. However, it is the most direct way to calibrate other nondestructive methods.

Figure 58 shows a high-resolution image of a quantum well consisting of the alternation of GaAs and (GaAl)As. The positions of the interfaces are indicated by arrows. The interfaces are flat to within two or three atomic layers; no interfacial dislocations are observed. The composition profiles across such interfaces can in principle be studied by measuring the brightness of the dots representing the atom columns [205].

For the study of semiconductor devices, diffraction contrast images made in a high-voltage electron microscope (≈ 1000 kV) are often useful since thicker specimens can be tolerated than in a
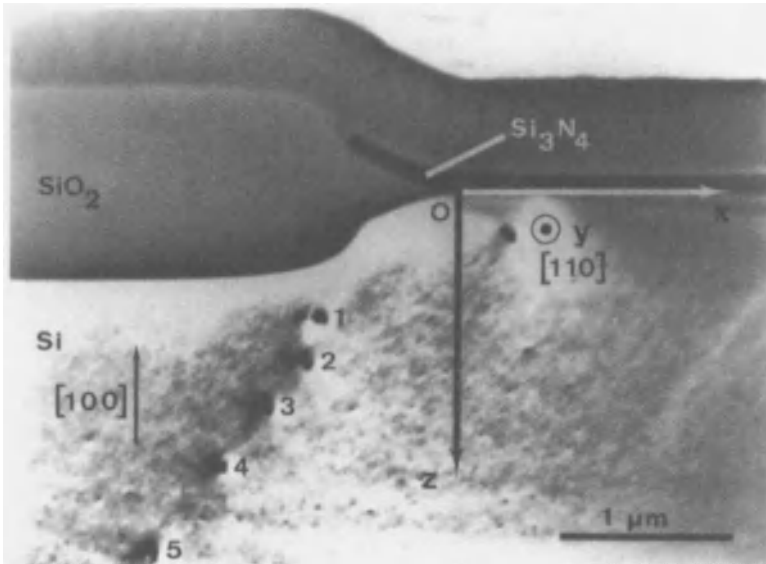
Figure 59. Cross section of field effect device
The procession of dislocation 1 ... 5 results from stresses set up by the oxidation process (Courtesy of J. Van Hellemont)

conventional microscope (≈ 200 kV). An example is shown in Figure 59, which is the cross section of a field effect device. The different layers have been indicated along with the orientation of the silicon substrate. The black dots numbered 1–5 are images of dislocations seen end on, which were nucleated by the stresses set up as a result of the oxidation process. From the spatial distribution of the dislocations the magnitude of the stress can be estimated.

During crystal growth, small oxide particles are often formed in the interior of silicon single crystals. Such particles cause a compressive stress on the surrounding matrix. This may lead to prismatic punching, whereby disks of self-interstitials limited by loops of perfect dislocations are emitted, thereby relieving the stresses. These loops can glide along cylindrical surfaces whose cross section is determined by the size and shape of the particle, and whose generators are parallel to the different glide vectors of silicon, which are of type 1/2 [110]. The geometry of the process can be established with diffraction contrast images (see Fig. 60).

Carbon Nanotubes, Onions. The preparation of fullerene-containing soot by the vaporization of graphite electrodes in an electric arc leads simultaneously to the formation of very fine *hollow carbon needles* or "tubules" with diameters as

low as 1 nm. Electron microscopy and electron diffraction have shown that these needles consist of concentric seamless graphene tubes (2–15 tubes) [206]. The most remarkable feature is the helical character of some of the tubes within a tubule, as deduced from fiber diffraction patterns (Fig. 61). The helical structure is a consequence of the stepwise increase in circumference of successive concentric tubes by πc ($c =$ lattice parameter of 2 H graphite). Since πc is not commensurate with the a-parameter of graphite the $c/2$ spacing between successive tubes and their seamless character can be reconciled only with the changing diameter if some of the tubes at least become helical. However, a detailed analysis of the fiber pattern has shown that in most tubules the majority of tubes are nonhelical [207].

Intense "in situ" electron irradiation of particles of carbon soot for tenths of minutes in the electron microscope produces spherical particles consisting of *onion-like* concentric spherical shells of graphene-like layers [208]. The microscope allows the concentric shell structure to be imaged during its formation.

High-T_c Superconductors. High-resolution electron microscopy and electron diffraction have recently been applied extensively in structural studies of high-T_c oxide superconductors. In particular, $YBa_2Cu_3O_{7-\delta}$ and the related 1–2–3
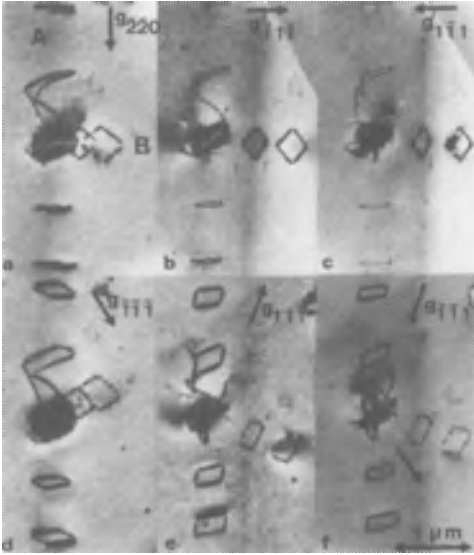
Figure 60. Oxide particle in the interior of a silicon single crystal
Stresses have been relieved by prismatic punching. The dislocation loops are imaged using different reflections (Courtesy of H. Bender)

Figure 61. Electron diffraction (fiber) pattern produced by a graphite nanotube showing that certain graphene tubes are helically wound, whereas others remain normal
Note the outward streaking of the diffraction spots
(Courtesy of X. F. Zhang)

family of compounds have been studied in great detail. Numerous studies of superconducting compounds can be found in *Physica C* from 1989 onward. The generic $1-2-3$ compound ($YBa_2Cu_3O_{7-\delta}$) has a structure derived from the triperovskite structure in which the succession of layers along the c-direction is $CuO-BaO-CuO_2-Y-CuO_2-BaO$, etc. The CuO layer consists of $Cu-O-Cu-O-$ chains parallel to the b_0-direction; this feature breaks the tetragonal symmetry and reduces the symmetry to orthorhombic. The average structure was determined by X-ray and neutron diffraction, but several important structural features were discovered by electron microscopy. The oxygen deficiency was found to be accommodated by the absence of oxygen in a fraction of the chains in the CuO layer [209]. The structure in which one of two chains is alternately free of oxygen, the so-called $2\,a_0$ (or ortho II) superstructure, has an ideal composition with $\delta = 0.5$. It occurs in two orientation variants, corresponding to the (110) reflection twin texture of the $1-2-3$ compound, which was also discovered by electron microscopy. The presence of the $2\,a_0$ structure was shown to be responsible for the 60 K plateau in the T_c versus composition curve [209]. Locally, structures with a period of $3\,a_0$ occur [210].

Much attention has been devoted to various substitutions in CuO layers, replacing copper ions by other metallic ions such as Fe, Co, etc., but also substituting by complex anions such as CO_3^{2-}, SO_4^{2-}, PO_4^{3-}, or NO_3^-. High-resolution electron microscopy has demonstrated that the substitution occurs in the CuO layers, and has allowed the resulting superstructures to be visualized. In the case of SO_4^{2-} substitution, for instance, the superstructure has been shown not to be commensurate with the basic $1-2-3$ lattice [211]. Figure 62 shows a view along the b_0-direction of the $1-2-3$ matrix. The prominent, bright dot squares reveal the SO_4^{2-}-containing rows that replace CuO chains. Their arrangement is not periodic and can best be described as resulting from a concentration wave with a wave vector that is not commensurate with the $1-2-3$ lattice along the a-direction.

Replacing copper by cobalt or gallium in the CuO layer results in the formation of corner-sharing chains of CoO_4 (GaO_4), tetrahedra along the [110] and [1$\bar{1}$0] directions of the $1-2-3$ matrix lattice [211]. The resulting structure is still orthorhombic, but with a diagonal unit cell ($a_0 \approx b_0 \approx a_p \sqrt{2}$), where a_p is the lattice parameter of the basic perovskite. Electron microscopy [212] showed that a superstructure is formed, which is localized in the $Co-O$ ($Ga-O$) layers, and in which alternating parallel CoO_4 (GaO_4) chains
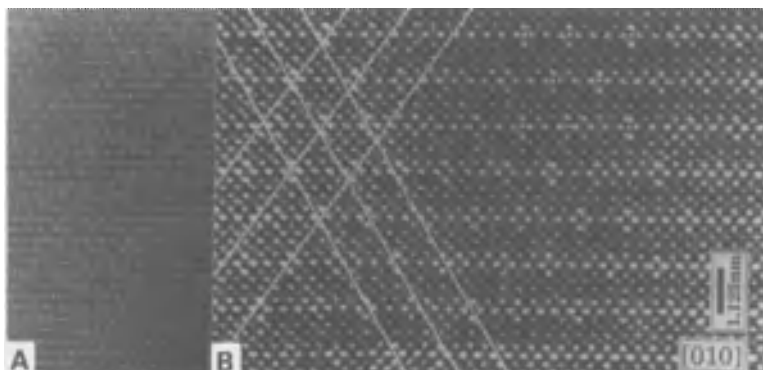
Figure 62. High-resolution image along the $[010]_0$ direction of the SO_4^{2-}-substituted $1-2-3$ ($YBa_2[Cu_{3-x}(SO_4)_x]O_{7-\delta}$) compound The SO_4^{2-}-containing rows parallel to [010] are marked as small crosses of prominent bright dots. (Courtesy of T. Krekels) [211]
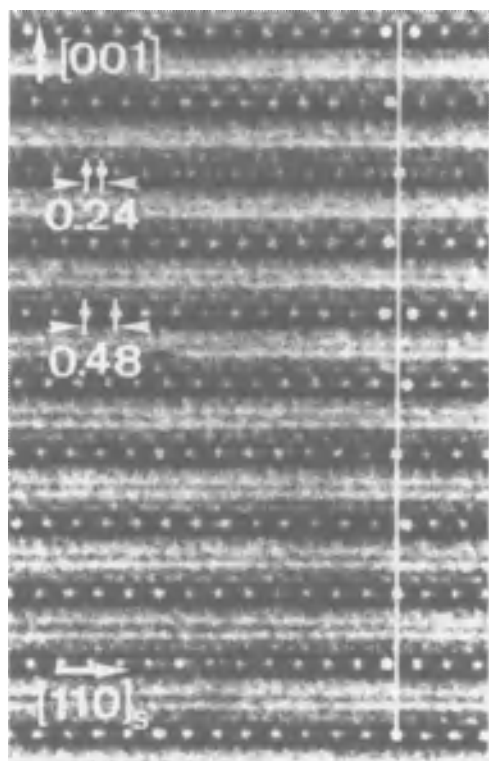
Figure 63. High-resolution image along the $[120]_p$ zone of the cobalt-substituted$1-2-3$ compound $YBa_2Cu_2CoO_7$ Dark dots reveal indirectly the arrangement of chains of corner-sharing CoO_4 tetrahedra that replace the CuO layers. (Courtesy of T. Krekels) [212]

have mirror-related configurations, leading to doubling of the lattice parameter perpendicular to the chain direction. Figure 63 is a view of the cobalt-containing material along a $[120]_p$ direction, which allows the period doubling along a_0, as well as the possibility of polytypism along c_0, to be observed.

Superionic Conductors. Superionic conductors often consist of a stable rigid framework forming large channels along which ions can move easily. The channels in such structures can be made visible in high-resolution images made along a zone parallel to the channel direction. Figure 64 shows such a high-resolution image of the natural mineral hollandite with idealized composition $Ba_xMn_8O_{16}$ [213]. Octagonal channels, formed by interconnected strings of edge and corner-sharing MnO_6 octahedra, are imaged as octagons of black dots. These channels are occupied by barium ions, which are also visible as black dots in the centers of the octagons.

Polymers. Although most organic materials usually deteriorate rapidly from ionization damage in the electron beam, meaningful observations can be made on certain polymers. In Figure 65, fiber patterns of a poly(-p-phenylene) stretched six to seven times are reproduced (PPV, an alternating copolymer of p-phenylene and acetylene) at two different photographic exposures to reveal the details of the pattern close to the origin [214]. The direction of stretching is indicated by arrows. The patterns of the left column (A) and (C) refer to the pristine material. On doping with $FeCl_3$ this material becomes a good electrical conductor. The effect of $FeCl_3$ doping on the diffraction pattern is
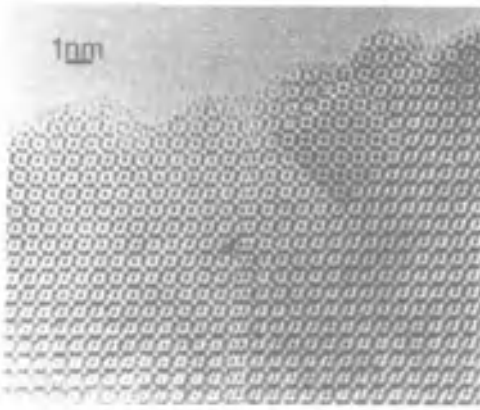
Figure 64. Structure of natural monoclinic hollandite $Ba_xMn_8O_{16}$ as viewed along the [010] zone
The image is compared with a simulated image in the inset [213].

Figure 66. Network of undissociated dislocations in the (0001) plane of zinc
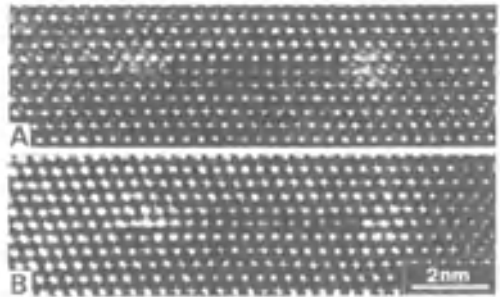
Figure 67. Dissociated dislocations in silicon viewed along the |110| zone in the high-resolution mode (Courtesy of H. Bender)

29.2.2.14.2. Defects

Dislocations. Dislocation configurations have been studied mainly by diffraction contrast bright-field imaging since this method is very sensitive to lattice strain fields. Under these conditions, dislocations appear as dark lines when two-beam diffraction conditions close to $s \approx 0$ are used. A network of dislocations in the basal plane (0001) of zinc is visible in Figure 66.

Images of dissociated dislocations (i.e., dislocation ribbons) have been used extensively as a means to deduce the stacking fault energy [125].

The separation between partial dislocations is usually measured on a weak beam image (see Section 29.2.2.7.2). The image, which is now a very narrow bright line on a darker background, is essentially kinematic. In the high-resolution image of Figure 67, dissociated dislocations in silicon are viewed along the [110] zone. Between the two partial dislocations an intrinsic stacking fault

Figure 65. Fiber texture pattern of a stretched PPV polymer A) and C) Pristine polymer; B) and D) After $FeCl_3$ doping (Courtesy of X. F. Zhang)

shown in the right column (B and D). Whereas in the undoped material the 001 diffraction vector encloses an obtuse angle with the row of hoo reflections, this angle becomes 90° in the doped material, which suggests that the symmetry changes from monoclinic to orthorhombic on doping.

Figure 68. High-resolution image of a stacking fault in 2H-wurtzite (Courtesy of H. Bender)

is present. Such configurations have been used to deduce stacking fault energies [125], [176].

Planar Defects. The diffraction contrast images of planar interfaces that intersect the foil surfaces consist of a set of fringes parallel to the closest surface with a depth period given by $1/\sigma_g$ *(which is equal to t_g for $s = 0$;* see Section 29.2.2.4.6). If the planar interface is parallel to the foil surfaces, which is often the case in cleaved foils, the faulted area exhibits only a brightness difference.

High-resolution images can reveal stacking faults directly, especially in close-packed structures. In such structures the image made along a zone parallel to the close-packed rows reveals the stacking directly. A stacking fault in a wurtzite crystal is reproduced in Figure 68.

Domain boundaries and twins in planes that intersect the foil surfaces are imaged as δ-fringes (Section 29.2.2.4.6) [155]. The extinction condition is now $\Delta s = 0$ (or $\Delta \bar{g} = 0$!) which is satisfied for the family of planes common to the two domains. When the interfaces are nearly perpendicular to the foil plane the main effect is domain contrast (i.e., a brightness difference in the domains on either side of the interface) as shown in Figure 69, where the same twin boundaries in the high-temperature superconductor $YBa_2Cu_3O_{7-}^{\delta}$ have been imaged in three different modes.

29.2.2.14.3. Small Particles

Like most other electron microscope techniques, transmission electron microscopy allows
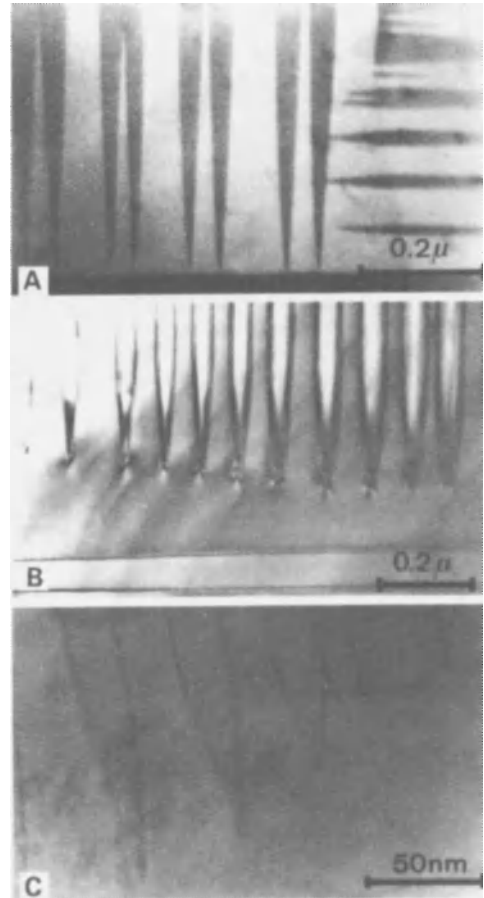
Figure 69. Twin boundaries in $YBa_2Cu_3O_{7-\gamma}$ (A) Domain contrast; (B) Interface contrast; (C) High-resolution imaging (Courtesy of H. W. Zandbergen)

the particle-size distributions and particle shapes of very fine powders to be studied. Such studies are important in several areas of research: including catalysis, magnetic recording, and photography. Transmission electron microscopy provides, in addition, the possibility of obtaining electron diffraction patterns of single particles as well as high-resolution images. In Figure 70 single particles of silver vapor deposited on an amorphous carbon substrate is shown. They are clearly not single crystals but aggregates of multiply twinned fcc crystallites having an overall icosahedral shape.

The topotactic dehydration reaction [215] transforming goethite into hematite was studied "in situ" at room temperature and shown to

Figure 70. Small silver particles consisting of an icosahedral aggregate of multiply twinned face-centered cubic crystals (Courtesy of C. Goessens)
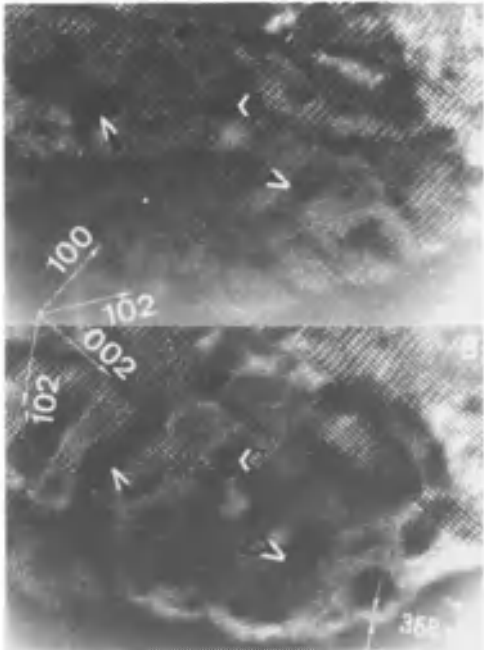
Figure 71. Small particles of twinned hematite formed in the electron microscope as a decomposition product of goethite [215]
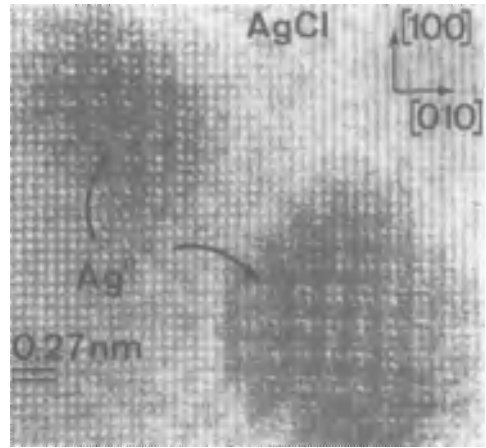
Figure 72. Silver chloride crystal covered by silver specks viewed along the cube zone

produce small twinned hematite crystals (Fig. 71). In the initial stages of the dehydration process, satellite spots appear in the diffraction pattern around the hematite spots; they were found to be due to a texture consisting of a periodic alternation of voids and hematite crystals, rather than to a long-period superstructure.

The crystal habit of the silver halide particles used in photographic emulsions is important since it is one of the parameters determining the photosensitivity of the film. The hexagonal and triangular (111) tabular crystals of AgCl were shown to be in fact multiply twinned on planes parallel to the habit plane. Depending on the parity of the number of twin lamellae the particle grows into either hexagonal or a triangular habit.

When silver halide crystals with a cube habit are exposed to electrons, photolytic metallic silver specks are formed "in situ" in an epitaxial relation to the silver halide substrate. This gives rise to moiré fringes parallel to the cube edges due to the difference in lattice spacing between AgCl and Ag (Fig. 72).

29.2.2.14.4. Surface Studies [216]

Crystal surfaces can be studied in a transmission electron microscope by different modes. It is possible to use an electron beam, in reflection under *grazing incidence*, magnifying the Bragg reflected beam due to planes roughly parallel to the surface. The surface of the sample must be almost parallel to the beam. The resulting image
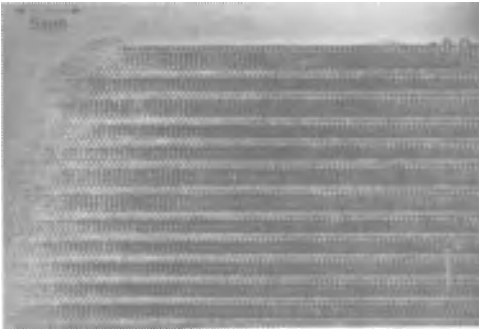
Figure 73. High-resolution image of the free cleavage surface of a superconductor $(Bi_2Sr_2Ca_1Cu_2O_x)$
Cleavage occurs between the two BiO layers [217]

Figure 74. Thin film of $La_{1-x}Ca_xMnO_3$ on a $SrTiO_3$ (STO) single-crystal substrate
a) Plane view, low-resolution image revealing antiphase boundaries
b) High-resolution image of a cross section specimen (courtesy O. Lebedev)

is highly distorted in the sense that the magnification in a direction perpendicular to the beam is much lower than the magnification along the projection of the electron beam. This Bragg reflection mode can be used to image surface steps and to detect surface superstructures due to reconstruction or to adsorbed layers by diffraction. A very clean vacuum is required.

A second mode consists of observing the surface profile in transmission with the *electron beam parallel to the surface*. In this way, surface relaxations and surface reconstruction can be observed together with atomic resolution of the substrate. An example of the cleavage surface of a high-T_c superconductor is reproduced in Figure 73. It allows the layer within the structure to be located along which cleavage takes place [217].

29.2.2.14.5. Thin Epitaxial Layers

The development of modern miniaturized electronic devices relies heavily on the use of thin epitaxial films of which the relevant physical properties sensitively depend on their thickness and on their microstructure. In recent years high-resolution electron microscopy has extensively been applied for the primary characterization of such films. In particular various modes of accommodating the misfit between a single-crystal substrate and the epitaxial films have been discovered by this method.

Routine characterization, which necessarily has to be based on nondestructive methods can be "calibrated" by means of HRTEM, which is a destructive, but highly informative method. Examples are shown in Figure 74.

29.2.2.14.6. "In situ" Studies [218]

By using different types of stages the specimen chamber of the microscope can be transformed into a small laboratory. Heating and cooling stages make it possible to study "in situ" the changes in crystal structure and in microtexture accompanying phase transitions, by observing (1) the appearance or disappearance of superstructure spots in the diffraction pattern, and (2) the fragmentation with translation and/or orientation domains in the direct space image.

Environmental cells can be used to study in situ solid–gas reactions, such as oxidation processes or decomposition reactions.

Plastic deformation can be studied and the dislocation propagation observed in real time in a straining stage.

The creation of radiation damage by ionization or by electron–atom collisions can be studied in medium- and high-voltage microscopes. The formation of agglomerates of point defects, such as dislocation loops and stacking fault tetrahedra, can be observed during irradiation with the image-forming electrons. (Light atoms, such as oxygen, are readily displaced by electrons of 400 kV.)

Radiation ordering and disordering in alloy systems have successfully been studied in high-voltage microscopes at different temperatures.
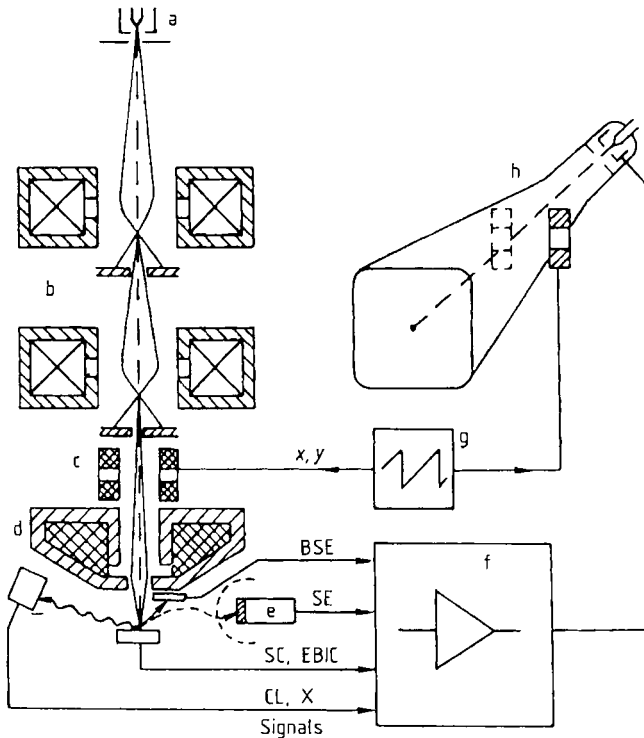
Figure 75. Schematic cross section of a scanning electron microscope (SEM)
a) Electron gun; b) Condenser lenses; c) Scan coils; d) Objective; e) Photomultiplier; f) Amplifier; g) Scan generator; h) Cathode-ray tube
BSE = Backscattered electrons; SE = Secondary electrons; SC = Specimen current; EBIC = Electron-beam-induced current; CL = Cathodoluminescence; X = X rays

29.2.3. Scanning Electron Microscopy

29.2.3.1. Introduction

In a scanning electron microscope (SEM) [219]–[221], a small electron probe 1–10 nm in diameter scans in a raster across the surface of the specimen (Fig. 75). The incident electrons are elastically and inelastically scattered by the specimen. *Elastic scattering* results in large scattering angles and zigzag electron trajectories. Therefore, a fraction of electrons can leave the specimen as backscattered electrons (BSE) (Fig. 76). The slowing down of electrons by *inelastic scattering* results in an electron range R. Electrons from the specimen atoms that are excited by inelastic scattering can leave the specimen as *secondary electrons* (SE) from a thin surface layer A_{SE} of ca. 1–10 nm. By convention, electrons in the energy spectrum (Fig. 77) with $E \leq 50$ eV are called SE. The secondary electrons consist of (1) SE1 excited by the primary electrons; (2) SE2 excited by BSE on their path through the surface; (3) SE3 are excited when BSE strike the lower polepiece; or flow (4) as SE4 through the polepiece bore (see Fig. 76). The ionization of inner atomic shells results in the emission either of characteristic X-ray quanta (X) or of Auger electrons (AE).

The *image* is formed by the signal of emitted secondary electrons, backscattered electrons, Auger electrons, absorbed speciment current (SC), or X-ray quanta, which modulate the intensity of a cathode-ray tube rastered in synchronism (Fig. 75). Conventional SEMs work with electron acceleration voltages of 5–30 kV, whereas a *low-voltage scanning electron microscope* (LVSEM) [222], [223] uses 0.5–5 kV.
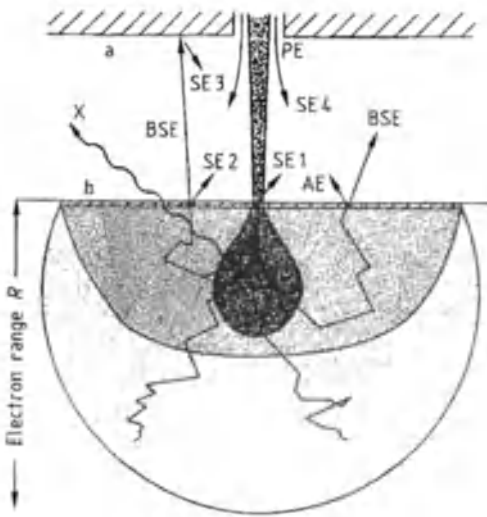
Figure 76. Excitation of backscattered electrons (BSE), Auger electrons (AE), and different groups of secondary electrons (SE) by primary electrons (PE)
Volumes of electron diffusion, BSE trajectories, and emitted X rays are shown with increasing density of gray levels
a) Polepiece; b) Specimen

Figure 77. Schematic energy spectrum of emitted electrons with secondary (SE), backscattered (BSE), elastically reflected (ERE), low-loss (LLE), and Auger electrons (AE)
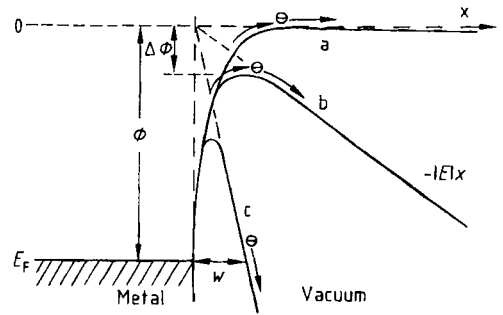
Figure 78. Potential at the cathode – vacuum interface for W and LaB_6 thermionic emission, $ZrO - W$ Schottky emission, and field emission guns
a) Thermionic emission; b) Schottky emission; c) Field emission
Φ = Work function; E_F = Fermi energy

29.2.3.2. Instrumentation

29.2.3.2.1. Electron Guns

The following types of electron emitters are used in SEM electron guns.

Thermionic cathodes consist of a directly heated tungsten hairpin cathode at $T_c = 2500 - 3000$ K, or an indirectly heated pointed rod of lanthanum or cerium hexaboride (LaB_6, CeB_6) at $1400 - 2000$ K. The electrons must overcome the work function Φ of 4.5 eV (W) or 2.7 eV (LaB_6) by thermal activation (Fig. 78, curve a). Between the cathode at the potential $-U$ and the grounded anode, a negatively biased Wehnelt electrode forms a crossover of diameter $20 - 50$ μm (W) or $10 - 20$ μm (LaB_6) as an effective electron source. The emitted electrons show an energy spread $\Delta E = 1 - 2$ eV (W) or $0.5 - 1$ eV (LaB_6). A measure of the quality of an electron gun is the axial gun brightness β:

$$\beta = j/\pi\alpha^2 \approx j_c E/\pi k T_c \tag{2.29}$$

where j is the current density, α the aperture angle, and E the electron energy. The axial gun brightness is constant for all points on the axis regardless

of lenses and aperture diaphragms (j_c and T_c are current density and temperature, respectively, at the cathode). It increases proportionally to the electron energy E, with $\beta \approx 10^5$ A cm^{-2} sr^{-1} (where sr denotes steradian) for tungsten at $E = 20$ keV and about ten times higher values for LaB_6.

Schottky emission cathodes consist of zirconium-doped tungsten tips, with a radius of about $0.5 - 1$ μm, coated with a ZrO layer. This layer decreases the work function from 4.5 to 2.7 eV. A Schottky emission cathode works with a higher electric field strength at the tip that decreases the work function by $\Delta\Phi$ (Schottky effect, Fig. 78, curve b) and concentrates the emission at the tip with a virtual electron source having a diameter of $15 - 20$ nm. However, the electrons still have to
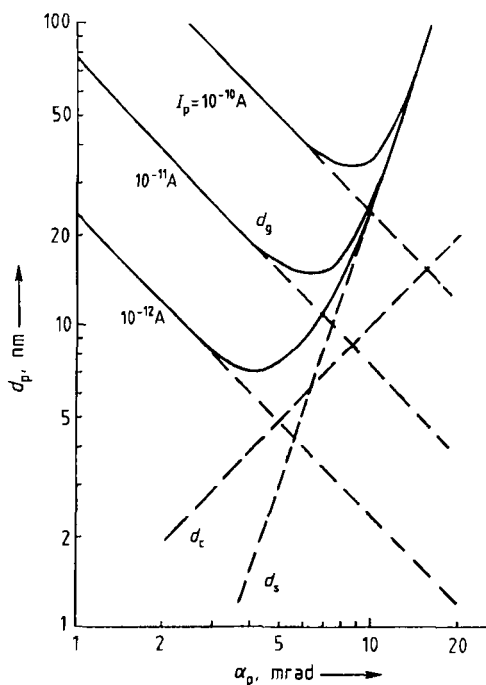
Figure 79. Double-logarithmic superposition of geometric probe size d_g for different probe currents I_p and aberration disks d_c and d_s versus the electron probe aperture α_p for a thermionic tungsten cathode $E = 20$ keV; $\beta = 7 \times 10^4$ A cm^{-2} sr^{-1}; $C_s = 50$ mm; $C_c = 20$ mm; $\Delta E = 1$ eV

overcome the work function by thermal activation. The energy spread is $0.3 - 1$ eV.

Field-emission guns (FEG) consist of tungsten tips with a radius of ca. $0.01 - 0.1$ μm. An intermediate anode at ca. $1 - 2$ kV extracts electrons by tunneling through the barrier of the work function, which is $w = 1 - 10$ nm in width (Fig. 78, curve c). Two types of FEG exist: (1) *cold FEGs* that work with the tip at room temperature, and the tip is flashed once a day, (2) *heated FEGs* work at 1800 K. The energy spread is only $0.2 - 0.4$ eV for cold and $0.5 - 0.7$ eV for heated FEGs. The gun brightness of FEGs $\beta \approx 10^7 - 10^8$ A cm^{-2} sr^{-1}.

29.2.3.2.2. Electron Probe Formation

The crossover of a thermionic gun or the virtual point sources of Schottky or field-emission guns are demagnified by two to three electromagnetic lenses (Fig. 75), so that a geometric probe size d_g is formed at the specimen with a diameter

of $1 - 10$ nm. A diaphragm of ca. $50 - 200$ μm diameter in front of the last lens limits the electron-probe aperture α_p. The electron-probe current I_p, d_g, and α_p cannot be changed independently because of the conservation of gun brightness β in Equation (2.29):

$$I_p \approx \frac{\pi}{4} d_g^2 j_p = \frac{\pi^2}{4} \beta d_g^2 \alpha_p^2 \qquad (2.30)$$

The geometric probe size d_g is superposed by aberration disks of spherical (d_s) and chromatic (d_c) aberration and diffraction (d_d):

$$d_s = 0.5 C_s \alpha_p^3; \quad d_c = C_c (\Delta E/E) \alpha_p; \quad d_d = 0.6 \lambda / \alpha_p \qquad (2.31)$$

with the aberration constants C_s and C_c, respectively; ΔE is the energy spread of the electron gun, and λ is the deBroglie wavelength

$$\lambda = h/mv = 1.23/\sqrt{E} \qquad (2.32)$$

with λ in nanometers and E in kiloelectronvolts. An astigmatism of the last probe-forming lens can be compensated using a stigmator, which consists of magnetic quadrupoles.

The (quadratic) superposition of d_g, d_d, d_s, and d_c results in an effective probe diameter d_p (Fig. 79), where d_g and d_s dominate for high electron energies and thermionic guns, and d_d and d_c dominate for low-voltage SEM and field-emission guns. This superposition results in a minimum probe size between 1 and 10 nm at optimum apertures of tens of milliradians. An anticipated increase in resolution (decrease of electron probe diameter) requires a compensation of the chromatic aberration by using a combination of electrostatic and magnetic multipoles.

The low optimum probe aperture α_p of the SEM results in a large depth of field D_f:

$$D_f = \delta / \alpha_p = \Delta / \alpha_p M \qquad (2.33)$$

where $\Delta = \delta M \approx 0.1$ mm is the resolution and M the magnification on the cathode-ray tube (CRT) screen. As a consequence, an SEM has a two-order-of-magnitude greater depth of field than a light microscope even at low magnification.

To decrease the aberration constants C_s and C_c, the focal length of the last probe-forming lens can be reduced by a stronger excitation of the magnetic lens, which means a short working distance between specimen and polepiece for high
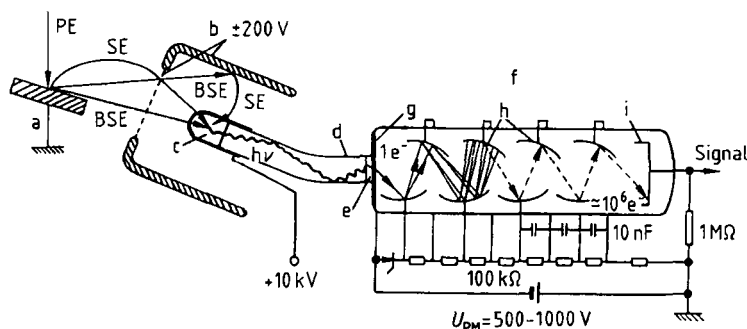
Figure 80. Everhart – Thornley detector for SE
a) Specimen; b) Collector grid and screen: c) Scintillator; d) Light pipe; e) Optical contact; f) Photomultiplier; g) Photocathode; h) Dynodes; i) Anode

resolution. The aberration constants can be decreased further if the specimen is positioned in the lens field; this makes through-lens detection of secondary electrons necessary.

29.2.3.2.3. Detectors

Secondary electrons are most frequently used for image formation and can be detected by a scintillator – photomultiplier combination (*Everhart – Thornley detector*, Fig. 80). A large fraction of slow SE is attracted by a positively biased grid (b) and accelerated to a yttrium – aluminum – garnet (YAG) scintillator or a P47 powder scintillator (c) biased at +10 kV. The photons generated are guided in a light pipe (d) to a photomultiplier (f). This SE detector shows a high signal-to-noise ratio and a large bandwidth up to 10 MHz. The collection field of an Everhart – Thornley detector mounted on one side of the specimen can disturb the electron probe in LVSEM. In this case, beam deflection and distortion can be avoided either by a combination of electrostatic and magnetic quadrupoles forming a Wien filter, by a retarding electrostatic lens, or by through-lens detection of SE.

Backscattered electrons have enough energy for direct production of a greater number of photons in a scintillator coupled via light pipe to a photomultiplier. However, the Everhart – Thornley detector cannot be used effectively for faster BSE because only a small fraction hits the scintillator area. Therefore, *annular or semiannular top detectors* collecting BSE with a high takeoff angle or

ring detectors with low takeoff, are applied to make the best use of the angular characteristics of BSE. Alternatively, *semiconductor detectors* can be used. The BSE produce a large number $n = E/E_i$ of electron – hole pairs, where $E_i = 3.6$ eV is the mean energy per excitation in silicon. These charge carriers can be separated in a $p - n$ junction and form an electron beam-induced current (EBIC). BSE can also be detected by the *conversion of BSE to SE3* at the polepiece and other parts of the specimen chamber, when a negatively biased electrode around the specimen retards SE1 and SE2 from the specimen.

Detectors for LVSEM. Both scintillator and semiconductor detectors show a signal decreasing linearly with decreasing electron energy, and a threshold energy of ca. 1 – 5 keV. Therefore, their application in LVSEM requires postacceleration. As a recent alternative, *microchannel plates* (MCPS) can be used for the detection of BSE. These consist of a slice from a boule of tightly packed, fused tubes of lead-doped glass with a 10 – 20 μm inner diameter and a resistance of $10^8 - 10^9$ Ω over their length. Annular disks 3 – 4 mm thick can be mounted below the polepiece. The incident electrons produce SE at the inner tube wall, which are accelerated by a continuous voltage drop along the tube with a bias of 1 keV and are multiplied inside the MCP tubes by a multiplier-like action. The anode plate and a preamplifier at a potential of 1 kV are electrically insulated by an optical decoupler or by transmitting a modulated 30-MHz signal.
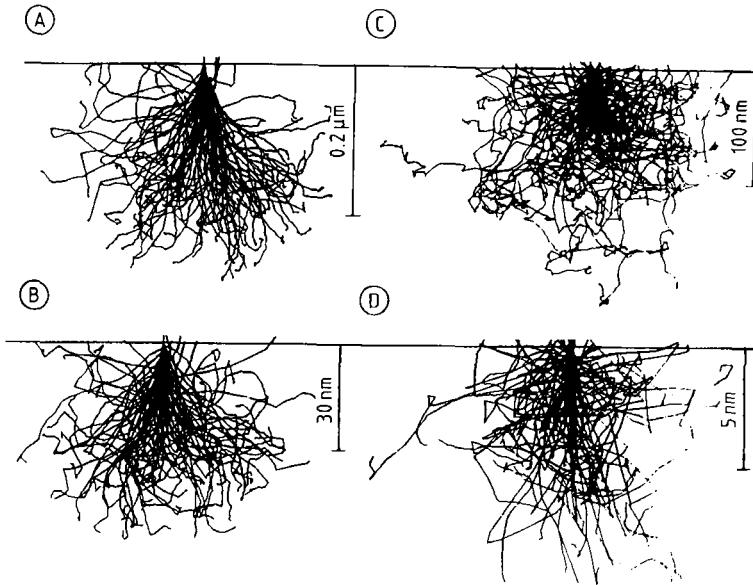
Figure 81. Monte Carlo simulations of 10- and 1-keV electron trajectories in carbon and gold
A) C, 10 keV, $\varrho = 1$ g/cm^3; B) C, 1 keV, $\varrho = 1$ g/cm^3; C) Au, 10 keV; D) Au, 1 keV

29.2.3.3. Electron – Specimen Interactions

29.2.3.3.1. Elastic and Inelastic Scattering

Elastic scattering of incident electrons results from the attractive Coulomb force of the nucleus screened by the atomic electron cloud with a differential Mott cross section of

$$\frac{d\sigma_M}{d\Omega} = r(\theta)\frac{eZ}{4\,(4\,\pi\,\varepsilon_0)^2\,m^2\,v^4}\frac{1}{\sin^4(\theta/2)} \qquad (2.34)$$

where $r(\theta)$ is the ratio between the Mott and Rutherford cross sections, Ω is the solid angle, Z the atomic number of the nucleus, ε_0 the vacuum permittivity, and m the mass of the electron [219]. The strong differences between Mott and Rutherford cross sections result from taking account of the spin – orbit coupling of electrons and solving the relativistic Dirac equation, whereas the Rutherford cross section is only a solution of the Schrödinger equation not containing the spin.

Inelastic scattering results in an excitation of electrons of the solid and a corresponding energy loss ΔE of the incident electron. Information about the differential inelastic cross section $d^2\sigma/d\Omega\,d(\Delta E)$ can be obtained from dielectric theory or experimental electron energy loss spectra (EELS; see Section 29.2.2.12.2) of high-energy electrons [224], [225].

A series of inelastic scattering processes with statistical energy losses results in a slowing down of electrons, which can be described by a mean energy loss per unit path length (Bethe stopping power S)

$$S = \left|\frac{dE_m}{ds}\right| = \frac{2\,\pi\,e^4\,N_A\,\rho\,Z}{(4\,\pi\,\varepsilon_0)^2\,A\,E}\ln(1.166E/J) \qquad (2.35)$$

with the mean ionization potential $J \approx 12.5\,Z$. As energy decreases, fewer subshells are ionized, which can be described by a parabolic decrease of $1/S$ below $E/J = 6.3$.

29.2.3.3.2. Electron Diffusion

The decrease of mean electron energy E_m with increasing path length s of electron trajectories due to Equation (2.35) results in a mean total path (Bethe range R_B) that increases with increasing Z of the material. In low-Z material with less frequent large-angle scattering, R_B and the practical range R are equal, whereas in high-Z material with more frequent scattering, the trajectories are strongly curled and $R < R_B$, as demonstrated by Monte Carlo simulations in Figure 81. The range
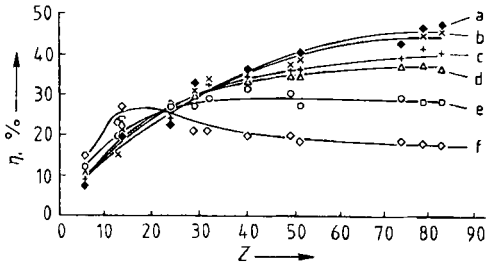
Figure 82. Backscattering coefficient η as a function of atomic number Z for different electron energies E
a) $E = 5$ keV; b) $E = 4$ keV; c) $E = 3$ keV; d) $E = 2$ keV; e) $E = 1$ keV; f) $E = 0.5$ keV

$$R \approx 6.6E^{5/3} \tag{2.36}$$

is to a large extent independent of material, when R is measured in units of mass thickness ($\mu g/cm^2$) and E in keV.

29.2.3.3.3. Emission of Secondary and Backscattered Electrons

The *backscattering coefficient* η describes the fraction of *primary electrons* leaving the specimen with an energy reduced by energy losses. Emitted electrons with energies of ≤ 50 eV are called secondary electrons (see Section 29.2.3.1) and are described by the *secondary electron yield* δ. Both quantities depend on electron energy E, atomic number Z of the specimen, and surface tilt angle ϕ ($\phi = 0$: normal incidence), and show characteristic energy and angular distributions that are important to the discussion of image formation in using BSE and SE signals.

Figure 82 shows the dependence of the *backscattering coefficient* η on atomic number Z for different electron energies E. For $E > 5$ keV, η increases monotonically with increasing atomic number. For multicomponent targets

$$\eta = \Sigma c_i \eta_i \tag{2.37}$$

shows a best fit to experiments, where c_i represents the mass fractions. This dependence of η on Z is responsible for the atomic number (compositional) contrast of the BSE signal for $E = 5 - 30$ keV (see Section 29.2.3.4.1). The backscattering coefficient η is approximately independent of E in the range $10 - 100$ keV. Below 5 keV, η decreases for $Z > 30$ and increases for $Z < 30$
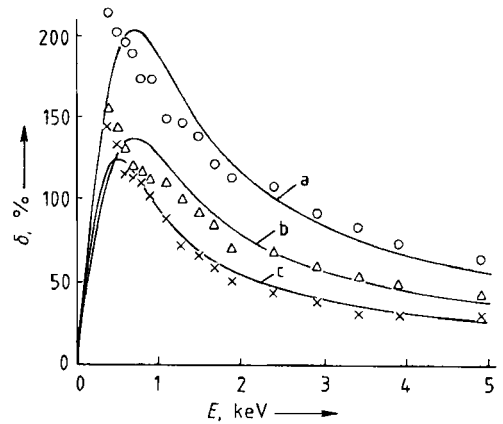
Figure 83. Decrease of SE yield δ with increasing electron energy E at normal incidence for
a) Au; b) Cu; c) C

with decreasing E. The reason for this is the Mott elastic cross section (Eq. 2.34). An increasing tilt angle ϕ of the specimen ($\phi = 0$: normal incidence) results in an increase of η.

The *secondary electron yield* δ shows a maximum for primary energies E of a few hundred electron volts and decreases $\sim E^{-0.8}$ for higher energies (Fig. 83). For $E > 5$ keV, the SE yield contains the contributions of SE1 excited by the primary electrons (PE) (Fig. 75) and of SE2 excited by the BSE. For $E < 5$ keV SE1 and SE2 can hardly be distinguished because the exit depths of SE and BSE have the same order of magnitude. The dependence of δ on the surface tilt angle ϕ can be approximated by $\delta \sim \sec^n \phi$, where the exponent n decreases from 1.3 (Be) to 1.1 (Al) and 0.65 (Au).

29.2.3.3.4. Specimen Charging and Damage

Insulating specimens show a *negative charging* by adsorbed electrons, when the total electron yield $\sigma = \eta + \delta$ is less than unity beyond a critical energy E_2 where σ changes from values larger than to values smaller than unity. Therefore, a conductive coating is necessary for working at high electron energies. With decreasing energy, σ can become greater than unity below E_2. In this case, more electrons leave the specimen as SE or BSE than are absorbed. The specimens becomes *positively charged* by a few volts only because SE of low energies will be retarded. Therefore, insulating specimens can be observed at low energies

Figure 85. Example of material contrast of an Al – Ag eutectic alloy recorded with 30-keV electrons

Figure 84. SE image of an etched structure (Y) on silicon with
A) 10-keV: B) 1-keV electrons

without coating, although rough specimens can still show negative charging when, for example, SE cannot escape from holes.

Temperature-sensitive specimens can be damaged by *heat generation*. The increase of surface temperature at the electron probe decreases with decreasing energy, and such specimens can be observed better by LVSEM. Biological specimens are damaged by ionization, which results in a loss of mass and finally a polymerized carbon-enriched conglomerate within a layer of the order of the electron range R.

29.2.3.4. Image Formation and Analysis

29.2.3.4.1. Topographic and Material Contrast

Surface topography can be imaged with the SE signal, where the contrast is generated by the dependence of δ on the tilt angle ϕ of a surface element, and by shadowing effects caused by reduced SE collection from surfaces with normals opposite to the direction of SE collection or inside

holes and trenches. This results in a contrast nearly equivalent to a light illumination from the detector, which, however, is disturbed by the diffusion contrast caused by SE2 generated by BSE at a greater distance from the electron impact. This typically results in bright zones near edges with a width of about the range R. Therefore, a low-voltage SEM at 1 – 5 keV decreases the diffusion effect and shows a better topography, as demonstrated in Figure 84. With in-lens operation and through-lens collection of SE by an Everhart – Thornley detector (ETD) inside the last lens, the SE signal becomes independent of azimuth, which is an advantage for the metrology of integrated circuits, but a disadvantage for recognizing the topography and distinguishing elevations and indentations, for example.

The dependence of δ and especially of η on atomic number Z (Fig. 82) leads to an atomic number or compositional contrast that can be used for the discrimination of phases with different mean atomic numbers, as demonstrated by a eutectic alloy shown in Figure 85.

29.2.3.4.2. Electron Channeling Effects

Electron waves entering a single crystal propagate as Bloch waves, which results in channeling effects and an orientation dependence of all electron – specimen interactions that are concentrated at the nuclei, such as large-angle scattering for backscattering or inner-shell ionization for X-ray and Auger electron emission.

When rocking an incident electron beam, the backscattering coefficient is modulated by a few percent. This results in an electron channeling
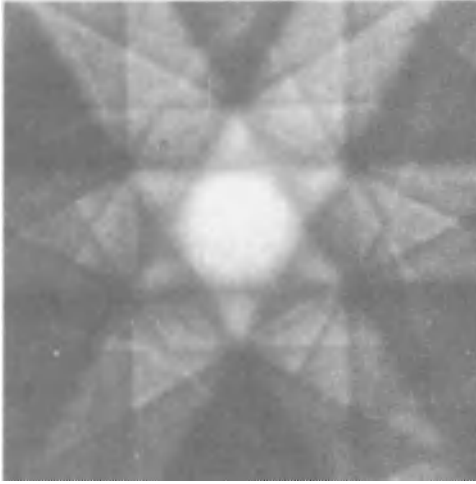
Figure 86. Electron channeling pattern of a (111) silicon surface recorded with 20-keV electrons
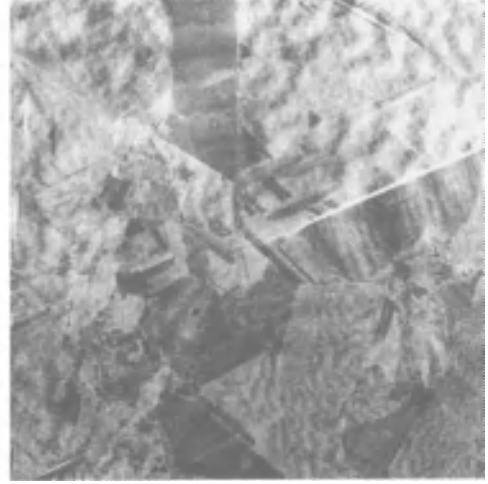
Figure 87. Superposed crystal orientation and magnetic contrast type I on polycrystalline cobalt

pattern (ECP) (Fig. 86) containg excess (bright) or defect (dark) Kikuchi lines (see Section 29.2.2.3.2) that are the intersections of the Kossel cones of apex $90° - \theta_B$ and an axis normal to the lattice planes of distance d. Therefore, such a pattern contains the crystal symmetry, and the distances of opposite Kikuchi lines include an angle $2\theta_B$, which fulfills the Bragg condition

$$2d\sin\theta_B = \lambda \tag{2.8}$$

Scanning a polycrystalline specimen results in changes of brightness forming the crystal orientation or channeling contrast (Fig. 87).

Another type of channeling pattern is the electron backscattering pattern (EBSP) in which the intensity modulation can be observed on a fluorescent screen as the dependence of backscattering on takeoff direction. An EBSP has the advantage of covering a large angular range of about $\pm 30°$, whereas the rocking for ECP can be realized only with maximum $\pm 2 - 3°$.

29.2.3.4.3. Imaging and Measurement of Surface Potentials

A positively biased or charged part of the specimen retards low-energy SE and appears darker than the surrounding parts at ground potential. Negatively biased or charged parts appear brighter because more SE are repelled and can reach the

collection field of the ETD. This results in the voltage contrast. However, no unique relation exists between brightness and bias because the SE signal depends on the surrounding potentials (near-field effect) and the collection efficiency of an ETD (far-field effect).

For quantitative measurement of surface potentials U_s on integrated circuits, a high extraction field strength (400 – 1000 V/cm) is needed at the surface, which decreases the influence of near fields by neighboring potentials. The potential is measured by the shift in the SE spectrum after passage through a spectrometer of the retarding or deflection type [226].

29.2.3.4.4. Imaging of Magnetic Fields

Magnetic fields of the specimen can act on primary, secondary, and backscattered electrons by the Lorentz force $F = e\,v \times B$. The *magnetic contrast type I* is caused by the deflection of SE by external magnetic fields, which can be observed for magnetic recording media and uniaxial ferromagnetic materials. The best contrast is observed when only about half of the SE are collected. Then the signal intensity can change by $1 - 10\%$ for opposite directions of the stray field. Figure 87 shows the superposition of magnetic contrast type 1 and the channeling contrast on a polycrystalline cobalt specimen.

Magnetic contrast type 2 is caused by the deflection of BSE in internal magnetic fields. However, sufficient contrast can be observed only for tilt angles ϕ of $40-50°$, and the contrast increases with increasing energy, but is only a few per thousand for opposite directions of *B*. Another possibility for measuring external stray fields is the deflection of primary electrons, which pass at a short distance parallel to the surface.

29.2.3.4.5. Electron-Beam Induced Current

The mean number E/E_i of electron–hole pairs generated in semiconductors — with a mean formation energy of $E_i = 3.6$ eV in silicon, for example — normally recombine. The electric field inside depletion layers separates the charge carriers, and minority carriers can diffuse to the depletion layer and contribute to the charge collection I_{cc} or electron-beam induced current. Depletion layers can be formed by $p-n$ junctions parallel or perpendicular to the surface or by Schottky barriers formed by a nonohmic evaporated metal contact. Therefore, a scanning electron probe becomes a useful tool for qualitative and quantitative analysis of junctions and semiconductor parameters [227], which is demonstrated by the following examples:

1) Imaging the position and depth of depletion layers below conductive pads and passivation layers, by utilizing the increasing penetration depth (range) with increasing electron energy
2) Measuring the width of depletion layers and their increase with increasing reverse bias
3) Imaging sites of avalanche breakdown in depletion layers
4) Imaging lattice defects (dislocations, stacking faults) and dopant striations that actively influence the diffusion length, for example, by a Cottrell atmosphere of dopant atoms
5) Measuring the diffusion length *L* from a semilogarithmic plot of the EBIC signal $[\sim \exp(-x/L)]$ versus the distance *x* of the electron probe from a perpendicular $p-n$ junction, and measuring the lifetime τ with a signal $[\sim \exp(-t/\tau)]$ when chopping the electron beam
6) Measuring the surface recombination rate S from the variation of EBIC with increasing electron energy and depth of carrier formation

29.2.3.4.6. Cathodoluminescence

The excitation of light from materials stroked by electrons, known from fluorescent screens of TV tubes for example, is called cathodoluminescence (CL). In semiconductors with a direct band gap, electrons that are excited from the valence to the conduction band can recombine with the holes by emission of radiation, whereas semiconductors with an indirect band gap have a reduced probability of radiative recombination. In semiconductors and most inorganic materials, CL depends strongly on the concentration of dopants, which can either enhance CL by forming luminescence centers for radiative transitions or quench CL by forming centers of nonradiative transitions.

Organic specimens such as anthracene and plastic scintillator material, for example, can also show CL. Fluorescent dyes are used in biology for selective staining. However, all organic material is damaged by electron irradiation, and CL is quenched at incident charge densities a few orders of magnitude lower than those necessary for radiation damage of the crystal structure and loss of mass by ionization processes.

The CL signal can be detected by a photomultiplier or dispersively by using a spectrometer between specimen and detector. As in color TV, the signals from three detectors with color filters can be used for a real-color image [228].

29.2.3.4.7. Special Imaging Methods

The *specimen current mode* uses the current

$$I_s = I_p[1 - (\eta + \delta)] \tag{2.39}$$

to earth, which is complementary to the SE and BSE emission and can be used for imaging the compositional contrast of BSE or the magnetic contrast type 2.

In *environmental* SEM, the partial pressure of water or other gases can be increased near the specimen by differentially pumped diaphragms. This offers the possibility of observing wet specimens without drying, for example [229]. The production of ions in the gas also reduces the negative charging of the specimen by absorbed electrons in nonconductive material.

In the *thermal-wave acoustical mode*, the electron beam is chopped at frequencies in the 100-kHz to 5-MHz range and produces periodic specimen heating. The periodic changes in thermal expansion excite an acoustic wave that can be
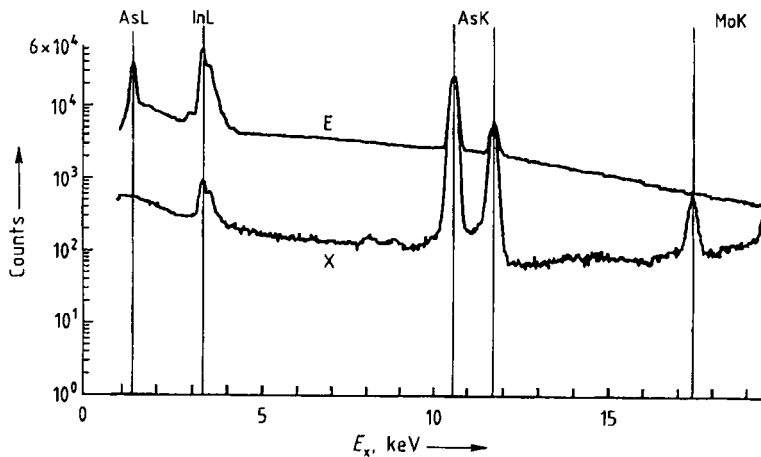
Figure 88. Energy-dispersive X-ray spectrum of InAs (logarithmic scale of counts per channel) recorded by excitation with 30-keV electrons (E) and X-ray fluorescence (X) using the molybdenum K radiation excited in a molybdenum foil target in front of the specimen

picked up by a piezoelectric crystal transducer of lead zirconate titanate.

29.2.3.5. Elemental Analysis

29.2.3.5.1. X-Ray and Auger Electron Emission

Either ionization of an inner shell and the subsequent filling of the vacancy by an electron from higher energy states result in the emission of an X-ray quantum, or the energy is transferred to another electron in the higher shell, which leaves the atom as an Auger electron (see also → Surface Analysis). The sum of the X-ray fluorescence yield ω and the Auger electron yield is unity, and ω decreases strongly from near unity to low values with decreasing Z and increasing X-ray series $n = K, L, M$. The X-ray quantum energies of a series increases approximately $\sim \sqrt{Z}$ (Moseley's law), which is basic for elemental analysis by X rays.

29.2.3.5.2. X-Ray Spectrometers

Two different types of X-ray spectrometer exist: wavelength- and energy-dispersive spectrometers (WDS and EDS, see Section 29.2.2.12.1), which can be used in either an X-ray microanalyzer (XRMA) or an SEM.

A *wavelength-dispersive spectrometer* can analyze the characteristic lines of all elements down to beryllium when special analyzing crystals with a large lattice-plane spacing are used for Bragg reflection. XRMA with electron energies between 10 and 60 keV can work with WDS for recording three different wavelengths simultaneously, whereas only one WDS is used in an SEM.

The electron–hole pairs produced in the *energy-dispersive spectrometer* are separated by a voltage drop of ca. 1 kV at the crystal and result in a charge pulse proportional to the X-ray quantum energy. This pulse is amplified by a charge-sensitive preamplifier and runs through a unit for pulse shaping so that it can be recorded in a multichannel analyzer (MCA). The number of counts in a channel that is proportional to the quantum energy is increased by one unit. This allows all quantum energies to be recorded simultaneously, and the growth of a spectrum can be followed on the screen of the MCA (Fig. 88). Therefore, EDS are often also used in XRMA to survey the emitted X-ray spectrum. The vacuum inside an EDS can be separated from the vacuum of the microscope by a 8–10-µm-thick beryllium foil, which absorbs the characteristic radiation of elements $Z \leq 11$ below $E_x = 1$ keV. An organic foil or a windowless detector can also be used to record the lines of carbon ($E_x = 250$ eV), oxygen, and nitrogen, for example.

WDS and EDS show the following characteristic differences. *WDS* requires very accurate orientation of the electron impact on the Rowland circle within an area a few micrometers in diameter. The specimen must be planar and is shifted mechanically to record a line scan or elemental

map. The electron probe can, by contrast, be scanned over an area of a few millimeters without any loss of efficiency when an EDS is used. WDS analyze only one line with a resolution of about 10 eV. The proportional counter allows a high counting rate of about 10 000 counts per second (cps). Therefore, high probe currents and a large number of counts per unit time can be recorded. The energy resolution of an *EDS* is only about 100 – 200 eV, and a ten times greater fraction of the continuum below the characteristic X-ray peaks is counted compared to WDS. This decreases the signal-to-background ratio in EDS. The poorer resolution of an EDS can also result in an overlap of neighboring X-ray lines, whereas all lines can normally be separated by a WDS. An EDS must work with lower probe currents because all quanta are counted and only a count rate of ca. 10 000 cps guarantees no overlap of sequential pulses. Pulse-rejection electronics are used when the time between two pulses is shorter than the pulse-shaping time necessary for the MCA.

29.2.3.5.3. X-Ray Microanalysis

For a quantitative XRMA [230], the number n_a of counted characteristic X-ray quanta of an element a of concentration c_a in an alloy or compound is compared with the number n_s from a pure elemental standard ($c_s = 1$) or a compound of known composition. Only an approximate value for c_a can be obtained from the k ratio

$$k = n_a/n_s \approx c_a/c_s \qquad (2.40)$$

The exact measurement of concentration needs a *ZAF correction method* (see p. 1097). This considers the differences in the stopping power and the backscattering of electrons between specimen and standard (*Z correction*), in the depth distribution of X-ray generation and *absorption* (*A*) of X rays recorded under a takeoff angle of 40 – 60°, and in the excitation of X-ray quanta of the element of interest by characteristic X rays and continuum of the matrix (*fluorescence F*).

Special correction programs must be used, for example, for tilted specimens, thin film coatings, small particles, and biological specimens.

29.2.3.5.4. Special X-Ray Techniques

When the specimen is scanned in a raster or along a line, the pulses of selected characteristic X-ray lines can produce an elemental distribution profile or map, respectively.

In a *single crystal*, the excited X rays are Bragg diffracted at the lattice plane, and their isotropic angular characteristics show defect and excess Kossel lines of apex angle $90° - \theta_B$, which can be used for accurate measurement of lattice parameters and strains when the Kossel pattern is recorded on a photographic emulsion.

X-ray fluorescence analysis can be applied in an SEM when generating X rays in a thin foil, which stops the electrons, but the transmitted X rays can excite X-ray fluorescence in the specimen, which is recorded by an energy-dispersive spectrometer. The background of the spectrum is much lower than that from direct electron excitation (Fig. 88, curve X versus curve E), which results in a better signal-to-background ratio for trace elements. However, only larger areas 0.1 – 1 mm in diameter can be analyzed.

The small source of X rays generated by an electron probe on a bulk target can also be used for X-ray projection microscopy.

29.3. References

[1] E. Abbe: "Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung," *Arch. Mikrosk. Anat.* **9** (1873) 413.

[2] D. Gerlach: *Das Lichtmikroskop*, 2nd ed., Thieme Verlag, Stuttgart 1985.

[3] H. Gundlach: *Mikroskope. Handbuch Biomedizinische Technik*, **vol. 4**, Springer Verlag, TÜV Rheinland, (1991) pp. 1 – 20.

[4] K. Michel: "Axiomat von Zeiss, ein Mikroskop mit neurem Konzept," *Zeiss. Inf.* **20** (1973).

[5] F. Muchel: ""ICS"-A New Principle in Optics," *Zeiss Inf.* **30** (1988) 20 – 27.

[6] K. Michel: *Die Grundzüge der Theorie des Mikroskops*. 3rd ed., Wissenschaftliche Verlagsgesellschaft, Stuttgart 1981.

[7] A. Köhler: "Mikrophotographische Untersuchungen mit ultraviolettem Licht," *Z. Wiss. Mikrosk. Mikrosk. Tech.* **21** (1904) 129 – 165, 273 – 404.

[8] A. Köhler: "Ein neues Beleuchtungsverfahren für mikrophotographische Zwecke," *Z. Wiss. Mikrosk. Mikrosk. Tech.* **10** (1893) 433 – 440.

[9] F. Zernike: "Das Phasenkontrastverfahren bei der mikroskopischen Beobachtung," *Phys. Z.* **36** (1935) 848.

[10] H. Gundlach: "Neuere Entwicklungen und Anwendungen in der Lichtmikroskopie," *Gynäkologie* **23** (1990) 328 – 335.

[11] G. Nomarski: "Microinterferometre Differentielle a Ondes Polarisees," *J. Phys. Radium* **16** (1955) 9.

[12] R. D. Allen et al.: "The Zeiss Nomarski Differential Interference Equipment for Transmitted-Light Microscopy," *Z. Wiss. Mikrosk. Mikrosk. Tech.* **69** (1969) 193 – 221.

[13] W. Lang: "Nomarski Differential Interference Contrast Microscopy," *Zeiss Inf.* **16** (1968) no. 70, 114–120.

[14] M. Francon: *Einführung in die neuen Methoden der Lichtmikroskopie,* Verlag Braun, Karlsruhe 1967.

[15] W. I. Patzelt: "Reflexionskontrast. Eine neue lichtmikroskopische Technik," *Mikrokosmos* **3** (1977) 78–81.

[16] S. Pentz et al.: "Darstellung des Adhäsionsverhaltens kultivierter Leberzellen an Glasoberflächen während der Mitose durch Reflexionskontrast-Mikroskopie," *Zeiss Inf.* **25** (1980) 91, 41–43.

[17] H. Gundlach et al.: "Immunfluorescance Microscopy with the New Zeiss Photomicroscope Axiophot," *Zeiss Inf.* **29** (1997) 36–39.

[18] A. J. Lacey: *Light Microscopy in Biology,* IRL Press, Oxford University, 1989.

[19] D. L. Taylor, M. Nederlof, F. Lanni, A. S. Waggoner: "The New Vision of Light Microscopy," *Am. Sci.* **80** (1992) 322–335.

[20] D. L. Taylor et al.: *Fluorescence Microscopy of Living Cells in Culture,* Academic Press, San Diego 1989.

[21] A. H. Coons et al.: "Immunological Properties of an Antibody Containing a Fluorescent Group," *Proc. Soc. Exp. Biol. Med.* **47** (1941) 200–202.

[22] J. S. Ploem: "The Use of a Vertical Illuminator with Interchangeable Dichroic Mirrors for Fluorescence Microscopy with Incident Light," *Z. Wiss. Mikrosk. Mikrosk. Tech.* **68** (1967) 129–143.

[23] L. Trapp: "Über Lichtquellen und Filter für die Fluoreszenzmikroskopie und über die Auflichtfluoreszenzmethode bei Durchlichtpräparaten," *Acta Histochemica* **7** (1965), pp. 327–338.

[24] W. W. Franke: "Different Intermediate-Sized Filaments Distinguished by Immunofluorescence Microscopy," *Proc. Nat. Acad. Sci. USA* **75** (1978) no. 10, 5034–5038.

[25] O. Greulich et al.: "The Light Microscope on its Way from an Analytical to a Preparative Tool," *J. Microsc. (Oxford)* **167** (1992) 127–151.

[26] P. Lichter et al.: "Analysis of Genes and Chromosomes by Nonisotopic in Situ Hybridization," *Gata* **8** (1991) no. 1, 24–35.

[27] Chr. Lengauer et al., "Chromosomal Barcodes produced by Multifluorescence in situ Hybridization with Multiple YAG Clones and Whole Chromosomes painting probes," in *Human Molecular Genetics,* 5th ed., vol. 2, Oxford University Press, pp. 505–512.

[28] M. Chalfie, Y. Tu, D. C. Prasher, "Green Fluorescent Protein as a Marker for Gene Expression", *Science* **263** (1994) 802–805.

[29] R. Heim, R. Y. Tsien, "Engineering Green Fluorescent Protein for Improved Brightness, Longer Wavelength and Fluorescence Resonance Energy Transfer", *Curr. Biol.* **6** (1996) 178–182.

[30] B. Herman, *Fluorescence Microscopy,* BIOS Scientific Publisher, Oxford, 1998.

[31] E. L. Elson, D. Magde, "Fluorescence Correlation Spectroscopy (I). Conceptual Basis and Theory", *Biopolymers* **13** (1974) 1–27.

[32] R. Rigler, J. Widengreen, Ü. Mets, "Interactions and Kinetics of Single Molecules as Observed by Fluorescence Correlation Spectroscopy", *Fluorescence Spectroscopy — New Methods and Applications,* Springer, Heidelberg, 1992, 13–24.

[33] M. Eigen, R. Rigler, "Sorting Single Molecules: Application to Diagnostics and Evolutionary Biotechnology", *Proc. Natl. Acad. Sci. USA* **91** (1994) 5740–5747.

[34] H. Utermöhl: "Neue Wege in der quantitativen Erfassung des Planktons," *Verh. intern. Verein Limnol.* **5** (1931) 567–596.

[35] H. Gundlach et al.: "Mikrokinematographie mit dem Mikroskop "Axiomat"," *Res. Film* **9** (1976) no. 1, 30–36.

[36] H. Gundlach: *Die Anwendung von inversen Mikroskopen in der Zell- und Entwicklungsbiologie,* Suppl. GIT, Verlag E. Giebler, Darmstadt 1981, pp. 9–15.

[37] H. A. Tritthart et al.: "The Zeiss Axiomat Applied to the Examination of the Morphology and Physiology of Myocardial Cells in Culture," *Zeiss Inf.* **25** 1980 no. 90, 10–13.

[38] M. Horster et al.: "Application of Differential Interference Contrast with Inverted Microscopes to the In-vitro Perfused Nephron," *J. Microsc. (Oxford)* **117** (1979) 375–379.

[39] F. K. Möllring: "Inverse Mikroskopie IM 35 und ICN 405 für Biologie, Medizin und Metallographie," *Zeiss Inf.* **23** (1977) 18–19.

[40] H. Piller: *Microscope Photometry,* Springer, Berlin 1977.

[41] H. Gundlach et al.: "Identification and Selective Micropreparation of Live Nuclear Components with the Zeiss IM 35 Inverted Microscope," *Zeiss Inf.* **25** (1980), no. 91, 36–40.

[42] H. Spring et al.: "DNA Contents and Numbers of Nucleoli and Pre-rRNA-Genes in Nuclei of Gemetes and Vegetative Cells of Acetabularia Mediterranea," *Exp. Cell. Res.* **114** (1978) 203–215.

[43] H. Spring et al.: "Transcriptionally Active Chromatinin in Loops of Lumpbrush Chromosomes at Physiological Salt Concentration as Revealed Electron Microscopy of Sections," *Eur. J. Cell Biol.* **24** (1981) 298–308.

[44] W. Ansorge et al.: "Performance of an Automated System for Capillary Microinjection into Living Cells," *J. Biochem. Biophys. Methods* **16** (1988) 283–292.

[45] R. D. Allen et al.: "Video-Enhanced-Contrast Polarization (AVEC-Pol) Microscopy: A New Method Applied to the Detection of Birefringence in the Motile Reticulopodial Network of Allogromia Laticollaris," *Cell Motil.* **1** (1981) 275–289.

[46] R. D. Allen et al.: "Video-Enhanced-Contrast, Differential Interference Contrast (AVEC-DIC) Microscopy: A New Method Capable of Analyzing Microtubule-Related Motility in the Reticulopodial Network of Allogromia Laticollaris," *Cell Motil.* **1** (1981) 291–302.

[47] S. Inoue: "Foundations of Confocal Scanning Imaging in Confocal Microscopy" in [60], pp. 1–14.

[48] S. Inoue et al.: "The Acrosonal Reaction of Thyone Sperm Head Visualized by High Resolution Video Microscopy," *J. Cell Biol.* **93** (1982) 812–819.

[49] J. H. Hayden et al.: "Detection of Single Microtubules in Living Cells: Particle Transport Can Occur in Both Directions Along the Same Microtubule," *J. Cell Biol.* **99** (1984) 1785–1793.

[50] T. Salmon et al.: "Video-Enhanced Differential Interference Contrast Light Microscopy," *Bio Techniques* **7** (1989) 624–633.

[51] D. J. Arndt-Jovin et al.: "Fluorescence Digital Imaging Microscopy in Cell Biology," *Science (Washington D.C.)* **230** (1985) 247–256.

[52] Y. Hiraoka et al.: "The Use of a Charge-Coupled Device for Quantitative Optical Microscopy of Biological Structures," *Science (Washington D.C.)* **238** (1987) 36–41.

[53] H. H. Sedlacek et al.: "The Use of Television Cameras Equipped with an Image Intensifier in the Immunofluorescence Microscopy," *Behring Inst. Mitt.* **59** (1976) 64–70.

[54] H. Gundlach et al.: *Electronic Photography-Technology, Systems and Applications in Microbiology. Proc. of Int. Symosium on Electronic Photography,* Cologne The Society for Imaging Sciene and Technology, Springfield, Virginia USA, 1992, pp. 60–65.

[55] M. Minsky: US 3013467, 1961.

[56] G. Cox: "Photons Under the Microscope. Marvin Minsky – the Forgotten Pioneer?," *Australian EM Newsletter* **38 no. 4,** (1993) 4–10.

[57] C. J. R. Sheppard: "15 Years of Scanning Optical Microscopy at Oxford," *Proc. Roy. Mic. Soc.* **25** (1990) 319–321.

[58] M. Petran, M. Hadravsky, M. D. Egger, R. Galambos: "Tandem-Scanning Reflected Light Microscope," *J. Opt. Soc. Am.* **58** (1968) 661–664.

[59] G. J. Brakenhoff, P. Blom, P. Barends: "Confocal Scanning Light Microscopy with High Aperture Immersion Lens," *J. Micros. (Oxford)* **117** (1979) 219–232.

[60] J. Pawley: *Handbook of Biological Confocal Microscopy.* 2nd ed.Plenum Press, New York 1995.

[61] C. J. R. Sheppard, A. Choundhury: "Image Formation in the Scanning Microscope," *Opt. Acta* **24** (1977) 1051–1073.

[62] C. J. R. Sheppard, T. Wilson: "Depth of Field in the Scanning Microscope," *Opt. Lett.* **3** (1978) 115–117.

[63] P. Wallen, K. Carlsson, K. Mossberg: "CLSM as a Tool for Studying the 3-D Morphology of Nerve Cells," in [109], pp. 110–143.

[64] T. Wilson: "Optical Sectioning in Confocal Fluorescence Microscopy," *J. Microsc. (Oxford)* **154** (1989) 143–156.

[65] S. Hell, E. Lehtonen, E. H. K Stelzer: "Confocal Fluorescence Microscopy: Wave Optics and Applications to Cell Biology", in [28], pp. 145–160.

[66] I. J. Cox, C. J. R. Sheppard, T. Wilson: "Superresolution in Confocal Fluorescent Microscopy," *Optik (Stuttgart)* **60** (1982) 391–396.

[67] T. Wilson (ed.): *Confocal Microscopy.* Academic Press, London 1990.

[68] C. J. R. Sheppard, M. Gu: "3-D Transfer Functions in Confocal Scanning Microscopy," in [109], pp. 251–280.

[69] M. Gu, *Principles of 3D Imaging in Confocal Microscopes,* World Scientific, Singapore, 1996.

[70] C. J. Cogswell, J. W. O'Bryan: "A High Resolution Confocal Transmission Microscope," *SPIE Proceedings* **1660** (1992) 503–511.

[71] G. J. Brakenhoff, K. Visscher: "Bilateral Scanning and Array Detectors," *J. Micros. (Oxford)* **165** (1990) 139–146.

[72] E.Gratten, M. J. van der Veen: "Laser Sources for Confocal Microscopy", in [60], pp. 69–98.

[73] K. S. Wells, D. R. Sandison, J. Strickler, W. W. Webb: "Quantitative Fluorescence Confocal Laser Scanning Microscopy," in [60], pp. 39–53.

[74] J. Art: "Photon Detectors for Confocal Microscopy," in [60], pp. 127–139.

[75] W. Denk, J. H. Strickler, W. W. Webb: "Two-photon laser scanning fluorescence microscopy", *Science (Washington D.C.)* **248** (1990) 73–76.

[76] M. Gu, X. Gan, A. Kisteman, M. G. Xu, *Appl. Phys. Lett.* **77** (2000) no. 10, 1551–1553.

[77] S. Lindek, E. H. K. Stelzer, S. W. Hell: "Two New High-Resolution Confocal Fluorescence Microscopies (4pi, theta) with One- and Two-Photon Excitation", in [60], pp. 445–458.

[78] M. D. Egger, M. Petran: "New Reflected Light Microscope for Viewing Unstained Brain and Ganglian Cells," *Science (Washington D.C.)* **157** (1987) 305–307.

[79] J. S. Deitch, K. L. Smith, J. W. Swann, J. N. Turner: "Parameters Affecting Imaging of HRP Reaction Product in the Confocal Scanning Laser Microscope," *J. Microsc. (Oxford)* **160** (1990) 265–278.

[80] J. P. Rigaut, S. Caravajal-Gonzalez, J. Vassy: "3-D Image Cytometry," in [109], pp. 205–237.

[81] B. R. Masters, A. Kriete, J. Kukulies: "Ultraviolet Confocal Fluorescence Microscopy of the In-Vitro Cornea: Redox Metabolic Imaging," *Applied Optics* **32** (1993) no. 4, 592–596.

[82] D. M. Shotton: "Confocal Scanning Optical Microscopy and its Applications for Biological Specimens," *J. Cell. Sci.* **94** (1989) 175–206.

[83] J. N. Turner: "Confocal Light Microscopy. Biological Applications (Special Issue)," *Electron Microsc. Tech.* **18** (1991) 1.

[84] R. P. Haugland: *Molecular Probes: Handbook of Fluorescent Probes and Research Chemicals,* Eugene Molecular Probes Inc., 1989.

[85] R. Y. Tsien, A. Waggoner: "Fluorophores for Confocal Microscopy: Photophysics and Photochemistry," in [60], pp. 169–178.

[86] A. Villringer, U. Dirnagl, K. Einhäupl: "Microscopical Visualization of the Brain in Vivo," in [109], pp. 161–181.

[87] R. Yuste, F. Lanni, A. Konnerth (eds.): *Imaging Neurons: A Laboratory Manual,* CSHL Press, Cold Spring Harbor, 2000.

[88] B. R. Masters: "Confocal Ocular Microscopy – a New Paradigm for Ocular Visualization," in [59], pp. 183–203.

[89] W. A. Mohler, J. G. White, "Stereo-4D Reconstruction and Animation from Living Fluorescent Specimens", *Biotechniques* **24** (1998) 1006–1012.

[90] R. G. Summer, P. C. Cheng: "Analysis of Embryonic Cell Division Patterns Using Laser Scanning Confocal Microscopy," in G. W. Bailey (ed.): *Proc. Annual meeting Electron Microsc. Soc. Am.* **47** (1989) 140–141.

[91] N. O'Rourke, S. E. Fraser: "Dynamic Changes in Optic Fiber Terminal Arbors Lead to Retinotopic Map Formation: An In-Vivo Confocal Microscopic Study," *Neuron* **5** (1990) 159–171.

[92] A. Kriete, H.-J. Wagner, "Computerized Spatio-temporal (4D) Representation in Confocal Microscopy: Application to Neuroanatomical Plasticity", *J. Micros. (Oxford)* **169** (1993) 27–31.

[93] I. T. Young, "Characterizing the Imaging Transfer Function", in D. L. Taylor, Y. L. Wang (eds.): *Methods in Cell Biology,* Academic Press, San Diego, 1989, pp. 1–45.

[94] R. A. Jarvis, "Focus Optimization Criteria for Computer Image Processing", *The Microscope* **24** (1976) 163–180.

[95] A. Kriete, "Image Quality Considerations in Computerized 2D and 3D Microscopies", in P. C. Cheng, T. H. Lin, W. H. Wu, J. L. Wu (eds.): *Multidimensional Microscopy,* Springer, New York, 1990, pp. 141–150.

[96] I. T. Young, Quantitative Microscopy, IEEE Eng. in Medicine and Biology, Jan./Feb. 1996, pp. 59 – 66.

[97] T. Visser, J. L. Oud, G. T. Brakenhoff: "Refractive Index and Axial Distance Measurements in 3-D Microscopy," *Optik (Stuttgart)* **90** (1991) 17 – 19.

[98] J.-A. Conchello, E. W. Hanssen: "Enhanced 3-D Reconstruction from Confocal Scanning Microscope Images. Deterministic and Maximum Likelihood Recon-strauctions," *App. Opt.* **29** (1990) no. 26, 3795 – 3804.

[99] T. J. Holmes, Y.-H. Liu: "Image Restoration for 2-D and 3-D Fluorescence Microscopy," in [109], pp. 283 – 323.

[100] D. A. Agard, J. W. Sedat: "Three-Dimensional Architecture of Polytene Nucleus," *Nature (London)* **302** (1983) 676 – 681.

[101] G. Wang, W. S. Liou, T. H. Lin, P. C. Cheng: "Image Restoration in Light Microscopy," in T. H. Lin, W. L. Wu, J. L. Wu (eds.): *Multidimensional Microscopy*, Springer Verlag, New York 1993, 191 –208.

[102] F. Meyer: "Mathematical Morphology: 2-D to 3-D," *J. Microsc.* **165** (1992) 5 – 28.

[103] P. C. Cheng et al.: "3-D Image Analysis and Visualization in Light Microscopy and X-Ray Micro-Tomography," in [109], pp. 361 – 398.

[104] K. Mossberg, U. Arvidsson, B. Ulfhake: "Computerized Quantification of Immunoflourescence Labelled Axon Terminals and Analysis of Co-Localization of Neurochemicals in Axon Terminals with A Confocal Scanning Laser Microscope," *J. Histochem Cyctochem.* **38** (1990) 179 – 190.

[105] B. Roysam et al.: "Unsupervised Noise Removal Algorithms for 3-D Confocal Fluorescence Microscopy," *Micron and Micros. Acta* **23** (1992) in press.

[106] M. Montag et al.: "Methodical Aspects of 3-D Reconstruction of Chromatin Architecture in Mouse Trophoblast Giant Nuclei," *J. Microsc. (Oxford)* **158** (1990) 225 – 233.

[107] V. Conan et al.: "Geostatistical and Morphological Methods Applied to 3-D Microscopy," *J. Microsc. (Oxford)* **166** (1992) 169 – 184.

[108] N. S.White: "Visualization Systems for Multidimensional CLSM Systems", in [60] pp. 211 – 254.

[109] A. Kriete: *Visualization in Biomedical Microscopies 3-D Imaging and Computer Applications*, VCH-Verlagsgesellschaft, Weinheim 1992.

[110] H. Chen, J. W. Sedat, J. A. Adard: "Manipulation, Display and Analysis of Three-Dimensional Biological Images," in J. Pawley (ed.): *Handbook of Biological Confocal Microscopy*, Plenum Press, New York 1990, pp. 141 – 150.

[111] J. S. Hersh: "A Survey of Modeling Representations and their Application to Biomedical Visualization and Simulation," Conf. VBC, IEEE Comp. Soc. Press, Los Alamitos 1990, pp. 432 – 441.

[112] D. P. Huijsmanns, W. H. Lamers, J. A. Los, J. Strackee: "Toward Computerized Morphometric Facilities: a Review of 58 Software Packages for Computer-Aided 3-D Reconstruction, Quantification, and Picture Generation from Parallel Serial Sections," *Anat. Rec.* **216** (1986) 449 – 470.

[113] A. Kriete, P. C. Cheng (eds.): "3-D Microscopy (Special Issue)," *Computerized Medical Imaging and Graphics*, vol. 17, Plenum Press, N.Y. 1993, p. 8.

[114] R. Ware, V. LoPresti: "Three-Dimensional Reconstruction from Serial Sections," *Comp. Graph.* **2** (1975) 325 – 440.

[115] J. K. Udupa, H. M. Hung: "Surface Versus Volume Rendering: a Comperative Assessment," *VBC '90*, Proceedings IEEE, Atlanta 1990, pp. 83 – 91.

[116] S. D. Roth: "Ray-Casting for Solid Modeling," *Comput. Graphics Image Processing* **18** (1982) 109 – 144.

[117] H. K. Tuy, L. T. Tuy: "Direct 2-D Display of 3-D Objects," *IEEE CG & A* **4** (1984) 29 – 33.

[118] G. Frieder, D. Gordon, R. A. Reynolds: "Back-to-Front Display of Voxel-Based Objects," *IEEE CG & A* **5** (1985) no. 1, 52 – 60.

[119] R. A. Debrin, L. Carpenter, P. Hanrahan: "Volume Rendering," *Computer Graphics* **22** (1988) no. 4, 65 – 74.

[120] M. Levoy: "A Hybrid Ray-Trace for Rendering Polygon and Volume Data," *IEEE CG & A* **3** (1990) 33 – 40.

[121] H. T. M. van der Voort, G. J. Brakenhoff, M. W. Baarslag: "Three-Dimensional Visualization Methods," *J. Microsc. (Oxford)* **153** (1989) no. 2, 123 – 132.

[122] P.-O. Forsgren: "Visualization and Coding in Three-Dimensional Image Processing," *J. Microsc. (Oxford)* **159** (1990) no. 2, 195 – 202.

[123] M. Levoy et al.: "Volume Rendering in Radiation Treatment Planning," *VBC '90*, Proceedings IEEE, Atlanta 1990, pp. 4 – 10.

[124] A. Kriete, N. Klein, L. C. Berger, "Data Compression in Microscopy: a Comparative Study", *Proc. SPIE*. **3605** (1999) 158 – 168.

General References

[125] S. Amelinckx: "The Direct Observation of Dislocations," Suppl. 6 in F. Seitz, D. Turnbull (eds.): *Solid State Physics*, Academic Press, London 1964.

[126] S. Amelinckx, R. Gevers, J. Van Landuyt (eds): *Diffraction and Imaging Techniques in Material Science*, North-Holland Publishing Company, Amsterdam 1970, 1978.

[127] F. R. N. Nabarro (ed.): *Dislocation in Solids*, North-Holland, Amsterdam 1979.

[128] P. B. Hirsch, R. B. Nicholson, A. Howie, D. W. Pashley, M. J. Whelan: *Electron Microscopy of Thin Crystals*. Butterworths, London 1965.

[129] H. Bethge, J. Heydenreich (eds.): *Electronenmikroskopie in der Festkörperphysik*, Springer Verlag, Berlin 1982.

[130] J. C. H. Spence: "Experimental High Resolution Electron Microscopy," *Monographs on the Physics and Chemistry of Materials*, Oxford Science Publications, Clarendon Press, Oxford 1981.

[131] G. Thomas: *Transmission Electron Microscopy of Metals*, John Wiley and Sons, New York 1962.

[132] J. C. H. Spence, J. M. Zuo: *Electron Microdiffraction*, Plenum Press, New York 1992.

[133] R. W. Cahn, P. Haasen, E. J. Kramer (eds.): *Materials Science and Technology*, vol. 2 A, VCH Verlagsgesellschaft, Weinheim 1992.

[134] D. C. Joy, A. D. Romig, Jr., J. I. Goldstein (eds.): *Principles of Analytical Electron Microscopy*, Plenum Press, New York 1986.

[135] L. Reimer: "Transmission Electron Microscopy," *Springer Series in Optical Sciences*, 4th ed., Springer Verlag, Berlin 1997.

[136] J. M. Cowley (ed.): "Electron Diffraction Techniques," vols. 1 and 2, *Monographs on Crystallography 3*, International Union of Crystallography, Oxford University Press, Oxford 1992.

[137] P. G. Merli, M. Vittori Antisari (eds.): *Electron Microscopy in Materials Science,* World Scientific, Singapore 1992.

Specific References

[138] S. Amelinckx, D. Van Dyck, J. Van Landuyt, G. Van Tendeloo (eds.): *Handbook of Microscopy.* VCH, Weinheim 1997.

[139] N. F. Mott, H. S. W. Massey: *The Theory of Atomic Collisions,* Clarendon Press, Oxford 1949.

[140] J. A. Ibers, B. K. Vainshtein: "Scattering Amplitudes for Electrons," in K. Londsdale (ed.): *International Tables for X-Ray Crystallography,* vol. 3, Kynoch, Birmingham 1962.

[141] in [133], p. 250.

[142] W. L. Bragg, *Nature (London)* **124** (1929) 125.

[143] P. P. Ewald, *Ann. Phys. (Leipzig)* **54** (1917) 519.

[144] in [133], p. 251.

[145] S. Takagi, *Acta Crystallogr.* **15** (1962) 1311.

[146] R. D. Heidenreich, *J. Appl. Phys.* **20** (1949) 993.

[147] S. Kikuchi, *Jpn. J. Phys.* **5** (1928) 23.

[148] in [125], p. 125.

[149] H. A. Bethe, *Ann. Phys. (Leipzig)* **87** (1928) 55.C. H. MacGillavry, *Physica (Amsterdam)* **7** (1940) 329.

[150] A. Howie, M. J. Whelan, *Proc. Roy. Soc. London A,* **263** (1961) 217.A. Howie, M. J. Whelan, *Proc. Roy. Soc. London A,* **267** (1962) 206.

[151] H. Yoshioka, *J. Phys. Soc. Jpn.* **12** (1957) 628.H. Hashimoto, A. Howie, M. J. Whelan, *Proc. Roy. Soc. London A* **269** (1962) 80.

[152] G. Bormann, *Z. Phys.* **42** (1941) 157.G. Bormann, *Z. Phys.* **127** (1950) 297.

[153] H. Hashimoto, M. Mannami, T. Naiki, *Philos. Trans. R. Soc. London A* **253** (1961) 459.J. W. Menter, *Proc. Roy. Soc. London A* **236** (1956) 119.

[154] G. Van Tendeloo, S. Amelinckx, *Acta Crystallogr. Sect. A: Cryst. Phys. Diffr. Theor. Gen. Crystallogr.* **A 30** (1974) 431.

[155] S. Amelinckx, J. Van Landuyt, in [126], p. 107.H. Hashimoto, M. J. Whelan, *J. Phys. Soc. Jpn.* **18** (1963) 1706.P. B. Hirsch, A. Howie, M. J. Whelan, *Philos. Trans. R. Soc. London A* **252** (1960) 499.]P. B. Hirsch, A. Howie, M. J. Whelan, *Philos. Mag.* **7** (1962) 2095.

[156] C. M. Drum, M. J. Whelan, *Philos. Mag.* **11** (1965) 205.]J. Van Landuyt, R. Gevers, S. Amelinckx, *Phys. Status Solidi* **7** (1964) 519.

[157] G. A. Bassett, J. W. Menter, D. W. Pashley, *Proc. Roy. Soc. London A* **246** (1958) 345.]D. W. Pashley, J. W. Menter, G. A. Bassett, *Nature (London)* **179** (1957) 752.

[158] H. Hashimoto, R. Uyeda, *Acta Crystallogr.* **10** (1957) 143.

[159] R. Gevers, *Philos. Mag.* **7** (1963) 769.

[160] D. J. H. Cockayne, 1. L. E. Ray, M. J. Whelan, *Philos. Mag.* **20** (1969) 1265.]D. J. H. Cockayne, M. J. Jenkins, I. L. E. Ray, *Philos. Mag.* **24** (1971) 1383.]R. de Ridder, S. Amelinckx, *Phys. Status Solidi B* **43** (1971) 541.

[161] S. Amelinckx, P. Delavignette, *J. Appl. Phys.* **33** (1962) 1458.

[162] P. Humble, in [126], p. 315.P. Humble, *Aust. J. Phys.* **21** (1968) 325.A. K. Head, *Aust. J. Phys.* **20** (1967) 557.A. K. Head, P. Humble, L. M. Clarebrough, A. T. Morton, G. T. Forwood: "Computed Electron Micrographs and Defect Identification," in: S. Amelinckx,

P. Gevers, J. Nihoul (eds.): *Defects in Crystalline Solids,* vol. 7, North-Holland, Amsterdam 1973.

[163] in [133], p. 58.

[164] C. Boulesteix, J. Van Landuyt, S. Amelinckx, *Phys. Status Solidi A* **33** (1976) 595.

[165] G. Van Tendeloo, J. Van Landuyt, S. Amelinckx, *Phys. Status Solidi A* **33** (1976) 723.M. Snijkers, R. Serneels, P. Delavignette, R. Gevers, S. Amelinckx, *Chryst. Lattice Defects* **3** (1972) 99.

[166] R. Gevers, J. Van Landuyt, S. Amelinckx, *Phys. Status Solidi* **11** (1965) 689.M. J. Goringe, U. Valdré, *Proc. R. Soc. London A* **295** (1966) 192.

[167] M. F. Ashby, L. M. Brown, *Philos. Mag.* **8** (1963) 1083, 1649.

[168] J. W. Steeds, E. Carlino, in [137], p. 279.

[169] P. Goodman, *Acta Crystallogr. Sect. A: Cryst. Phys. Diffr. Theor. Gen. Crystallogr.* **31** (1975) 793;31 (1975) 804.M. Tanaka, *J. Electron Microsc. Tech.* **13** (1989) 27.J. W. Steeds, R. Vincent, *J. Appl. Crystallogr.* **16** (1983) 317.

[170] B. F. Buxton, J. A. Eades, J. W. Steeds, G. M. Rackham, *Philos. Trans. R. Soc. London* **A 281,** 171.J. Eades, M. Shannon, B. Buxton in O. Johari (ed.): *Scanning Electron Microscopy,* I.I.I.R. Institute, Chicago 1983, p. 83.

[171] M. Tanaka, M. Terauchi, T. Kaneyama: *Convergent Beam Electron Diffraction I and II,* Jeol Ltd., Tokyo.

[172] N. S. Blom, E. W. Schapink, *J. Appl. Crystallogr.* **18** (1985) 126.

[173] J. Gjønnes, A. F. Moodie, *Acta Crystallogr.* **19** (1965) 6567.P. Goodman, G. Lempfuhl, *Z. Naturforsch. A* **19** (1964) 818.

[174] W. Kossel, G. Möllestedt, *Ann. Phys. Leipzig* **36** (1939) 113.C. H. MacGillavray, *Physica (Amsterdam)* **7** (1940) 329.

[175] G. R. Grinton, J. M. Cowley, *Optik (Stuttgart)* **34** (1971) 221.J. M. Cowley, S. Iijima: "The Direct Observation of Crystal Structures," in H. R. Wenk (ed.): Electron Microscopy in Mineralogy. Springer Verlag, Berlin 1976, p. 123.

[176] S. Amelinckx in [133], p. 5.

[177] O. Scherzer, *J. Appl. Phys.* **20** (1949) 20.

[178] E. Abbe, *Arch. Mikrosk. Anat.* **9** (1873) 413.

[179] D. van Dyck, M. Op de Beeck, *Proc. Int. Conf. Electron Microsc.* **12th** (1990) vol. 1, 64. D. van Dyck, W. Coene, *Ultramicroscopy* **15** (1989) 29. D. van Dyck, M. Op de Beeck, *Proc. Int. Conf. Electron Microsc.* **12th** (1990) 26.

[180] D. Van Dyck: "High Resolution Electron Microscopy" in [138] vol. I, Chap. 1.1.2, p. 353.

[181] J. M. Cowley, A. F. Moodie, *Acta Crystallogr.* **10** (1957) 609.

[182] D. van Dyck et al. in W. Krakow, M. O'Keefe (eds.): *Computer Simulation of Electron Microscope Diffraction and Images,* The Minerals, Metals and Materials Society, 1989, p. 107.

[183] J. M. Cowley: "Scanning Transmission Electron Microscopy" in [138] vol. II, Chap. 2.2, p. 563.

[184] S. J. Pennycook, D. E. Tesson, P. D. Nellist, M. F. Chisholm, N. D. Browning: "Scanning Transmission Electron Microscopy: Z-Contrast" in [138] vol. II, Chap. 2-3, p. 595.

[185] H. G. J. Moseley, *Philos. Mag.* **26** (1913) 1024;27 (1914) 703.

[186] E. L. Hall, in [133], p. 147.

[187] M. H. Jacobs, J. Baborovska: *Electron Microscopy,* The Institute of Physics, London 1972, p. 136.

K. F. J. Heinrich: *Electron Beam X-Ray Microanalysis*, Van Rostrand, New York 1981.

[188] S. J. B. Reed: *Electron Microprobe Analysis*, Cambridge University Press, London 1975.

[189] J. C. Spence, H. Tafto, *J. Microsc. (Oxford)* **130** (1983) 147.

[190] C. Colliex: "Electron Energy Loss Spectrometry Imaging" in [138] vol. 1, Chap. 1.3, p. 425.

[191] B. Jouffrey, in [137], p. 363.

[192] R. D. Leapman, L. A. Grunes, P. L. Fejes, J. Silcox, in B. K. Teo, D. C. Joy (eds.): *EXALFS Spectroscopy*, Plenum Press, New York 1981, p. 217.

[193] G. Zanchi, J. P. Perez, J. Sevely, *Optik* **43** (1945) 495.W. Pejas, H. Rose, *Electron Microscopy*, Microscopic Society of Canada, Toronto 1978, p. 44.
H. T. Pearce-Percy, D. Krahl, J. Jeager in D. G. Brandon (ed.): *Electron Microscopy*, **vol. 1**, Tal International, Jerusalem 1976, p. 348.

[194] G. Van Tendeloo, M. Op de Beeck, S. Amelinckx, J. Bohr, W. Krätschmer, *Europhys. Lett.* **15** (1991) no. 3, 295.

[195] S. Kuypers et al., *J. Solid State Chem.* **73** (1988) 192.

[196] S. Amelinckx, *Chim. Scr.* **14** (1978/1979) 197.

[197] J. van Landuyt, R. de Ridder, R. Gevers, S. Amelinckx, *Mater. Res. Bull.* **5** (1970) 353.

[198] G. van Tendeloo, S. Amelinckx, *Phys. Status Solidi A* **43** (1977) 553;**49** (1978) 337;**65** (1981) 431.

[199] S. Muto, G. Van Tendeloo, S. Amelinckx, *Philos. Mag.*, in press.

[200] G. van Tendeloo et al., *J. Phys. Chem.* **96** (1992) 7424.

[201] G. Van Tendeloo et al., *Europhys. Lett.* **21** (1993) 329. M. A. Verheijen et al., *Chem. Phys.* **166** (1992) 287.

[202] D. Schechtman, I. Blech, D. Gratias, J. W. Cahn, *Phys. Rev. Lett.* **53** (1984) 1951.

[203] C. Janot: "Quasicrystals, A Primer," *Monographs on the Physics and Chemistry of Materials*, Oxford Science Publishers, Oxford 1992.

[204] R. Penrose, *Bull. Inst. Math. Appl.* **10** (1974) 266.

[205] A. Ourmazd, D. W. Taylor, J. Cunningham, C. W. Tu, *Phys. Rev. Lett.* **62** (1989) no. 8, 933.A. Ourmazd, R. H. Baumann, M. Bode, Y. Kim, *Ultramicroscopy* **34** (1990) 237.

[206] S. Iijima, *Nature (London)* **354** (1991) 56.S. Iijima, T. Ichihashi, Y. Ando, *Nature (London)* **356** (1992) 776.

[207] X. F. Zhang et al., *J. Cryst. Growth* **130**, (1993) p. 36 ff. (1993).

[208] D. Ugarte, *Nature (London)* **359** (1992) 707.

[209] G. Van Tendeloo, H. W. Zandbergen, S. Amelinckx, *Solid State Commun.* **63** (1987) no. 5, 389–393;**63** (1987) no. 7, 603–606.H. W. Zandbergen, G. Van Tendeloo, T. Okabe, S. Amelinckx, *Phys. Status Solidi A* **103** (1987) 45–72.

[210] T. Krekels et al., *Appl. Phys. Lett.* **59** (1991) 3048.T. Krekels et al., *Solid State Commun.* **79** (1991) 607–614.

[211] T. Krekels et al., *Physica C* **210** (1993), 439–446.

[212] T. Krekels et al., *J. Solid State Chem.* (1993), in press.

[213] L. C. Nistor, G. Van Tendeloo, S. Amelinckx, *J. Solid State Chem.* **105** (1993), 313–335.

[214] X. F. Zhang, personal communication.

[215] F. Watari, J. van Landuyt, P. Delavignette, S. Amelinckx, *J. Solid State Chem.* **29** (1979) 137–150.F. Watari, P. Delavignette, S. Amelinckx, *J. Solid State Chem.* **29** (1979) 417–427.

[216] K. Yagi, in [136], vol. 2, p. 261.K. Yagi, *J. Appl. Crystallogr.* **20** (1987) 147.T. Hasegawa et al., *The Structure of Surfaces II*, Springer Verlag, Berlin 1987, p. 43.

[217] H. W. Zandbergen et al., *Physica C* **158** (1989) 155.H. W. Zandbergen, W. A. Groen, F. C. Mijlhoff, G. Van Tendeloo, S. Amelinckx, *Physica C* **151** (1988) 325.

[218] H. Fujita (ed.): *In-situ Experiments with High Voltage Microscope*, Research Center for High Voltage Electron Microscopy, Osaka 1985.

[219] L. Reimer: "Scanning Electron Microscopy. Physics of Image Formation and Microanalysis," *Springer Ser. in Opt. Sciences*, vol. 45, 2nd ed., Springer Verlag, Berlin 1998.

[220] O. C. Wells: *Scanning Electron Microscopy*, McGraw-Hill, New York 1974.

[221] D. B. Holt, M. D. Muir, P. R. Grant, I. M. Boswarva: *Quantitative Scanning Electron Microscopy*, Academic Press, London 1974.

[222] J. B. Pawley: "LVSEM for High Resolution Topographic and Density Contrast Imaging," *Adv. Electron. Electron. Phys.* **83** (1992) 203–274.

[223] L. Reimer: *Image Formation in Low-Voltage Scanning Electron Microscopy*, SPIE Press, Bellingham 1993.

[224] H. Raether: "Excitation of Plasmons and Interband Transitions by Electrons," *Springer Tracts Mod. Phys.* **88**, Springer, Berlin 1980.

[225] R. F. Egerton: *Electron Energy-Loss Spectroscopy in the Electron Microscope*, Plenum Publishing, New York 1986.

[226] E. Menzel, E. Kubalek, *Scanning* **5** (1983) 151–171.

[227] D. B. Holt, D. C. Joy: *SEM Microcharacterization of Semiconductors*, Academic Press, London 1989.

[228] G. V. Saparin: "Cathodoluminescence" in P. W. Hawkes, U. Valdrè (eds.): *Biophysical Electron Microscopy*, Academic Press, London 1990, pp. 451–478.

[229] G. D. Danilatos: "Foundations of Environmental Scanning Electron Microscopy," *Adv. Electron. Electron. Phys.* **78** (1988) 1–102.

[230] K. F. J. Heinrich: *Electron-Beam X-Ray Microanalysis*, Van Nostrand-Reinhold, New York 1981.

30. Techniques for DNA Analysis

Wilson J. Wall, Kidderminster, UK

30.1. Introduction

DNA can be viewed as a biological polymer made from very simple building blocks, but resulting in large-scale complexity. It is the complexity which lends this remarkable molecule its power to control cells, tissues, and ultimately organisms. It is also this complexity which is looked at when DNA is analyzed, for whatever reason and in whatever way. The length of the molecule gives a sense of scale. For example, each human chromosome contains DNA varying between 1.4 and 7.3 cm in length, depending on the chromosome. It should always be remembered, however, that this is a biological molecule and manipulation of it can have profound results, both physically and ethically.

30.1.1. DNA Structure

DNA is made up of a chain of nucleotides, with each nucleotide being made up of a deoxyribose sugar, a phosphate group, and a base. Variety within DNA stems from the bases, these can be either adenine or guanine (purines), or, cytosine or thymidine (pyrimidine). The phosphates and sugars provide the external backbone, from which the bases project. The backbone is constructed of sugars attached to each other by a phosphate ester being formed between the 3'-hydroxyl group of one sugar with the 5'-phosphate group on the next sugar, the DNA molecule is therefore polarized 5' to 3'.

The double helix structure is formed by two strands of DNA pairing in opposite polarities. The phosphate-sugar backbone is on the outside with the bases projecting into the middle. While the components of the individual strands are held together by covalent bonds, the two strands are held

Adenine   Thymine
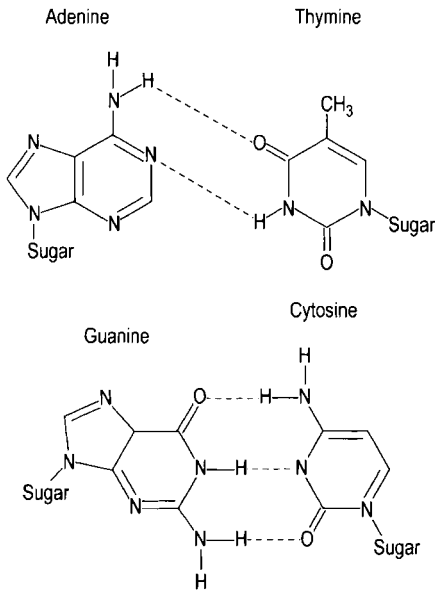
Guanine

Cytosine

Figure 1. Normal base pairing found in DNA between adenine and thymine and between cytosine and guanine

Table 1. DNA content in pictograms of a range of species showing the wide variation that is found

Common name	Species	Nuclear DNA content, pg
Saccharomyces cerevisiae	Yeast	0.026
Drosophila	Fruit fly	0.1
Mus musculus	Mouse	2.5
Homo sapiens	Man	3.7
Avena sativa	Bread wheat	18.1
Protopterus	Lungfish	50
Fritillaria davisii	Lily	98.4

together by hydrogen bonds formed between the bases.

There are four bases associated with DNA, which are generally referred to by their initial letters. These are adenine (A), cytosine (C), thymidine (T), and guanine (G). Throughout this chapter the initial letters will be used, rather than the complete name. These bases key together from opposite strands in a very particular way. A will only pair with T and G will only pair with C. The association between these bases is controlled by the formation of hydrogen bonds, as shown in Figure 1.

It is the delicacy of the hydrogen bonds compared with covalent bonds which lends both functional performance and elegance to DNA. The conformation of the double stranded helix can be visualized as what you would see when viewing a right handed spiral staircase, with 10 steps (bases) per complete revolution and a diameter of approximately 2 nm. As a consequence of this structure it is not possible to separate the two strands without unwinding the helix.

It is a useful convention when dealing with DNA to ignore the sugars and phosphates and merely refer to the bases of a single strand. This is because the exact bonding of one strand with another allows us to determine, precisely, the complementary sequence of any given single strand.

Consequentially, when a DNA sequence is written out, it is conventional only to list the base sequence of a single strand, the other one being explicitly determined. It should always be remembered that genetic complexity is not directly related to DNA content of a cell. This can be broadly seen in Table 1.

The huge differences in DNA content are generally associated with variations in the amount of repeated sequences present. The complexity of the genome, that it the proportion of repeats compared with a unique sequence in any given organism is determined using Cot curves, Cot being the concentration (Co) of DNA in moles of nucleotide per liter × renaturation time (t) in seconds. The calculations for working out Cot values are complex, but are always presented in a uniform way. The values are plotted on a log scale resulting in a curved plot whose slope gives an indication of the level of repetition present in the DNA sample. Cot curves are based on the observation that DNA can be denatured by heat and will reanneal when the temperature is dropped. This observation also forms an important part of many other methods of DNA analysis.

By cutting DNA into short sections of about 400 – 600 bp (base pairs) and then denaturing these fragments the highly repetitious sequences will reanneal first, simply because the repeat sequences are more likely to come into contact with complementary sequence areas. Therefore the sequences are complementary. The more complex the genome is the longer it will take low copy, or single-copy sequences to find the right partner.

30.1.2. Structure of DNA in Life

DNA does not sit in cells in an uncontrolled manner. It is regulated in both its position and expression. These are carried out by proteins, ge-

netically silent, but of huge importance. The proteins involved in structural control of DNA are some of the most highly conserved found in the living world. While it is not the remit of this chapter to detail the manner in which DNA is controlled within the cell, it is worth noting the first-order structure of chromatin, this being the term used to describe the DNA – protein complex.

Using osmotic shock it is possible to unwind chromatin to show the fundamental unit as a beaded string of about 10 nm in diameter. The beads are histone proteins with the DNA double helix wrapped around them. Histone proteins are almost exclusively associated with nuclear DNA. There are essentially five different histones associated with DNA, designated H1, H2A, H2B, H3, and H4. These have been highly conserved throughout evolutionary time, indicating their fundamental importance to the integrity and control of DNA. Changes to histone proteins by methylation, phosphorylation, etc., have effects on the charge of the protein, which can alter the interaction between histone and DNA. For example, acetylation of histone H4 results in the nucleosome core unfolding and is associated with transcriptionally active regions of DNA.

Beyond this level of organization the DNA – histone complex is repeatedly coiled such that the whole strand becomes shorter and shorter and more tightly controlled. This process is much like repeatedly winding an elastic band, but under much more controlled conditions. The ultimate step results in a chromosome, which during certain stages of the cell cycle is sufficiently well condensed to be visible with a light microscope, and using specific staining techniques, individual chromosomes can be identified and compared. Visualizing chromosomes is relatively easy in eukaryotes, that is higher organisms with a clearly defined nucleus. For prokaryotes, however, such as bacteria, the system is slightly different in that nucleic acid is not generally associated with histone proteins and the single chromosome is often circular. This reflects the simplicity of the genome. For example, humans, who do not have the largest genome in the animal kingdom contain approximately 3000 million base pairs, while a gut bacteria *Escherichia coli* contains 4 720 000 base pairs and a virus which infects *E. coli* with the designation ΦX174 contains 5386 bases of single-stranded DNA.

30.2. Primary Molecular Tools for DNA Analysis

With DNA being such a large molecule it is important to be able to manipulate it effectively. This can take any form of manipulation from the entire chromosome, or to just very small pieces of DNA of particular interest. To make these manipulations various enzymes are used which are a part of a cell's own ability to alter the activity of DNA and in some cases protect the cell from invasion by pathogens.

Enzymes used in DNA analysis come under a range of headings, depending on their type of activity. Some are highly specific in their action, some less so. It is therefore important to choose the right tool for the investigation being carried out.

The most important enzyme types used in molecular biology are:

- Exonucleases
- Endonucleases
- Polymerases
- Ligases
- Methylases

These will be dealt with separately and their function and activity explained. It is important to remember that all these enzymes occur naturally and serve very precise functions in the cell. Later on specific, practical, examples of their use will be described. First it would be worth looking at the conventions of nomenclature covering some of these enzymes. Exonucleases and endonucleases are collectively referred to as DNases, or more fully deoxyribonucleases. Enzymes which can degrade both DNA and RNA (RNA is a variety of nucleic acid which is chemically similar to DNA but serving a different biological function) are called simply nucleases. If a DNase cuts only one strand of a double stranded molecule it is called a nicking enzyme, and the resultant single stranded break is a nick. One such enzyme is Dnase I. These enzymes seem to be important in releasing tension in the double helix as it is unwound during replication.

30.2.1. Exonucleases

These enzymes degrade DNA from the ends and cannot themselves generally introduce breaks within a strand. For example, mung bean nuclease

Table 2. A range of enzymes and their cutting sites

Enzyme name	Restriction site
Nae I	5'...GCC↓GGC...3' 3'...CGG↑CCG...5'
Nco I	5'...C↓CATGG...3' 3'...GGTAC↑C...5'
Hae III	5'...GG↓CC...3' 3'...CC↑GG...5'
Mbo I	5'...↓GATC...3' 3'......CTAG↑...5'
Msp I	5'...C↓CGG...3' 3'...GGC↑C...5'

a nuclease, as the name implies, extracted from Mung beans will progressively remove bases from both the 5' and 3' end of DNA where there is an overhang, resulting in two blunt ends, that is, strands of equal length of complementary sequences. In contrast, exonuclease III preferentially degrades recessed 3' ends, that is, where the complementary strand extends beyond the 3' end. If the 3' end overhangs more than 4 bases the enzyme cannot attack it. This property can be exploited by first nicking the DNA at specific sites with an enzyme and then using exonuclease III to create a single strand gap propagating in a specific direction.

30.2.2. Endonucleases

These are enzymes which cut DNA internally, rather than at the ends and are more correctly called restriction endonucleases. The activity of endonucleases can vary widely from enzyme to enzyme. Several hundred are now known, with very different operating conditions and sites at which they restrict, or cut, the DNA. Endonucleases recognize short sequences along the DNA, which may occur once every few hundred, or once every few thousand bases along the molecule. The frequency with which cutting sites can be found depends largely upon either how specific the sequence has to be, or how long the sequence has to be before it is recognized by the enzyme.

Most endonucleases recognize sites between 4 and 6 bases long and will either cut the two strands symmetrically, or the cut can be staggered. A range of such restriction endonucleases and their restriction sites are shown in Table 2.

The names are generally derived from the initials of the organism in which the particular enzyme was found, so for example all the restriction enzymes beginning *Eco* are derived from

Escherichia coli. It is not possible to derive the species originator simply from the enzyme name.

Using any given endonuclease on a specific piece of DNA will always render the same number and size of fragments. With large genomes there can be variation between individuals, but with very small genomes, as found in viruses where there is no sexual reproduction, the variation in restriction sites between individuals is effectively zero. So using a restriction endonuclease on such material would result in fragments of an exactly known size. These could then be used to compare with unknown fragments to give information about the DNA under analysis. An example of this is the digestion of a small plasmid of 4361 base pairs, designated pBR322, which is usually found in *E. coli* . Digestion of this with the endonuclease *Msp* I always results in 26 fragments ranging in size from 622 base pairs (4.04×10^5 Da) to only 9 base pairs (0.06×10^5 Da).

30.2.3. Polymerases

These are enzymes which can take individual nucleotides and attach them to a progressively lengthening chain of DNA. In living cells they perform the essential function of DNA replication. This is also exploited in vitro to produce multiple copies of specific regions of DNA. In cells, control of polymerase is undertaken by a battery of associated enzymes. All known polymerases replicate in the $5' \Rightarrow 3'$ direction. When carried out in the laboratory a different approach has to be taken to make full use of the technique.

There are many different types of DNA polymerase, most of which are, like most enzymes, unstable at high temperatures. These enzymes are of a specific value in some circumstances. For example, T7 DNA polymerase comes from bacteriophage T7, which attacks *E. coli* by taking over cellular control and replicating its own DNA. This enzyme can be very useful for replicating long stretches of DNA quickly. Sometimes a short section of DNA needs to be replicated, but more than once, and for this thermostable enzymes are needed. The first of these was extracted from a deep sea species *Thermus aquaticus* and called *Taq* . Later on other forms of heat stable DNA polymerases were found from other species living around deep sea thermal vents. These later additions to the list of useable enzymes tend to be more accurate in their replication, making fewer errors. Examples of these include a DNA polymerase
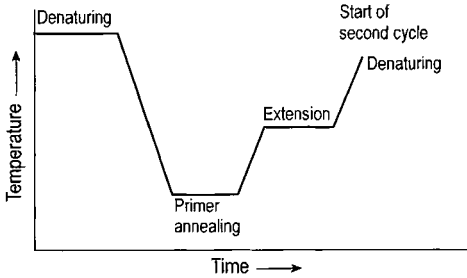
Figure 2. The Polymerase Chain Reaction temperature cycle. The exact temperatures depend upon the type of primers, AT rich primers denature more easily than GC rich primers

from *Thermococcus litoralis*, which lives in a marine environment at 98 °C and one extracted from a species of *Pyrococcus* recovered from a deep sea vent at 2010 m and capable of growth at 104 °C.

The value of these thermostable enzymes is that they can easily be used in a Polymerase Chain Reaction (PCR). This is a technique in which short sections of DNA are replicated exponentially. By knowing two short sequences of DNA on either side of the target DNA, which itself can be of a completely unknown sequence, it is possible to produce single stranded primers, complementary to the two sequences. At this stage the target DNA is double stranded, held together by hydrogen bonds. Exposure to heat will denature the DNA and result in single-stranded DNA. By cooling the sample at this stage, to a temperature at which the DNA will reanneal, some complementary strands will rejoin, but many of the primers will attach to the single stranded DNA. Increasing the temperature at this point to below the denaturing temperature allows the DNA polymerase to extend the strand by adding on bases that are also present in the solution. This cycle is repeated until enough target DNA is present for further analysis [1].

PCR is one of the most powerful methods of producing material for analysis from very small samples. If PCR was 100 % efficient then after 20 cycles there would be 2^{20} times the number of starting molecules, each one with a primer at one end. It is these primers which can be tagged if required for later detection. Every primer detected, which is greater than the original primer mass, implies the presence of an amplified product. After approximately 25 – 30 cycles the enzyme limits the reaction and so, in conjunction with other factors, this results in a realistic 10^6 fold increase in target DNA. One such limiting factor is that as PCR proceeds and more complementary DNA is

produced, so there is an increasing tendency for complementary strands to anneal, to the exclusion of the primers. Figure 2 illustrates the temperature cycles and the increase in amplified product.

30.2.4. Ligases

These are enzymes which are used to reattach strands of DNA together. Ligases come in many forms, some only able to work on single-stranded DNA or RNA, while some are able to work on duplex strands. Another major difference between ligases is the ability to reseal a single-strand nick created by a nicking enzyme. Although some ligases will only reattach blunt ends, single base overhangs or cohesive ends, that is two single-strand overhangs with exactly complementary sequences, all ligases operate in the same way. The enzyme catalyzes the formation of a phosphodiester bond between juxtaposed 5'-phosphate and 3'-hydroxyl termini of two adjacent oligonucleotides.

30.2.5. Methylases

These are enzymes which chemically modify DNA residues. In higher eukaryotes this usually takes the form of attachment of a methyl group to cytosine to form 5-methylcytosine. Usually the methylated residues are in C–G pairs. The level of methylation of eukaryote DNA shows a strong correlation with transcriptional inactivity. Lower eukaryotes and prokaryotes are also able to methylate adenine residues. There are no known methylases that work on guanine.

Methylases have recognition sites in which they work, so it is generally possible to tailor the methylation to a specific area of DNA. The value of this stems from methylated residues being resistant to attack from restriction endonucleases. This allows the bacteria carrying endonucleases for protection against invading viruses to protect their own DNA by methylation. It is this aspect of methylases which is important in DNA analysis.

30.3. Methods of DNA Detection

Once a DNA sequence of interest has been produced by PCR, or extracted from a cell, there are several further stages which can be used to determine much more about the DNA. There are

broadly two aspects here which can be investigated, size and sequence. Often, for practical purposes, sizing alone is sufficient to detect, say, a mutation of medical importance [2]. In some cases, as we will see later, sizing of some specific fragments may be enough to allow determination of a sequence.

Sequencing is generally undertaken only when absolutely necessary because of cost and time constraints. The most well known of the sequencing efforts is the Human Genome Mapping Project (HGMP). This is intended to sequence the entire human genome, but it should not be imagined that the story ends there. The reading frame of much of the DNA is still unknown, it is also unknown what most of the sequences are for, if anything. Consequently, knowing a sequence is only the start.

There are several different ways in which DNA can be detected and sized. The majority of which involve separation of different sized DNA fragments under the influence of an electric current — electrophoresis. The medium for gel electrophoresis varies from simple agarose to polyacrilamide.

The earliest method introduced for detecting DNA fragments after electrophoresis, was the Southern blot. It involves transfer of the various discrete fragments to a solid support, where they are immobilized. This transfer is accomplished by capillary action in SSC (sodium chloride – sodium citrate) buffer. The capillary action transfers the DNA to a nylon or other artificial membrane, after which it is fixed by heating to approximately 80 °C. The transfer takes place on an horizontal bed, so the relative positions of the DNA does not alter. Reacting the DNA on the blot with probes, which are complimentary sequences to those which are under investigation, yields a series of bands that can be taken as representing the positions of those known, probe, sequences of DNA. These probes are not irreversibly bound to the membrane, unlike the DNA under investigation, so they can be removed and a different probe, or probes, hybridized.

Broadly, there are two different forms of probe, multilocus probes (MLP) and single locus probes (SLP). The difference between these two is that multilocus probes hybridize at several different sites on the genome, while single locus probes only hybridize to a single site. For this reason an MLP will give a multi-banded result, but an SLP will give a maximum of only two bands/probe/individual.

The reason for there being a maximum of two bands is that all higher organisms carry two copies of every gene, one from each parent. When using SLPs these may be both the same, homozygous, resulting in only a single band. However, when the two alleles are different, heterozygous, an SLP will produce two bands. Whichever of these systems, SLP or MLP, is used, indeed, whenever a gel separation system is used, an internal size standard is required. This takes the form of either an allelic ladder or DNA digested by a specific enzyme yielding fragments of known sizes. Whatever the sizing fragment type which is used, the same system applies to the comparison of fragment sizes. This is generally referred to as the Local Southern method. It takes into account the tendency of DNA fragments that are similar in mass, to migrate in a generally linear fashion relative to each other. It therefore requires at least three DNA mass markers which are close in size to the fragment of interest so that an accurate comparison can be made. It is important that the DNA mass markers are closely associated with the fragments under investigation because as they depart in mass the calculation of mass by comparison becomes increasingly inaccurate.

This technique of comparison of size of DNA fragments against a standard is still the best and most accurate method of determining the size of a piece of DNA when it is either fixed on a membrane, or travelling through a gel. This extends to other techniques which do not require the use of a membrane blot to transfer DNA to a solid substrate.

While the original method of DNA analysis required the transfer to a membrane of DNA that had been separated by gel electrophoresis before further analysis could be undertaken, more recent developments have rendered this unnecessary except in specific areas of research. However, automated systems still require a method of sizing which is both reliable and repeatable, and so DNA fragments of known mass are still the method of choice.

Techniques for detection of DNA on gels has also changed. The original method required the use of ^{32}P, which has a half-life of approximately 14 days. It would be substituted for a non-radioactive phosphoros atom in the DNA probe. When hybridized to the probe and exposed to X-ray film a darkening would take place on the film at the radioactive sites. Besides the short half-life there are two other major considerations which have spawned a considerable field of research into the non-isotopic detection of nucleic acids. These are the potential long-term health hazards and the

problem of safe disposal of residual radioactive material from the assay. Most of the development of chemical detection techniques was based around immunoassay tests and then developed further.

Non-isotopic labeling to detect nucleic acid – probe hybrids can be carried out using either direct or indirect labeling methods [3]. Direct methods involve attaching a detectable label directly to the target, usually the probe, by a covalent bond. This is the technique now employed in machine based systems, but on gels it is still common to use an indirect method. This involves a hapten, a low molecular weight molecule usually of less than 1000, attached to the probe which can then be detected using a specific binding protein, these may be antibody based. Although the binding strategy must attempt to avoid involving the hydrogen bonds, which are essential for binding of probes to the target DNA, this is not always practicable and may partially compromise the efficacy of the system.

The sensitivity of hybridization assays is a function of the detection limit of the label, some of which are not only as sensitive as radiolabelled assays, but are much faster in exposure times to X-ray film. There are four broad categories of detection for non-isotopic labels. These are bioluminescence, colorimetry, electrochemiluminescence, and fluorescence.

30.3.1. Bioluminescence

This is essentially the same as chemiluminescence found in nature, for example in fireflies and certain types of fungi and bacteria. They involve the use of a luciferin substrate and an enzyme, luciferase. Because of the versatility and sensitivity of these systems, they can be used with either film or charge coupled device (CCD) cameras.

30.3.2. Colorimetry

Systems using colorimetry tend to be less sensitive than luminescent systems, but they do have some advantages. By producing a precipitate associated with the DNA – probe hybrid they are ideal to produce a permanent record of the distribution on a membrane.

30.3.3. Electrochemiluminescence

This is a very specialized technique in which an electrochemical reaction produces an excited state, after which decay to a ground state produces light. It can be readily appreciated that the need for both specialist electrochemical generation and detection facilities renders this sort of system both expensive and specialist.

30.3.4. Fluorescence

This is the system of choice now generally employed in automated equipment. Straightforward fluorescence detection of DNA on gels, by either film or CCD devices, is difficult because of intrinsic problems of noise. However, specific fluorochromes have enabled systems to be developed which utilize different methods of detection, rather than simply overlaying a gel with a photoreactive paper.

Fluorescence detection of specific dyes at specific wavelengths has revolutionized not only the detection of target alleles, but also personal identification, paternity details, and DNA sequencing. This technology has been implemented in many different systems using a range of equipment. A common method for detection of specific fragments is to attach a dye to either an oligonucleotide primer, for use in a PCR reaction, or a probe. By using different fluorochromes for different oligonucleotide primers it is possible to run several PCR reactions simultaneously in the same reaction vessel. Detection in these systems is carried out by inducing fluorescence with a laser, the data being captured and processed automatically as the DNA fragments pass in front of the laser. In this manner an image can be created from the data which broadly corresponds to the older autoradiograph image created from Southern blots and probes. However, this is not actually the most convenient method of data display. The graphic display in Figure 3 would be the more usual method of showing results of analysis.

By using four different dyes, red, blue, green and yellow, large numbers of reactions can be carried out in a single tube. It is also possible to put in size standards, which makes accurate sizing of fragments very much more precise. Figure 4 shows the emission spectra of the various dyes in commercial use. Even if two different fragments are the same molecular weight, if they fluoresce at different wavelengths they can be sep-
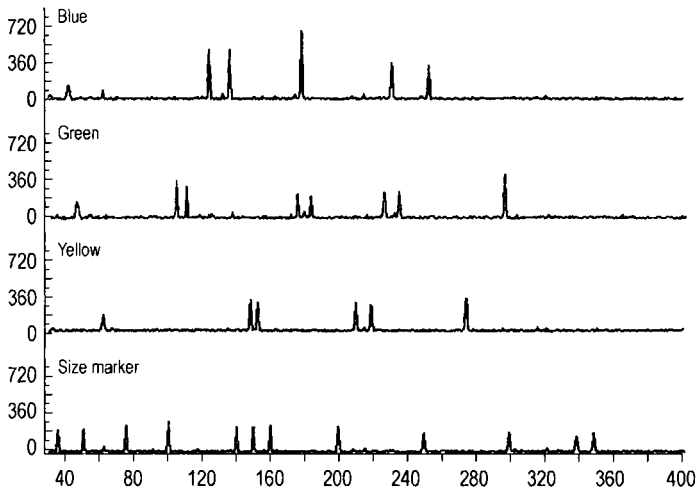
Figure 3. Graphic display of output from an automated DNA analyzer. By using loci of widely differing sizes in the same lane, they can use the same laser activated dye without risk of confusion. This is a composite image of data generated on the basis of when each fragment passes the laser window. It is therefore not a static image but based on a dynamic time scale. The larger molecular weight fragments are towards the right-hand side of the display. It should be noted that there is still a need for the size maker, at the bottom, which fluoresces red. Since all the alleles are separable, they can be run in a single column, the image being worked out by the associated computer. Image courtesy of PE Applied Biosystems

arated using filters. It should be noted that many of these dyes are proprietary products so the exact chemical constitution has not been released.

So far these methods have broadly been applicable to slab gels, but more recently capillary electrophoresis for DNA has been introduced. The main advantage of this system is speed and reproducibility. Arrays of capillaries are used instead of a single gel, but in general terms the technology is very similar and the output identical.

As new technologies are introduced there is a move away from gel based systems towards more exotic methods of DNA analysis. One such is the DNA chip. This is a technology still in its infancy, but potentially of great power and speed. DNA chips utilize the negative charge on DNA to move it about and concentrate it in target areas. These target areas have probes chemically bound to the chip. After several cycles of DNA binding to the chip, each with a different probe, the same process is repeated, but this time with the DNA to be analyzed. By concentrating the DNA around the probes, hybridization rapidly takes place. Detection of the bound DNA is carried out using a laser and arrays to detect the resulting fluoresence. As can be readily appreciated, although potentially very useful, the difficulty is in making the chips flexible enough to be applicable to any but the most specialist of areas of research and diagnos-

tics, where mass screening would make the system commercially viable.

One of the most useful techniques to help in understanding DNA and genes is DNA sequencing. This is quite different to the measurement of sizes of fragments, sequencing determines the precise arrangement of bases making up the DNA. This has received considerable public exposure with the advent of the Human Genome Mapping Project. The HGMP is an attempt to sequence the entire human genome by a combination of public and private collaboration. Although the implementation of new fluorescence based technology has speeded up and automated the process of DNA sequencing, the method used is still broadly based on the Sanger–Coulson method of chain terminating nucleotides. It is, therefore, really the detecting methods that have changed, rather than the underlying science. The alternative method of sequencing, chain degradation, has not been so widely employed in modern systems.

The chain terminating method of DNA sequencing requires single-stranded DNA, because this technique involves the production of a second complementary strand. Consequently, the DNA to be sequenced is normally cloned into a single-stranded vector such as a plasmid. This is done by cutting the gene of interest out, using nucleases and then ligating it back into the plasmid. By
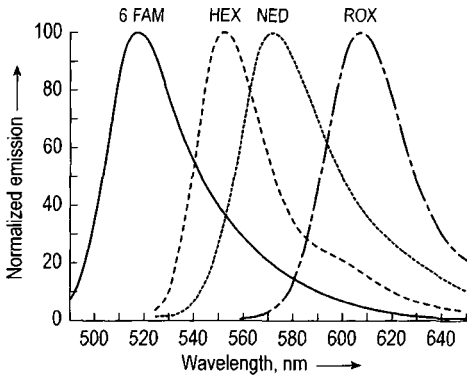
Figure 4. The emission spectra of four commonly used fluorescent dyes in automated DNA analyzers. Image courtesy of PE Applied Biosystems

having single-stranded DNA, both to be sequenced and in the vector, it is possible to produce a short oligonucleotide primer which will anneal to the plasmid and allow extension of the DNA using *the gene under investigation as the template for* the growing PCR product. The use of a short primer is necessary because most of the DNA polymerases which are used in sequencing reactions need a double stranded region from which to start synthesis.

When undertaking a chain terminating sequencing reaction, there has to be each of the four deoxynucleotides which can be incorporated into the growing strand. These are dATP, dTTP, dGTP, and dCTP. At this point the original method would require four reactions to be carried out separately. One with the addition of dideoxyATP (ddATP), one with ddTTP, one with ddGTP, and one with ddCTP. These are the chain terminators. Dideoxynucleotides can be incorporated into a growing chain as easily as a deoxynucleotide, but will block further strand synthesis. This blockage happens because the dideoxynucleotides lack the hydroxyl group at the 3'-position of the sugar, where the next nucleotide would normally be attached.

If ddCTP is added to the mixture, strand *growth will cease when it is incorporated opposite* a G on the template strand. However, by adjusting the concentrations of the reagents in the reaction mixture, termination will not always take place at the first template G, because an ordinary dCTP may be added by the enzyme instead. So a number of different fragments of different lengths are produced from this one reaction, each terminated

by a ddCTP. When it was normal practice to radiolabel the dideoxynucleotides, they had to be run in separate reactions and run in adjacent columns on a sequencing gel, which was then blotted, autoradiographed, and the bands read horizontally across the gel lanes, with only one band in any position down the gel representing the base at that position on the strand. Sequencing gels are very thin polyacrilamide containing urea to denature the strands.

The modern trend is towards labeling the dideoxynucleotides with a fluorescent dye, each one fluorescing at a different wavelength [3]. Besides allowing the sequence reading part of the experiment to be carried out automatically, it also means that the four different reactions can be carried out in one tube, and with the sequence reading carried out in one lane of a gel [4]. This automation of processes helps to minimize errors which would otherwise appear. As can be appreciated, since the reaction is competitive between deoxynucleotides and dideoxynucleotides, there is a tendency for more low molecular weight products to be formed than high molecular weight ones. There is also a limit to the length of strand which can be accurately sequenced because as they get longer the percentage molecular weight difference between strands differing in length by a single base becomes progressively smaller. Figure 5 shows diagramatically how cycle sequencing is carried out using a circular plasmid and a thermostable DNA polymerase, so that the reaction can be set up and left to run without intervention until the results are transferred to an electrophoresis system.

In Figure 6 a traditional sequencing gel is represented using four lanes of a sequencing gel. It can be seen that as the molecular weight of the fragments gets larger, towards the top of the gel, they move more slowly and become increasingly difficult to separate. Using fluorescent dyes allows all the products to be run in the same gel lane. Although the DNA fragments still tend to "bunch" as their molecular weight increases, it is easier to separate the different products because of their different fluorescent wavelengths.

30.3.5. Denaturing Gradient Gel Electrophoresis (DGGE)

It is possible to determine the presence or absence of a DNA polymorphism without actually knowing anything about the nature of the poly-

Figure 5. The circular plasmid, containing the DNA to be sequenced, short oligonucleotide primers complementary to a known section of DNA, and labeled nucleotides are reacted together. The primer attaches to the plasmid and then polymerase enzymes extend the primer until incorporation of a dideoxynucleotide terminates chain extension. Denaturing the products, followed by anealing of more primer results in accumulation of artificially produced single strands of DNA complementary to the original template. Overall there will be DNA strands varying throughout the total area of interest in steps of only one base, allowing for step by step sequencing

Figure 6. A sequencing gel produced from four separate reactions in a chain termination experiment. Note how these are read sequentially across the gel and as the molecular weights get larger band separation becomes more difficult. With different colored fluorescent dyes these reactions can be carried out in the same tube and sequenced on the same track of a gel, or simultaneously on a single capillary electrophoresis device

morphism itself. This is done using DNA which has been through a PCR to produce enough material to be analyzed. The detection relies on the fact that DNA heteroduplexes, which differ from each other by only a single base pair, not of content, but of type, have measurably differing characteristics affecting the transition from double-stranded to

single-stranded DNA. Since single-stranded DNA travels at a slower rate than double-stranded DNA when electrophoresed through polyacrilamide gels, a denaturing gradient will enhance separation.

By having a gradient of a DNA denaturing agent through the gel, as the similar, but not identical, double-stranded DNA passes down the gel it will melt at different points, at which point the passage through the gel slows down and the original heteroduplexes can be seen to be different. Of course, it is not possible to say in what way the heteroduplexes are different, only that they are. The denaturing gradient can be introduced in a number of ways, such as formamide, urea, or a temperature gradient.

30.3.6. Single-Strand Conformation Polymorhism (SSCP)

Similar to DGGE, single-strand conformation polymorphism (SSCP) is a method which can be used to detect mutations and polymorphisms without characterizing them. SSCP analysis utilizes the principle that electrophoretic mobility is dependent on size, and shape of the molecule as well as the electric charge. Ideally DNA to be analyzed using SSCP is less than about 400 base pairs. By running a PCR across the area to be tested, enough DNA can be produced to analyze. The analysis is carried out by denaturing the PCR products and

then rapidly cooling them. This stops most of the single-stranded DNA reassociating with complementary sequences, but does allow the single-strands to fold into a conformation determined by the sequence itself. A difference of only a single base will alter the conformation and therefore the mobility within the gel, which is of a native, rather than a denaturing, type. Because the conformation can be affected by ion concentration and temperature it is important that the system is highly uniform throughout the procedure. This will mean controlling the temperature of the gel during the experiment as well as the solvent and ion concentrations.

30.3.7. Random Amplified Polymorphic DNA (RAPD)

Random amplified polymorphic DNA is in some ways a misleading name. What it does allow is detection of polymorphisms without any knowledge of the polymorphism, but also without any knowledge of the DNA sequence of the target organism. Indeed, the results can be of limited value because of this, but there are specific applications to which this technique can be put with great success. RAPD analysis requires the use of a single short oligonucleotide, usually about 10 bases long, the sequence of which is chosen at random. The primer will often anneal to genomic DNA during PCR at several different places, giving rise to several bands on subsequent electrophoresis, but if conditions are maintained between experiments the number of bands should remain the same for any individual.

The amplified fragments are highly dependent on the primer sequence and the origin of genomic DNA being looked at. So the results may be of interest, but without further analysis it is not possible to say what form the polymorphisms take or where in the genome they reside. One area of particular interest for RAPD use is in agriculture where it might be necessary to determine the sex of a plant in dioecious species, that is species where the male and female flowers occur on different plants. This is useful to stop the planting of more pollinating plants than is necessary, but, as is the case with date palms, until the trees mature and flower it is not possible to determine which trees are male and which female. So by using RAPD analysis differences may come to light between the sexes which can be exploited to determine the sex of the plant while still a seedling.

30.3.8. Short Tandem Repeat (STR) Analysis

Short tandem repeats (STR) are widely distributed throughout the genomes of virtually every higher organism that has been investigated. STRs are arrays of variable numbers of short repeated sequences. They are sometimes referred to as microsatellites being very small compared with other satellite DNA. Satellite DNA can be defined as DNA from a eukaryote that separates as a distinct fraction during gradient centrifugation in caesium chloride. The satellite fraction is made up of material of a distinct composition, having a preponderance of A+T or G+C. Some satellite DNA, such as the Alu sequence found in humans is made up of 300 bp repeats, but in STRs the repeat may be only 1, 2, 3, 4, 5, or 6 bases long. The repeat can be made up of any combination of bases, repeated tandemly.

STRs can be found scattered through the genome in most, if not all, genes. They are to be found in intron DNA, that is the interspersed noncoding sections which lie between the exons, or coding sequences. While it is true that introns are not transcribed, it would be incorrect to call them junk DNA. Although not as yet clear why, it seems that introns are necessary to maintain the integrity of the gene. They are also important in evolution, allowing changes to take place within the genome without compromising the metabolism of the cell.

The importance of STR analysis stems from the observation that while some STRs are constant in their repeat number across entire populations, others are highly variable in repeat number. This variability has been shown in all species looked at. It has also been shown that not all the different alleles occur at the same frequency in any given species. So for a specified STR some alleles will be commoner than others, the distribution frequently approximating to normal, with very small and very large repeat numbers being rarest. Because this is not always the case we can surmise that some other mechanism than simple random processes are affecting repeat numbers of STRs.

It is the inherent variability of STR repeats which makes them of value [5]. By using a PCR reaction across the STR, given that we know the amount of flanking DNA, the fragment can be accurately sized in terms of the number of repeats present. This means that an STR result is expressed not in terms of base pairs, but as a single figure denoting the number of repeats. STR anal-

ysis is one of the simplest of PCR processes as analysis of results can be carried out using small agarose gel electrophoresis systems. This would include a sizing ladder in a second lane. The entire gel can then be stained with ethidium bromide and fluoresced using ultraviolet radiation. This approach requires that unless widely separated by size only a single STR is analyzed per experiment. The alternative is to make use of the automated systems which can cope with several different STRs at a time using several different fluorescent markers.

STRs have several important uses for the biologist, among them are personal identification, which we will look at later and population studies of animal and plant species. This latter function of STR analysis is especially useful in endangered species where there is small residual genetic variation. Since STRs are inherited in a strictly Mendelian fashion, with mutation rates generally less than 1 in 10000 meiotic events, it is possible to determine the level of inbreeding among a group. STRs can also be used to help breeding programs of endangered species by formulating family trees and aiding captive breeding to optimize outbreeding and reduce inbreeding suppression of fitness. Mutation of STRs most often result in a change in complete repeat number, but this is not always so, with occasional single bases being added in the middle of an STR. It has been suggested that mutation events only serve to increase the size of the STRs, but simply because reductions have not been seen it does not mean they do not occur. It is quite obvious that a continuos increase in STR size would eventually result in physical restraints on the genome as well as a reduced ability of the cell to function metabolically.

30.3.9. Single Nucleotide Polymorphism (SNP) Detection

Single nucleotide polymorphisms (SNPs) are represented in the genome as a single base change in a gene, or other DNA being looked at, although they may be considered less informative than other mutations that have several different variants at a single loci. This is because an SNP will normally be biallelic, that is it is found in either one of two forms only. However, with SNPs being found at a rate across the human genome varying from one every 500 base pairs to one every 5000 base pairs, depending upon which area of the genome is being looked at, there are potentially anything up to

3 million SNPs in the human genome. This makes them potentially very useful in a range of areas, especially population structure, pharmacogenetics, and complex genetic traits. These last conditions come under the general heading of polygenic and multifactorial inheritance. Examples of this include a surprising range of conditions that have not generally been thought of in the past to be genetically controlled because the number of genes involved and the interaction of these with the environment made it impossible using classical genetic techniques to analyze them. Such conditions as some forms of heart disease and a persons height can be included in complex genetic traits.

SNP analysis is another PCR based technique. It can be done in a variety of different ways but the most straightforward, although not the method with the greatest throughput, involves a modification of the sequencing reaction. By creating an initial PCR reaction which runs for approximately 100 bp across the known SNP, sufficient material can be generated for the determination of the nature of the SNP. This first PCR has to be carried out to make sure that enough target DNA is available. The next stage uses a primer immediately adjacent to the SNP, and dideoxynucleotides attached to a fluorescent marker. As soon as the next nucleotide is incorporated, which is in the SNP position, the sequence is terminated. When these products are run out on a gel there will be either one fluorescent band of a single color, if the individual is homozygous for that SNP, or two bands of different colors if the individual is heterozygous. Although the two products are essentially the same length in base pairs the difference in mass with the attached dyes is sufficient for them to be separated. If for any reason they are not, then electrophoresis mobility modifiers can be used. It can be appreciated that the first PCR reaction is necessary to produce enough material for the second round fluorescence to be detectable.

30.3.10. Mitochondrial DNA Analysis

Mitochondria are small self replicating subcellular organelles found in the cytoplasm of all eukaryotes. However, the situation is rather more complicated than that. Mitochondria are essential as they are the site of oxidative phosphorylation, the energy producing process in eukaryotes, and although they have their own DNA (mtDNA), only a minority of mitochondrial proteins are ac-

tually coded on the mitochondrial genome, most are produced by nuclear DNA. So although self-replicating, they are not autonomous. Generally, mitochondria also lack intron sequences within their genes.

The mitochondrial genome varies from about 15–20 kb of double stranded DNA with approximately 50 genes sequences. At this point it is worth considering mitochodria in general because it can impinge on interpretation of analysis of mtDNA results. Mitochondria are generally thought to be inherited only from the maternal line. This has been shown to be incorrect, but in a way which appears contradictory. In humans most of the mitochondrial genome originates from the maternal side: the ovum. However, it has been demonstrated that the mitochondria of the sperm can, and does, enter the ovum on fertilization. The question then arises as to how much of the mtDNA in the zygote comes from a maternal origin and how much is paternal. Generally speaking most testable mtDNA comes from the maternal side, although the sperm does in fact contribute approximately 100 mtDNA genomes to the ovum on fertilisation in mammals. Recent studies have shown that there is a selective destruction of the invading sperm mitochondria soon after fertilization.

Studies of three dimensional reconstructions of cells using serial sections and electron microscopy suggest that mitochondria either freely associate with each other or are, in any given cell, single entities. Certainly in yeasts it has been shown that mitochondria fuse and bud off from each other quite frequently. Either way, every cell contain several copies of the mtDNA genome which can be tested [6].

Being generally accepted as maternal in origin, mtDNA has the potential to enable us to track through generations via the maternal pathway of inheritance, rather than paternal. This differentiation has great significance in some applications [7].

Mitochodrial DNA comes in the form of a distorted circle, forming a three dimensional structure as it forms hydrogen bonds with itself. In humans mtDNA is generally about 16 570 base pairs long, but it is not the entire sequence which is interrogated when mtDNA is analyzed. It is a short section of highly variable sequence which is looked at, again by PCR.

The reason mtDNA analysis can be so valuable is that unlike nuclear DNA, mtDNA can be found in robust tissues such as bone shafts and hair shafts, which contain no nuclear DNA at all. These tissues, which are tough and resistant to decay, can harbor useful DNA signatures long after the soft tissues have decayed along with the nuclear DNA and become of no use as to the determination of origin of a sample.

It is the robust nature of mtDNA and relative ease with which it can be copied using a PCR reaction which makes it ideal for investigating ancient samples, such as mummified corpses and bones from ancient burial sites. It has also found favor for determining phylogeographic variability in ancient societies. It is not generally necessary to sequence the entire mtDNA because some areas are intrinsically highly variable. One such is the mitochondrial DNA control region with a high variation rate in the 360 nucleotides at the 5'-end of the sequence and to a lesser degree 200 nucleotides at the 3'-end of the control region. Other methods focus on a section of 780 bases of a variable region, which shows a variation rate of approximately 1 %. By concentrating on 12 sites within this region it becomes easier to produce a result and yet maintains the discriminatory power of the test. Mitochondrial DNA analysis has wide applications from familial identification to questioning species origin in food materials.

30.3.11. DNA Analysis and Bioinformatics

It has become ever more apparent that to analyze DNA from any organism increasingly requires fast and reliable computers. The sheer scope of trying to analyze something of the size of, say, the CF gene which spans over 250 kb can be enormous. It would also be extremely difficult without computers to have any chance of finding out whether a sequence has already been defined amongst all the available sequence data. It is often no longer possible to publish database information as tables or lists, the data are just too much for such a manual system to cope with [8].

When looking for something more precise than just sequence data, perhaps data on mutation rates or sequence variability, it is normal to employ a search engine. Although search engines designed for scientists are slowly becoming available, commercial search engines do not well serve the need of scientists. This is essentially because the software uses crawler programs, which index a page, and then jumps to a linked page and goes through the process again. Commercial web pages often have embedded words in HTML code for their

Table 3. Useful starting points for finding genetic data on the World Wide Web

Adress	Type Of Data Contained
http://www.gdb.org/hugo/	Home page of Human Genome Organization
http://www.expasy.ch/sprots/sprot-top.html	Home page of the Swiss-Prot database containing data relating to protein function and variation as well as domain structures
http://www.ncbi.nim.nih.gov/omin	Home page of the database Online Mendelian Inheritance in Man. This is an annotated listing of traits, both diseases and bening polymorphisms which are inherited in a Mendelian fashion. It also contains data regarding mitochondrial dissorders
http://www.incyte.com./products/pathoseq/pathoseq.html	This is a commercial database which contains data relating to sequence data of pathogenic organisms, primarily bacteria and fungi
http://www.ncbi.nim.nih.gov/genbank/genbankoverview.html	A primary depository for human sequence data, currently containing over 4.5 million sequences covering approximately 3 billion base pairs
http://www.ncbi.nim.nih.gov/snp/	A primary depository for SNPdata, currently containing in excess of 2.4 million SNPs

page, or use metatags, words invisible but present on the home page.

The unaudited nature of many web pages, both commercial and scientific is also a potential source of problems. If a laboratory sets up a web page that allows publication of any results they have, two things may occur, neither of which are desirable. The first is that erroneous data may be assumed correct when they are not and the second is that even if correct and valuable they may still be overlooked because of the construction of the web page. This second possibility is all the more distressing since the web has a primary aim of disseminating information as widely as possible. A primary feature of DNA analysis in all its forms is the need for different groups to be able to compare their results with other people working in the same area. Without unified databases this is not easily possible and can almost negate the potential value of having data in an electronic form.

Some large databases containing scientific data have the ability to check new data for consistency and contradiction with what is already available. This is a very valuable facility and underlines the need for well edited and definitive databases. Although there are good starting points for searching a particular area of data or data types, there are also some areas where fractioned data are so widely distributed that it becomes extremely difficult to navigate around the subject. There are for example, more than 150 different sites devoted to human mutations. If it were a simple case of the data being widely distributed, that would make the system complicated enough to deal with, but for many years nomenclature also varied so that it was not always entirely obvious if the same mutation was being referred to. Such was the potential for

confusion that the Human Genome Organization (HUGO) has created the Mutation Database Initiative which has produced a consensus on nomenclature and computing conventions so that human mutation data can be cross-referenced with other types of data dealing with both other types of human genetic data and also data from other species. Table 3 lists some useful starting points for data mining.

30.4. Applications of DNA Analysis

One of the earliest applications of DNA analysis came about indirectly, through the agency of gene mapping [9]. Ordering genes on a chromosome or section of DNA can be carried out with no knowledge of the DNA, gene or gene product, such maps have been constructed since the beginning of the twentieth century by breeding experiments. Creation of such maps became ever more sophisticated to the point where a classical map gave way to data regarding the fundamental chemical nature of DNA.

It is not only by sequencing strands of DNA that we can gain information and knowledge from the genetic code. Indeed, it is well recognized that even when the entire human genome has been sequenced, this will only be the very first stage in a long story of research. Knowing a DNA sequence alone gives no indication of the reading frame, the length of the gene and perhaps more importantly, the nature of the protein product. This last point is particularly significant because even if an amino acid sequence is known from a gene, it may not be obvious what the final tertiary struc-

ture might be, and consequently what the protein does. A good example of this is the characterization of the cystic fibrosis gene, where the gene was known before the gene product or its action was known.

Also, it may not be necessary to know the entire sequence of a gene to produce a valid analysis of its significance and activity. Taking the case of cystic fibrosis (CF), this is the commonest severe autosomal recessive condition in Caucasian populations. Approximately 5 % of people are carriers, which results in close to 1 in 2000 live births being affected. For a long time it had been possible to predict the probability of carrier status from a simple family history, but it was not possible to be more precise than that for many years, until it became possible to use linkage analysis to associate specific DNA polymorphisms within a family. This is a process where a marker, that is any determinable DNA sequence, is usually transmitted with the disease gene of interest by simply being in very close proximity to the disease gene. The logic is therefore that if an individual carries the measurable DNA then they will also carry a defective gene.

Using this idea the CF gene was broadly localized to a specific area on human chromosome 7. Later and progressively detailed analysis of the area produced a candidate gene. This was the first time a gene had been sought and found with no knowledge of the protein product for which it coded. It was only found later, after the protein sequence had been determined, that in 68 % of all CF mutations a single 3 bp deletion of the gene, resulting in the loss of a phenylalanine residue at position 508 was responsible for the problem. This mutation was designated ΔF508, Δ for delete, F for phenyanaline, and 508 for the amino acid position. Since then more than 170 different variant mutations of the gene have been described.

To analyze DNA mutations directly in the CF gene a technique can be used which is widely applicable in many areas of DNA analysis and mutation detection. This is the amplification refractory mutation system (ARMS).

30.4.1. General Principles of ARMS Analysis of DNA

A basic ARMS test requires the use of three primers. These cover the area where a known mutation occurs There is one common primer and two other primers, one which we can refer to as the wild type, that is the normal, fully functional one and one which includes the mutation. The length of the PCR products is not in itself important, what is important is the sequence at the end of the mutation detection primers. It is for this reason that the sequence of the gene being studied is known.

If a primer has a mismatch at the 3'-end it is most unlikely to produce a PCR product. So by having one wild type primer and one mutation primer two PCR reactions will produce either one product, the homozygous wild type or normal, alternatively the homozygous affected, or two products indicating carrier status. This mismatch specificity is maintained by using a PCR enzyme which does not have a 3' to 5' proof reading ability. If an enzyme is used with this capacity to proof read, no sense can be made of the results. Because the PCR reaction is indicating a specific finding, it is usual to run another, unrelated, PCR reaction in the same tube simply to be certain that the PCR has worked; a form of quality control.

Increased specificity of the reaction can be created by introducing deliberate mismatches close to the 3'-end of the primer. It should always be remembered that the nearer to the 3'-end the introduced mismatch is the greater the destabilization, so although specificity may be increased, product yield is likely to go down.

While it was necessary to use separate PCR reactions, when there was no way in which PCR products could be differentiated individual reactions had to be carried out. With the advent of fluorescent tagging of products this situation has changed. It is now possible, using only a limited number of different fluorescent wavelengths, to carry out ARMS testing in a single reaction for more than 10 different known mutations. This is done by using a combination of different common primers which results in a series of differently sized products. So the fluorescence might appear as red, blue, yellow, red, green, blue ... and so on, essentially the same fluorophore is used several times, but has to come within a sandwich of other fluorophores to be accurately sized.

Another technique, now widely used to construct tests for conditions like CF where there are a large number of possible mutations, all resulting in essentially the same phenotype, is the PCR – oligonucleotide ligation assay (PCR/OLA). In this technique multiplex PCR reactions are followed by a multiplex oligonucleotide ligation reaction which is made up of three probes for each biallelic site to be looked at. One probe is common and is

labeled with one of three fluorescent dye markers, the other two correspond to either the normal or mutant variant at that site. These interrogative probes are attached to various different numbers of electrophoresis mobility markers. These are generally pentaethyleneoxide units, which can be stacked to give products of the same number of nucleotides widely different mobility rates. So when the common probe is ligated to the specific probe it results in a unique combination of fluorescence and mobility. Using only four different fluorescent dyes, three for the probes and one as a lane size marker, it is possible to size a large number of different products in the same lane and determine the presence or absence of mutant or normal alleles at each site within the gene.

30.4.2. Analysis of Dynamic Mutations

There are a group of human conditions which have caused considerable consternation to those trying to define them clearly. These are the group of disorders now known to be characterized by dynamic mutations resulting in familial transmission, but in some conditions not necessarily in a clearly Mendelian fashion. In those that are clearly inherited in a dominant/recessive way, the age of onset may be extremely variable. This latter state could sometimes be mistakenly associated with closely linked polygenic inheritance or environmental factors. Although there have been relatively few of these conditions so far described, they are very important, either because of their frequency or severity. These include Fragile X syndrome, Huntington's Chorea and Myotonic dystrophy, where normal dominant inheritance of the condition is modified by a dynamic mutation in the 3'-untranslated region of the gene for myotonin protein kinase.

One of the worlds commonest genetic diseases and certainly the commonest form of familial mental retardation is Fragile X syndrome, affecting approximately 1 in 2500 children. The condition gets its name from the fragile site found at the end of the X chromosome, this being one of the two sex determining chromosomes, the other being the Y. Two X chromosomes results in a female phenotype but an X and a Y results in a male phenotype. For many years the only way of diagnosing this condition was cytogenetically, which was inaccurate, time consuming, and very expensive. Cytogenetic techniques gradually gave way to rather cumbersome pre-screening using molecular methods. Now, however, it is possible to gain a much clearer picture using PCR in a technique which is equally applicable to other dynamic mutations. Using Fragile X as our model we shall look at how dynamic mutations can be investigated.

Dynamic mutations are those which are represented not by a change in base composition, but a progressive lengthening of a sequence. This takes the form of progressive expansion of a trinucleotide repeat. In the case of Fragile X it is an increase in the number of polymorphic CCG repeats in the 5'-untranslated region of the gene. The number of repeats varies from about 6 to 55 in normal X chromosomes, but towards the upper limit of 55 the copy number becomes unstable between generations. Once over 55 repeats the chromosome is said to have a premutation, the rate of instability increases dramatically, so a dynamic mutation is one where once changed the DNA has an altered risk of further change. Once the number of copies goes beyond 230 it is a full mutation. It is therefore useful to know how many copies are present in premutation conditions. This is done by PCR reactions across the area of the trinucleotide repeat. With very large numbers of repeats in these conditions it may sometimes be necessary to produce products which are more than 2.5 kb long. This can be extremely difficult as the possibility of damage to target DNA is in proportion to its length. For this reason it is important to make a mild extraction of the DNA and not risk shearing it when using pipettes. It is also worth simultaneously using an internal control to check the voracity of the PCR reaction. Figure 7 shows the results of sizing the PCR products from three normal individuals, homozygous, heterozygous, and what is commonly referred to as the N+1 heterozygote where there is only one repeat difference between the two alleles. The other peaks represent the ApoE check on PCR and the sizing ladder used to measure the number of repeats.

30.4.3. Using DNA Analysis to Determine Sex

We have seen how RAPD systems can be used to determine the sex of dioecious plants, but there are other times when knowing the sex of an organism is just as important, but for other reasons. We shall take two examples, one human and one animal.
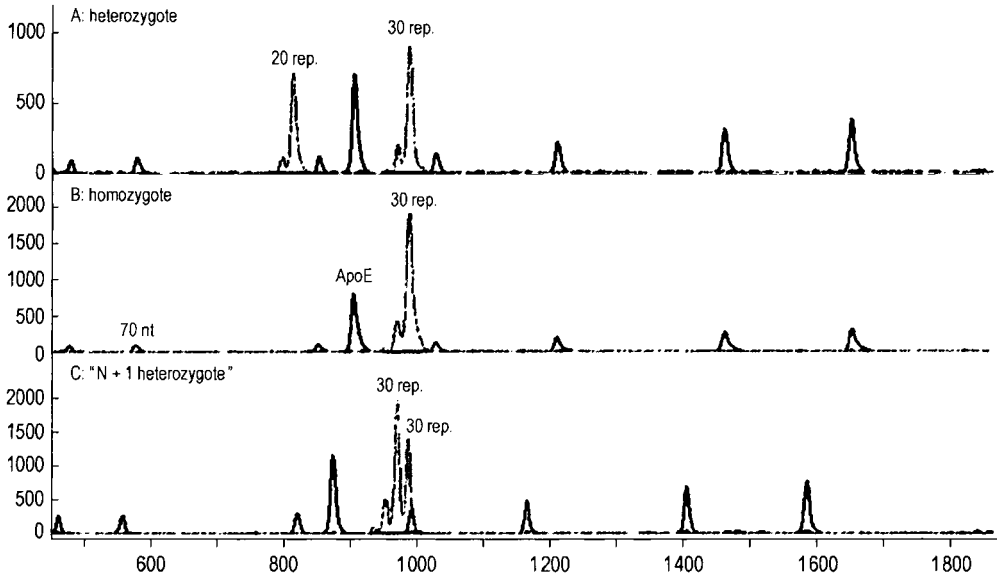
Figure 7. The automated output from measurement of repeat numbers in Fragile X analysis. The three traces represent heterozygote, homozygote, and N+1 heterozygote. The peak marked ApoE is a check that the PCR has run adequately and the other peaks are the sizing ladder. This technique is readily extended to other conditions involving this sort of dynamic repeat expansion. Image courtesy of PE Applied Biosystems

Humans differ chromosomally between the sexes in that females are the homogametic sex having two identical X chromosomes and males are the heterogametic sex having one X and one Y chromosome. There is a gene which codes for a protein associated with development of tooth enamel called amelogenin. By running a PCR across part of the gene two fragments are generated, one of 212 bases from the X chromosome and the other of 218 bases from the inactive gene on the Y chromosome. These are created by the same primers, so by running the products on a gel there will be two bands from a male and only one band from a female. Although the gene on the Y chromosome is inert, it does not seem to be susceptible to alteration, so this test is now routinely used on forensic samples to determine the sex of the originator of the sample.

Birds are quite different. Here it is the males which are homogametic and females heterogametic. Because of this difference the sex chromosomes are designated Z and W, with males being ZZ and females ZW. There is a highly conserved sequence on the sex chromosomes of birds which differ in only a minor way. By running a PCR reaction across the highly conserved region, enough product is produced for the second stage.

The difference between the two is a restriction site which can be cut by the restriction endonuclease *DdeI*. This operates by cutting at the sequence 5'CTNAG3'. The W chromosome produces a 73 bp fragment and the Z chromosome a 104 bp fragment, so there will be either one or two bands on a gel depending on the sex of the bird being tested. Because of the highly conserved nature of the tested sequence, it has been possible to use it in virtually any bird group other than the Ratites (emus, ostriches, etc). With so many bird species being identical between the sexes it has become important to be able to determine the sex of the birds in captivity so that breeding programs can make the best use of stocks [10].

30.4.4. Methods of Personal Identification

There have been many different methods of personal identification developed in the last 25 years based on DNA analysis. The more recently developed ones require more automated equipment but less human intervention, making interpretation of the results far less subjective. The first systems required relatively large quantities of material, usually blood, but as technology prog-

ressed so the amount of starting material has decreased so that it is now theoretically possible to determine with a good degree of certainty the generator of a sample. This is especially important in two areas of law, criminal cases and paternity disputes. Although there are specific methods used in such cases, this is more for continuity of service and agreement of acceptable methods, rather than one being better than another.

All methods which can discriminate between two sequences can in principle be used to discriminate between two individuals, but the level at which they can be relied upon does vary enormously. This reliability can stem just from factors such as inadequate discriminatory power of the test or subjective interpretation of the results. This latter situation being more important in earlier systems that were implemented rather than later ones.

The first methods of personal identification involved the use of mulitlocus (MLP) and single locus (SLP) probes. A sample of native DNA would be digested using an enzyme and then run out on a gel. This could then be probed to give several bands per individual in the case of MLPs, or one or two bands in the case of an SLP analysis. The major problem with these techniques is the size of the fragments which are probed. As the measured DNA can run into several kilobases, separation inevitably involves interpretation and agreement on methods of being able to say that a large band does in fact only comprise DNA of a particular type The possibility being that a band in what appears to be the position representing 14 kb, may in fact be 14.1 kb and therefore different, but not separable at this scale. It is these sorts of questions which helped to move techniques of personal identification along.

Although the ultimate personal identification method would be to sequence an individuals genome this is at the present time not a practical possibility. There are, however, two very good methods currently available which are routinely used. When used in this application they both utilize the same methodology. These methods are STR analysis and SNP. Both of these suffer from an intrinsically low power of discrimination. For example, while any given STR may have 10 or more possible alleles, they will not all be equally common, so a single STR may have a frequency within the population of 1 in 10, that is, you would expect 1 in 10 of the population to have that allele. Some may be much commoner than that. The way this is got round is by using a panel of STRs, either

singly or in one PCR reaction mixture. By multiplying the individual frequencies it is possible to generate a composite value for the probability of finding the total combination of STRs by chance, therefore allowing us to generate a likelihood ratio for that profile coming from a specific individual. With SNP analysis this procedure is taken slightly further, because for any given SNP the number of possible variants is both smaller and fixed, but the method is essentially the same, production of an accumulate probability by analysis of several different SNPs.

As can be appreciated, when comparing, say, material from a body with a reference sample, mutation rates are unimportant. But when looking at questions of paternity it is vital to remember that mutation from one generation to another is a natural consequence of genetic variation. So although these systems may be relatively stable, mutations do occur. This is one reason that the number of sites looked at is usually much bigger than in situations where a sample is compared directly with an individual [11]. Another reason for this increase is that it gives a greater degree of certainty in the production of a family tree because only half the loci in the genome of a child comes from the father, the other half originates from the mother.

A very specialized, and very limited, method which can be employed in personal identification is mtDNA analysis. This procedure is limited in that although it is possible to link a child to a mother, it is not possible to link a child to a father using this technique, and neither is it possible to separate siblings using mtDNA. It has proved very useful, however, where remains have been found and only mtDNA has survived, typically this might be after burial or fire. Under these circumstances mtDNA analysis can be used to help authorities re-unite the right remains with the right family.

30.4.5. Bacterial Contamination of Water Supplies

A practical and very useful application of DNA analysis is in detecting coliform bacteria in water supplies. The same techniques can be extended to check for pollution of surface water and beaches by sewage. The importance of this is can easily be understood by looking at the range of pathogenic microorganisms which can be found in fecal material and can from time to time appear in areas

where it is not wanted: *Salmonella* (causing typhoid fever), *Shigella* (giving rise to dysentery), and *Vibrio cholerae* (the causal agent of cholera). Although not associated in any way with fecal pollution another important contaminant is *Legionella pneumophila*, often present in air conditioning units and the cause of legionnaires disease. Indirect qualitative analysis of DNA from these organisms can be said to occur when an infection takes place. The toxins produced are proteins which are themselves transcribed from genes, but analytical techniques are now more sophisticated.

Although it has long been possible to detect these organisms, the traditional methods of microscopy and viable plate count are not necessary quantitative methods. Sensitivity is also inadequate with these systems, being approximately 10^4 cells/ml, which is far too low for sensitive issues of environmental contamination. A better and more sensitive way of assessing both the presence and quantity of these organisms is by the use of a PCR reaction coupled with a suitable probe. Sensitivity is now orders of magnitude greater than previously achievable, it being possible to detect far beyond the limits required by issues of public health. Thus it is possible to detect as little as 1 – 10 cells/ml, which equates to between 1 and 10 fg of DNA depending on the target organism.

For these environmentally important species a panel of very precise probes are used after DNA extraction and PCR of target DNA. The very high specificity of this technique is essential as there are frequently several different species in a sample, most of which will be benign. Because the detection is based on a presence/absence assay, the PCR products are not run out on a gel, but dot blotted onto a suitable membrane, usually nylon or one of the many proprietary brands available. Once immobilized onto the membrane a probe is used which is labeled at the 5'-end. These dot blots are then probed using the complementary sequence to the target. Various probes are used for this, depending upon the target. For *Legionella pneumophila* the probe is based on the macrophage infecting protein (*mip*), and for total coliform bacteria it is based upon a section of the *lacZ* gene of *Escherichia coli*.

30.4.6. Adulteration of Food Stuffs

Perhaps surprisingly, food labeling is not always as accurate as might be supposed, sometimes

accuracy has to be checked. Any question which can be asked as to whether, for example, if a tin of tuna is tuna, if meat products are from the species that is stated on the product, whether for human consumption or pet food. It is even possible to determine whether pasta is correctly made from Durum wheat or from bread wheat. Durum wheat is a hard wheat, while bread wheat is soft, and this considerably alters the stability of pasta when it is cooked.

Given the importance of checking foods for adulteration this is one of the most rapidly developing areas of applied DNA analysis. The constraints are similar to environmental testing, such that it is essential to have an unequivocal standard against which to measure the result. This may seem self-evident, but if you are trying to decide whether a food is what it says it is, there is little point in comparing your results against a standard for which you do not have the provenance, or chain of custody. Mistakes on these lines have happened.

Generally, the methods of checking for authenticity of species origin are based on DNA targets within the cytochrome B gene. This is located on the mitochondrial DNA and has a high species specificity. Although autoclaving material is usually regarded as an adequate decontaminating procedure for material being used in DNA analysis involving PCR, this may not necessarily be so. For example, it is possible to determine the species of animal from which a pet food is produced, even though it has been rigorously cooked at high temperature and pressure. Similarly the checking of canned tuna for authenticity is important. This is because "tuna" covers a range of fish in several different genera, so to be sure that a can labelled "skipjack", really is of the genus *Katsuwonus* rather than *Thunnus* (tunny) or *Germo* (the albacore) it becomes necessary to use DNA analysis. In these cases, as with any cooked product, it is generally possible to produce a repeatable result, unless the food stuff has been cooked at high temperature in an acidic liquor, after which results are unreliable and frequently non-existent.

Whenever food adulteration is suspected, it should be remembered that PCR is a technique of extraordinary sensitivity. Any result must be quantified, the possibility of accidental contamination must always be considered.

30.5. References

[1] B. Budowle, R. Chakraborty, A. M. Giusti, A. J. Eisenberg, R. C. Allen, *Am. J. Hum. Genetics* **48** (1991) 137–144.

[2] C. M. Hearne, S. Ghsoh, J. A. Todd, *TIG* **8** (1992) 288–294.

[3] L. Kricka (ed.): *Nonisotopic DNA Probe Techniques.* Academic Press, London 1992.

[4] K. M. Sullivan, S. Pope, P. Gill, J. M. Robertson, *PCR Methods and Applications* **2** (1992) 34–40.

[5] C. P. Kimpton, N. J. Oldroyd, S. K. Watson, R. Frazier, P. E. Johnson, et al., *Electrophoresis* **17** (1996) 1283–1293.

[6] S. Anderson, A. T. Bankier, B. G. Barrell, M. H. L. deBruijn, A. R. Coulson, et al., *Nature (London)* **290** (1981) 457–465.

[7] E. Hagelberg, J. B. Clegg, *Proc. R. Soc. London* **B 252** (1993) 163–170.

[8] S. M. Maurer, R. B. Firestone, C. R. Scriver, *Nature (London)* **405** (2000) 117–120.

[9] D. Chadwick, G. Cardew (eds.): *Variation in the Human Genome,* Ciba Foundation Symposium 197, John Wiley and Sons, Chichester 1996.

[10] R. Griffiths, B. Tiwari, *Nature (London)* **375** (1995) 454.

[11] R. L. Alford, H. A. Hammond, I. Coto, C. T. Caskey, *Am. J. Hum. Genet.* **55** (1994) 190–195.

Index