

MATERIALS CHARACTERIZATION

Introduction to Microscopic and Spectroscopic Methods

MATERIALS CHARACTERIZATION

Introduction to Microscopic and Spectroscopic Methods

Yang Leng

Hong Kong University of Science and Technology



John Wiley & Sons (Asia) Pte Ltd

Copyright © 2008

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop, # 02-01,
Singapore 129809

Visit our Home Page on www.wiley.com

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as expressly permitted by law, without either the prior written permission of the Publisher, or authorization through payment of the appropriate photocopy fee to the Copyright Clearance Center. Requests for permission should be addressed to the Publisher, John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop, #02-01, Singapore 129809, tel: 65-64632400, fax: 65-64646912, email: enquiry@wiley.com.sg

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The Publisher is not associated with any product or vendor mentioned in this book. All trademarks referred to in the text of this publication are the property of their respective owners.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Other Wiley Editorial Offices

John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons Canada Ltd, 6045 Freemont Blvd, Mississauga, ONT, L5R 4J3, Canada

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Leng, Yang

Materials characterization / Yang Leng.

p. cm.

Includes bibliographical references and index.

ISBN 978-0-470-82298-2 (cloth)

1. Materials. 2. Materials—Analysis. I. Title.

TA403.L4335 2008

620.1'1—dc22

2008001782

ISBN 978-0-470-82298-2 (HB)

Typeset in 10/12pt Times by Thomson Digital, Noida, India.

Printed and bound in Singapore by Markono Print Media Pte Ltd, Singapore.

This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Contents

Preface	xi
1 Light Microscopy	1
1.1 Optical Principles	1
1.1.1 Image Formation	1
1.1.2 Resolution	3
1.1.3 Depth of Field	5
1.1.4 Aberrations	6
1.2 Instrumentation	8
1.2.1 Illumination System	9
1.2.2 Objective Lens and Eyepiece	13
1.3 Specimen Preparation	15
1.3.1 Sectioning	15
1.3.2 Mounting	16
1.3.3 Grinding and Polishing	19
1.3.4 Etching	22
1.4 Imaging Modes	25
1.4.1 Bright-Field and Dark-Field Imaging	25
1.4.2 Phase Contrast Microscopy	26
1.4.3 Polarized Light Microscopy	30
1.4.4 Nomarski Microscopy	34
1.4.5 Fluorescence Microscopy	37
1.5 Confocal Microscopy	38
1.5.1 Working Principles	39
1.5.2 Three-Dimensional Images	41
References	43
Questions	43
2 X-ray Diffraction Methods	45
2.1 X-ray Radiation	45
2.1.1 Generation of X-rays	45
2.1.2 X-ray Absorption	48
2.2 Theoretical Background of Diffraction	49
2.2.1 Diffraction Geometry	49
2.2.2 Diffraction Intensity	56

2.3	X-ray Diffractometry	58
2.3.1	Instrumentation	59
2.3.2	Samples and Data Acquisition	61
2.3.3	Distortions of Diffraction Spectra	64
2.3.4	Applications	66
2.4	Wide Angle X-ray Diffraction and Scattering	71
2.4.1	Wide Angle Diffraction	71
2.4.2	Wide Angle Scattering	74
	References	76
	Questions	76
3	Transmission Electron Microscopy	79
3.1	Instrumentation	79
3.1.1	Electron Sources	81
3.1.2	Electromagnetic Lenses	83
3.1.3	Specimen Stage	85
3.2	Specimen Preparation	86
3.2.1	Pre-Thinning	86
3.2.2	Final Thinning	87
3.3	Image Modes	89
3.3.1	Mass-Density Contrast	90
3.3.2	Diffraction Contrast	92
3.3.3	Phase Contrast	96
3.4	Selected Area Diffraction	101
3.4.1	Selected Area Diffraction Characteristics	101
3.4.2	Single-Crystal Diffraction	103
3.4.3	Multi-Crystal Diffraction	108
3.4.4	Kikuchi Lines	108
3.5	Images of Crystal Defects	111
3.5.1	Wedge Fringe	111
3.5.2	Bending Contours	113
3.5.3	Dislocations	115
	References	118
	Questions	118
4	Scanning Electron Microscopy	121
4.1	Instrumentation	121
4.1.1	Optical Arrangement	121
4.1.2	Signal Detection	123
4.1.3	Probe Size and Current	124
4.2	Contrast Formation	129
4.2.1	Electron–Specimen Interactions	129
4.2.2	Topographic Contrast	130
4.2.3	Compositional Contrast	133
4.3	Operational Variables	134
4.3.1	Working Distance and Aperture Size	134

4.3.2 Acceleration Voltage and Probe Current	137
4.3.3 Astigmatism	138
4.4 Specimen Preparation	138
4.4.1 Preparation for Topographic Examination	139
4.4.2 Preparation for Micro-Composition Examination	142
4.4.3 Dehydration	142
References	143
Questions	144
5 Scanning Probe Microscopy	145
5.1 Instrumentation	145
5.1.1 Probe and Scanner	147
5.1.2 Control and Vibration Isolation	148
5.2 Scanning Tunneling Microscopy	148
5.2.1 Tunneling Current	148
5.2.2 Probe Tips and Working Environments	149
5.2.3 Operational Modes	149
5.2.4 Typical Applications	150
5.3 Atomic Force Microscopy	152
5.3.1 Near-Field Forces	152
5.3.2 Force Sensors	154
5.3.3 Operational Modes	155
5.3.4 Typical Applications	161
5.4 Image Artifacts	165
5.4.1 Tip	165
5.4.2 Scanner	167
5.4.3 Vibration and Operation	168
References	169
Questions	169
6 X-ray Spectroscopy for Elemental Analysis	171
6.1 Features of Characteristic X-rays	171
6.1.1 Types of Characteristic X-rays	173
6.1.2 Comparison of K, L and M Series	175
6.2 X-ray Fluorescence Spectrometry	176
6.2.1 Wavelength Dispersive Spectroscopy	179
6.2.2 Energy Dispersive Spectroscopy	183
6.3 Energy Dispersive Spectroscopy in Electron Microscopes	186
6.3.1 Special Features	186
6.3.2 Scanning Modes	187
6.4 Qualitative and Quantitative Analysis	189
6.4.1 Qualitative Analysis	189
6.4.2 Quantitative Analysis	191
References	195
Questions	195

7 Electron Spectroscopy for Surface Analysis	197
7.1 Basic Principles	197
7.1.1 X-ray Photoelectron Spectroscopy	197
7.1.2 Auger Electron Spectroscopy	198
7.2 Instrumentation	201
7.2.1 Ultra-High Vacuum System	201
7.2.2 Source Guns	202
7.2.3 Electron Energy Analyzers	204
7.3 Characteristics of Electron Spectra	206
7.3.1 Photoelectron Spectra	206
7.3.2 Auger Electron Spectra	208
7.4 Qualitative and Quantitative Analysis	209
7.4.1 Qualitative Analysis	209
7.4.2 Quantitative Analysis	219
7.4.3 Composition Depth Profiling	221
References	222
Questions	223
8 Secondary Ion Mass Spectrometry for Surface Analysis	225
8.1 Basic Principles	226
8.1.1 Secondary Ion Generation	226
8.1.2 Dynamic and Static SIMS	229
8.2 Instrumentation	230
8.2.1 Primary Ion System	230
8.2.2 Mass Analysis System	234
8.3 Surface Structure Analysis	237
8.3.1 Experimental Aspects	238
8.3.2 Spectrum Interpretation	239
8.4 SIMS Imaging	244
8.4.1 Generation of SIMS Images	244
8.4.2 Image Quality	245
8.5 SIMS Depth Profiling	245
8.5.1 Generation of Depth Profiles	246
8.5.2 Optimization of Depth Profiling	246
References	250
Questions	250
9 Vibrational Spectroscopy for Molecular Analysis	253
9.1 Theoretical Background	253
9.1.1 Electromagnetic Radiation	253
9.1.2 Origin of Molecular Vibrations	255
9.1.3 Principles of Vibrational Spectroscopy	257
9.1.4 Normal Mode of Molecular Vibrations	259
9.1.5 Infrared and Raman Activity	261

9.2	Fourier Transform Infrared Spectroscopy	267
9.2.1	Working Principles	267
9.2.2	Instrumentation	269
9.2.3	Fourier Transform Infrared Spectra	271
9.2.4	Examination Techniques	273
9.2.5	Fourier Transform Infrared Microspectroscopy	276
9.3	Raman Microscopy	279
9.3.1	Instrumentation	280
9.3.2	Fluorescence Problem	283
9.3.3	Raman Imaging	284
9.3.4	Applications	285
9.4	Interpretation of Vibrational Spectra	290
9.4.1	Qualitative Methods	290
9.4.2	Quantitative Methods	297
	References	299
	Questions	300
10	Thermal Analysis	301
10.1	Common Characteristics	301
10.1.1	Thermal Events	301
10.1.2	Instrumentation	303
10.1.3	Experimental Parameters	303
10.2	Differential Thermal Analysis and Differential Scanning Calorimetry	305
10.2.1	Working Principles	305
10.2.2	Experimental Aspects	309
10.2.3	Measurement of Temperature and Enthalpy Change	312
10.2.4	Applications	315
10.3	Thermogravimetry	319
10.3.1	Instrumentation	321
10.3.2	Experimental Aspects	322
10.3.3	Interpretation of Thermogravimetric Curves	326
10.3.4	Applications	328
	References	331
	Questions	331
Index		333

Preface

This book serves as a textbook for introductory level courses on materials characterization for university students, and also a reference book for scientists and engineers who need basic knowledge and skills in materials characterization. After teaching courses in materials characterization for several years, I strongly feel that a comprehensive single-volume book covering state-of-the-art techniques in materials characterization is clearly needed. The book is based on my teaching notes for an introductory undergraduate course on materials characterization.

This book covers two main aspects of materials characterization: materials structures and chemical analysis. The first five chapters cover commonly used techniques for microstructure analysis including light microscopy, transmission and scanning electron microscopy and scanning probe microscopy. X-ray diffraction is introduced in Chapter 2, even though it is not a microscopic method, because it is the main technique currently used to determine crystal structures. The basic concepts of diffraction introduced in Chapter 2 are also important for understanding analytical methods of transmission electron microscopy. The final five chapters of the book mainly cover techniques for chemical analysis. These chapters introduce the most widely used chemical analysis techniques for solids: fluorescence X-ray spectroscopy, X-ray energy dispersive spectroscopy for microanalysis in electron microscopes, popular surface chemical analysis techniques of X-ray photoelectron spectroscopy (XPS) and secondary ion mass spectroscopy (SIMS), and the molecular vibrational spectroscopy methods of Fourier transform infrared (FTIR) and Raman. Thermal analysis is a rather unique technique and hard to categorize as it is neither a microscopic nor a spectroscopic method. I include thermal analysis in this book because of its increasing applications in materials characterization.

I have tried to write the book in an easy-to-read style by keeping theoretical discussion to a minimal level of mathematics and physics. Only are theories introduced when they are required for readers to understand related working principles of characterization techniques. Technical aspects of preparing specimens and operating instruments are also introduced so as to help readers understand what should be done and what should be avoided in practice. For most engineers and scientists, the interpretation and analysis of characterization outputs are even more important than the technical skills of characterization. Thus, the book provides a number of examples for each characterization technique to help readers interpret and understand analysis outputs. In addition, problems are provided at the end of each chapter for those using the book as a course text.

Note that there are paragraphs printed in italic font in Chapters 2, 3, 4, 5, 6, 7, 9 and 10. These paragraphs provide more extensive knowledge, and are not essential in order for readers to understand the main contents of this book. Readers may skip these paragraphs if they choose.

Finally, I gratefully acknowledge the considerable help I have received from my colleagues as well as my graduate students at the Hong Kong University of Science and Technology during preparation of the manuscript. In particular, the following people helped me to review the manuscript and provided valuable comments: Jiaqi Zheng for Chapter 2, Ning Wang for Chapter 3, Gu Xu for Chapter 4, Jie Xhie for Chapter 5, Lutao Weng for Chapters 7 and 8 and Borong Shi for Chapter 9. Jingshen Wu, who shares teaching duties with me in the materials characterization course, provided valuable suggestions and a number of great micrographs. My graduate students Xiong Lu and Renlong Xin, and especially Xiang Ge, have drawn and reproduced many figures of this book. Also, I am particularly grateful to my friend Martha Dahlen who provided valuable suggestions in English writing and editing.

Yang Leng

The Solutions Manual for this book can be found at <http://www.wiley.com/go/leng>

1

Light Microscopy

Light or optical microscopy is the primary means for scientists and engineers to examine the microstructure of materials. The history of using a light microscope for microstructural examination of materials can be traced back to the 1880s. Since then, light microscopy has been widely used by metallurgists to examine metallic materials. Light microscopy for metallurgists became a special field named *metallography*. The basic techniques developed in metallography are not only used for examining metals, but also are used for examining ceramics and polymers. In this chapter, light microscopy is introduced as a basic tool for microstructural examination of materials including metal, ceramics and polymers.

1.1 Optical Principles

1.1.1 Image Formation

Reviewing the optical principles of microscopes should be the first step to understanding light microscopy. The optical principles of microscopes include image formation, magnification and resolution. Image formation can be illustrated by the behavior of a light path in a compound light microscope as shown in Figure 1.1. A specimen (*object*) is placed at position A where it is between one and two focal lengths from an *objective lens*. Light rays from the object firstly converge at the objective lens and are then focused at position B to form a magnified inverted image. The light rays from the image are further converged by the second lens (*projector lens*) to form a final magnified image of an object at C.

The light path shown in Figure 1.1 generates the real image at C on a screen or camera film, which is not what we see with our eyes. Only a real image can be formed on a screen and photographed. When we examine microstructure with our eyes, the light path in a microscope goes through an *eyepiece* instead of projector lens to form a *virtual image* on the human eye retina, as shown in Figure 1.2. The virtual image is inverted with respect to the object. The virtual image is often adjusted to be located as the minimum distance of eye focus, which is conventionally taken as 250 mm from the eyepiece. A modern microscope is commonly equipped with a device to switch between eyepiece and projector lens for either recording images on photographic film or sending images to a computer screen.

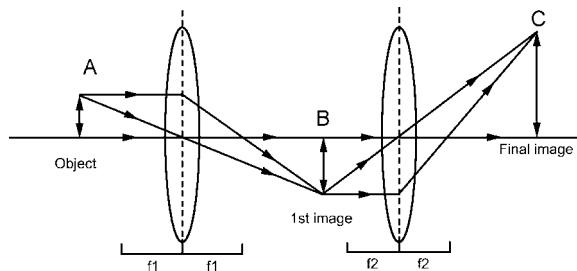


Figure 1.1 Principles of magnification in a microscope.

Advanced microscopes made since 1980 have a more complicated optical arrangement called ‘infinity-corrected’ optics. The objective lens of these microscopes generates parallel beams from a point on the object. A tube lens is added between the objective and eyepiece to focus the parallel beams to form an image on a plane, which is further viewed and enlarged by the eyepiece.

The magnification of a microscope can be calculated by linear optics, which tells us the magnification of a convergent lens M .

$$M = \frac{v - f}{f} \quad (1.1)$$

where f is the focal length of the lens and v is the distance between the image and lens. A higher magnification lens has a shorter focal length as indicated by Equation 1.1. The total

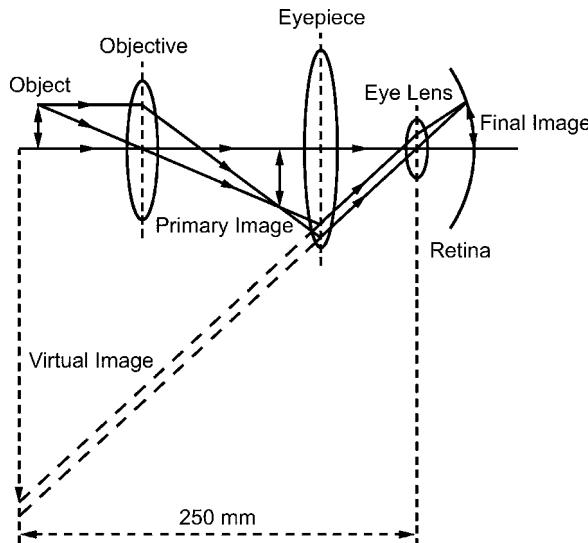


Figure 1.2 Schematic path of light in a microscope with eyepiece. The virtual image is reviewed by a human eye composed of eye lens and retina.

magnification of a compound microscope as shown in Figure 1.1 should be the magnification of the objective lens multiplied by that of the projector lens.

$$M = M_1 M_2 = \frac{(v_1 - f_1)(v_2 - f_2)}{f_1 f_2} \quad (1.2)$$

When an eyepiece is used, the total magnification should be the objective lens magnification multiplied by eyepiece magnification.

1.1.2 Resolution

We naturally ask whether there is any limitation for magnification in light microscopes because Equation 1.2 suggests there is no limitation. However, meaningful magnification of a light microscope is limited by its *resolution*. Resolution refers to the minimum distance between two points at which they can be visibly distinguished as two points. The resolution of a microscope is theoretically controlled by the diffraction of light.

Light diffraction controlling the resolution of microscope can be illustrated with the images of two self-luminous point objects. When the point object is magnified, its image is a central spot (the *Airy disk*) surrounded by a series of diffraction rings (Figure 1.3), not a single spot. To distinguish between two such point objects separated by a short distance, the Airy disks should not severely overlap each other. Thus, controlling the size of the Airy disk is the key to controlling resolution. The size of the Airy disk (d) is related to wavelength of light (λ)

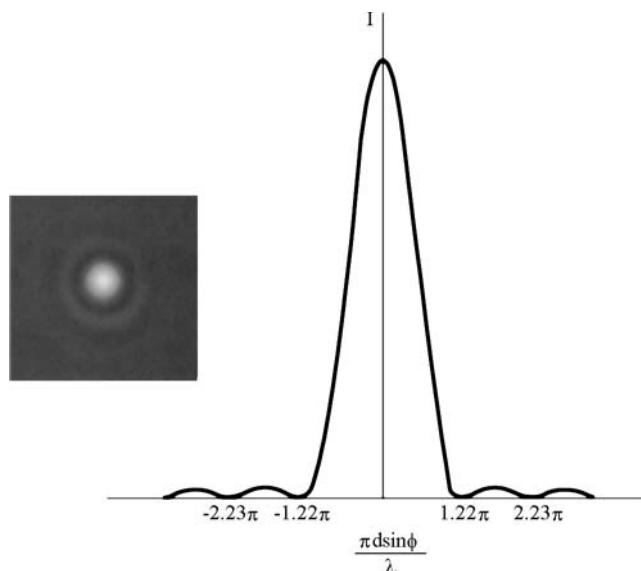


Figure 1.3 A self-luminous point object and the light intensity distribution along a line passing through its center.

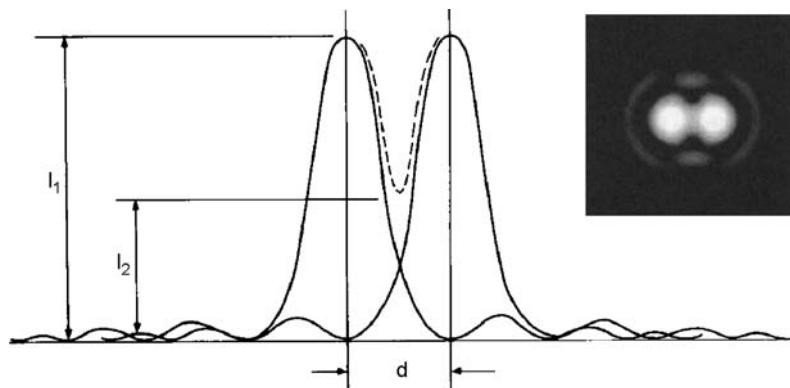


Figure 1.4 Intensity distribution of two Airy disks with a distance $\frac{d}{2}$. I_1 indicates the maximum intensity of each point and I_2 represents overlap intensity.

and angle of light coming into the lens. The resolution of a microscope (R) is defined as the minimum distance between two Airy disks that can be distinguished (Figure 1.4). Resolution is a function of microscope parameters as shown in the following equation.

$$R = \frac{d}{2} = \frac{0.61\lambda}{\mu \sin \alpha} \quad (1.3)$$

where μ is the refractive index of the medium between the object and objective lens and α is the half-angle of the cone of light entering the objective lens (Figure 1.5). The product, $\mu \sin \alpha$, is called the *numerical aperture (NA)*.

According to Equation 1.3, to achieve higher resolution we should use shorter wavelength light and larger NA . The shortest wavelength of visible light is about 400 nm,

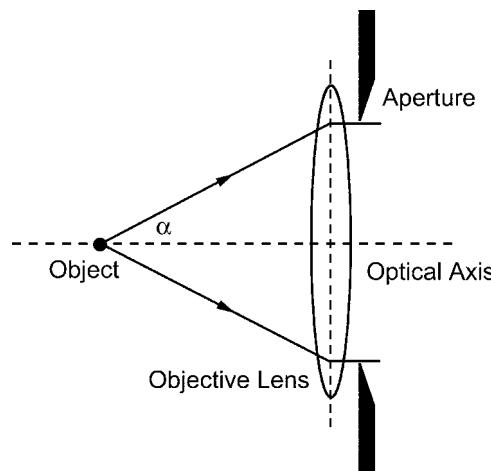


Figure 1.5 The cone of light entering an objective lens showing α is the half angle.

while the NA of the lens depends on α and the medium between the lens and object. Two media between object and objective lens are commonly used: either air for which $\mu = 1$, or oil for which $\mu \approx 1.5$. Thus, the maximum value of NA is about 1.5. We estimate the best resolution of a light microscope from Equation 1.3 as about $0.2 \mu\text{m}$.

Effective Magnification

Magnification is meaningful only in so far as the human eye can see the features resolved by the microscope. Meaningful magnification is the magnification that is sufficient to allow the eyes to see the microscopic features resolved by the microscope. A microscope should enlarge features to about 0.2 mm, the resolution level of the human eye. Thus, the *effective magnification* of light microscope should approximately be $M_{\text{eff}} = 0.2 \div 0.2 \times 10^3 = 1.0 \times 10^3$.

A higher magnification than the effective magnification only makes the image bigger, may make eyes more comfortable during observation, but does not provide more detail in an image.

Brightness and Contrast

To make a microscale object in a material specimen visible, high magnification is not sufficient. A microscope should also generate sufficient *brightness* and *contrast* of light from the object. Brightness refers to the intensity of light. In a transmission light microscope the brightness is related to the numerical aperture (NA) and magnification (M).

$$\text{Brightness} = \frac{(NA)^2}{M^2} \quad (1.4)$$

In a reflected light microscope the brightness is more highly dependent on NA .

$$\text{Brightness} = \frac{(NA)^4}{M^2} \quad (1.5)$$

These relationships indicate that the brightness decreases rapidly with increasing magnification, and controlling NA is not only important for resolution but also for brightness, particularly in a reflected light microscope.

Contrast is defined as the relative change in light intensity (I) between an object and its background.

$$\text{Contrast} = \frac{I_{\text{object}} - I_{\text{background}}}{I_{\text{background}}} \quad (1.6)$$

Visibility requires that the contrast of an object exceeds a critical value called the *contrast threshold*. The contrast threshold of an object is not constant for all images but varies with image brightness. In bright light, the threshold can be as low as about 3% while in dim light the threshold is greater than 200%.

1.1.3 Depth of Field

Depth of field is an important concept when photographing an image. It refers to the range of position for an object in which image sharpness does not change. As illustrated in Figure 1.6, an object image is only accurately in focus when the object lies in a plane within a certain

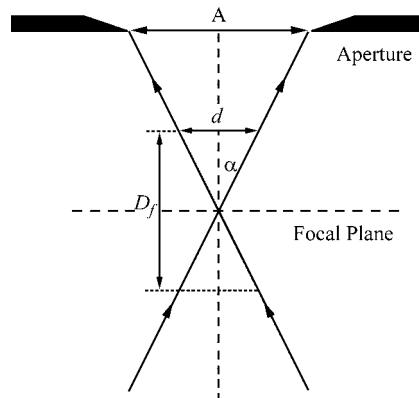


Figure 1.6 Geometric relation among the depth of field (D_f), the half angle entering the objective lens (α) and the size of Airy disk (d).

distance from the objective lens. The image is out of focus when the object lies either closer to or farther from the lens. Since the diffraction effect limits the resolution R , it does not make any difference to the sharpness of the image if the object is within the range of D_f shown in Figure 1.6. Thus, the depth of field can be calculated.

$$D_f = \frac{d}{\tan \alpha} = \frac{2R}{\tan \alpha} = \frac{1.22\lambda}{\mu \sin \alpha \tan \alpha} \quad (1.7)$$

Equation 1.7 indicates that a large depth of field and high resolution cannot be obtained simultaneously; thus, a larger D_f means a larger R and worse resolution. We may reduce angle α to obtain a better depth of field only at the expense of resolution. For a light microscope, α is around 45° and the depth of field is about the same as its resolution.

We should not confuse *depth of field* with *depth of focus*. Depth of focus refers to the range of image plane positions at which the image can be viewed without appearing out of focus for a fixed position of the object. In other words, it is the range of screen positions in which images can be projected in focus. The depth of focus is M^2 times larger than the depth of field.

1.1.4 Aberrations

The aforementioned calculations of resolution and depth of field are based on the assumptions that all components of the microscope are perfect, and that light rays from any point on an object focus on a correspondingly unique point in the image. Unfortunately, this is almost impossible due to image distortions by the lens called *lens aberrations*. Some aberrations affect the whole field of the image (*chromatic* and *spherical aberrations*), while others affect only off-axis points of the image (*astigmatism* and *curvature of field*). The true resolution and depth of field can be severely diminished by lens aberrations. Thus, it is important for us to have a basic knowledge of aberrations in optical lenses.

Chromatic aberration is caused by the variation in the refractive index of the lens in the range of light wavelengths (*light dispersion*). The refractive index of lens glass is greater for

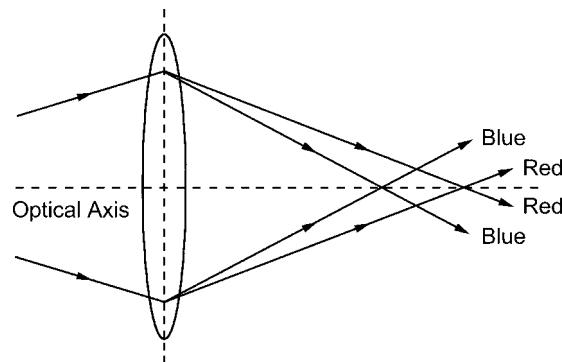


Figure 1.7 Paths of rays in white light illustrating chromatic aberration.

shorter wavelengths (for example, blue) than for longer wavelengths (for example, red). Thus, the degree of light deflection by a lens depends on the wavelength of light. Because a range of wavelengths is present in ordinary light (white light), light cannot be focused at a single point. This phenomenon is illustrated in Figure 1.7.

Spherical aberration is caused by the spherical curvature of a lens. Light rays from a point on the object on the optical axis enter a lens at different angles and cannot be focused at a single point, as shown in Figure 1.8. The portion of the lens farthest from the optical axis brings the rays to a focus nearer the lens than does the central portion of the lens.

Astigmatism results when the rays passing through vertical diameters of the lens are not focused on the same image plane as rays passing through horizontal diameters, as shown in Figure 1.9. In this case, the image of a point becomes an elliptical streak at either side of the best focal plane. Astigmatism can be severe in a lens with asymmetric curvature.

Curvature of field is an off-axis aberration. It occurs because the focal plane of an image is not flat but has a concave spherical surface, as shown in Figure 1.10. This aberration is especially troublesome with a high magnification lens with a short focal length. It may cause unsatisfactory photography.

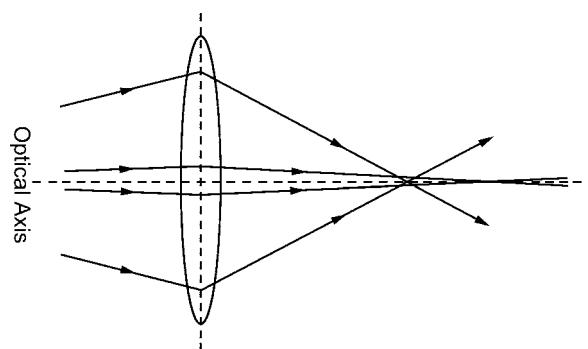


Figure 1.8 Spherical aberration.

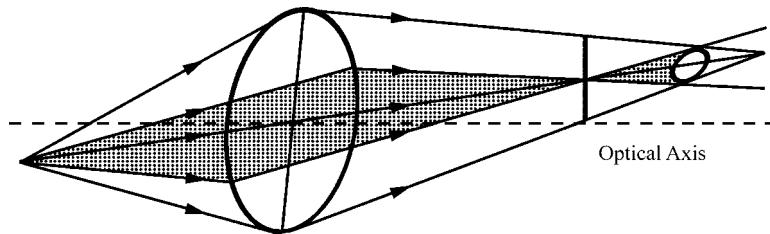


Figure 1.9 Astigmatism is an off-axis aberration.

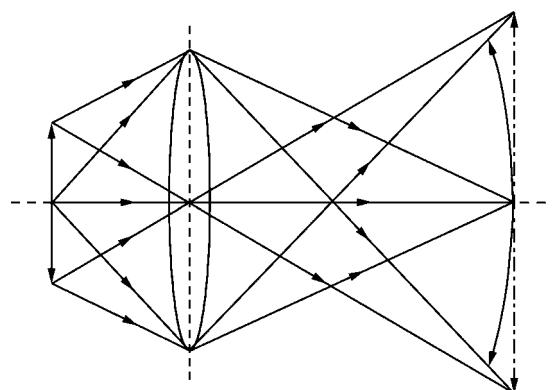


Figure 1.10 Curvature of field is an off-axis aberration.

There are a number of ways to compensate for and/or reduce lens aberrations. For example, combining lenses with different shapes and refractive indices corrects chromatic and spherical aberrations. Selecting single wavelength illumination by the use of filters helps eliminate chromatic aberrations. We expect that the extent to which lens aberrations have been corrected is reflected in the cost of the lens. It is a reason that we see huge price variation in microscopes.

1.2 Instrumentation

A light microscope includes the following main components:

- Illumination system;
- Objective lens;
- Eyepiece;
- Photomicrographic system; and
- Specimen stage.

A light microscope for examining material microstructure can use either transmitted or reflected light for illumination. *Reflected light microscopes* are the most commonly used for metallography, while *transmitted light microscopes* are typically used to examine transparent

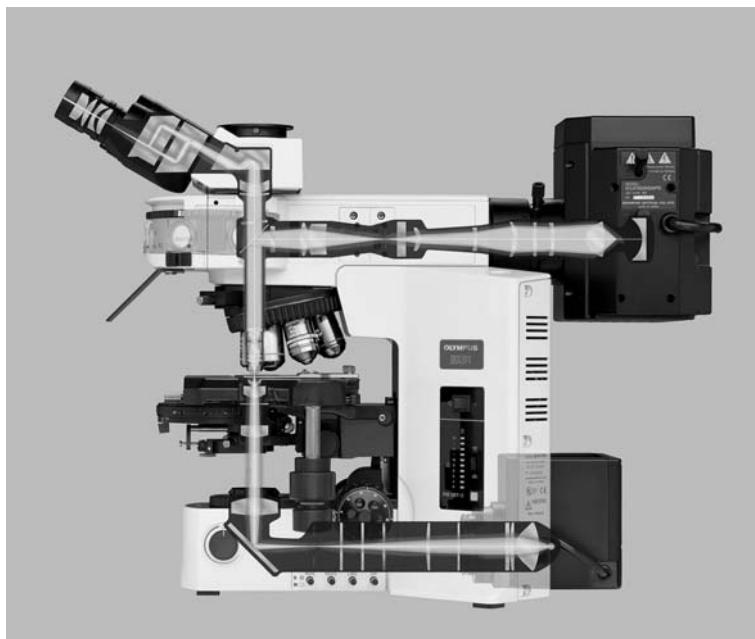


Figure 1.11 An Olympus light microscope used for material examination. The microscope includes transmission and reflection illumination systems. (This image is courtesy of Olympus Corporation.)

or semi-transparent materials, such as certain types of polymers. Figure 1.11 illustrates the structure of a light microscope for materials examination.

1.2.1 Illumination System

The illumination system of a microscope provides visible light by which the specimen is observed. There are three main types of electric lamps used in light microscopes:

1. Low-voltage tungsten filament bulbs;
2. Tungsten–halogen bulbs; and
3. Gas discharge tubes.

Tungsten bulbs provide light of a continuous wavelength spectrum from about 300 to 1500 nm. The *color temperature* of the light, which is important for color photography, is relatively low. Low color temperature implies warmer (more yellow–red) light while high color temperature implies colder (more blue) light. Tungsten–halogen bulbs, like ordinary tungsten bulbs, provide a continuous spectrum. The light is brighter and the color temperature is significantly higher than ordinary tungsten bulbs. The high filament temperature of tungsten–halogen bulbs, however, needs a heat filter in the light path and good ventilation. Gas discharge tubes filled with pressurized mercury or xenon vapor provide extremely high brightness. The more commonly used tubes are filled with mercury, of which the arc has a

discontinuous spectrum. Xenon has a continuous spectrum and very high color temperature. As with tungsten–halogen bulbs, cooling is required for gas discharge tubes.

In a modern microscope, the illumination system is composed of a light lamp (commonly a tungsten–halogen bulb), a *collector lens* and a *condenser lens* to provide integral illumination; such a system is known as the *Köhler system*. The main feature of the Köhler system is that the light from the filament of a lamp is firstly focused at the front focal plane of the condenser lens by a collector lens. Then, the condenser lens collects the light diverging from the source and directs it at a small area of the specimen being examined. The Köhler system provides uniform intensity of illumination on the area of specimen. The system generates two sets of *conjugate focal planes* along the optic axis of a microscope as shown in Figure 1.12. One set of focal planes is for illuminating rays; these are known as the *conjugate aperture planes*. Another set comprises the image-forming rays called the *conjugate field planes*. During normal microscope operation, we see only the image-forming rays through the eyepiece. We can use the illuminating rays to check the alignment of the optical system of the microscope.

There are two important controllable diaphragms in the Köhler system: the *field diaphragm* and the *aperture diaphragm*. The field diaphragm is placed at a focal plane for image formation. Its function is to alter the diameter of the illuminated area of the specimen. When the condenser is focused and centered, we see a sharp image of the field diaphragm with the image of specimen

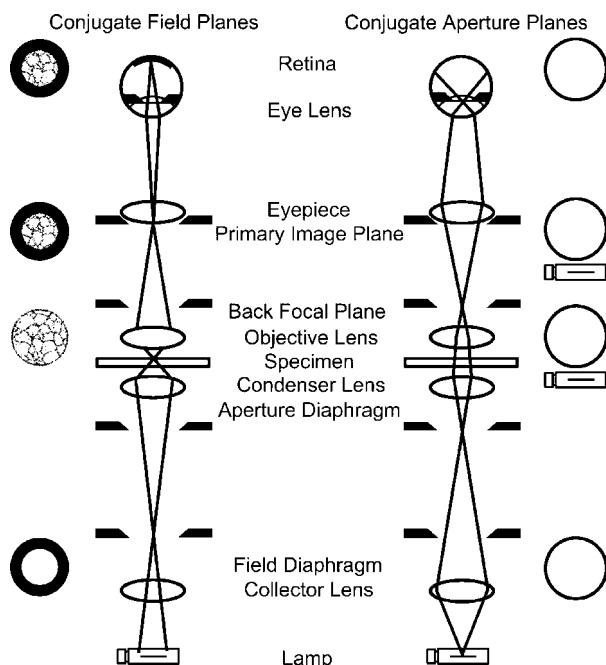


Figure 1.12 Two sets of conjugate focal planes in the Köhler system illustrated in a transmitted light microscope. Image-forming rays focus on the field planes and illuminating rays focus on the aperture planes. The far left-hand and far right-hand parts of the diagram illustrate the images formed by image-forming rays and illuminating rays, respectively. (Reproduced with permission from D.B. Murphy, *Fundamentals of Light Microscopy and Electronic Imaging*, Wiley-Liss. © 2001 John Wiley & Sons Inc.)

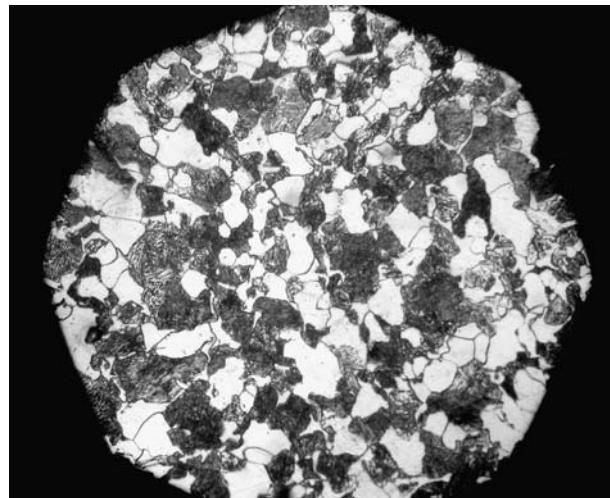


Figure 1.13 Image of the field diaphragm with an image of the specimen. Magnification 100 \times .

(Figure 1.13). The field diaphragm restricts the area of view and blocks scattering light that could cause glare and image degradation if they entered the objective lens and eyepiece. The aperture diaphragm is placed at a focus plane of illuminating rays. Its function is to control α , and thus affect the image resolution and depth of field (Section 1.2). We cannot see the aperture diaphragm with the image of specimen. Figure 1.14 illustrates the influence of the aperture diaphragm on the image of a specimen.

The main difference between transmitted light and reflected light microscopes is the illumination system. The Köhler system of reflected light illumination (*epi-illumination*) is illustrated in Figure 1.15 in which a *relay lens* is included. The illuminating rays are reflected by a semi-transparent reflector to illuminate the specimen through an objective lens. There is no difference in how reflected and transmitted light microscopes direct light rays after the rays leave the specimen. There may be a difference in the relative position of the field and aperture diaphragms (Figure 1.12). However, the field diaphragm is always on the focal plane of the image-forming rays while the aperture diaphragm is on a focal plane of the illuminating rays.

Light filters are often included in the light path of illumination systems, even though they are not shown in Figures 1.12 and 1.15. Light filters are used to control wavelengths and intensity of illumination in microscopes in order to achieve optimum visual examination for photomicrography. *Neutral density (ND) filters* can regulate light intensity without changing wavelength. *Colored filters* and *interference filters* are used to isolate specific colors or bands of wavelength. The colored filters are commonly used to produce a broad band of color, while the interference filters offers sharply defined bandwidths. Colored filters are used to match the color temperature of the light to that required by photographic films. Selected filters can also increase contrast between specimen areas with different colors. *Heat filters* absorb much of the infrared radiation which causes heating of specimen when a tungsten–halogen bulb is used as light source.

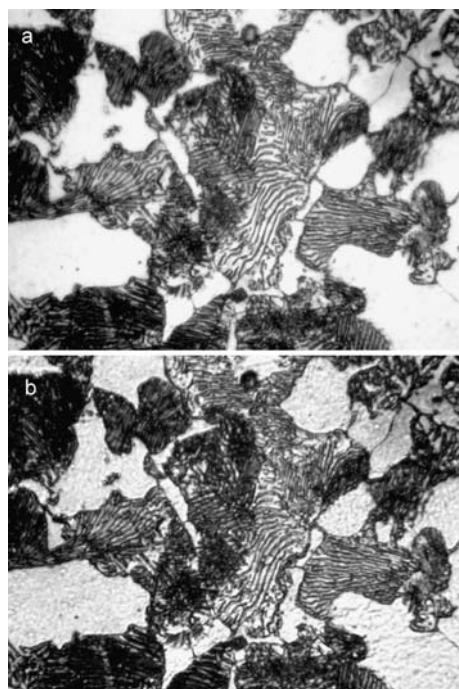


Figure 1.14 Effect of aperture diaphragm on specimen image when: (a) the aperture is large; and (b) the aperture is small. Magnification 500 \times .

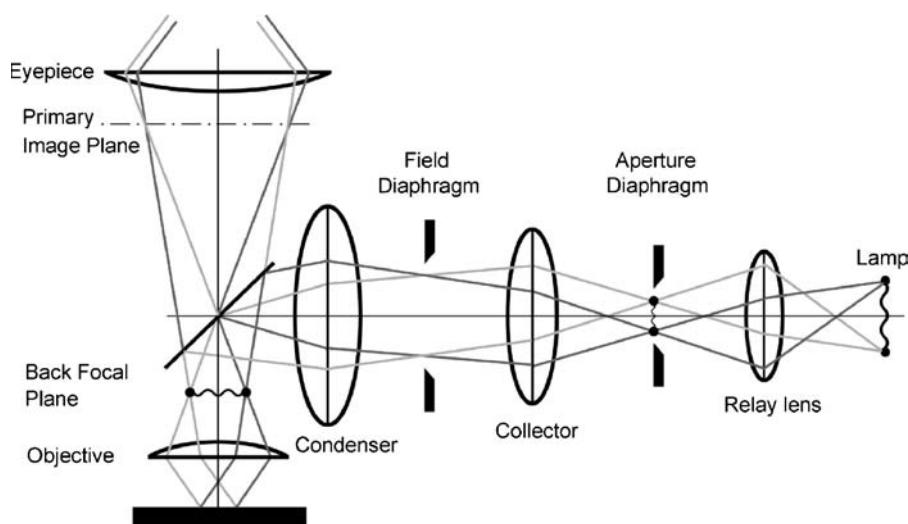


Figure 1.15 Illumination system of a reflected light microscope with illuminating rays.

1.2.2 Objective Lens and Eyepiece

The objective lens is the most important optical component of a light microscope. The magnification of the objective lens determines the total magnification of the microscope because eyepieces commonly have a fixed magnification of $10\times$. The objective lens generates the primary image of the specimen, and its resolution determines the final resolution of the image. The numerical aperture (*NA*) of the objective lens varies from 0.16 to 1.40, depending on the type of lens. A lens with a high magnification has a higher *NA*. The highest *NA* for a dry lens (where the medium between the lens and specimen is air) is about 0.95. Further increase in *NA* can be achieved by using a lens immersed in an oil medium. The oil immersion lens is often used for examining microstructure greater than $1000\times$ magnification.

Classification of the objective lens is based on its aberration correction capabilities, mainly chromatic aberration. The following lenses are shown from low to high capability.

- Achromat;
- Semi-achromat (also called ‘fluorite’); and
- Apochromat.

The achromatic lens corrects chromatic aberration for two wavelengths (red and blue). It requires green illumination to achieve satisfactory results for visual observation and black and white photography. The semi-achromatic lens improves correction of chromatic aberration. Its *NA* is larger than that of an achromatic lens with the same magnification and produces a brighter image and higher resolution of detail. The apochromatic lens provides the highest degree of aberration correction. It almost completely eliminates chromatic aberration. It also provides correction of spherical aberration for two colors. Its *NA* is even larger than that of a semi-achromatic lens. Improvement in quality requires a substantial increase in complexity of the lens structure, and costs. For example, an apochromatic lens may contain 12 or more optical elements.

The characteristics of an objective lens are engraved on the barrel as shown in Figure 1.16. Engraved markings may include the following abbreviations.

- ‘FL’, ‘FLUOR’ or ‘NEOFLUOR’ stands for ‘fluorite’ and indicates the lens is semi-achromatic;
- ‘APO’ indicates that the lens is apochromatic;
- If neither of the above markings appears, then the lens is achromatic;
- ‘PLAN’ or ‘PL’ stands for ‘planar’ and means the lens is corrected for curvature of field, and thus generates a flat field of image;
- ‘DIC’ means the lens includes a Wollaston prism for *differential interference contrast* (Section 1.4.4);
- ‘PH’ or ‘PHACO’ means the lens has a phase ring for phase contrast microscopy (Section 1.1.4.2); and
- ‘number/number’ indicates *magnification/numerical aperture*. Thus, ‘40/0.75’ means the lens has a magnification of $40\times$ and a numerical aperture of 0.75.

The eyepiece is used to view the real primary image formed by the objective lens. In some cases it also completes the correction of aberrations. The eyepiece allows a glass disc with an



Figure 1.16 Engraved markings on the barrel of an objective lens. (Reproduced with permission from D.B. Murphy, *Fundamentals of Light Microscopy and Electronic Imaging*, Wiley-Liss. © 2001 John Wiley & Sons Inc.)

etched graticule to be inserted into the optical path. The graticule serves as a micrometer for measurement. The eyepiece has either a helical thread or a sliding mount as a focusing mechanism. Importantly, the focusing mechanism of an eyepiece provides a ‘parfocal’ adjustment of the optics so that the same focal plane examined by the eye will be in focus on the film plane of the camera mounted on microscope. Thus, focusing the eyepiece is a necessary step before photographing images in a microscope.

We can summarize the methods for achieving optimum resolution and depth of field in light microscopy. While both resolution and depth of field are crucial for achieving high quality images, one often is achieved at the expense of the other. Thus, compromises must be made while using good judgment.

Steps for Optimum Resolution

- Use an objective lens with the highest *NA* possible;
- Use high magnification;
- Use an eyepiece compatible with the chosen objective lens;
- Use the shortest possible wavelength light;
- Keep the light system properly centered;
- Use oil immersion lenses if available;
- Adjust the field diaphragm for maximum contrast and the aperture diaphragm for maximum resolution and contrast; and
- Adjust brightness for best resolution.

Steps to Improve Depth of Field

- Reduce *NA* by closing the aperture diaphragm, or use an objective lens with lower *NA*;
- Lower the magnification for a given *NA*;

- Use a high-power eyepiece with a low-power, high-NA objective lens; and
- Use the longest possible wavelength light.

1.3 Specimen Preparation

The microstructure of a material can only be viewed in a light microscope after a specimen has been properly prepared. Metallurgists have developed extensive techniques and accumulated knowledge of metal specimen preparations for over a century. In principle, we can use these techniques to examine not only metallic materials but also ceramics and polymers; in practice, certain modifications are needed and a certain degree of caution must be exercised. The main steps of specimen preparation for light microscopy include the following.

- Sectioning;
- Mounting;
- Grinding;
- Polishing; and
- Etching.

1.3.1 Sectioning

Sectioning serves two purposes: generating a cross-section of the specimen to be examined; and reducing the size of a specimen to be placed on a stage of light microscope, or reducing size of specimen to be embedded in mounting media for further preparation processes. The main methods of sectioning are abrasive cutting, electric discharge machining and microtomy, which is mainly for polymer specimens.

Cutting

Abrasive cutting is the most commonly used method for sectioning materials. Specimens are sectioned by a thin rotating disc in which abrasive particles are supported by suitable media. The abrasive cutoff machine is commonly used for sectioning a large sample. The machine sections the sample with a rapidly rotating wheel made of an abrasive material, such as silicon carbide, and bonding materials such as resin and rubber. The wheels are consumed in the sectioning process. Abrasive cutting requires cooling media in order to reduce friction heat. Friction heat can damage specimens and generate artifacts in the microstructure. Commonly used cooling media consist of water-soluble oil and rust-inhibiting chemicals. The abrasive cutoff machine can section large specimens quickly but with poor precision.

More precise cutting can be achieved by a diamond saw or electric discharge machine (Figure 1.17). The diamond saw is a precision abrasive cutting machine. It sections specimens with a cutting wheel made of tiny diamond particles bonded to a metallic substrate. A cooling medium is also necessary for diamond saw cutting. Electrically conductive materials can be sectioned by an electric discharge machine (EDM). Cutting is accomplished by an electric discharge between an electrode and the specimen submerged in a dielectric fluid. EDM is particularly useful for materials that are difficult to be sectioned by abrasive cutting. EDM may produce significant changes at the machined surface because the electric discharge melts the solid in the cutting path. The solidified material along a machining path must be carefully removed during further preparation processes.

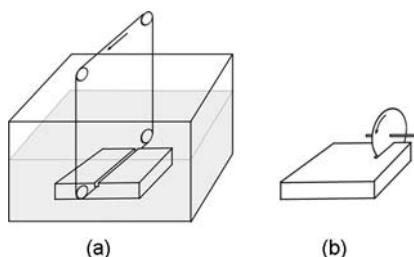


Figure 1.17 Specimen sectioning by: (a) wire cutting with electric discharging; and (b) diamond saw sectioning.

Microtomy

Microtomy refers to sectioning materials with a knife. It is a common technique in biological specimen preparation. It is also used to prepare soft materials such as polymers and soft metals. Tool steel, tungsten carbide, glass and diamond are used as knife materials. A similar technique, *ultramicrotomy*, is widely used for the preparation of biological and polymer specimens in transmission electron microscopy. This topic is discussed in Chapter 3.

1.3.2 Mounting

Mounting refers to embedding specimens in mounting materials (commonly thermosetting polymers) to give them a regular shape for further processing. Mounting is not necessary for bulky specimens, but it is required for specimens that are too small or oddly shaped to be handled or when the edge of a specimen needs to be examined in transverse section. Mounting is popular now because most automatic grinding and polishing machines require specimens to have a cylindrical shape. There are two main types of mounting techniques: *hot mounting* and *cold mounting*.

Hot mounting uses hot-press equipment as shown in Figure 1.18. A specimen is placed in the cylinder of a press and embedded in polymeric powder. The surface to be examined faces the bottom of the cylinder. Then, the specimen and powder are heated at about 150 °C under constant pressure for tens of minutes. Heat and pressure enable the powder to bond with the specimen to form a cylinder. Phenolic (bakelite) is the most widely used polymeric powder for hot mounting. Hot mounting is suitable for most metal specimens. However, if the microstructure of the material changes at the mounting temperature, cold mounting should be used.

In cold mounting, a polymer resin, commonly epoxy, is used to cast a mold with the specimen at ambient temperature. Figure 1.19a shows a typical mold and specimens for cold mounting. Figure 1.19b demonstrates the casting of epoxy resin into the mold in which the specimen surface to be examined is facing the bottom. A cold mounting medium has two constituents: a fluid resin and a powder hardener. The resin and hardener should be carefully mixed in proportion following the instructions provided. Curing times for mounting materials vary from tens of minutes to several hours, depending on the resin type. Figure 1.20 shows the specimens after cold mounted in various resins.

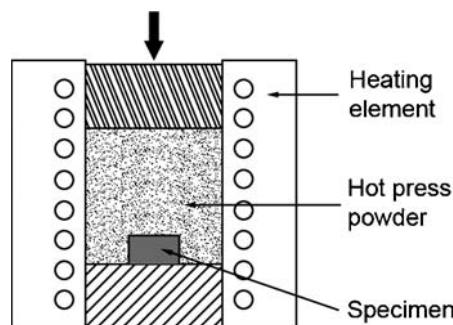


Figure 1.18 Internal arrangement of a hot mounting press.

An important issue in the selection of a mounting material is hardness compatibility with the specimen. Generally, plastics used for embedding are not as hard as the specimen, particularly when the specimens are of metallic or ceramic. Too great a difference in hardness can cause inhomogeneous grinding and polishing, which in turn may generate a rough, rather than sharp

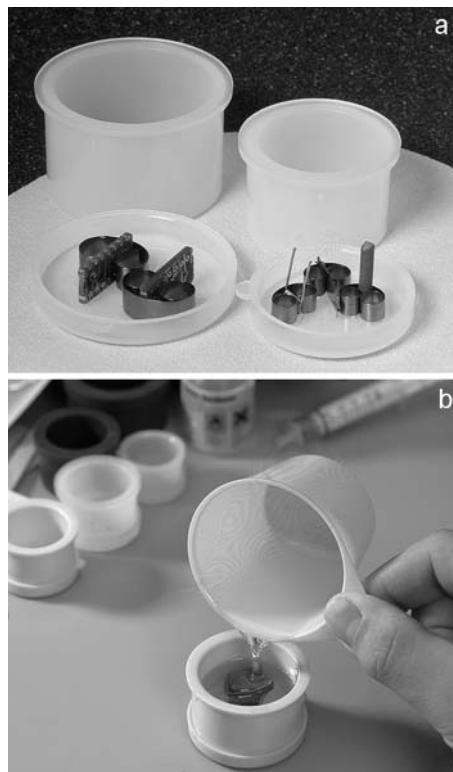


Figure 1.19 Cold mounting of specimens: (a) place specimens on the bottom of molds supported by clamps; and (b) cast resin into the mold. (Reproduced with permission of Struers A/S.)

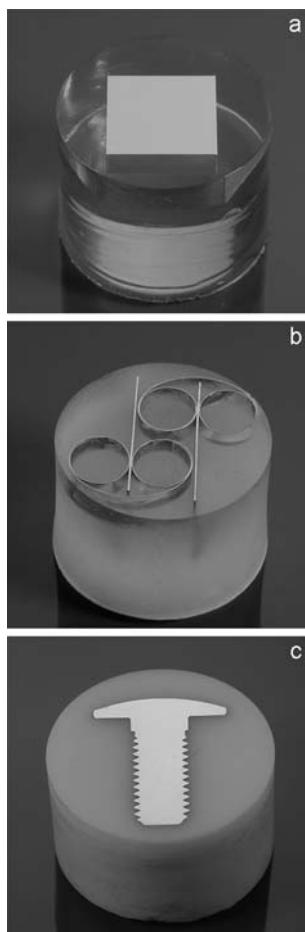


Figure 1.20 Cold mounted specimens: (a) mounted with polyester; (b) mounted with acrylic; and (c) mounted with acrylic and mineral fillers. (Reproduced with permission of Struers A/S.)

edge on the specimen. A solution to this problem is to embed metal beads with a specimen to ensure that the grinding surface has a more uniform hardness.

There are a number of other mounting techniques available but they are less widely used. The most simple is *mechanical clamping*, in which a thin sheet of the specimen is clamped in place with a mechanical device. *Adhesive mounting* is to glue a specimen to a large holder. *Vacuum impregnation* is a useful mounting method for porous specimens and ceramics. It removes air from the pores, crevices and cracks of specimens, and then replaces such empty space in the specimen with epoxy resin. Firstly, a specimen is ground with grit paper to flatten the surface to be examined. The specimen is placed with the surface uppermost inside the mold in a vacuum chamber. Then, the chamber is evacuated for several minutes before filling the mold with epoxy. The vacuum is maintained for a few minutes and then air is allowed to enter the chamber for a curing period.

1.3.3 Grinding and Polishing

Grinding refers to flattening the surface to be examined and removing any damage caused by sectioning. The specimen surface to be examined is abraded using a graded sequence of abrasives, starting with a coarse grit. Commonly, abrasives (such as silicon carbide) are bonded to abrasive paper. Abrasive paper is graded according to particle size of abrasives such as 120-, 240-, 320-, 400- and 600-grit paper. The starting grit size depends on the surface roughness and depth of damage from sectioning. Usually, starting grade is 240 or 320 grit after sectioning with a diamond saw or EDM. Both *hand grinding* and *machine grinding* are commonly used.

Grinding

We can perform hand grinding with a simple device in which four belts of abrasive paper (240-, 320-, 400- and 600-grit) are mounted in parallel as shown in Figure 1.21. Running water is supplied to cool specimen surfaces during hand grinding. Grinding produces damage that must be minimized by subsequent grinding with finer abrasives. The procedure is illustrated in Figure 1.22. In particular, two things must be followed to ensure optimal results. First, specimens are rinsed with running water to remove surface debris before switching grinding belts; and second, specimens are rotated 90° from the previous orientation. Rotation ensures that grinding damage generated by a coarse grit is completely removed by a subsequent finer grit. Thus, at the end of any grinding step, the only grinding damage present must be from that grinding step. Damage from the final grinding step is removed by polishing.

Automatic grinding machines have become very popular because they reduce tedious work and are able to grind multiple specimens simultaneously. Also, machine grinding produces more consistent results. A disc of abrasive paper is held on the surface of a motor-driven wheel. Specimens are fixed on a holder that also rotates during grinding. Modern grinding machines control the speeds of the grinding wheel and the specimen holder independently. The direction of rotation of the holder and the compressive force between specimen and grinding wheel



Figure 1.21 Hand grinding using a simple hand grinding device. (Reproduced with permission of Buehler Ltd.)

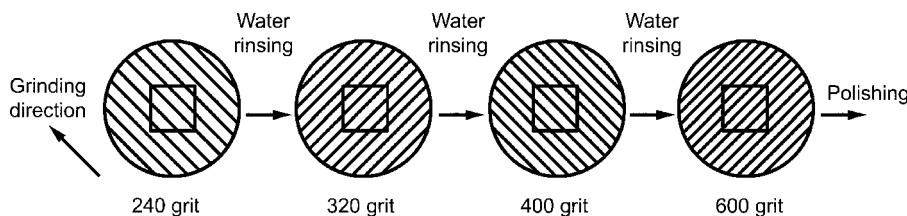


Figure 1.22 Hand grinding procedure.

can also be altered. The machines usually use running water as the cooling medium during grinding to avoid friction heat and to remove loose abrasives that are produced continuously during grinding.

Polishing

Polishing is the last step in producing a flat, scratch-free surface. After being ground to a 600-grit finish, the specimen should be further polished to remove all visible scratches from grinding. Effects of grinding and polishing a specimen surface are shown in Figure 1.23. Polishing generates a mirror-like finish on the specimen surface to be examined. Polishing is commonly conducted by placing the specimen surface against a rotating wheel either by hand or by a motor-driven specimen holder (Figure 1.24). Abrasives for polishing are usually diamond paste, alumina or other metal oxide slurries. Polishing includes coarse and fine polishing. Coarse polishing uses abrasives with a grit size in the range from 3 to 30 μm ; 6- μm diamond paste is the most popular. The abrasive size for fine polishing is usually less than 1 μm . Alumina slurries provide a wide range of abrasive size, ranging down to 0.05 μm .

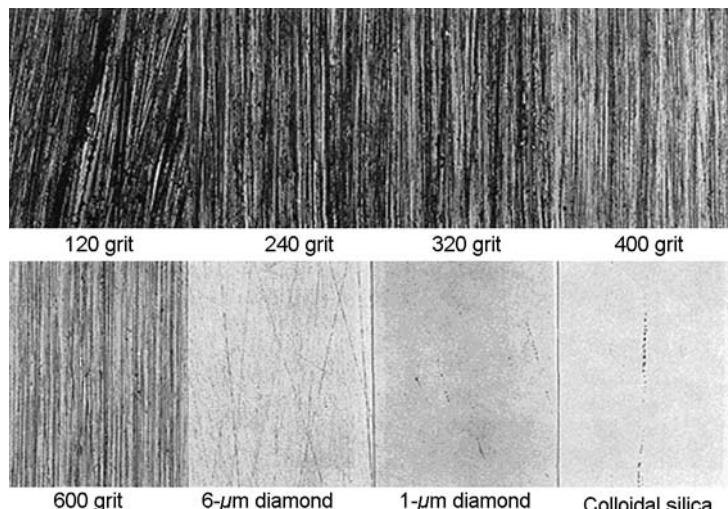


Figure 1.23 Sample of specimen surfaces after grinding and polishing with abrasives of different grits and size. (Reproduced with permission of ASM International®. All Rights Reserved. www.asminternational.org. G.F. Vander Voort, *Metallography Principles and Practice*, McGraw-Hill, New York. © 1984 ASM International®.)



Figure 1.24 Polishing on a rotating wheel with a mechanical sample holder. (Reproduced with permission of Struers A/S.)

A *polishing cloth* covers a polishing wheel to hold the abrasives against the specimen during polishing. Polishing cloths must not contain any foreign matter that may scratch specimen surfaces. Polishing cloths should also be able to retain abrasives so that abrasives are not easily thrown out from the wheel. For coarse polishing, canvas, nylon and silk are commonly used as they have little or no nap. For fine polishing, medium- or high-nap cloths are often used; one popular type consists of densely packed, vertical synthetic fibers.

When hand polishing using a polishing wheel, we should not push the specimen too hard against the wheel as excessive force will generate plastic deformation in the top layer of a polished surface. We should also rotate the specimen against the rotation direction of wheel. Without rotation, artifacts of comet tailing will appear on the polished surfaces as shown in Figure 1.25. After each polishing step, the surface should be cleaned in running water with cotton or tissue, followed by alcohol or hot air drying. Alcohol provides fast drying of surfaces without staining.

Electrolytic polishing is an alternative method of polishing metallic materials. A metal specimen serves as the anode in an electrochemical cell containing an appropriate electrolyte. The surface is smoothed and brightened by the anodic reaction in an electrochemical cell when the correct combination of bath temperature, voltage, current density and time are used. The advantage of this method over conventional polishing is that there is no chance of plastic deformation during the polishing surface. Plastic deformation in the surface layer of specimens can be generated by compression and shear forces arising from conventional polishing methods. Plastic deformation from polishing may generate artifacts in microstructures of materials.

The aforementioned methods of specimen preparation, except microtomy, are regarded as an important part of metallography. These methods are also used for non-metallic materials

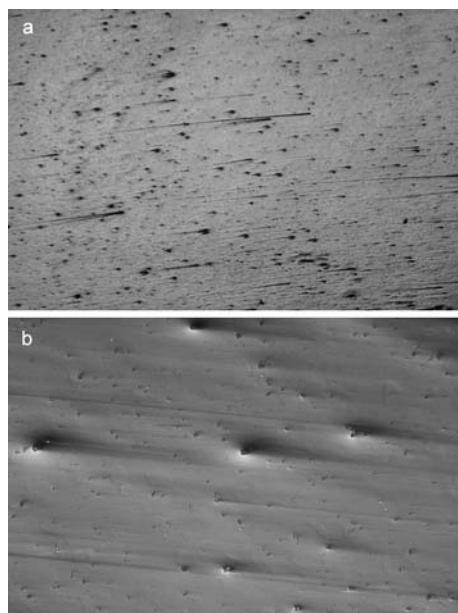


Figure 1.25 Comet tailing generated by polishing on specimen surface: (a) bright-field image; and (b) Nomarski contrast image. (Reproduced with permission of Struers A/S.)

such as ceramics, composites and polymers. However, various precautions must be taken in consideration of each material's particular characteristics. For example, ceramic materials are brittle. To avoid fracture they should be mounted and sectioned with a slow-speed diamond saw. Ceramic materials also should be carefully sectioned with a low-speed diamond saw. Composite materials may exhibit significant differences in mechanical properties between the reinforcement and matrix. These specimens require light pressure and copious cooling during grinding and polishing. Polymeric materials can be examined by either reflected or transmitted light microscopes. For reflected light microscopy, specimen preparation is similar to that of metals. For transmitted light microscopy, a thin-section is required. Both surfaces of the thin-section should be ground and polished. This double-sided grinding and polishing can be done by mounting the specimen in epoxy, preparing one surface, mounting that polished surface on a glass slide, and finally grinding and polishing the other side.

1.3.4 Etching

Chemical etching is a method to generate contrast between microstructural features in specimen surfaces. Etching is a controlled corrosion process by electrolytic action between surface areas with differences in electrochemical potential. Electrolytic activity results from local physical or chemical heterogeneities which render some microstructural features anodic and others cathodic under specific etching conditions. During etching, chemicals (*etchants*) selectively dissolve areas of the specimen surface because of the differences in the electrochemical potential by electrolytic action between surface areas that exhibit differences. For example, grain boundaries in polycrystalline materials are more severely attacked by etchant, and thus are

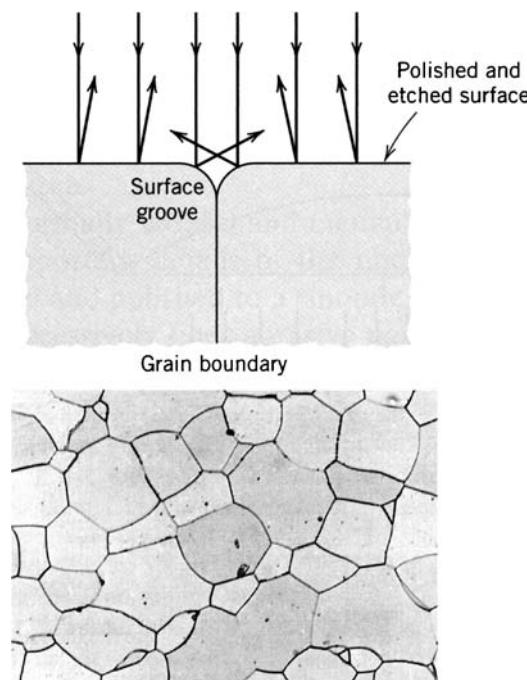


Figure 1.26 Contrast generated by etching grain boundaries in light microscope: (a) reflection from different parts of a surface (Reproduced with permission from W.J. Callister Jr., *Materials Science and Engineering: An Introduction*, 7th ed., John Wiley & Sons Inc., New York. © 2006 John Wiley & Sons Inc.); and (b) micrograph of grain boundaries which appear as dark lines. (Contribution of the National Institute of Standards and Technology.)

revealed by light microscopy because they reflect light differently as illustrated in Figure 1.26a and appear as the dark lines shown in Figure 1.26b. Also, grains are etched at different rates because of differences in grain orientation (certain crystallographic planes are more subject to etching), resulting in crystal faceting. Thus, the grains show different brightness. Etching a specimen that has a multi-phase microstructure will result in selective dissolution of the phases.

Many chemical etchants are mixtures of acids with a solvent such as water. Acids oxidize atoms of a specimen surface and change them to cations. Electrons released from atoms of specimen surfaces are combined with hydrogen to form hydrogen gas. For more noble materials, etchants must contain oxidizers (such as nitric acid, chromic acid, iron chloride and peroxides). Oxidizers release oxygen, which accepts electrons from atoms of the specimen surface. Table 3.1 lists some commonly used etchants, their compositions and applications.

Etching can simply be performed by immersion or swabbing. For immersion etching, the specimen is immersed in a suitable etchant solution for several seconds to several minutes, and then rinsed with running water. The specimen should be gently agitated to eliminate adherent air bubbles during immersion. For swab etching, the polished surface of a specimen is wiped with a soft cotton swab saturated with etchant. Etching can also be assisted with direct electric current, similar to an electrolytic polishing, using the specimen as an anode and insoluble

Table 1.1 Common Etchants for Light Microscopy

Material	Composition ^a	Procedure
Al and alloys	<i>Keller's reagent</i> 2.5 ml HNO ₃ , 1.5 ml HCl 1.0 ml HF, 95 ml water	Immerse 10–20 s
Fe and steels	<i>Nital</i>	Immerse few seconds to 1 minute
Fe and steels	1–10 ml HNO ₃ in 90–99 ml methanol <i>Picral</i>	Immerse few seconds to 1 minute
Stainless steels	4–10 g picric acid, 100 ml ethanol <i>Vilella's Reagent</i> 1 g picric acid, 5 ml HCl, 100 ml ethanal	Immerse for up to 1 minute
Cu and alloys	2 g K ₂ Cr ₂ O ₇ , 8 ml H ₂ SO ₄ , 4 drops HCl, 100 ml water	Add the HCl before using; immerse 3 to 60 seconds
Ti and alloys	10 ml HF, 5 ml HNO ₃ , 85 ml water	Swab 3–20 seconds
Mg and alloys	1 ml HNO ₃ , 75 ml ethylene glycol, 25 ml water	Swab 3 to 120 seconds
Zn and alloys	<i>Palmerton's Reagent</i>	Immerse up to 3 minutes; rinse in 20% aq. CrO ₃
Co and alloys	40 g CrO ₃ , 3 g Na ₂ SO ₄ , 200 ml water 15 ml HNO ₃ , 15 ml acetic acid, 60 ml HCl, 15 ml water	Age 1 hour before use; immerse for up to 30 seconds
Ni and alloys	50 ml HNO ₃ , 50 ml acetic acid	Immerse or swab 5–30 seconds; use hood, do not store
Al ₂ O ₃	15 ml water, 85 ml H ₃ PO ₄	Boil 1–5 minutes
CaO and MgO	Concentrated HCl	Immerse 3 seconds to few minutes
CeO ₂ , SrTiO ₃ , Al ₂ O ₃ and ZrO–ZrC	20 ml water, 20 ml HNO ₃ , 10 ml HF	Immerse up to 15 minutes
Si ₃ N ₄	Molten KOH	Immerse 1–8 minutes
SiC	10 g NaOH, 10 g K ₃ Fe(CN) ₆ in 100 ml water at 110 °C	Boil to dryness
Polyethylene (PE)	Xylene	Immerse 1 minute at 70 °C
poly(acrylonitrile butadiene styrene) (ABS); high-impact polystyrene (HIPS); and poly(phenylene oxide) (PPO)	400 ml H ₂ SO ₄ , 130 ml H ₃ PO ₄ , 125 ml water, 20 g CrO ₃	Immerse 15–180 seconds
Polypropylene (PP)	6M CrO ₃	Immerse 96 hours at 70 °C
Phenol formaldehyde	75 ml dimethyl sulfoxide, 25 ml HNO ₃	Immerse 4 hours at 75–80 °C

^a The names of reagents are given in italics.

material (such as platinum) as the cathode in an electrochemical cell filled with electrolyte. The electrochemical reaction on the anode produces selective etching on the specimen surface. Since electrochemical etching is a chemical reaction, besides choosing a suitable etchant and electrolyte, temperature and time are the key parameters to avoiding under-etching and over-etching of specimens.

We may also use the method of *tint etching* to produce color contrast in microstructure. Tint etchants, usually acidic, are able to deposit a thin (40–500 nm) film such as an oxide or sulfide on specimen surfaces. Tint etching require a very high-quality polished surface for best results. Tint etching can also be done by *heat tinting*, a process by which a specimen is heated to a relatively low temperature in air. As it warms, the polished surface is oxidized. The oxidation rate varies with the phase and chemical composition of the specimen. Thus, differences in the thickness of oxidation films on surfaces generate variations in color. Interference colors are obtained once the film reaches a certain thickness. Effectiveness of heat tinting depends on the material of specimens: it is effective for alloy steels and other non-ferrous metals and carbides, but not for carbon or low alloy steels.

1.4 Imaging Modes

The differences in properties of the light waves reflected from microscopic objects enable us to observe these objects by light microscopy. The light wave changes in either amplitude or phase when it interacts with an object as illustrated in Figure 1.27. The eye can only appreciate amplitude and wavelength differences in light waves, not their phase difference. The most commonly used examination modes, *bright-field* and *dark-field imaging*, are based on contrast due to differences in wave amplitudes. The wave phase differences have to be converted to amplitude differences through special optical arrangements such as in the examination modes of *phase contrast*, *polarized light* and *Nomarski contrast*. This section introduces commonly used modes of light microscopy for materials characterization.

1.4.1 Bright-Field and Dark-Field Imaging

Bright-field imaging is the predominant mode for examining microstructure. Dark-field imaging is also widely used to obtain an image with higher contrast than in bright-field. Figure 1.28

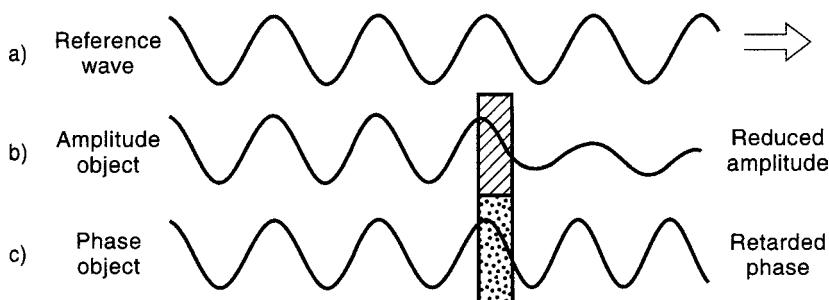


Figure 1.27 (a) Reference wave; (b) amplitude difference; and (c) phase difference generated by objects. (Reproduced with permission from D.B. Murphy, *Fundamentals of Light Microscopy and Electronic Imaging*, Wiley-Liss. © 2001 John Wiley & Sons Inc.)

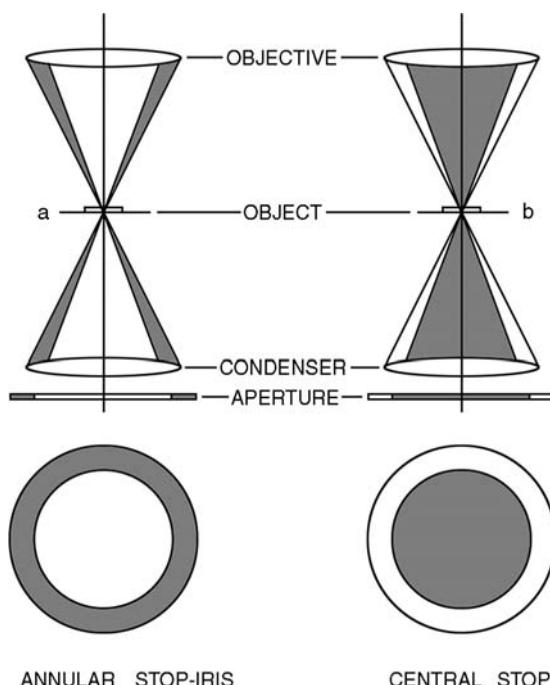


Figure 1.28 (a) Bright-field illumination; and (b) dark-field illumination in transmitted mode. Shaded areas indicate where the light is blocked.

illustrates the difference in optical arrangement between these modes in transmitted illumination. In bright-field imaging, the specimen is evenly illuminated by a light source. Dark-field imaging requires that the specimen is illuminated by oblique light rays. There is a central stop in the light path to block the central portion of light rays from illuminating the specimen directly. Thus, the angle of the light rays illuminating the specimen is so large that light from the specimen cannot enter the objective lens unless it is scattered by microscopic objects. The dark field in reflected illumination is also realized using a central stop in the light path (Figure 1.29), similar to that of transmitted illumination. The light rays in a ring shape will be further reflected in order to illuminate a specimen surface with an oblique angle. Figure 1.30 shows the comparison between bright- and dark-field images of an identical field in a high-carbon steel specimen under a reflected light microscope. Microscopic features such as grain boundaries and second phase particles appear self-luminous in the dark field image as shown in Figure 1.30.

1.4.2 Phase Contrast Microscopy

Phase contrast is a useful technique for specimens such as polymers that have little inherent contrast in the bright-field mode. In the technique, a phase change due to light diffraction by an object is converted to an amplitude change. This conversion is based on interference

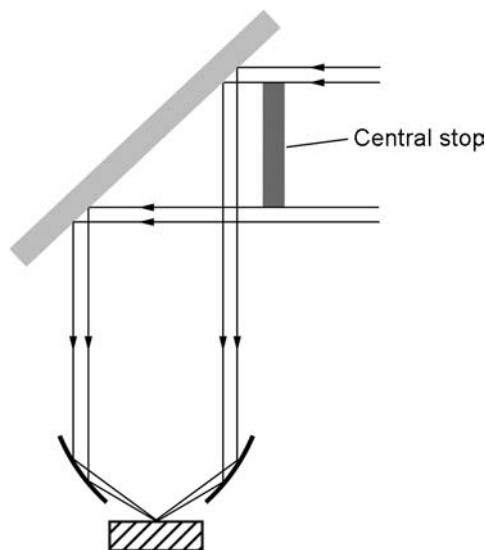


Figure 1.29 Dark-field illumination in a reflected light microscope.

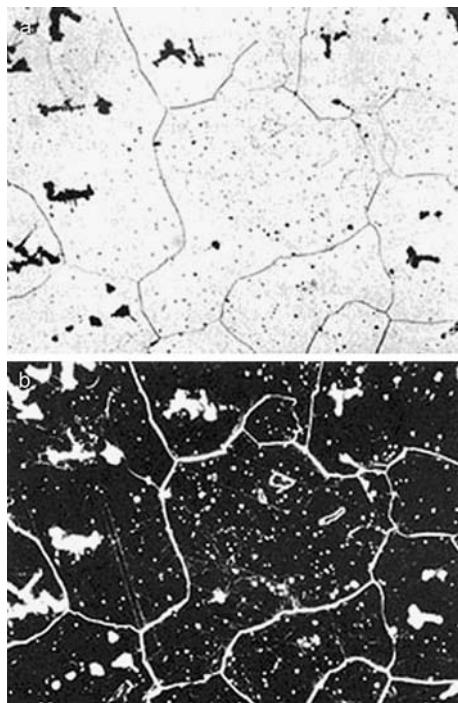


Figure 1.30 Comparison between: (a) bright-field; and (b) dark-field images of AISI 1080 high carbon steel. In addition to grain boundaries and oxide particles, annealing twins are revealed in the dark-field image. (Reproduced with permission of ASM International®. All Rights Reserved. www.asminternational.org. G.F. Vander Voort, *Metallography Principles and Practice*, McGraw-Hill, New York. © 1984 ASM International®.)

phenomenon of light waves as illustrated in Figure 1.31. Constructive interference occurs when combining two same-wavelength waves that do not have a phase difference between them. However, completely destructive interference occurs when combining two waves with phase difference of a half wavelength ($\frac{\lambda}{2}$).

Figure 1.32 illustrates generation of phase contrast by a special optical arrangement in a transmitted light microscope. This optical arrangement creates completely destructive interference when light is diffracted by an object in the specimen. A *condenser annulus*, an opaque black plate with a transparent ring, is placed in the front focal plane of the condenser lens. Thus, the specimen is illuminated by light beams emanating from a ring. The light beam that passes through a specimen without diffraction by an object (the straight-through light beam) will pass the ring of a *phase plate* placed at the back focal plane of the objective lens. The phase plate is a plate of glass with an etched ring of reduced thickness. The ring with

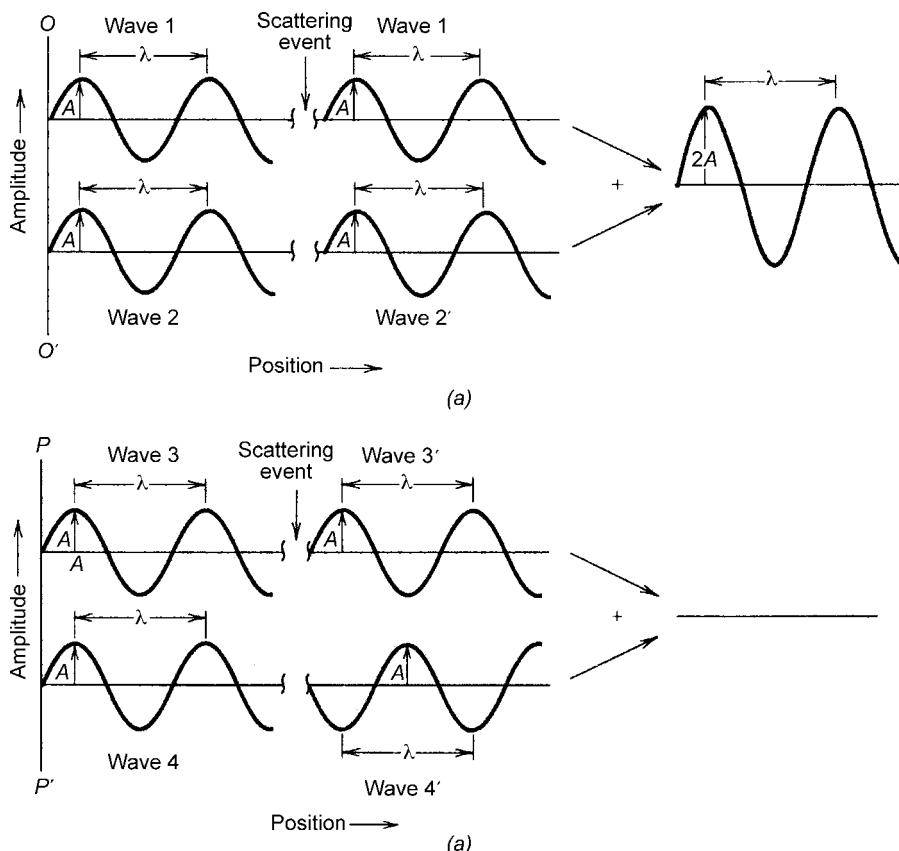


Figure 1.31 Illustration of interference between waves: (a) constructive interference; and (b) completely destructive interference. (Reproduced with permission from W.J. Callister Jr., *Materials Science and Engineering: An Introduction*, 7th ed., John Wiley & Sons Inc., New York. © 2006 John Wiley & Sons Inc.)

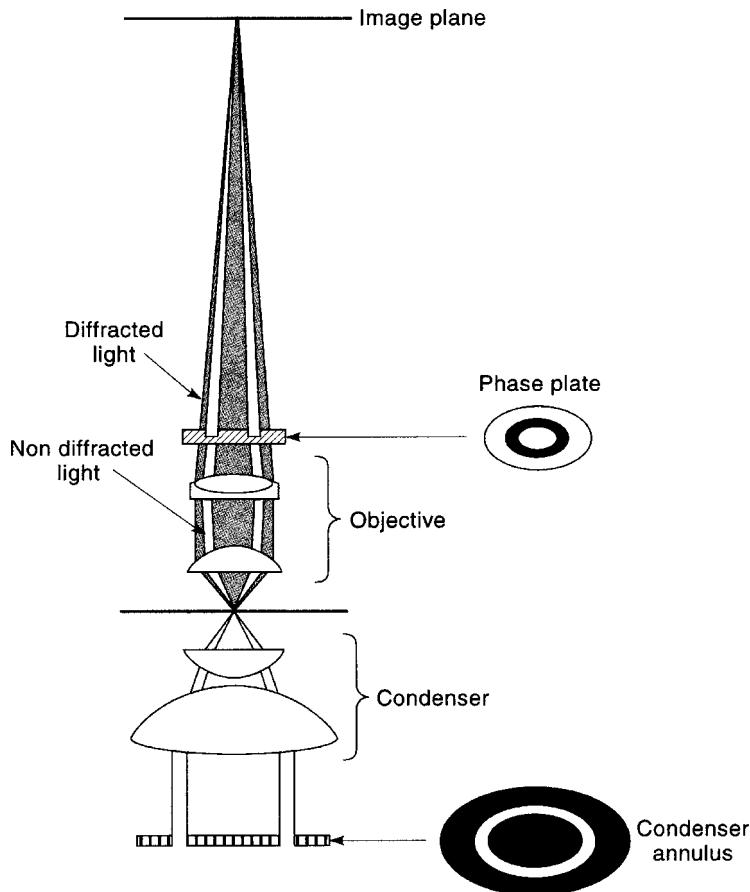


Figure 1.32 Optical arrangement of phase contrast microscopy. Shading marks the paths of diffracted light. (Reproduced with permission from D.B. Murphy, *Fundamentals of Light Microscopy and Electronic Imaging*, Wiley-Liss. © 2001 John Wiley & Sons Inc.)

reduced thickness in the phase plate enables the waves of the straight-through beam to be advanced by $\frac{\lambda}{4}$. The light beam diffracted by the object in the specimen cannot pass through the ring of the phase plate but only through the other areas of the phase plate. If the diffracted beam is delayed by $\frac{\lambda}{4}$ while passing through the object, a total ($\frac{\lambda}{2}$) difference in phase is generated.

When the straight-through beam and diffracted beam recombine at the image plane, completely destructive interference occurs. Thus, we expect a dark image of the object in phase contrast microscopy. Variation in phase retardation across the specimen produces variations in contrast. Figure 1.33 shows image differences between bright field and phase contrast images of composites in transmission light microscopy. In reflected light microscopy, phase contrast can also be created with a condenser annulus and phase plate similar to those in transmitted light microscopy.

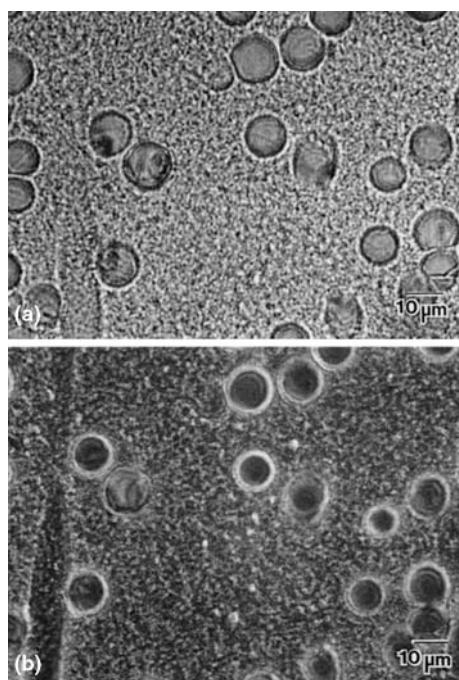


Figure 1.33 Comparison of light transmission images of glass fiber-reinforced polyamide in: (a) bright-field; and (b) phase contrast modes. (Reproduced with kind permission of Springer Science and Business Media from L.C. Swayer, and D.T. Grubb, *Polymer Microscopy*, Chapman & Hall, London. © 1996 Springer Science.)

1.4.3 Polarized Light Microscopy

Polarized light is used to examine specimens exhibiting optical anisotropy. Optical anisotropy arises when materials transmit or reflect light with different velocities in different directions. Most materials exhibiting optical anisotropy have a non-cubic crystal structure. Light, as an electromagnetic wave, vibrates in all directions perpendicular to the direction of propagation. If light waves pass through a polarizing filter, called a *polarizer*, the transmitted wave will vibrate in a single plane as illustrated in Figure 1.34. Such light is referred to as *plane polarized light*. When polarized light is transmitted or reflected by anisotropic material, the polarized light vibrates in a different plane from the incident plane. Such polarization changes generate the contrast associated with anisotropic materials.

Figure 1.35 illustrates the interaction between polarized light and an anisotropic object. For a transparent crystal, the optical anisotropy is called double refraction, or *birefringence*, because refractive indices are different in two perpendicular directions of the crystal. When a polarized light ray hits a birefringent crystal, the light ray is split into two polarized light waves (*ordinary wave* and *extraordinary wave*) vibrating in two planes perpendicular to each other. Because there are two refractive indices, the two split light rays travel at different velocities, and thus exhibit phase difference. These two rays, with a phase difference vibrating in two perpendicular planes, produce polarized light because the electric vectors of their waves are

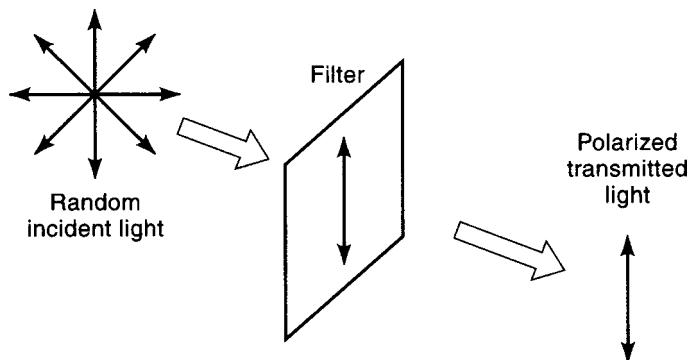


Figure 1.34 Formation of plane polarized light by a polarizing filter. (Reproduced with permission from D.B. Murphy, *Fundamentals of Light Microscopy and Electronic Imaging*, Wiley-Liss. © 2001 John Wiley & Sons Inc.)

added (Figure 1.36). The resultant polarized light is called *elliptically polarized light* because the projection of resultant vectors in a plane is elliptical.

If the two polarized-light waves have a phase difference of $\frac{\lambda}{4}$, the projection of resultant light is a spiraling circle. If the two polarized-light waves have a phase difference of $(\frac{\lambda}{2})$, the

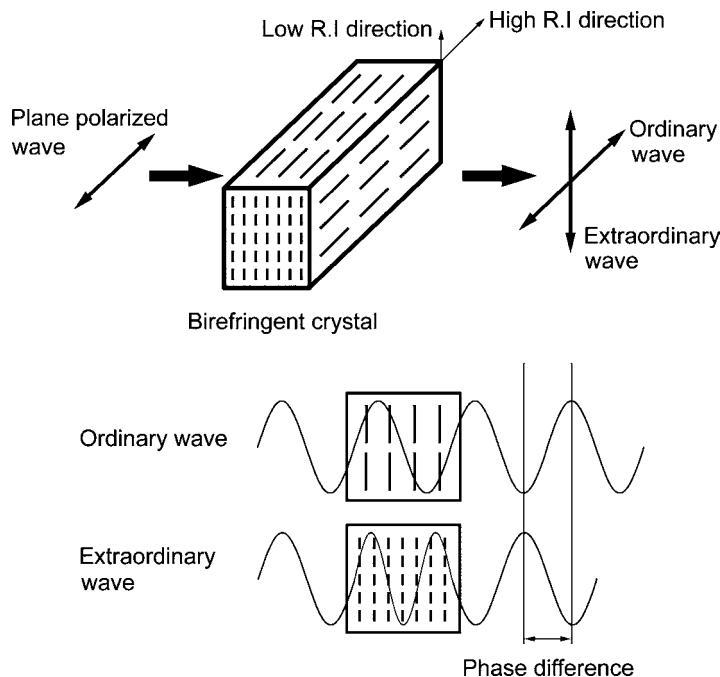


Figure 1.35 Interaction between plane-polarized light and a birefringent object. Two plane-polarized light waves with a phase difference are generated by the materials. RI, refractive index.

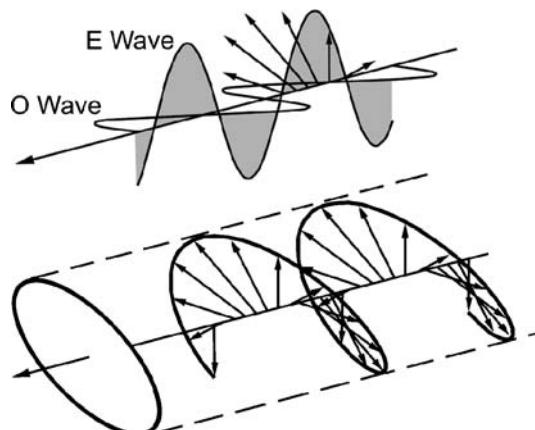


Figure 1.36 Resultant polarized light by vector addition of two plane-polarized light waves from a birefringent object, the ordinary (O) wave and extraordinary (E) wave. (Reproduced with permission from D.B. Murphy, *Fundamentals of Light Microscopy and Electronic Imaging*, Wiley-Liss. © 2001 John Wiley & Sons Inc.)

projection of resultant light is linear (45° from two perpendicular directions of polarized-light wave planes). If the two polarized-light waves have another phase difference, the projection of resultant ray is a spiraling ellipse.

Differences in the resultant light can be detected by another polarizing filter called an *analyzer*. Both polarizer and analyzer can only allow plane polarized light to be transmitted. The analyzer has a different orientation of polarization plane with respect to that of the polarizer. Figure 1.37 illustrates the amplitude of elliptically polarized light in a two-dimensional plane and the light amplitude passing through an analyzer which is placed 90° with respect to the polarizer (called the *crossed position* of the polarizer and analyzer).

Anisotropic materials are readily identified by exposure to polarized light because their images can be observed with the polarizer and analyzer in the crossed position. Such situations are illustrated in Figure 1.37 as the phase difference between ordinary and extraordinary wave are not equal to zero and equals $m\lambda$ (m is an integer). Understandably, when examining anisotropic materials with polarized light, rotation of the analyzer through 360° will generate two positions of maximum and two positions of minimum light intensity.

Polarized light can enhance the contrast of anisotropic materials, particularly when they are difficult to etch. It can also determine optical axis, demonstrate *pleochroism* (showing different colors in different directions) and examine the thickness of anisotropic coatings from its cross-sections. Figure 1.38 demonstrates that polarized light reveals the grains of pure titanium, which cannot be seen in the bright-field mode. Figure 1.39 shows a polarized light micrograph of high density polyethylene that has been crystallized. Polarized light revealed fine spherulite structures caused by crystallization.

An isotropic material (either with cubic crystal structure or amorphous) cannot change the plane orientation of polarizing light. When a polarized light wave leaves the material and passes through the analyzer in a crossed position, the light will be extinguished. This situation is equivalent to the case of a resultant ray having a phase difference of zero or $m\lambda$. However,

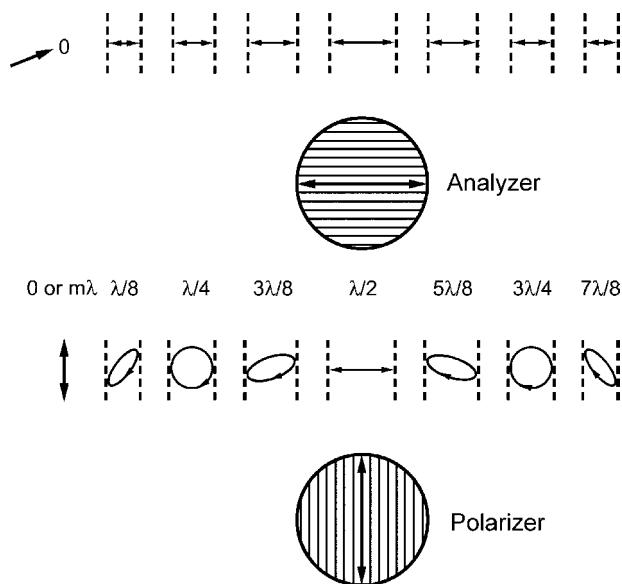


Figure 1.37 Intensity change of polarized light passing through an analyzer when the elliptically polarized light changes. The polarizer and analyzer are in a crossed position. (Reproduced with permission from D.B. Murphy, *Fundamentals of Light Microscopy and Electronic Imaging*, Wiley-Liss. © 2001 John Wiley & Sons Inc.)

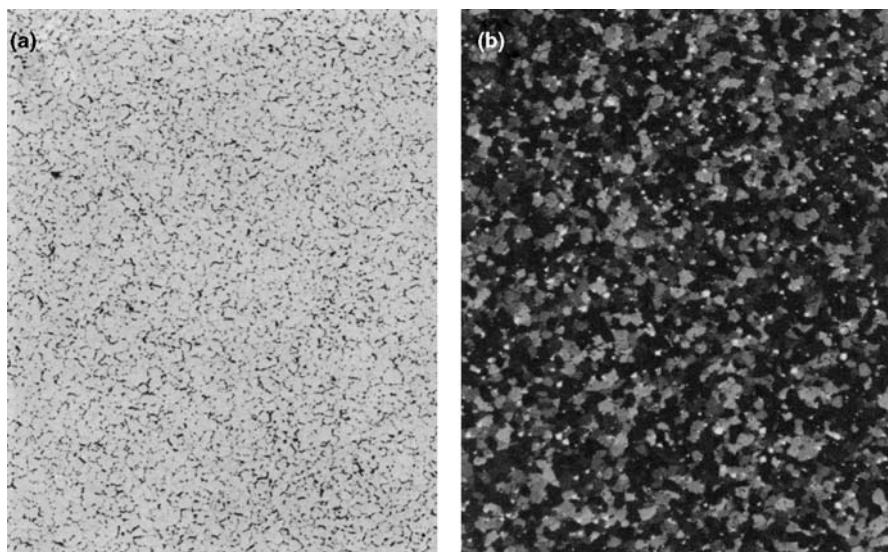


Figure 1.38 Example of using polarized-light microscopy for pure titanium specimen: (a) bright-field image; and (b) polarized light image in which grains are revealed. (Reproduced from G.F. Vander Voort, *Metallography Principles and Practice*, McGraw-Hill, New York. © N. Gendron, General Electric Co.)

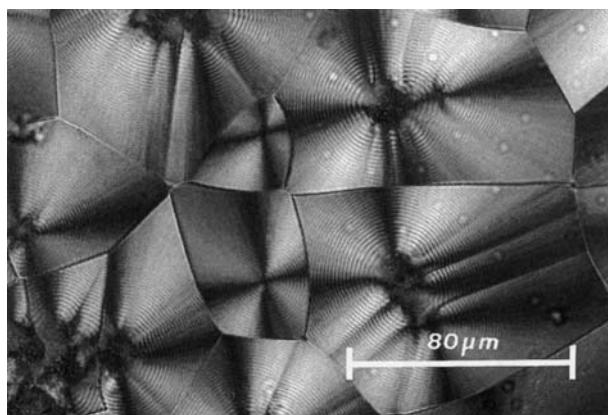


Figure 1.39 Polarized light micrograph of crystallized high density polyethylene (HDPE). (Reproduced with permission from J. Scheirs, *Compositional and Failure Analysis of Polymers: A Practical Approach*, John Wiley & Sons Ltd, Chichester. © 2000 John Wiley & Sons Ltd.)

isotropic materials can be examined in the polarized light mode when optical anisotropy is introduced into the materials or on their surfaces. For example, if an isotropic transparent crystal is elastically deformed, it becomes optically anisotropic. A thick oxide film on isotropic metals also makes them sensitive to the direction of polarized light because of double reflection from surface irregularities in the film.

1.4.4 Nomarski Microscopy

Nomarski microscopy is an examination mode using *differential interference contrast* (DIC). The images that DIC produces are deceptively three-dimensional with apparent shadows and a relief-like appearance. Nomarski microscopy also uses polarized light with the polarizer and the analyzer arranged as in the polarized light mode. In addition, double quartz prisms (*Wollaston prisms* or *DIC prisms*) are used to split polarized light and generate a phase difference.

The working principles of Nomarski microscopy can be illustrated using the light path in a transmitted light microscope as illustrated in Figure 1.40. The first DIC prism is placed behind the polarizer and in front of the condenser lens, and the second DIC prism is placed behind the objective lens and in front of the analyzer. The two beams created by the prism interfere coherently in the image plane and produce two slightly displaced images differing in phase, thus producing height contrast. The first DIC prism splits polarized light beam from the polarizer into two parallel beams traveling along different physical paths. If a specimen does not generate a path difference between the two parallel beams as shown by two left-side beam pairs in Figure 1.40, the second DIC prism recombines the pairs and produces linearly polarized light with the same polarization plane as it was before it was split by the first DIC prism. Thus, the analyzer in the crossed position with a polarizer will block light transmission. However, if a specimen generates a path difference in a pair of beams, as shown

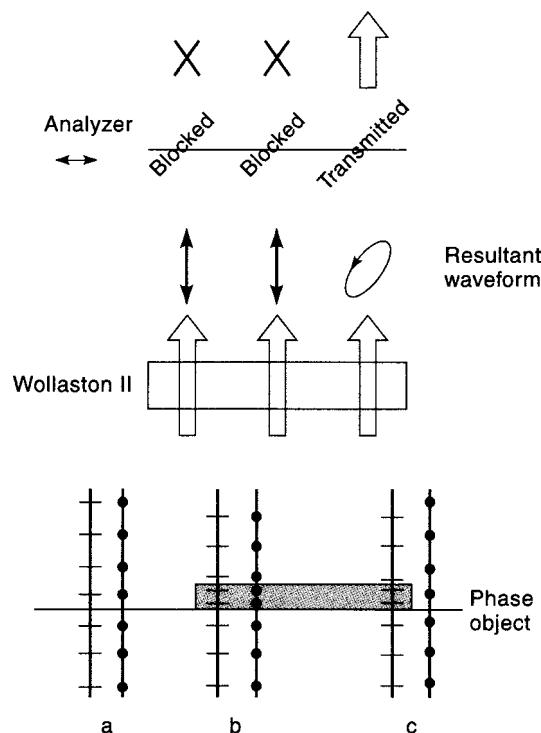


Figure 1.40 Nomarski contrast generation using polarized light. The first differential interference contrast (DIC) prism (not shown) generates two parallel polarized beams illuminating the specimen. The second DIC prism recombines two beams. Elliptically polarized light is generated by the second DIC prism when phase difference between the two beams is induced by an object: (a) both polarized beams do not pass through a phase object; (b) both beams pass through a phase object; and (c) one of the two beams passes through a phase object. (Reproduced with permission from D.B. Murphy, *Fundamentals of Light Microscopy and Electronic Imaging*, Wiley-Liss. © 2001 John Wiley & Sons Inc.)

by the right-side beam pair, the recombined pair produced by the second DIC prism will be elliptically polarized light. The analyzer cannot block such light and a bright area will be visible.

The optical arrangement of Nomarski microscopy in a reflected light microscope is similar to that of a transmitted light microscope, except that there is only one DIC prism serving both functions of splitting the incident beam and recombining reflected beams as illustrated in Figure 1.41. The DIC prism is commonly integrated in the barrel of the objective lens, indicated by 'DIC' marked on the barrel surface (Section 1.2.2).

Figure 1.42 compares bright-field and Nomarski images using a carbon steel specimen as an example. Contrast enhancements are achieved in the Nomarski micrographs. The Nomarski image appears three-dimensional and illuminated by a low-angle light source. However, the image does not necessarily represent real topographic features of a surface because the shadows and highlights that result from phase differences may not correspond to low and high relief on

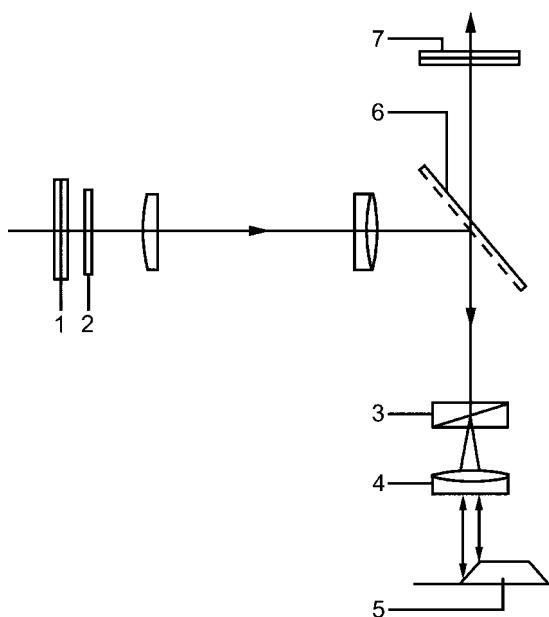


Figure 1.41 Optical arrangement of Nomarski microscopy in reflected light illumination. 1, polarizer; 2, $\frac{\lambda}{4}$ -plate; 3, DIC prism; 4, objective lens; 5, specimen; 6, light reflector and 7, analyzer.

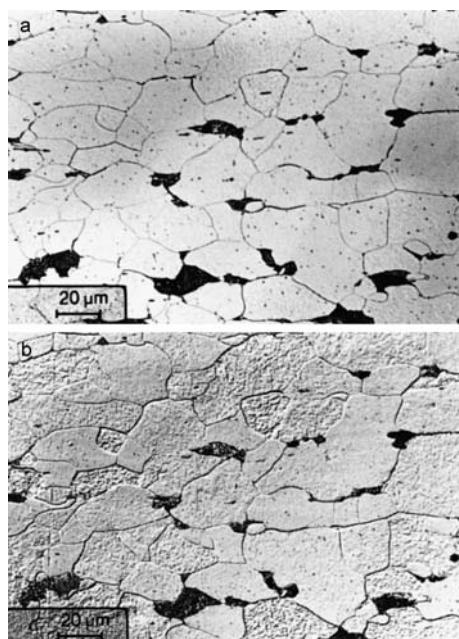


Figure 1.42 Effects of Nomarski contrast on carbon steel micrographs: (a) bright field; and (b) Nomarski contrast of the same field. (Reproduced with permission of Gonde Kiessler.)

the surface, particularly in transmitted light microscopy. The reason is that the phase differences generated in Nomarski microscopy may result from differences either in the optical path or in refractive index.

1.4.5 Fluorescence Microscopy

Fluorescence microscopy is useful for examining objects that emit fluorescent light. Fluorescence is an optical phenomenon; it occurs when an object emits light of a given wavelength when excited by incident light. The incident light must have sufficient energy, that is, a shorter wavelength than that light emitting from the object, to excite fluorescence. While only a small number of materials exhibit this capability, certain types of materials can be stained with fluorescent dyes (*fluorochromes*). The fluorochromes can selectively dye certain constituents in materials, called *fluorescent labeling*. Fluorescent labeling is widely used for polymeric and biological samples.

Fluorescence microscopy can be performed by either transmitted or reflected illumination (epi-illumination). Reflected light is more commonly used because it entails less loss of excited fluorescence than transmitted light. Figure 1.43 illustrates the optical arrangement for

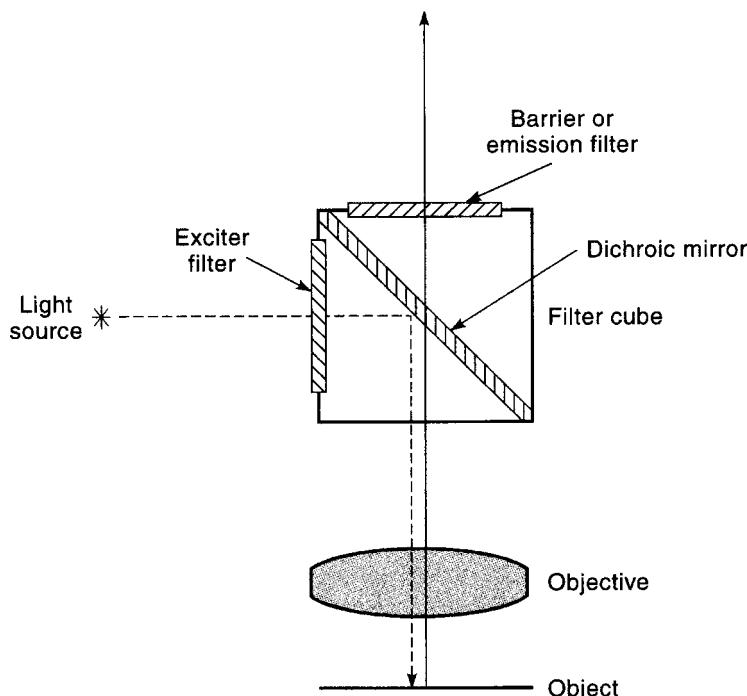


Figure 1.43 Optical arrangement for fluorescence microscopy with epi-illumination. The dotted line indicates the path of excitation light, and the solid line indicates the path of fluorescent light. (Reproduced with permission from D.B. Murphy, *Fundamentals of Light Microscopy and Electronic Imaging*, Wiley-Liss. © 2001 John Wiley & Sons Inc.)

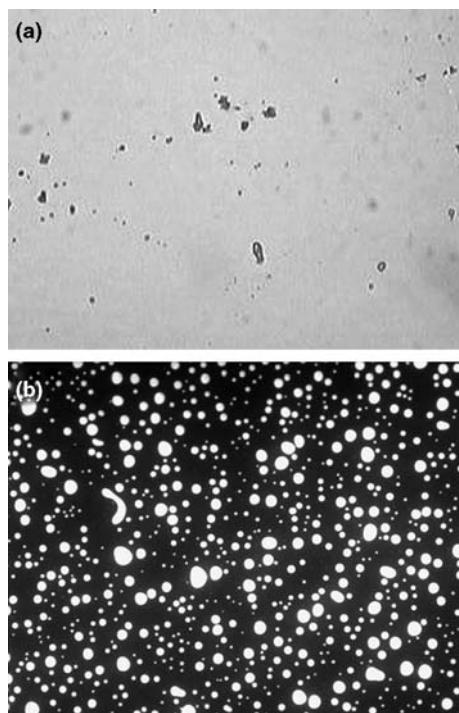


Figure 1.44 Micrographs of asphalt–polyolefin elastomer (POE) blend obtained with transmitted light microscopy: (a) bright-field image that cannot reveal the two phases in the blend; and (b) a fluorescence-labeled image that reveals two-phase morphology. (Reproduced with permission of Jingshen Wu.)

fluorescence microscopy with epi-illumination. A high pressure mercury or xenon light can be used for generating high intensity, short wavelength light. The light source should be ultraviolet, violet or blue, depending on the types of fluorochromes used in the specimen. A fluorescence filter set, arranged in a cube as shown in Figure 1.43, includes an *exciter filter*, *dichroic mirror* and a *barrier filter*. The exciter filter selectively transmits a band of short wavelengths for exciting a specific fluorochrome, while blocking other wavelengths. The dichroic mirror reflects short wavelength light to objective lens and specimen, and also transmits returning fluorescent light toward the barrier filter. The barrier filter transmits excited fluorescent light only, by blocking other short wavelength light. Figure 1.44 shows an example of fluorescence microscopy used for examining a polymer blend. Fluorescence labeling reveals the dispersed polymer particles in an asphalt matrix, which cannot be seen in the bright-field image.

1.5 Confocal Microscopy

Confocal microscopy is a related new technique that provides three-dimensional (3D) optical resolution. Image formation in a confocal microscope is significantly different from a conventional light microscope. Compared with a conventional compound microscope, a modern confocal microscope has two distinctive features in its structure: a laser light source and a

scanning device. Thus, the confocal microscope is often referred to as the *confocal laser scanning microscope* (CLSM). The laser light provides a high-intensity beam to generate image signals from individual microscopic spots in the specimen. The scanning device moves the beam in a rectangular area of specimen to construct a 3D image on a computer.

1.5.1 Working Principles

The optical principles of confocal microscopy can be understood by examining the CLSM optical path that has reflected illumination as illustrated in Figure 1.45. The laser beam is focused as an intense spot on a certain focal plane of the specimen by a condenser lens, which is also serves as an objective lens to collect the reflected beam. A pinhole aperture is placed at a confocal plane in front of the light detector. The reflected beam from the focal plane in a specimen becomes a focused point at the confocal plane. The pinhole aperture blocks the reflected light from the out-of-focal plane from entering the detector. Only the light signals

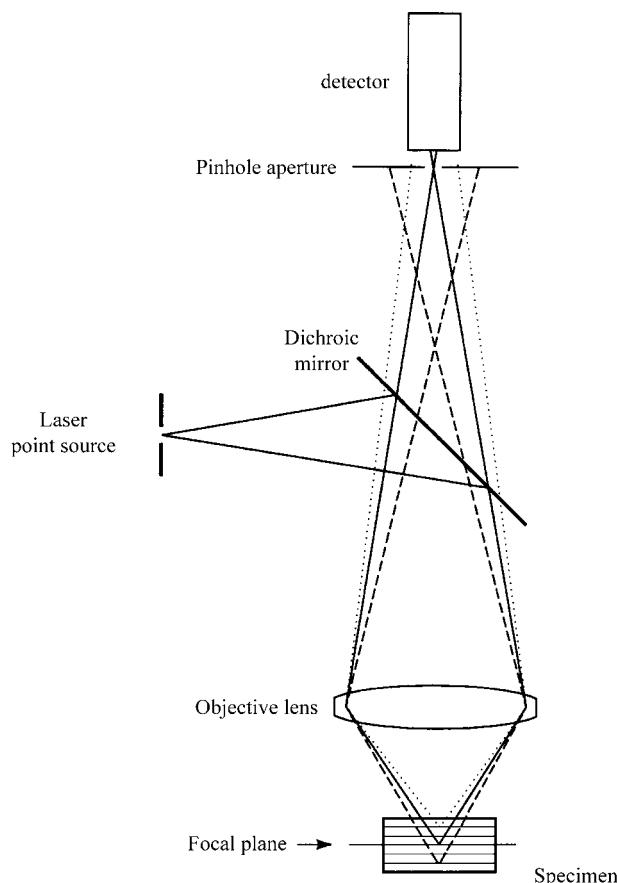


Figure 1.45 Optical path in the confocal microscope. (Reproduced with permission from D.B. Murphy, *Fundamentals of Light Microscopy and Electronic Imaging*, Wiley-Liss. © 2001 John Wiley & Sons Inc.)

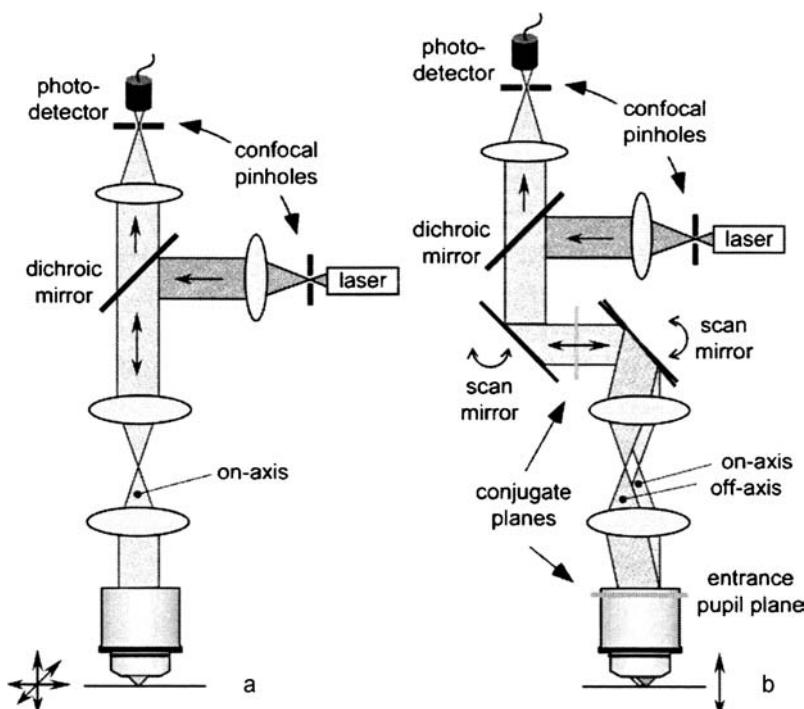


Figure 1.46 Two scanning methods in the confocal microscope: (a) specimen scanning; and (b) laser scanning. (Reproduced with permission from M. Müller, *Introduction to Confocal Fluorescence Microscopy*, 2nd ed., SPIE Press, Bellingham, Washington. © 2006 Michiel Müller)

from the focal point in the specimen are recorded each time. Since the pinhole aperture can block a large amount of reflected light, high-intensity laser illumination is necessary to ensure that sufficient signals are received by the detector. The detector is commonly a photomultiplier tube (PMT) that converts light signals to electric signals for image processing in a computer.

To acquire an image of the focal plane, the plane has to be scanned in its two lateral directions ($x-y$ directions). To acquire a 3D image of a specimen, the plane images at different vertical positions should also be recorded. A scanning device moves the focal laser spot in the $x-y$ directions on the plane in a regular pattern called a *raster*. After finishing one scanning plane, the focal spot is moved in the vertical direction to scan a next parallel plane.

Figure 1.46 illustrates two different methods of scanning in CLSM: specimen and laser scanning. Specimen scanning was used in early confocal microscopes. The specimen moves with respect to the focal spot and the optical arrangement is kept stationary as shown in Figure 1.46a. The beam is always located at the optical axis in the microscope so that optical aberration is minimized. The main drawback of this method is the low scanning speed. Laser scanning is realized by two scanning mirrors rotating along mutually perpendicular axes as shown in Figure 1.46b. The scan mirror can move the focal spot in the specimen by sensitively changing the reflecting angle of mirror. Changing the vertical position of the spot is still achieved by moving the specimen in the laser scanning method.

The resolution of the confocal microscope is mainly determined by the size of focal spot of the laser beam. High spatial resolution about $0.2\text{ }\mu\text{m}$ can be achieved. A 3D image of specimen

with thickness of up to $200\text{ }\mu\text{m}$ is also achievable, depending on the opacity of specimen. Most confocal microscopes are the fluorescence type. The microscopic features under the specimen surface are effectively revealed when they are labeled with fluorescent dyes. Therefore, the major use of confocal microscopy is confocal fluorescence microscopy in biology.

1.5.2 Three-Dimensional Images

The technique of confocal microscopy can be considered as optical sectioning. A three-dimensional (3D) image is obtained by reconstructing a deck of plane images. Figure 1.47

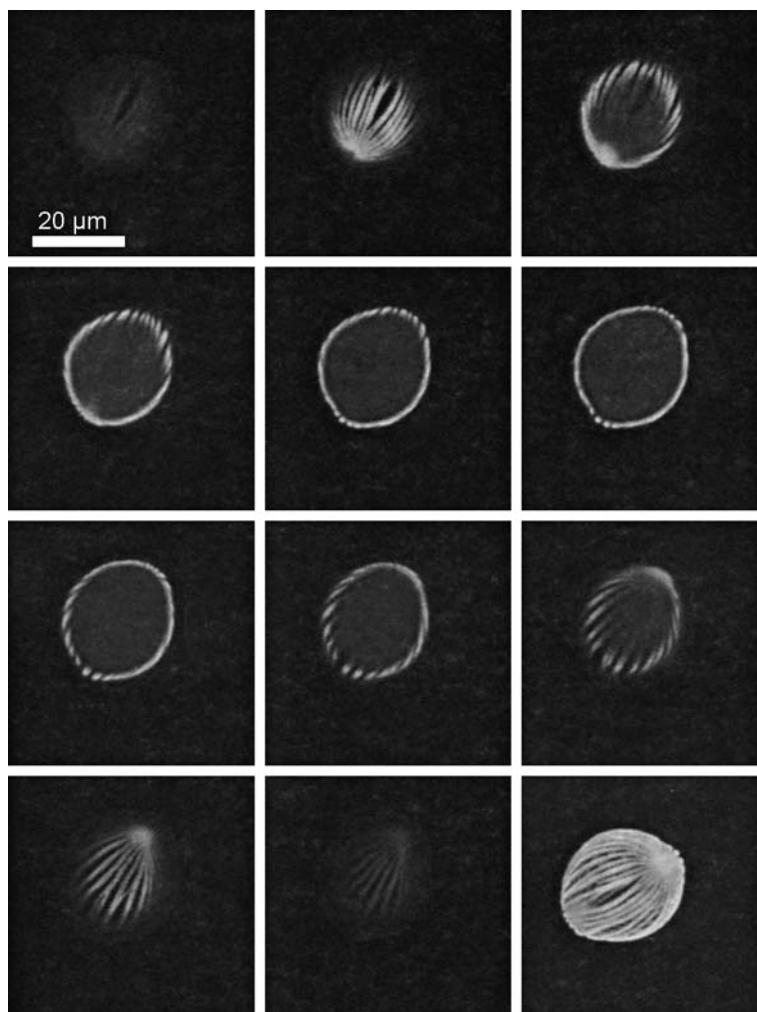


Figure 1.47 Confocal fluorescence micrographs of a *Spathiphyllum* pollen grain. The optical section sequence is from left to right and top to bottom. The bottom right image is obtained by 3D reconstruction. (Reproduced with permission from M. Müller, *Introduction to Confocal Fluorescence Microscopy*, 2nd ed., SPIE Press, Bellingham, Washington. © 2006 Michiel Müller)

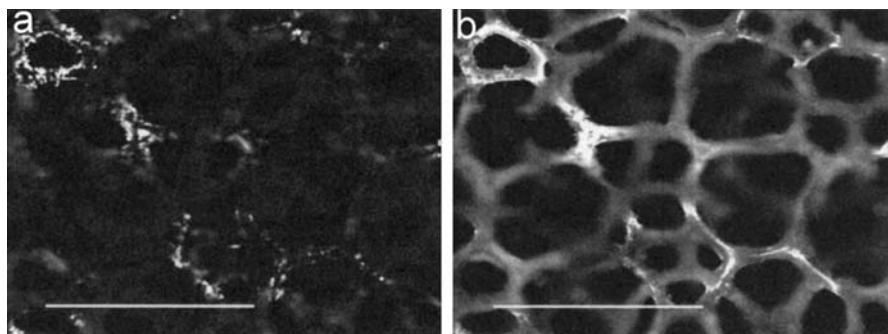


Figure 1.48 Confocal micrographs of polyurethane foam with labeled particulates: (a) highlights of particulates; and (b) locations of particulates on the foam surfaces. The scale bar is 1 mm. (Reproduced with permission from A.R. Clarke and C.N. Eberhardt, *Microscopy Techniques for Materials Science*, Woodhead Publishing Ltd, Cambridge UK. © 2002 Woodhead Publishing Ltd.)

shows an example of how a 3D image is formed with confocal fluorescence microscopy. A biological specimen, a *Spathiphyllum* pollen grain, was fluorescently labeled with acridine orange. In total, 80 sections of the pollen grain were imaged for 3D reconstruction. The vertical distance between sections is $1.1\text{ }\mu\text{m}$. Figure 1.47 shows 11 of 80 plane images and the reconstructed 3D image.

Although its major applications are in biology, confocal microscopy can also be used for examining the surface topography and internal structure of semi-transparent materials. Figure 1.48 shows an example of using confocal fluorescent microscopy to examine particulate penetration through an air filter made of polymer foam. The particulates were fluorescently labeled and their locations in the polymer foam are revealed against the polymer foam surfaces.

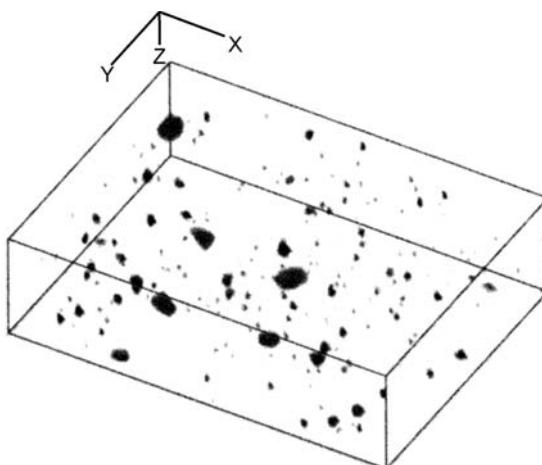


Figure 1.49 3D confocal fluorescent micrograph of silica particles in low density polyethylene. (Reproduced with permission from A.R. Clarke and C.N. Eberhardt, *Microscopy Techniques for Materials Science*, Woodhead Publishing Ltd, Cambridge UK. © 2002 Woodhead Publishing Ltd.)

Figure 1.49 shows an example of using confocal fluorescent microscopy to reveal microscopic features in a specimen. The specimen is low density polyethylene (LDPE) containing fluorescently labeled silica particles. The particle size and distribution in the polymer matrix can be clearly revealed by 3D confocal microscopy. Thus, confocal microscopy provides us a new dimension in light microscopy for materials characterization, even though its applications in materials science are not as broad as in biology.

References

- [1] Bradbury, S.S. and Bracegirdle, B. (1998) *Introduction to Light Microscopy*, BIOS Scientific Publishers.
- [2] Rawlins, D.J. (1992) *Light Microscopy*, BIOS scientific Publishers.
- [3] Murphy, D.B. (2001) *Fundamentals of Light Microscopy and Electronic Imaging*, Wiley-Liss.
- [4] Vander Voort, G.F. (1984) *Metallography Principles and Practice*, McGraw-Hill Book Co., New York.
- [5] Callister Jr., W.J. (2006) *Materials Science and Engineering: An Introduction*, 7th ed., John Wiley & Sons, New York.
- [6] Clarke, A.R. and Eberhardt, C.N. (2002) *Microscopy Techniques for Materials Science*, Woodhead Publishing Ltd, Cambridge UK.
- [7] Hemsley, D.A. (1989) *Applied Polymer Light Microscopy*, Elsevier Applied Science, London.
- [8] Swayer, L.C. and Grubb, D.T. (1996) *Polymer Microscopy*, Chapman & Hall, London.
- [9] American Society for Metals (1986) *Metals Handbook*, Volume 10, 9th edition, American Society for Metals, Metals Park.
- [10] Scheirs, J. (2000) *Compositional and Failure Analysis of Polymers: A Practical Approach*, John Wiley & Sons, Ltd, Chichester.
- [11] Müller, M. (2006) *Introduction to Confocal Fluorescence Microscopy*, 2nd edition, The International Society for Optical Engineering, Bellingham, Washington.

Questions

- 1.1 Three rules govern light path for a simple lens:
 1. A light ray passing through the center of a lens is not deviated;
 2. A Light ray parallel with optic axis will pass through the rear focal point; and
 3. A ray passing through the front focal point will be refracted in a direction parallel to the axis. Sketch the light paths from object to image in a single lens system in following situations.
(a) $a < f$; (b) $a = f$; (c) $2f > a > f$; (d) $a = 2f$; and (e) $a > 2f$, where a is the distance of object from lens and f is the focal length of lens. By this exercise, you may understand that $2f > a > f$ is necessary to obtain a real magnified image.
- 1.2 The working distance between the specimen and objective lens is determined by magnifications of the lens. Estimate the difference in the working distance for objective lens with power of $5\times$, $20\times$ and $50\times$.
- 1.3 Calculate the resolution and the depth of field of the objective lenses of a light microscope, listed below. The refractive index of vacuum is 1, and that of air can be treated as 1. Assume blue light is used in the microscope.

Magnification/NA

$5\times/0.13$

$10\times/0.25$

$20\times/0.40$

$50\times/0.70$

$100\times/0.90$

- 1.4 Compare resolution and depth of field of light microscopes and electron microscopes. The wavelength of electrons is 0.0037 nm (100 kV) and the angle α of an electron microscope is 0.1 radians.
- 1.5 We have samples of an annealed Al alloy and an annealed plain carbon steel to be examined. The polishing area of samples is about 5 mm \times 3 mm.
- To avoid plastic deformation in the surface layer, what are the maximum compression forces that should be used for polishing the samples of annealed Al alloy and plain carbon steel, respectively?
 - If normal compression force cannot cause plastic deformation, what kind of loading on the samples more likely causes plastic deformation?
Yield strength of annealed Al alloy = 150 MPa
Tensile strength of annealed Al alloy = 400 MPa
Yield strength of annealed plain carbon steel = 400 MPa
Tensile strength of annealed plain steel = 700 MPa
- 1.6 To obtain a 400 \times magnification image we may choose a 40 \times objective lens with a 10 \times projector lens, or 20 \times objective lens with a 20 \times projector lens. What are the differences in their image quality?
- 1.7 Describe a specimen preparation procedure for examining:
- a cross-section of a metal needle;
 - a coating layer on metal substrate;
 - lead-tin solder; and
 - polyethylene blended with another crystalline polymer.
- 1.8 Which parts of a specimen will be highlighted under dark-field illumination in an optical microscopic study if the specimen is a polycrystalline metal with a few ceramic particles?
- 1.9 Why do we rotate the analyzer when examining microstructure with polarized light?
- 1.10 Why do we say that the Nomarski contrast may not provide a real three-dimensional image?
- 1.11 Why is fluorescence microscopy more commonly used in biological and polymer specimens than in metals and ceramics?

2

X-ray Diffraction Methods

X-ray diffraction methods are the most effective methods for determining the crystal structure of materials. Diffraction methods can identify chemical compounds from their crystalline structure, not from their compositions of chemical elements. It means that the different compounds (or phases) that have the same composition can be identified. Diffraction methods include X-ray diffraction, electron diffraction and neutron diffraction. X-ray diffraction by crystals was discovered in 1912, and since then it has been the most extensively studied and used technique for materials characterization. This chapter introduces X-ray diffraction methods. The theoretical background of diffraction discussed in this chapter is also applied to other types of diffraction methods. X-ray diffraction methods can be classified into two types: spectroscopic and photographic. The spectroscopic technique known as the *X-ray powder diffractometry*, or simply *X-ray diffractometry*, is the most widely used diffraction method and the main technique discussed in this chapter. Photographic techniques are not widely used as diffractometry in modern laboratories. One reason is that spectroscopic methods can replace most photographic methods. However, photographic methods are used to determine unknown crystal structures. These methods will not be discussed in this chapter because they require extensive knowledge of crystallography.

2.1 X-ray Radiation

2.1.1 Generation of X-rays

X-rays are short-wavelength and high-energy beams of electromagnetic radiation. X-ray energy is characterized either by wavelength or photon energy. X-rays are produced by high-speed electrons accelerated by a high-voltage field colliding with a metal target. Rapid deceleration of electrons on the target enables the kinetic energy of electrons to be converted to the energy of X-ray radiation. The wavelength of X-ray radiation (λ) is related to the acceleration voltage of electrons (V) as shown in the following equation.

$$\lambda = \frac{1.2398 \times 10^3}{V} \text{ (nm)} \quad (2.1)$$

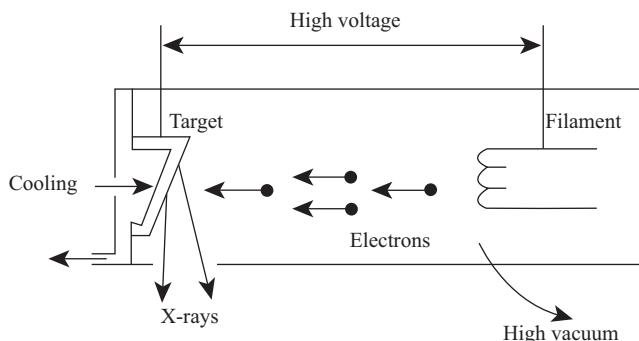


Figure 2.1 X-ray tube structure. X-ray radiation from the anode is guided by the windows of the tube to produce an X-ray beam for diffraction.

To generate X-rays, we need a device called the *X-ray tube*. Figure 2.1 illustrates the structure of an X-ray tube containing a source of electrons and two metal electrodes in a vacuum tube. The high voltage maintained across these electrodes rapidly draws the electrons to the anode (target). X-rays are produced at the point of impact on the target surface and radiated in all directions. There are windows to guide X-rays out of the tube. Extensive cooling is necessary for the X-ray tube because most of the kinetic energy of electrons is converted into heat; less than one percent is transformed into X-rays.

The X-ray tube generates the X-ray radiation with a range of wavelengths starting from a minimum λ , called *continuous X-rays* or *white X-rays*, shown as the background of radiation spectrum in Figure 2.2. The minimum λ (with maximum radiation energy) of continuous X-rays is determined by the maximum acceleration voltage of electrons in the X-ray tube according to Equation 2.1. For example, the minimum λ is 0.062 nm at an acceleration voltage of 20 kV. Figure 2.2 shows that there are sharp intensity maxima at certain wavelengths superimposed on a continuous X-ray spectrum. These intensity maxima are the *characteristic X-rays*. X-ray diffraction methods commonly require a source with single-wavelength (monochromatic) X-ray radiation. The monochromatic radiation must come from the characteristic X-rays that are generated by filtering out other radiations from the spectrum.

The physical principles of characteristic X-ray generation are schematically illustrated in Figure 2.3. When an incident electron has sufficient energy to excite an electron in the inner shell of an atom to a higher energy state, the vacancy left in the inner shell will be filled by an electron in an outer shell. As the electron falls to the inner shell, energy will be released by emitting an X-ray with a specific wavelength or photons with specific energy.

For example, a K -shell vacancy can be filled by an electron from either the L shell or M shell, which results in emitting the characteristic $K\alpha$ or $K\beta$ X-rays, respectively (Figure 2.2). Similarly, when a vacancy in the L shell is filled by an electron of the M shell, the $L\alpha$ X-rays will be emitted. The probability of an L shell electron filling the K shell vacancy is much higher than that of an M shell electron. Thus, the intensity of $K\alpha$ X-rays is higher than that of $K\beta$. More precisely, $K\alpha$ contains two characteristic lines: $K\alpha_1$ and $K\alpha_2$, with the latter wavelength being slightly longer than the former. This phenomenon results from the subshell structure of the L shell as indicated by L_1 and L_2 and L_3 . $K\alpha_1$ is the radiation when electrons fall from the L_3 to K shell; $K\alpha_2$ is generated when electrons fall from the L_2 to K shell (Figure 2.3).

Image Not Available In The Electronic Edition

$K\alpha_1$, $K\alpha_2$ and $K\beta$ are the three strongest characteristic X-rays used for diffraction radiation. The wavelength difference between $K\alpha_1$ and $K\alpha_2$ are so small that they are not always resolved as separate radiation. For example, wavelengths generated by a copper target are approximately the following.

$$\lambda_{K\alpha_1} = 0.15406 \text{ nm}$$

$$\lambda_{K\alpha_2} = 0.15444 \text{ nm}$$

$$\lambda_{K\beta} = 0.13922 \text{ nm}$$

We often simply call $K\alpha_1$ and $K\alpha_2$ radiations the $K\alpha$ doublet. The $K\alpha$ doublet is the most widely used monochromatic X-rays for diffraction work. Table 2.1 lists the commonly used $K\alpha$ doublet, of which the wavelength is an average of $K\alpha_1$ and $K\alpha_2$.

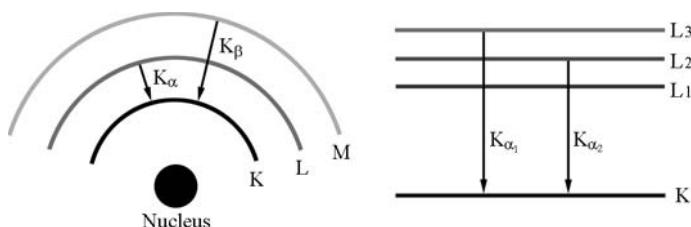


Figure 2.3 Schematic illustration of characteristic X-ray radiation.

Table 2.1 Characteristic X-rays of Anode Materials

Anode Material	Atomic Number	$K\alpha(\text{nm})$	Critical Excitation Potential (keV)	Optimum Voltage (kV)
Cr	24	0.22291	5.99	40
Fe	26	0.1937	7.11	40
Cu	29	0.1542	8.98	45
Mo	42	0.0710	20.00	80

2.1.2 X-ray Absorption

To obtain monochromatic $K\alpha$, a filter system is necessary to filter out the continuous X-rays and other characteristic X-rays, which are not generated with $K\alpha$ from the X-ray tube. X-ray filters can be made from material that strongly absorbs X-rays other than $K\alpha$. Generally, materials exhibit various abilities to absorb X-rays. X-ray absorption by materials is a function of the *linear absorption coefficient* (μ) and mass density (ρ). The X-ray intensity (I) passing through an absorption layer with thickness x is expressed by the following equation.

$$I_x = I_0 e^{-(\mu/\rho)\rho x} \quad (2.2)$$

Equation 2.2 expresses the exponential in terms of $(\mu\rho^{-1})$, which is called the *mass absorption coefficient*. It is independent of physical state of solid or liquid. The mass absorption coefficient varies with chemical element; an element of higher atomic number has a higher $\mu\rho^{-1}$ value than an element of lower atomic number.

The mass absorption coefficient generally decreases with decreasing wavelength of X-ray radiation. For example, Figure 2.4 shows the change in $\mu\rho^{-1}$ of barium in a wavelength range of X-ray radiation. Figure 2.4 also shows that the $\mu\rho^{-1}$ curve sharply jumps at certain wavelengths, marked as L_{III} , L_{II} , L_I and K . A jump-up point in the curve is called the *absorption edge* of the mass absorption coefficient. The reason $\mu\rho^{-1}$ increases sharply at certain wavelengths is that energies of X-ray radiation, represented by the wavelengths, are sufficient to excite certain

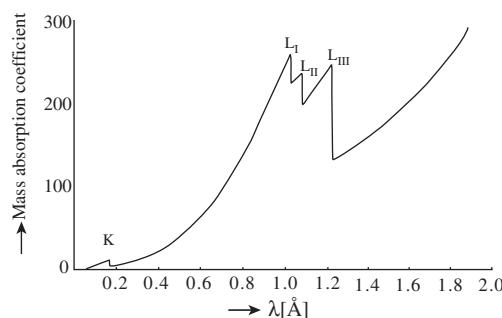


Figure 2.4 X-ray absorption curve of barium. (Reproduced with permission from R. Jenkins and R.L. Snyder, *Introduction to X-ray Powder Diffractometry*, John Wiley & Sons Inc., New York. © 1996 John Wiley & Sons Inc.)



Image Not Available In The Electronic Edition

characteristic X-ray radiations in barium. The absorption edges of L_{III} , L_{II} , L_I and K represent the excitation of L_{III} , L_{II} , L_I and K characteristic X-rays in barium, respectively.

The feature of absorption edge can be used for X-ray radiation filtering. The X-ray filtering mechanism is illustrated in Figure 2.5. We may select a filter material of which the absorption edge is located at the wavelength slightly shorter than that of $K\alpha$ radiation. The filter material can effectively absorb $K\beta$ and continuous X-rays with wavelengths shorter than the absorption edge as shown in Figure 2.5.

2.2 Theoretical Background of Diffraction

2.2.1 Diffraction Geometry

Bragg's Law

X-rays are electromagnetic waves, as are visible light, but the X-ray wavelength is much shorter than visible light, only in the order of 0.1 nm. X-ray diffraction methods are based on the phenomenon of wave interferences as introduced in Section 1.4.2 of Chapter 1. Two light waves with the same wavelength and traveling in the same direction can either reinforce or cancel each other, depending on their phase difference. When they have a phase difference of $n\lambda$ (n is an integer), called 'in phase', constructive interference occurs. When they have a phase difference of $n\lambda/2$, called 'completely out of phase', completely destructive interference occurs.

X-ray beams incident on a crystalline solid will be diffracted by the crystallographic planes as illustrated in Figure 2.6. Two in-phase incident waves, beam 1 and beam 2, are deflected by two crystal planes (A and B). The deflected waves will not be in phase except when the following relationship is satisfied.

$$n\lambda = 2d \sin \theta \quad (2.3)$$

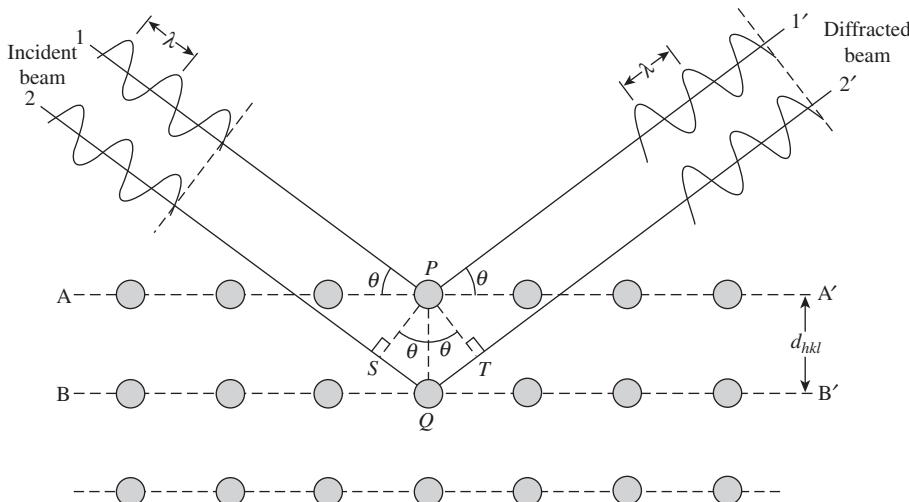


Figure 2.6 Bragg diffraction by crystal planes. The path difference between beams 1 and 2 is $SQ + QT = 2 PQ \sin\theta$. (Reproduced with permission from W.J. Callister Jr., *Materials Science and Engineering: An Introduction*, 7th ed., John Wiley & Sons Inc., New York. © 2006 John Wiley & Sons Inc.)

Equation 2.3 is the basic law of diffraction called *Bragg's Law*. Bragg's Law can be simply obtained by calculating the path differences between the two beams in Figure 2.6. The path difference depends on the incident angle (θ) and spacing between the parallel crystal planes (d). In order to keep these beams in phase, their path difference ($SQ + QT = 2d \sin\theta$) has to equal one or multiple X-ray wavelengths ($n\lambda$).

We are able to obtain information on the spacing between atomic planes of a crystal when constructive interference is detected at a given incident angle and a wavelength of the incident beam, based on Bragg's Law. Knowing the spacings of crystallographic planes by diffraction methods, we can determine the crystal structure of materials. For example, the plane spacing of cubic crystal relates to the lattice parameter (a) by the following equation.

$$d_{hkl} = \frac{a}{\sqrt{h^2 + k^2 + l^2}} \quad (2.4)$$

The Miller indices (hkl) represent a series of parallel planes in a crystal with spacing of d_{hkl} . Combining Equations 2.3 and 2.4, we obtain the following relationship between diffraction data and crystal parameters for a cubic crystal system.

$$\sin^2\theta = \frac{\lambda^2}{4a^2}(h^2 + k^2 + l^2) \quad (2.5)$$

Equation 2.5 does not directly provide the Miller indices of crystallographic planes. We need to convert $(h^2 + k^2 + l^2)$ to (hkl) or $\{hkl\}$. This should not be difficult for low-index planes of cubic systems. For example, when $(h^2 + k^2 + l^2)$ is equal to 1, the plane index must be $\{001\}$; when it is equal to 2, the index must be $\{110\}$. This type of calculation is often not necessary

because the relationship between deflection angle and the Miller indices of most crystalline materials, for a given λ , have been determined and published by the International Centre for Diffraction Data.

Reciprocal Lattice

A crystallographic plane (hkl) is represented as a light spot of constructive interference when the Bragg conditions (Equation 2.3) are satisfied. Such diffraction spots of various crystallographic planes in a crystal form a three-dimensional array that is the *reciprocal lattice* of crystal. The reciprocal lattice is particularly useful for understanding a diffraction pattern of crystalline solids. Figure 2.7 shows a plane of a reciprocal lattice in which an individual spot (a lattice point) represents crystallographic planes with Miller indices (hkl).

A reciprocal lattice is in an imaginary reciprocal space that relates to the corresponding crystal lattice in real space. A direction in the crystal lattice is defined by a vector \mathbf{r}_{uvw} with unit vectors \mathbf{a} , \mathbf{b} and \mathbf{c} in real space.

$$\mathbf{r}_{uvw} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c} \quad (2.6)$$

where u , v and w are integers for indexing crystallographic directions. Similarly, we can define a direction in reciprocal lattice by a vector \mathbf{d}_{hkl}^* with the reciprocal unit vectors \mathbf{a}^* , \mathbf{b}^* and \mathbf{c}^*

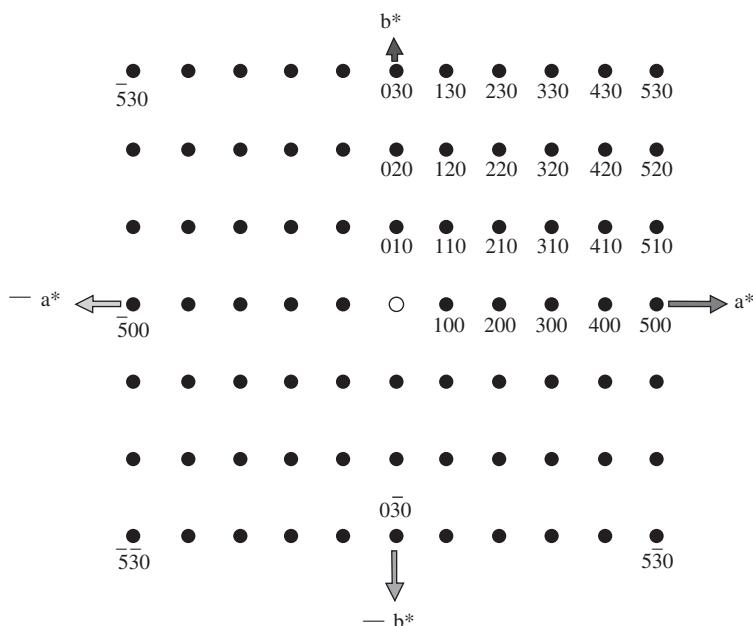


Figure 2.7 A reciprocal lattice plane of a simple tetragonal crystal. The indices of lattice points are those of the crystallographic plane which the points represent. (Reproduced with permission from R. Jenkins and R.L. Snyder, *Introduction to X-ray Powder Diffractometry*, John Wiley & Sons Inc., New York. © 1996 John Wiley & Sons Inc.)

in reciprocal space.

$$\mathbf{d}_{hkl}^* = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \quad (2.7)$$

A dimension in reciprocal space is a reciprocal of the dimension in real space (with a factor of unity). The magnitude of a vector \mathbf{d}_{hkl}^* in a reciprocal lattice equals the reciprocal of plane spacing (d_{hkl}) in real space.

$$|\mathbf{d}_{hkl}^*| = \left| \frac{1}{d_{hkl}} \right| \quad (2.8)$$

The relationship between a reciprocal lattice and its real crystal structure is highlighted as follows:

- A crystal structure always has a unique reciprocal lattice structure (as defined by Equations 2.7 and 2.8);
- Each lattice point in reciprocal space represents a crystal plane (hkl). The direction of \mathbf{d}_{hkl}^* is perpendicular to the crystal plane (hkl);
- The relationship between the unit vectors of true and reciprocal lattices should be

$$\mathbf{a}^* \perp \mathbf{b} \text{ and } \mathbf{c}, \mathbf{b}^* \perp \mathbf{c} \text{ and } \mathbf{a}, \mathbf{c}^* \perp \mathbf{a} \text{ and } \mathbf{b}, \mathbf{a}\mathbf{a}^* = \mathbf{b}\mathbf{b}^* = \mathbf{c}\mathbf{c}^* = 1$$

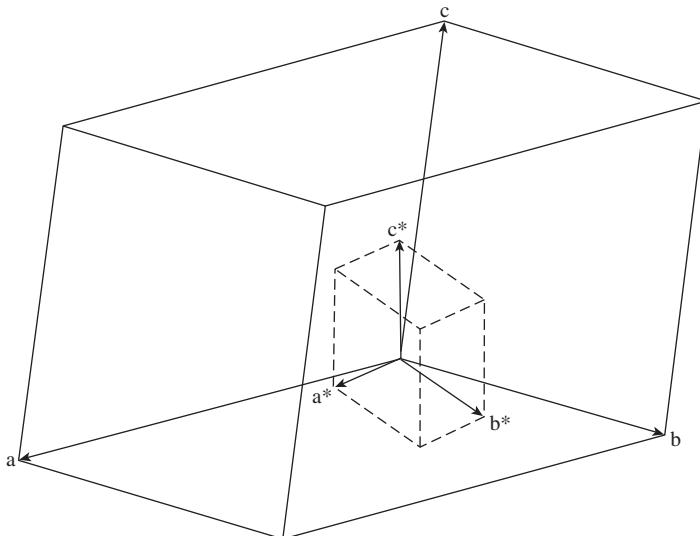


Figure 2.8 Relationship between directions of unit vectors in a real crystal lattice (\mathbf{a} , \mathbf{b} , \mathbf{c}) and its reciprocal lattice (\mathbf{a}^* , \mathbf{b}^* , \mathbf{c}^*). \mathbf{a}^* must be perpendicular to \mathbf{b} and \mathbf{c} because \mathbf{a}^* is perpendicular to the (100) plane. Generally, \mathbf{a}^* is not parallel to \mathbf{a} except in orthogonal systems.

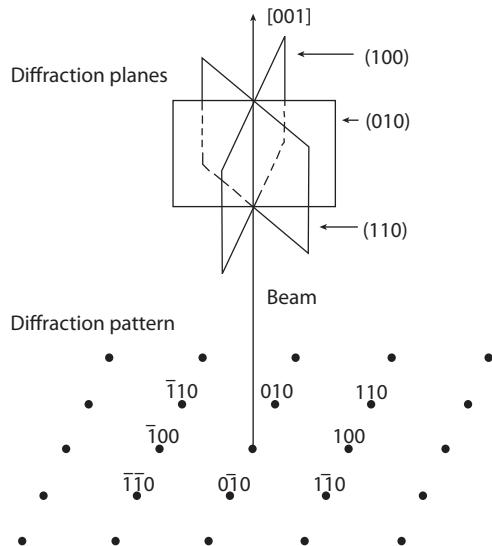


Figure 2.9 Relationship between a reciprocal lattice plane and real space directions. The reciprocal lattice plane is composed of diffraction spots from crystal planes $(hk\bar{0})$. The zone axis [001] is perpendicular to the reciprocal lattice plane.

- It is not generally true that $\mathbf{a}^* \parallel \mathbf{a}$, $\mathbf{b}^* \parallel \mathbf{b}$ and $\mathbf{c}^* \parallel \mathbf{c}$ unless the crystal is orthogonal ($\mathbf{a} \perp \mathbf{b}$, $\mathbf{b} \perp \mathbf{c}$ are $\mathbf{c} \perp \mathbf{a}$). For example, Figure 2.8 shows an example that \mathbf{a}^* is perpendicular to the (100) plane and \mathbf{a} is parallel to the [100] directions. The [100] directions in real space are not perpendicular to the (100) planes if [100] are not perpendicular to [010] and [001]; and
 - Crystal planes belonging to one crystal zone (a group of crystallographic planes of which normal directions are perpendicular to one direction called the zone axis direction) form one reciprocal lattice plane. The zone axis direction is perpendicular to that reciprocal lattice plane. For example, Figure 2.9 shows an example of the relationship between crystal zone [001] and the reciprocal lattice plane. The reciprocal lattice plane is composed of the diffraction spots of crystallographic planes belonging to the [001] zone.

Ewald Sphere

Bragg's Law describes the necessary conditions to detect crystal planes by diffraction. The conditions can also be graphically expressed by the *Ewald sphere* method using the concept of the reciprocal lattice. Figure 2.10 illustrates the construction of the Ewald sphere. The Ewald sphere is an imaginary sphere with a radius of λ^{-1} in reciprocal space. The center of the Ewald sphere is located at a crystal to be examined. The incident beam is represented by a line that passes through a crystal at the sphere center and targets the origin of the reciprocal lattice located on the surface of the sphere (*CO*). The diffraction beam is represented by line connect the sphere center and a lattice point of reciprocal lattice (*CB*). The angle between lines *CO* and *CB* should be twice that of θ defined in Figure 2.6. Changing the diffraction angle (2θ) can be represented by rotation of the reciprocal lattice around its origin, point *O* in Figure 2.10. The Bragg conditions are satisfied when a lattice point touches the Ewald sphere surface. It can

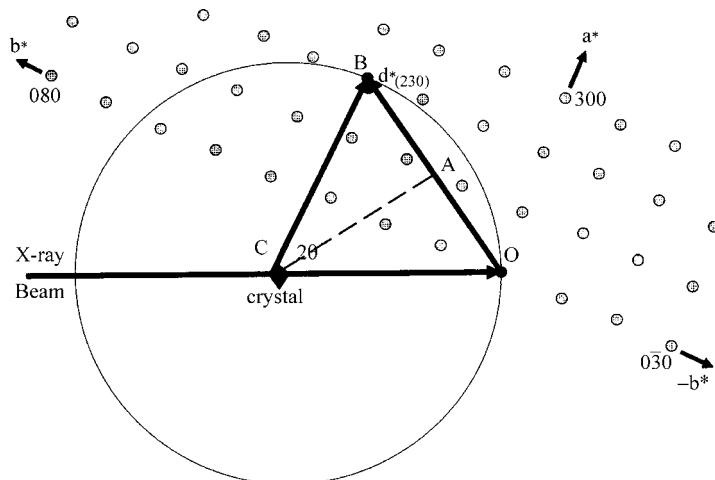


Figure 2.10 Construction of the Ewald sphere. The Bragg conditions for plane (230) are satisfied when its reciprocal lattice point touches the surface of the Ewald sphere. (Reproduced with permission from R. Jenkins and R.L. Snyder, *Introduction to X-ray Powder Diffractometry*, John Wiley & Sons Inc., New York. © 1996 John Wiley & Sons Inc.)

be proved that these geometric relationships represented in the Ewald sphere are equivalent to Bragg's Law, using Figure 2.10 as an example.

$$\begin{aligned}
 \text{Since} \quad OC &= \frac{1}{\lambda}, \quad OA = \frac{d_{230}^*}{2} \\
 \text{Hence} \quad \sin\theta &= \frac{OA}{OC} = \frac{d_{230}^*/2}{1/\lambda} \\
 \text{Recall} \quad d_{230} &= \frac{1}{d_{230}^*} \\
 \text{Thus} \quad \lambda &= 2d_{230} \sin\theta.
 \end{aligned}$$

This equivalence allows us to detect diffraction from crystallographic planes when the corresponding reciprocal lattice points touch the Ewald sphere surface. Formation of diffraction patterns from a single crystal can be well illustrated by the Ewald sphere as shown in Figure 2.11, which is commonly seen by electron diffraction in a transmission electron microscope (TEM). The electron wavelength in TEM is even shorter than X-rays (~ 0.0037 nm with acceleration voltage of 100 kV). For an incident beam with such a short wavelength (λ), the radius of the Ewald sphere (λ^{-1}) is huge (~ 100 times larger than the value of d^* for a low Miller index plane). Thus, the surface of the Ewald sphere is flat compared with the unit vectors of the reciprocal lattice for crystals. The relatively flat surface of the Ewald sphere touches a number of reciprocal lattice points, and these points form a diffraction pattern on the screen of a TEM as shown in the lower part of Figure 2.11.

The Ewald sphere also predicts a ring type diffraction pattern in a polycrystalline material, which is considered as an aggregate of the crystals with all possible orientations in three-dimensional space. We may consider that placing a polycrystalline solid at the center of the Ewald sphere is equivalent to a case of a single crystal by rotating the reciprocal lattice around

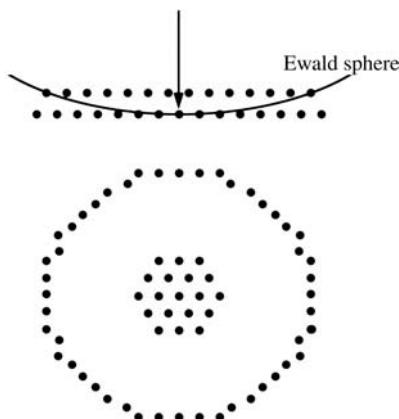


Figure 2.11 Formation of a single crystal diffraction pattern in transmission electron microscopy. The short wavelength of electrons makes the Ewald sphere flat. Thus, the array of reciprocal lattice points in a reciprocal plane touches the sphere surface and generates a diffraction pattern on the TEM screen. The outer ring may be visible when the Ewald sphere surface touches the reciprocal plane above the original plane.

its origin in all directions of reciprocal space. Rotation of a reciprocal lattice in all directions reveals all possible crystallographic planes satisfying the Bragg conditions. Figure 2.12 illustrates the diffraction pattern of a polycrystalline solid recorded by the method of powder X-ray photography. The polycrystalline diffraction patterns shown at the right-hand side of Figure 2.12 are called the *Debye rings* recorded on a negative film with a central hole.

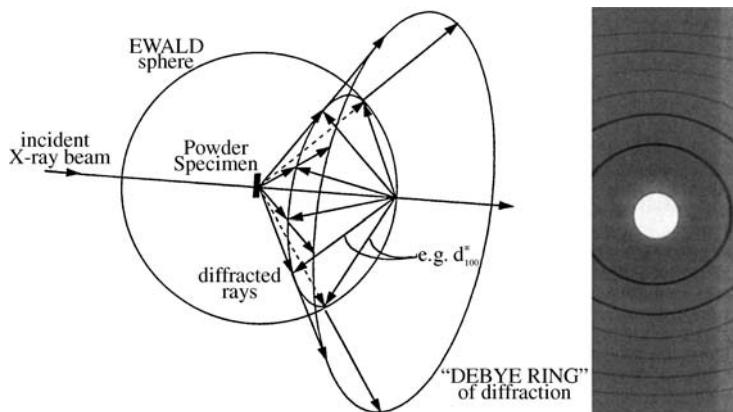


Figure 2.12 The Ewald sphere method illustrates the ring pattern of diffraction from a powder specimen. The Debye ring recorded by the Hull–Debye–Scherrer method results from randomly oriented crystals in the powder specimen, in which reciprocal lattice points of (hkl) touch the Ewald sphere surface in various directions to form individual rings. It is equivalent to rotating a reciprocal lattice along an incident beam axis. (Reproduced with permission from R. Jenkins and R.L. Snyder, *Introduction to X-ray Powder Diffractometry*, John Wiley & Sons Inc., New York. © 1996 John Wiley & Sons Inc.)

2.2.2 Diffraction Intensity

In the previous section, we were only concerned with diffraction conditions stated in Bragg's Law. Satisfying such conditions does not guarantee that we can detect or see diffraction from crystallographic planes because there is an issue of diffraction intensity. Diffraction intensity may vary among planes even though the Bragg conditions are satisfied. This section does not extensively discuss diffraction intensity, but only introduces important factors affecting the intensity of the X-ray diffraction.

X-ray diffraction by a crystal arises from X-ray scattering by individual atoms in the crystal. The diffraction intensity relies on collective scattering by all the atoms in the crystal. In an atom, the X-ray is scattered by electrons, not nuclei of atom. An electron scatters the incident X-ray beam to all directions in space. The scattering intensity is a function of the angle between the incident beam and scattering direction (2θ). The X-ray intensity of electron scattering can be calculated by the following equation.

$$I(2\theta) = \frac{I_0}{r^2} K \frac{1 + \cos^2(2\theta)}{2} \quad (2.9)$$

I_0 is the intensity of the incident beam, r is the distance from the electron to the detector and K is a constant related to atom properties. The last term in the equation shows the angular effect on intensity, called the *polarization factor*. The incident X-ray is unpolarized, but the scattering process polarizes it.

The total scattering intensity of an atom at a certain scattering angle, however, is less than the simple sum of intensities of all electrons in the atom. There is destructive interference between electrons due to their location differences around a nucleus, as illustrated in Figure 2.13a.

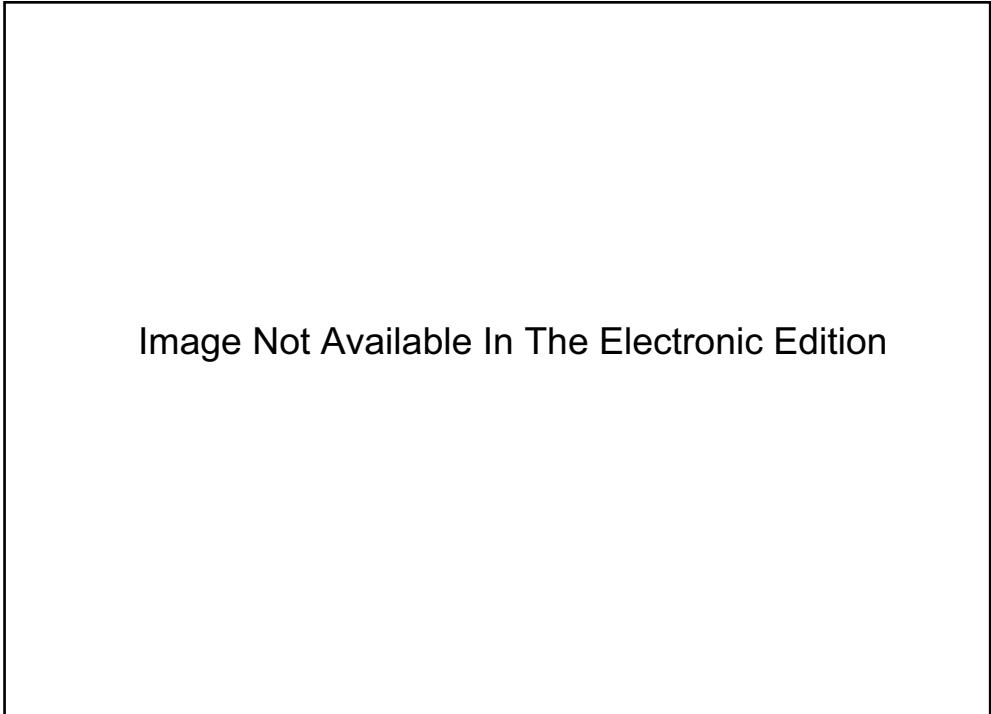


Image Not Available In The Electronic Edition

Destructive interference that results from the path difference between X-ray beams scattered by electrons in different locations is illustrated in Figure 2.13b. Figure 2.13b shows an example of a copper atom, which has 29 electrons. The intensity from a Cu atom does not equal 29 times that of a single electron in Equation 2.9 when the scattering angle is not zero. The *atomic scattering factor* (f) is used to quantify the scattering intensity of an atom.

$$f = \frac{\text{amplitude of the wave scattered by one atom}}{\text{amplitude of the wave scatter by one electron}} \quad (2.10)$$

The atomic scattering factor is a function of scattering direction and X-ray wavelength as shown in Figure 2.13b for copper atom. The scattering intensity of an atom is significantly reduced with increasing 2θ when the X-ray wavelength is fixed.

Structure Extinction

The most important aspect of intensity calculations is the *structure extinction*. This term refers to interference between scattering atoms in a unit cell of a crystal. If there is more than one atom in a unit cell (for example, a face-centered cubic lattice has four atoms in a unit cell), intensity from certain crystallographic planes can become extinct due to interference between atoms on different planes. Figure 2.14 illustrates this concept by comparing diffraction from two unit cells that have the same size. When the Bragg conditions are satisfied for (100) plane diffraction, and thus, the path difference ABC equals one wavelength, diffraction from the (001) plane of the base-centered crystal (Figure 2.14a) can be seen. However, diffraction from the (001) plane of a body-centered crystal cannot be seen because the path difference DEF between rays $1'$ and $3'$ is exactly equal to half that of the rays $1'$ and $2'$, ABC , or one-half wavelength (Figure 2.14b). Thus, rays $1'$ and $3'$ are completely out of phase and annul each other. Similarly, ray $2'$ will be annulled by the plane below it. This means that no diffraction from the (001) plane in a body-centered crystal can be detected. Note that diffraction of (002) plane in a body-centered crystal can be detected at another angle that it makes DEF equal to one wavelength.

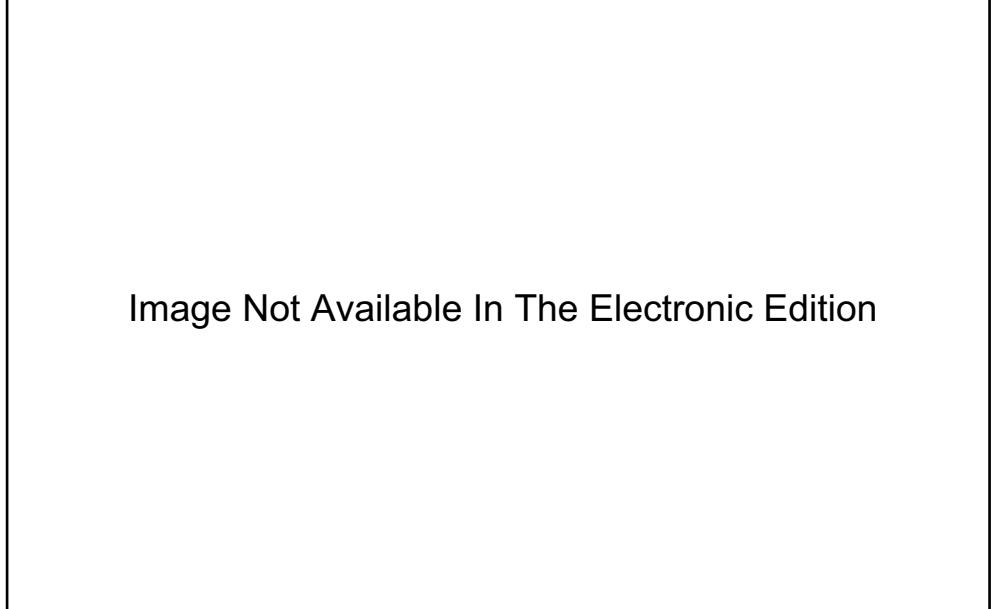


Image Not Available In The Electronic Edition

Table 2.2 Structure Extinction Rules for Crystallographic Plane (hkl)

Lattice Type	Diffraction possibly present	Diffraction necessarily absent
Simple	all	none
Base-centered	h and k unmixed ^a	h and k mixed ^a
Body-centered	$(h+k+l)$ even	$(h+k+l)$ odd
Face-centered	h, k , and l unmixed	h, k , and l mixed

^aApplies to a cell with (00l) face centered.

The structure extinction of intensity can be calculated by the *structure factor* (F). Suppose that there are N atoms per unit cell, and the location of atom n is known as u_n, v_n, w_n and its atomic structure factor is f_n . The structure factor for (hkl) plane (F_{hkl}) can be calculated.

$$F_{hkl} = \sum_n^N f_n \exp [2\pi i(hu_n + kv_n + lw_n)] \quad (2.11)$$

For example, a body-centered cubic (BCC) unit cell has two atoms located at [0,0,0] and [0.5,0.5,0.5]. We can calculate the diffraction intensity of (001) and (002) using Equation 2.11.

$$\begin{aligned} F_{001} &= f \exp [2\pi i(0 + 0 + 1 \times 0) + f \exp [2\pi i(0 \times 0.5 + 0 \times 0.5 + 1 \times 0.5)] \\ &= f + f \exp [\pi i] = 0 \\ F_{002} &= f \exp [2\pi i(0 + 0 + 2 \times 0) + f \exp [2\pi i(0 \times 0.5 + 0 \times 0.5 + 2 \times 0.5)] \\ &= f + f \exp [2\pi i] = 2f \end{aligned}$$

This structure factor calculation confirms the geometric argument of (001) extinction in a body-centered crystal. Practically, we do not have to calculate the structure extinction using Equation 2.11 for simple crystal structures such as BCC and face-centered cubic (FCC). Their structure extinction rules are given in Table 2.2, which tells us the detectable crystallographic planes for BCC and FCC crystals. From the lowest Miller indices, the planes are given as following.

BCC (1 1 0), (2 0 0), (2 1 1), (2 2 0), (3 1 0), (2 2 2), (3 2 1), ...

FCC (1 1 1), (2 0 0), (2 2 0), (3 1 1), (2 2 2), (3 3 1), ...

For a unit cell with more than one type of atom, the calculation based on Equation 2.11 will be more complicated because the atomic scattering factor varies with atom type. We need to know the exact values of f for each type of atom in order to calculate F . Often, reduction of diffraction intensity for certain planes, but not total extinction, is seen in crystals containing multiple chemical elements.

2.3 X-ray Diffractometry

X-ray diffractometry (XRD), the most widely used X-ray diffraction technique in materials characterization, was originally used for examining the crystal structure of powder samples; thus, traditionally it is called *X-ray powder diffractometry*. In fact, polycrystalline aggregate

solids other than powder are commonly examined by this technique. The XRD instrument is called an X-ray diffractometer. In the diffractometer, an X-ray beam of a single wavelength is used to examine polycrystalline specimens. By continuously changing the incident angle of the X-ray beam, a spectrum of diffraction intensity versus the angle between incident and diffraction beam is recorded. Diffractometry enables us to identify the crystal structure and quality by analyzing then comparing the spectrum with a database containing over 60,000 diffraction spectra of known crystalline substances.

2.3.1 Instrumentation

The basic function of a diffractometer is to detect X-ray diffraction from materials and to record the diffraction intensity as a function of the diffraction angle (2θ). Figure 2.15 demonstrates the geometrical arrangement of X-ray source, specimen and detector. The X-ray radiation generated by an X-ray tube passes through special slits which collimate the X-ray beam. These *Soller slits* are commonly used in the diffractometer. They are made from a set of closely spaced thin metal plates parallel to the figure plane of Figure 2.15 to prevent beam divergence in the director perpendicular to the figure plane. A divergent X-ray beam passing through the slits strikes the specimen. The specimen is usually in the form of a flat plate that is supported by specimen table (not shown in Figure 2.15). X-rays are diffracted by the specimen and form a convergent beam at receiving slits before they enter a detector.

The diffracted X-ray beam needs to pass through a monochromatic filter (or a monochromator) before being received by a detector. Commonly, the monochromatic filter is placed in the diffracted beam path instead of the incident beam path. This arrangement can suppress wavelengths other than $K\alpha$ radiation and also decrease background radiation originating within the specimen. Modern diffractometers commonly use a monochromator made from a graphite crystal that is designed to diffract a single wavelength based on Bragg's Law.

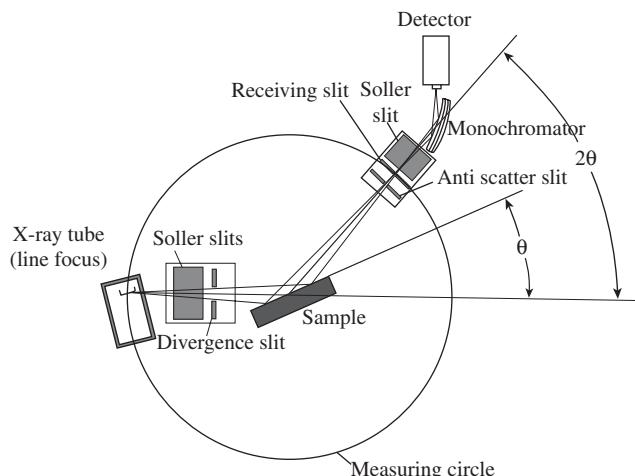


Figure 2.15 Geometric arrangement of X-ray diffractometer.

Relative movements among the X-ray tube, specimen and the detector ensure the recording of diffraction intensity in a range of 2θ . Note that the θ angle is not the angle between the incident beam and specimen surface; rather it is the angle between the incident beam and the crystallographic plane that generates diffraction. Diffractometers can have various types of geometric arrangements to enable collection of X-ray data. The majority of commercially available diffractometers use the *Bragg–Brentano arrangement*, in which the X-ray incident beam is fixed, but a sample stage (not shown in Figure 2.15) rotates around the axis perpendicular to the figure plane of Figure 2.15 in order to change the incident angle. The detector also rotates around the axis perpendicular to the plane of the figure but its angular speed is twice that of the sample stage in order to maintain the angular correlation of θ – 2θ between the sample and detector rotation.

The Bragg–Brentano arrangement, however, is not suitable for thin samples such as thin films and coating layers. The technique of *thin film X-ray diffractometry* uses a special optical arrangement for detecting the crystal structure of thin films and coatings on a substrate. The incident beam is directed to the specimen at a small glancing angle (usually $<1^\circ$) and the glancing angle is fixed during operation and only the detector rotates to obtain the diffraction signals as illustrated in Figure 2.16. Thin film X-ray diffractometry requires a parallel incident beam, not a divergent beam as in regular diffractometry. Also, a monochromator is placed in the optical path between the X-ray tube and the specimen, not between the specimen and the detector. The small glancing angle of the incident beam ensures that sufficient diffraction signals come from a thin film or a coating layer instead of the substrate.

System Aberrations

In ideal conditions, an X-ray beam focuses on a specimen and the intensity of diffraction from the specimen is detected accurately at the 2θ angle. The focusing arrangement in a diffractometer, however, cannot ensure true focusing due to features of the X-ray source and specimen. Thus, errors due to *parafocusing* always exist. The errors of parafocusing are referred to as:

- Axial divergence error;
- Flat-specimen error; and
- Specimen transparency error.

Errors due to parafocusing can be illustrated using the Bragg–Brentano arrangement of the diffractometer as shown in Figure 2.17. The X-ray source is fixed at F, but the receiving slit (RS) coupled with the detector rotates along the axis perpendicular to the plane of the figure on

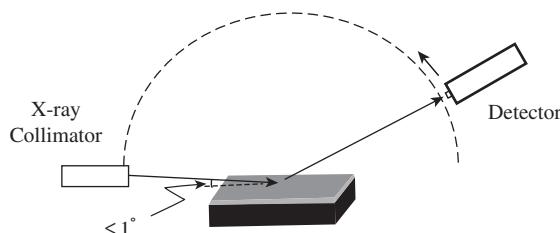


Figure 2.16 Optical arrangement for thin film diffractometry.

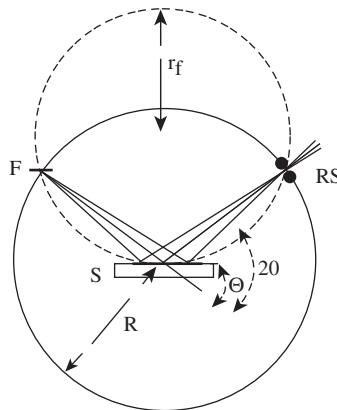


Figure 2.17 System aberrations of the X-ray diffractometer due to axial-divergence and flat-specimen errors. (Reproduced with permission from R. Jenkins and R.L. Snyder, *Introduction to X-ray Powder Diffractometry*, John Wiley & Sons Inc., New York. © 1996 John Wiley & Sons Inc.)

a circle centered at the specimen surface with radius R (the *goniometer circle*). The surface of the specimen rotates along the same axis at half the rotation speed of RS, remaining tangential to a focusing circle with radius of r_f . This arrangement of rotations ensures the scan a range of 2θ angles. As shown in Figure 2.17, the incident X-ray beam has a certain degree of divergence; part of the specimen surface is not located on the focusing circle and the atomic planes inside the specimen are not located on the focusing circle. These aforementioned conditions generate system errors which affect the peak positions and widths in the diffraction spectrum.

Besides conventional X-ray diffractometry, high resolution X-ray diffractometry (HRXRD) was developed in recent years. HRXRD is mainly used for examining the crystal quality of thin films and epitaxial layers on single-crystal substrates. The main features of the HRXRD instrument are that highly monochromatic X-rays are used (only $K\alpha_1$, not $K\alpha_2$) and the specimen stage is able to rotate with three degrees of freedom. Such features enable us to obtain highly accurate information including crystal orientation of a thin film with respect to its substrate, lattice mismatch of an epitaxial layer, and elastic strain in thin coatings.

2.3.2 Samples and Data Acquisition

Sample Preparation

The X-ray diffractometer was originally designed for examining powder samples. However, the diffractometer is often used for examining samples of crystalline aggregates other than powder. Polycrystalline solid samples and even liquids can be examined. Importantly, a sample should contain a large number of tiny crystals (or grains) which randomly orient in three-dimensional space because standard X-ray diffraction data are obtained from powder samples of perfectly random orientation. Relative intensities among diffraction peaks of a non-powder sample can be different from the standard because perfect randomness of grain orientation in solid samples is rare. The best sample preparation methods are those that allow analysts to obtain the desired information with the least amount of treatment because chemical contamination may occur during the sample treatments.

Table 2.3 Penetration Depths of Cu K α Radiation with Different Incident Angles

Material	$\frac{\mu}{\rho}$	ρ (g cm $^{-3}$)	20° (μm)	40° (μm)	60° (μm)
MoO ₃	92.7	4.71	1.8	2.7	3.6
CaCO ₃	39.9	2.71	7.4	11	15
Aspirin	7.0	1.40	82	121	161

Data taken from R. Jenkins and R.L. Snyder, *Introduction to X-ray Powder Diffractometry*, John Wiley & Sons Inc., New York. © 1996 John Wiley & Sons Inc.

Another sample-related issue is that X-ray penetration into matter is limited because of X-ray absorption in matter. A certain depth from the surface can only be examined. Table 2.3 lists typical penetration depths for three types of materials. Obviously, X-rays penetrate deeper in materials composed of lighter elements than those of heavy elements.

Acquisition and Treatment of Diffraction Data

A diffractometer records changes of diffraction intensity with 2θ. The diffractometer records the diffraction intensity starting from a low 2θ and ending at a high 2θ. Figure 2.18 shows the X-ray diffraction spectrum for a powder mixture of Si and Al₂O₃. A number of intensity peaks located at different 2θ provide a ‘fingerprint’ for a crystalline solid such as Al₂O₃. Each peak represents diffraction from a certain crystallographic plane. Matching an obtained peak spectrum with a standard spectrum enables us to identify the crystalline substances in a specimen.

Spectrum matching is determined by the positions of the diffraction peak maxima and the relative peak intensities among diffraction peaks.

In a modern diffractometer, data acquisition and treatment are mainly done by computer. To obtain adequate and data of sufficient quality for identifying a material, the following factors are important:

- Choice of 2θ range: commonly a range of about 5–65° is necessary with an unknown material;
- Choice of X-ray radiation: copper K α is the most popular radiation source used in diffractometry because of its short wavelength. Table 2.4 lists the detection limitations of X-ray radiations commonly used in diffractometry; and

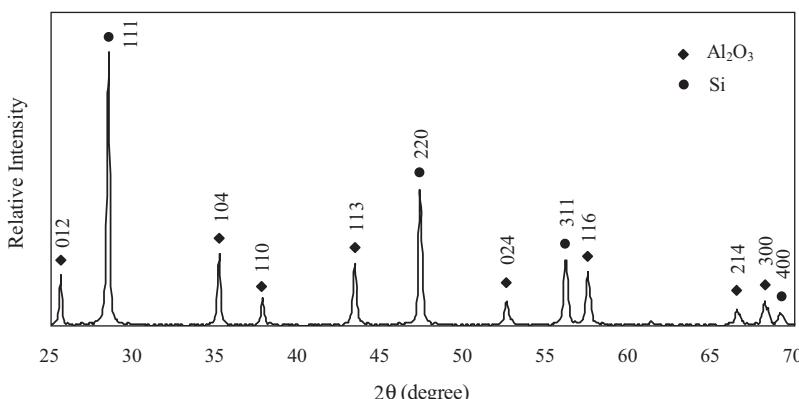


Figure 2.18 The experimental diffraction pattern of a silicon and Al₂O₃ mixture. Numbers with three digits mark the Miller indices of corresponding crystallographic planes.

Table 2.4 Maximum and Minimum Measurable Plane Spacing with Characteristic Radiation

Radiation	Wavelength (nm)	Maximum d (nm)	Minimum d (nm)
CrK α	0.2291	6.56	0.116
CoK α	0.1790	5.13	0.091
CuK α	0.1542	4.42	0.078
MoK α	0.0709	2.03	0.036

- Choice of step width for scanning: scanning in a 2θ range is not continuous, but in progressive steps. Choosing the step width is a trade-off between accuracy in peak location and peak intensity. While it might seem that scanning in narrow steps would be more accurate, in fact small step widths may lead to peak shifts. On the other hand, a large step width may lead to suppression of peak intensity. The rule of thumb is to use about 10–20 individual data points above the *full-width at half-maximum* (FWHM) as B marked in Figure 2.19. For example, for a typical peak that has a width of about $0.1\text{--}0.3^\circ$ in 2θ , a step width of 0.02° in 2θ is reasonable.

To obtain a satisfactory diffraction spectrum, the following treatments of diffraction data treatment should be done:

- Smooth the spectrum curve and subtracting the spectrum background;
- Locate peak positions; and
- Strip K α_2 from the spectrum, particularly for peaks in the mid-value range of 2θ .

Recorded diffraction peaks always have a finite width and are not single lines as illustrated in Figure 2.19. The location of a peak position can be determined as the point of a peak curve where the first derivative equals zero. To ensure the accuracy of a peak position, an

Image Not Available In The Electronic Edition

external standard made by a suitable material with known peak positions should be used to calibrate the acquired 2θ value. Comparing known peak positions of the standard with peaks acquired in the diffractometer, we can detect system errors ($\Delta 2\theta$) of the diffractometer. The $\Delta 2\theta$ error against a 2θ range should be corrected before analyzing a spectrum of sample to be examined.

2.3.3 Distortions of Diffraction Spectra

Preferential Orientation

A diffraction spectrum is considered a ‘fingerprint’ for identifying examined materials. An acquired spectrum should be compared with the standard from an X-ray diffraction database. The standard spectrum has been collected from a pure powder sample with fine powder size, usually several micrometers in diameter. We would expect to match both peak locations and relative peak intensities between the acquired spectrum and the standard if the sample is the same substance as the standard. Matching relative intensities, however, is not always possible even when the examined specimen has the same crystalline structure as the standard. One reason for this discrepancy might be the existence of preferential crystal orientation in the sample. This often happens when we examine coarse powder or non-powder samples. Preferential orientation of crystals (grains in solid) may even make certain peaks invisible.

Crystallite Size

Ideally, a Bragg diffraction peak is a line without width as shown in Figure 2.19b. In reality, diffraction from a crystal specimen produces a peak with a certain width as shown in Figure 2.19a. The peak width can result from instrumental factors and the size effect of the crystals. Small crystals cause the peak to be widened due to incompletely destructive interference. This phenomenon can be understood by examining the case illustrated in Figure 2.20.

A diffraction peak is generated at the Bragg angle θ_B satisfying the Bragg conditions. In a diffractometer, however, the incident X-ray beam is not perfectly parallel, and the beam includes a range of incident angles from θ_1 to θ_2 which deviate slightly from θ_B . θ_1 and θ_2 should not generate completely constructive interference as θ_B does, but may not generate completely destructive interference either. We know from Bragg’s Law that completely destructive interference needs diffraction from two planes producing exactly one-half wavelength difference of path. If the paths of rays with angle θ_1 scattered by the first two planes differ only slightly from an integral number of wavelengths, a plane lying deep within the crystal can produce the exact path difference of one-half wavelength. In this case, no diffraction other than θ_B should be visible. However, such a plane lying deep within the crystal may not exist if a crystal is thin in the direction perpendicular to the plane. Thus, such incompletely destructive interference generated by X-rays with incident angles θ_1 and θ_2 will produce a certain level of diffraction intensities around the Bragg angle θ_B in the spectrum, and result in widening of the diffraction peak. Peak width is inversely related to crystal size; that is, peak width increases with decreasing crystal particle size, as shown in Figure 2.21.

Residual Stress

Any factor that changes the lattice parameters of crystalline specimens can also distort their X-ray diffraction spectra. For example, residual stress in solid specimens may shift the diffraction peak position in a spectrum. Residual stress generates strain in crystalline materials by stretching or compressing bonds between atoms. Thus, the spacing of crystallographic planes

Image Not Available In The Electronic Edition

changes due to residual stress. According to Bragg's Law, the Bragg angles should either decrease or increase when spacing of the crystallographic plane changes. Peak shifts in spectra occur when there is residual stress in a sample, as illustrated in Figure 2.22. The tensile stress of increasing the spacing shifts a peak to lower 2θ while compression stress of decreasing the

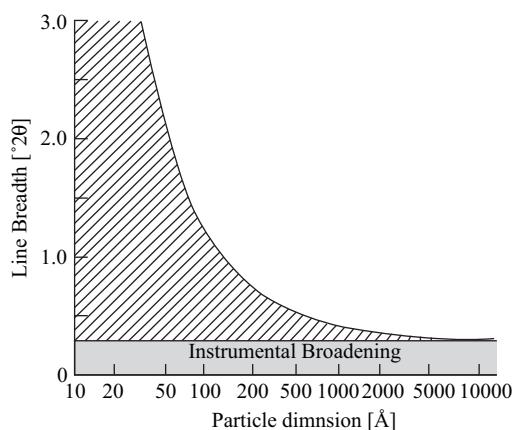


Figure 2.21 Change of peak line width with crystal size. (Reproduced with permission from R. Jenkins and R.L. Snyder, *Introduction to X-ray Powder Diffractometry*, John Wiley & Sons Inc., New York. © 1996 John Wiley & Sons Inc.)

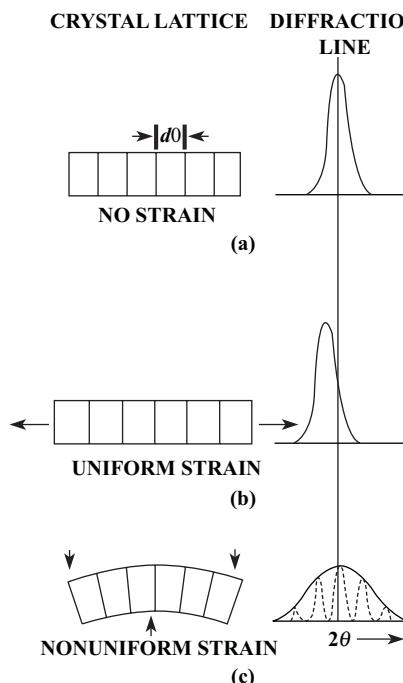


Figure 2.22 Effects of residual stress and strain on diffraction peak position and shape: (a) no strain; (b) uniform strain; and (c) non-uniform strain. (Reproduced with permission from R. Jenkins and R.L. Snyder, *Introduction to X-ray Powder Diffractometry*, John Wiley & Sons Inc., New York. © 1996 John Wiley & Sons Inc.)

spacing shifts a peak to higher 2θ in the spectrum. In fact, X-ray diffraction is an effective tool to examine residual stress in solids.

2.3.4 Applications

Crystal Phase Identification

Identification of crystalline substance and crystalline phases in a specimen is achieved by comparing the specimen diffraction spectrum with spectra of known crystalline substances. X-ray diffraction data from a known substance are recorded as a *powder diffraction file* (PDF). Most PDFs are obtained with $\text{CuK}\alpha$ radiation. Standard diffraction data have been published by the International Centre for Diffraction Data (ICDD), and they are updated and expanded from time to time. For one crystalline substance, there may be more than one file. The most recently updated file is recommended for phase identification. The early PDFs may contain errors in data obtained experimentally. More recently published PDFs are either obtained by more accurate experimental measurements or by theoretical calculation. A specimen to be identified should be in a powder form for most accurate matching. When we need to identify the crystal structure of a specimen that cannot be prepared as powder, matches of peak positions and relative intensities might be less than perfect. In this case, other information about the specimen such as chemical composition should be used to make a judgment.

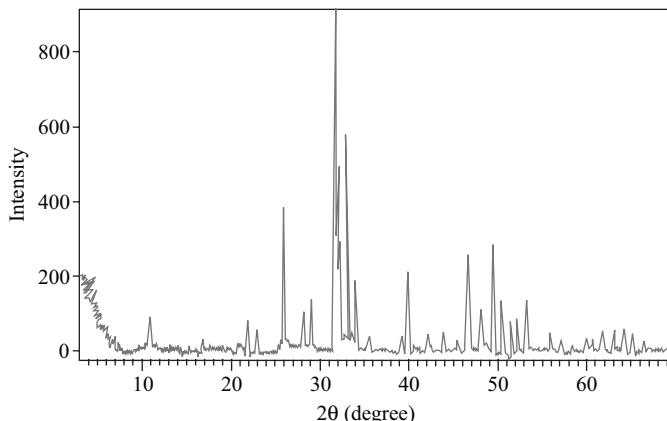


Figure 2.23 Diffraction pattern of a powder specimen of hydroxyapatite.

In a modern diffractometer, computer software performs the search–match task. All PDFs of the ICDD can be stored in computer. A search–match program can find all the possible matches for a specimen. The software algorithms also can identify more than one crystalline phase in a specimen. It searches the recorded pattern with its background, and adds candidate crystalline phases together to compose, rather than decompose, an observed multiphase pattern.

Figure 2.23 shows the diffraction spectrum of a powder sample of calcium phosphate after subtracting background. With assistance of a computer, we can identify the peak positions in the spectrum and search for a possible match between the spectrum and a PDF data file. Additional chemical information is often used to help in the search process. For example, this specimen contents Ca, P and O. The computer quickly searches for a compound containing Ca, P and O. It finds a match between the diffraction spectrum of a sample with data for hydroxyapatite (Figure 2.24). There are two important parameters in a standard data file shown in Figure 2.24: the position of diffraction (2θ) and relative intensities of peaks ($\frac{I}{I_1}$), or *int-f* in the PDF. I_1 is the peak intensity with the maximum value in a spectrum. The highest *int-f* value is 999 which should be read as 0.999 in the relative intensity. The PDF may also list the corresponding *d-spacing* of peaks, which are the true crystal properties.

Figure 2.25 shows a nearly perfect match between the pattern and diffraction data of hydroxyapatite. The powder sample shows 35 peak matches in the 2θ range of $10\text{--}60^\circ$. Generally, the relative intensities of peaks also match well with the standard. For example, the highest intensity is at the peak of $2\theta = 31.77^\circ$ and the second highest peak at $2\theta = 32.90^\circ$ in the acquired spectrum. Both of their 2θ values and relative intensities match those in the PDF for hydroxyapatite. The relative intensities of the peaks, however, do not exactly match all those in the PDF. The matches for lower intensity peaks between the spectrum and the standard file are relatively poorer, even though a sample of fine powder was examined.

Section 2.2.3 described factors that can affect relative intensities of peaks. These include the preferential orientation of crystal grains, residual stress and other crystal defects. Eventually, we must rely on human judgment to make the final identification. In this case, the matches of relative intensities are very good in general and no other calcium phosphate PDFs can provide matches as good as hydroxyapatite. Thus, hydroxyapatite can be positively identified.

74-0566 Quality:C	Ca10[P O4]6 [OH]2
CAS Number:	Calcium Hydroxide Phosphate
Molecular Weight: 1004.64	Ref: Calculated from ICSD using POWD-12++ [1997]
Volume[CD]: 529.09	Ref: Sudarsanan, K., Young, R.A., Acta Crystallogr., Sec. B, 25, 1534 [1969]
D _x : 3.153 D _m :	
Sys:Hexagonal Lattice:Primitive S.G.:P6 ₃ /m[176] Cell Parameters: a 9.424 b c 6.879	
I/Icor: 1.06 Rad: CuK _α Lambda: 1.54060 Filter: d-sp: calculated ICSD #: 026205	
Mineral Name Hydroxyapatite	
	Fixed Slit Intensity
	d [Å]
	Int-f h k l
8.1614	170 1 0 0 1.7197
5.2598	46 1 0 1 1.6828
4.7120	23 1 1 0 1.6444
4.0807	65 2 0 0 1.6322
3.8874	63 1 1 1 1.6108
3.5096	23 2 0 1 1.5881
3.4395	357 0 0 2 1.5847
3.1695	88 1 0 2 1.5815
3.0847	160 2 1 0 1.5706
2.9146	98 2 1 1 1.5423
2.7781	514 1 1 2 1.5312
2.7204	612 3 0 0 1.5050
2.6299	210 2 2 0 1.5020
2.5298	39 3 0 1 1.4746
2.3560	50 2 2 2 1.4658
2.2564	50 2 1 2 1.4536
2.2635	205 1 3 0 1.4502
2.2289	18 2 2 1 1.4336
2.2075	4 1 0 3 1.4065
2.1501	56 1 3 1 1.3890
2.1337	11 3 0 2 1.3693
2.0618	45 1 1 3 1.3602
2.0403	11 4 0 0 1.3566
1.9990	35 2 0 3 1.3484
1.9561	7 4 0 1 1.3417
1.9437	281 2 2 2 1.3344
1.8908	122 1 3 2 1.3297
1.8723	40 2 3 0 1.3206
1.8402	313 2 1 3 1.3163
1.8066	161 3 2 1 1.3068
1.7809	116 1 4 0 1.3037
1.7548	117 4 0 2 1.2958
1.7548	117 3 0 3 1.2839
	Int-f h k l
	1.0939

Figure 2.24 A powder diffraction file (PDF) of hydroxyapatite. (Reproduced with permission from the International Centre for Diffraction Data, ICDD.)

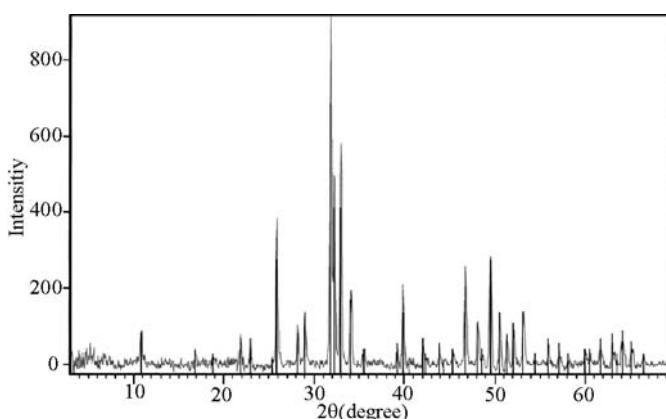


Figure 2.25 Match between the diffraction pattern and diffraction lines in the hydroxyapatite PDF.

Quantitative Measurement

Quantitative analysis can be done to determine relative amounts of compounds or phases in a sample of compound/phase mixtures. Quantitative analysis of a diffraction spectrum is based on the following factor. The intensity of the diffraction peaks of a particular crystalline phase in a phase mixture depends on the weight fraction of the particular phase in the mixture. Thus, we may obtain weight fraction information by measuring intensities of peaks assigned to a particular phase. Generally, we may express the relationship as the following.

$$I_i = \frac{KC_i}{\mu_m} \quad (2.12)$$

where the peak intensity of phase i in a mixture is proportional to the phase weight fraction (C_i), and inversely proportional to the linear absorption coefficient of the mixture (μ_m). K is a constant related to experimental conditions including wavelength, incident angle and the incident intensity of the X-ray beam. The relationship between intensity and weight fraction is not truly linear because μ_m itself is a function of C_i .

On the other hand, measurement of peak intensities is not simply measurement of peak heights. First, the background should be extracted from the peaks. Second, the integrated intensity should be measured for quantitative analysis. It means that the total area under a peak should be measured. Three basic types of peaks are encountered when attempting quantitative phase analysis (Figure 2.26):

1. Peak height is proportional to peak area. This simplest situation allows estimation of peak intensity by measuring peak height;
2. Peak height is not proportional to peak area. This common situation needs integration of peak area; and
3. Peak area is overlapped by other peaks. This also frequently occurs and profile-fitting methods have to be used for peak area measurement.

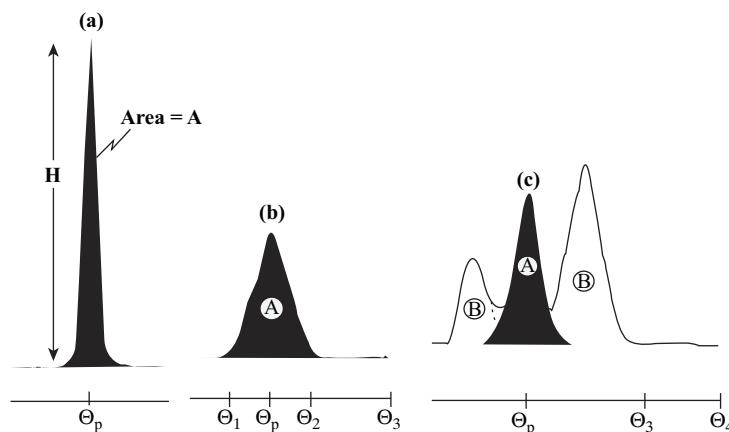


Figure 2.26 Three types of peak shapes in X-ray diffraction pattern: (a) peak height proportional to peak area; (b) peak height not proportional to peak area; and (c) peak area overlapped by other peaks. (Reproduced with permission from R. Jenkins and R.L. Snyder, *Introduction to X-ray Powder Diffractometry*, John Wiley & Sons Inc., New York. © 1996 John Wiley & Sons Inc.)

There are three main methods for quantitative analysis:

1. *External standard method*;
2. *Internal standard method*; and
3. *Direct comparison method*.

The external standard method needs peak intensities of a particular phase in its pure form. The internal standard method needs a series of standard mixtures in which the particular phase contents are known. The direct comparison method is of the most interest and is easy to use because it does not have either of these restrictive requirements; that is, it does not require samples of pure substances or standard mixtures of known phase content. The principles of the direct comparison method are introduced as follows.

The direct comparison method enables us to estimate the weight fraction of a particular phase by comparing peak intensities between phases in a mixture. For example, we have a mixture of phases α and β . We may rewrite Equation 2.12 for a peak intensity of α phase in a mixture.

$$I_\alpha = \frac{K_2 R_\alpha C_\alpha}{\mu_m} \quad (2.13)$$

where constant K is divided into K_2 and R_α . K_2 is related to the X-ray source only. R_α is a parameter related to the diffraction angle and crystal properties of α phase, which can be evaluated using the following equation.

$$R = \left(\frac{1}{v^2} \right) \left[|F_{hkl}|^2 p \left(\frac{1 + \cos^2 2\theta}{\sin^2 \theta \cos \theta} \right) \right] (e^{-2M}) \quad (2.14)$$

R is a function of incident angle (θ), F_{hkl} is the structure factor for a specific crystal plane, v is unit cell volume of crystal and temperature factor is e^{-2M} , and p is the multiplicity factor of a crystal, which is the number of crystal planes that have the same plane spacing. For a cubic crystal, the p value of {001} is six and p value of {111} is eight, because there six and eight planes in the plane family, respectively. The temperature factor can be extracted from Figure 2.27.

Image Not Available In The Electronic Edition

Similarly, we have a peak intensity of β phase in the mixture.

$$I_\beta = \frac{K_2 R_\beta C_\beta}{\mu_m} \quad (2.15)$$

Under identical conditions of X-ray source and incident intensity, constant K_2 should be the same. Thus, the following relationship holds.

$$\frac{I_\alpha}{I_\beta} = \frac{R_\alpha C_\alpha}{R_\beta C_\beta} \quad (2.16)$$

Equation 2.16 simply tells us how to obtain the ratio of weight fractions of α and β phases from their ratio of peak intensity, as long as the R values of individual peaks in the mixture are calculated. We can also apply this method to a mixture containing more than two phases, because the following condition always holds.

$$\sum_i C_i = 1$$

This method does not require powder samples. Thus, it is convenient for estimating phase contents in polycrystalline solids.

2.4 Wide Angle X-ray Diffraction and Scattering

Wide angle X-ray diffraction (WAXD) or *wide angle X-ray scattering* (WAXS) is the technique that is commonly used to characterize crystalline structures of polymers and fibers. The technique uses monochromatic X-ray radiation to generate a transmission diffraction pattern of a specimen that does strongly absorb X-rays, such as organic or inorganic specimens containing light elements. Wide angle X-ray diffraction or scattering refers to X-ray diffraction with a diffraction angle $2\theta > 5^\circ$, distinguishing it from the technique of small angle X-ray scattering (SAXS) with $2\theta < 5^\circ$. SAXS will not be discussed in Chapter 2 because its complexity in theoretical treatment is beyond the scope of this book.

2.4.1 Wide Angle Diffraction

Figure 2.28 illustrates the optical arrangement of WAXD. The incident X-ray beam is formed after X-rays pass through a monochromator and a collimator (not shown in Figure 2.28). The highly monochromatic and parallel X-ray beam strikes a sample and generates a diffracted beam at angle $2\theta_i$ with respect to the transmitted beam. The diffracted beams are recorded as an image on a film or by an area detector that is placed at distance D from the specimen and perpendicular to the incident beam direction. Usually, there is a beam-stop (made from lead) placed in front of the film or detector center to prevent the transmitted beam from striking the film or area detector directly. Thus, WAXD is a photographic method, differing from the spectroscopic methods of diffractometry. Small angle scattering uses the same optical arrangement as WAXD, except that D is longer in order to record scattering from small 2θ angles.

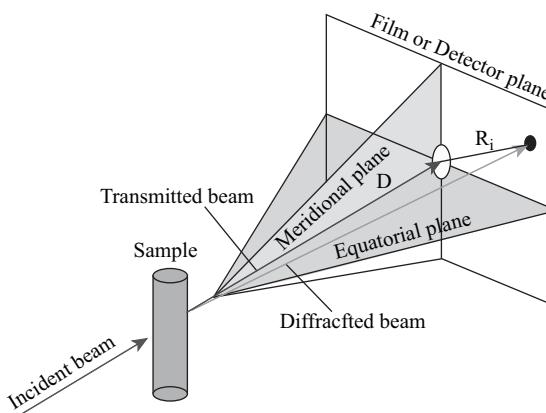


Figure 2.28 Diffraction geometry of wide angle X-ray diffraction. (WAXD)

The Bragg angle θ_i is determined by a simple geometric relationship.

$$\theta_i = \frac{1}{2} \tan^{-1}(R_i/D) \quad (2.17)$$

where R_i is the distance between the transmitted beam and the diffracted beam on the recording film plane as shown in Figure 2.28. The d -spacing of the crystallographic plane generating diffraction is calculated according to Bragg's Law.

$$d = \frac{\lambda}{2\sin[\{\tan^{-1}(R_i/D)\}/2]} \quad (2.18)$$

The accuracy of d -spacing measurement relies on the measurement of R and D . A reliable method for determining D is to calculate it from Equation 2.18 with a known d -spacing of crystal, not from a direct measurement. R_i is commonly obtained by measuring $2R_i$ on a film in order to avoid the problem of determining the center of transmitted beam on the film.

One of the main reasons for using WAXD for polymer analysis is that crystalline polymers are frequently presented in the form of fibers with preferential crystal orientation with respect to the fiber axis; for example, the $a-b$ crystallographic plane is perpendicular to the fiber axis that is along a vertical direction in Figure 2.28. WAXD analysis can readily reveal the preferential orientation of polymer crystals for a diffraction pattern recorded on the film. Figure 2.29 illustrates the generation of such a diffraction pattern of highly oriented polymer crystals using the Ewald sphere. The orientation of crystals with $a-b$ planes perpendicular to a fiber axis can be represented by the reciprocal lattice of the crystals with c^* perpendicular to the incident beam. The $a-c$ or $b-c$ planes of crystals can be randomly oriented in the fiber. Collective reciprocal lattices of the crystals with alignment of $a-b$ plane normal along the fiber axis but with random orientation of $a-c$ and $b-c$ planes are equivalent to rotating the reciprocal lattices around the c^* axis. These collective reciprocal lattices form a series of co-centered rings as shown in Figure 2.29.

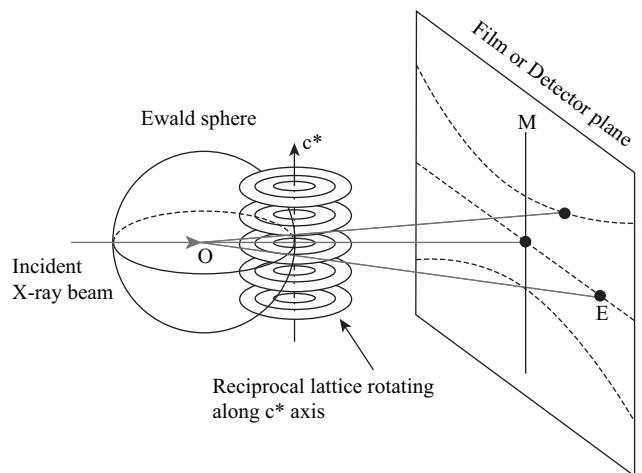


Figure 2.29 A WAXD pattern of highly oriented polymer crystals with their reciprocal lattices and the Ewald sphere.

As illustrated in Figure 2.10, the intersections between rings of reciprocal lattices and the surface of the Ewald sphere generate diffraction. Consequently, the oriented crystals in fiber generate a diffraction pattern as shown in the film plane of Figure 2.29. The diffraction spots on the equatorial plane represent the diffraction of $[hk0]$ planes. A line of diffraction spots above or below the equatorial plane should represent the diffraction of $(hk1)$ planes or $(hk\bar{1})$ planes, respectively. The diffraction principles of oriented crystals discussed here apply to all type of samples through which X-rays can be transmitted, even though the main applications of WAXD are for crystalline polymers and proteins (special types of polymers). The most famous application to proteins is probably the determination of the DNA double helix structure.

Figure 2.30 shows an example of a WAXD pattern of Kevlar fibers. The fiber axis is perpendicular to the incident X-ray beam and aligned in the meridional plane. The $[hkl]$ diffraction

Image Not Available In The Electronic Edition

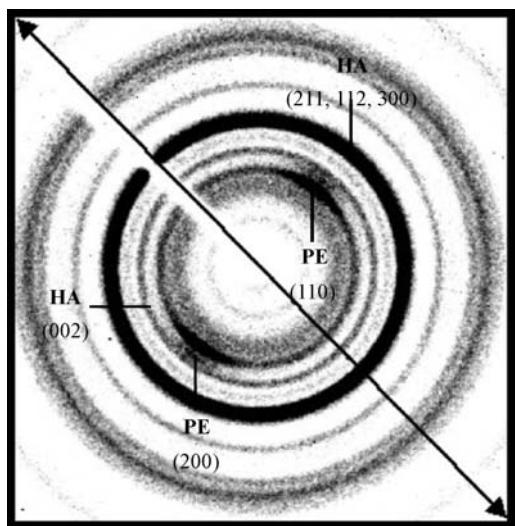


Figure 2.31 A WAXD pattern of a composite sample containing ultra-high molecular weight polyethylene (PE) fibrils and nanoparticles of hydroxyapatite (HA). The arrow indicates the drawing direction in the composite sample.

spots are mainly located on the meridional plane with high symmetry with respect to the equatorial plane. Figure 2.31 shows another example of WAXD pattern of a composite with ultra-high molecular weight polyethylene (UHMWPE) fibrils and nano-sized ceramic particulates. The composite sample has been mechanically drawn in order to increase alignment of UHMWPE fibrils. The composite sample is perpendicular to the incident beam but the fibrils are aligned with a drawing direction marked on the pattern. The WAXD pattern revealed the preferential orientation of UHMWPE crystals and the perfectly random orientation of ceramic particles. A digital area detector provides a high-contrast WAXD pattern as shown in Figure 2.31.

2.4.2 Wide Angle Scattering

The terms ‘wide angle X-ray scattering’ and ‘wide angle X-ray diffraction’ are often used without clear distinction because there is little difference in their instrumentation. However, ‘scattering’ is a more general term than ‘diffraction’ that refers to constructive interference of scattered rays from crystallographic planes. Wide angle X-ray scattering (WAXS) is considered to be an appropriate technique for examining both crystalline and non-crystalline materials. WAXS is particularly useful for polymers because a non-crystalline (amorphous) structure is the common characteristic of polymers, even for crystalline polymers because they are not totally crystalline.

An amorphous material can generate a diffraction spectrum in a 2θ range as shown schematically in Figure 2.32. For an amorphous polymer, sharp diffraction peaks are absent from its diffraction spectrum, only showing a wide halo peak. Although there is a lack of diffraction peaks from crystallographic planes, an amorphous polymer may exhibit scattering intensity variations because of fluctuations in electron density over a certain length scale.

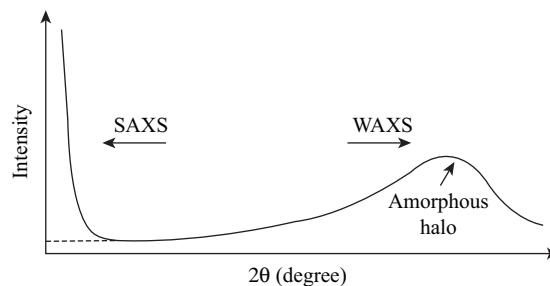


Figure 2.32 X-ray scattering by amorphous materials showing the difference in range of small angle and wide angle scattering.

The fluctuations in electron density of polymers result from heterogeneity of their mass and chemical distributions at a microscopic level. WAXS is used to detect microscopic features less than about 10 nm in length, compared with SAXS that is often used to examine microscopic features greater than about 50 nm.

WAXS is commonly presented with scattering intensity curves over a range of *scattering vector magnitude* as shown in Figure 2.33, or two-dimensional (2D) scattering intensity maps as shown in Figure 2.34. The scattering vector magnitude is equal to $4\pi(\sin\theta)\lambda^{-1}$. The scattering vector magnitude is approximately proportional to length of R_i in Figure 2.28; however, it is independent of the distance between the sample and detector plane. We may interpret a WAXS curve or map as the scattering intensity variation along the radial direction in the detector plane (Figure 2.28).

Figure 2.33 shows an example of WAXS curves for a polymer (polytetrafluoroethylene). It is a crystalline polymer at room temperature and its crystalline structure is revealed by the sharp

Image Not Available In The Electronic Edition

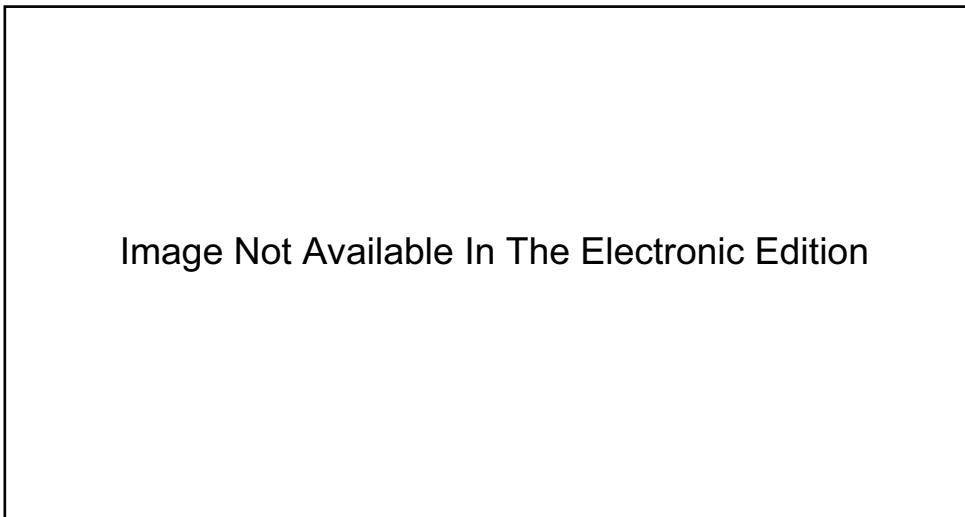


Image Not Available In The Electronic Edition

peaks as shown by the dashed line in Figure 2.33. The polymer becomes an amorphous molten phase at elevated temperature; consequently, the peaks disappear and become a broad ‘hilly’ curve as shown by the solid line. Figure 2.34 shows an example of WAXS presentation with a 2D map (Figure 2.34a) and also scattering intensity curves along two perpendicular directions (Figure 2.34b). The WAXS intensities of extruded non-crystalline polyetheretherketone (PEEK) indicate a certain degree of microstructural alignment along the extrusion direction. The alignment is revealed by the difference in the scattering intensity at about 1.3 \AA^{-1} in equatorial and meridional directions. The difference is shown in both the intensity contours in the 2D map and the intensity curves along equatorial and meridional directions.

References

- [1] Cullity, B.D. and Stock, S.R. (2001) *Elements of X-Ray Diffraction*, 3rd edition, Prentice-Hall, Inc., New Jersey.
- [2] Callister, W.D., (2000) *Materials Science and Engineering: An Introduction*, 7th edition, John Wiley & Sons.
- [3] von Heimendahl, M., (1980) *Electron Microscopy of Materials*, Academic Press, New York.
- [4] Jenkins, R. and Snyder, R.L. (1996) *Introduction to X-Ray Powder Diffractometry*, John Wiley & Sons, Inc., New York.
- [5] Lifshin, E., (1999) *X-ray Characterization of Materials*, Wiley-VCH, Verlag GmbH, Weinheim.
- [6] Eberhart, J.P. (1991) *Structural and Chemical Analysis of Materials*, John Wiley & Sons, Ltd, Chichester.
- [7] Als-Nielsen, J. and McMorrow, D. (2001) *Elements of Modern X-ray Physics*, John Wiley and Sons, Ltd, New York, 2001.
- [8] Baltá-Calleja, F.J. and Vonk, C.G. (1989) *X-ray Scattering of Synthetic Polymers*, Elsevier, Amsterdam.

Questions

- 2.1 Calculate the 2θ values in the X-ray diffraction spectrum for aluminum (FCC). Assume the X-ray radiation is $\text{CuK}\alpha_1$.
- 2.2 Calculate the 2θ values in the X-ray diffraction spectrum for iron (BCC). Assume the X-ray radiation is $\text{CuK}\alpha_1$.

- 2.3 Calculate the shifts in 2θ values in the diffraction spectrum in Question 2.1 by $\text{CuK}\alpha_2$ radiation.
- 2.4 Show that a relationship between real and reciprocal lattice is: $\mathbf{r} \cdot \mathbf{d}^* = uh + vk + wl = m$ (m is an integer).
- 2.5 Sketch a two-dimensional plane of the reciprocal lattice for a face-centered cubic (FCC) structure with a plane normal as [001].
- 2.6 Sketch a two-dimensional plane of the reciprocal lattice for a body-centered cubic (BCC) structure with a plane normal as [001].
- 2.7 Calculate the structure factor of extinction for a hexagonal close-packed (HCP) structure. There are two atoms per unit cell at (000) and at $(\frac{1}{3}, \frac{2}{3}, \frac{1}{2})$ respectively.
- 2.8 Why does the background intensity of an X-ray spectrum decrease with 2θ ?
- 2.9 Is the peak density (number of peaks per 2θ angle) higher at low 2θ or at high 2θ in X-ray diffraction spectra? Why? Use a cubic crystal as an example to interpret this phenomenon.
- 2.10 Explain possible reasons to cause discrepancies between the peak positions of an X-ray spectrum and its standard data.
- 2.11 Explain possible reasons to cause discrepancies between the peak intensities of an X-ray spectrum and its standard data.

3

Transmission Electron Microscopy

Electron microscopes generate images of material microstructures with much higher magnification and resolution than light microscopes. The high resolution of electron microscopes results from short wavelengths of the electrons used for microscope illumination. The wavelength of electrons in electron microscopes is about 10,000 times shorter than that of visible light. The resolution of electron microscopes reaches the order of 0.1 nm if lens aberrations can be minimized. Such high resolution makes electron microscopes extremely useful for revealing ultrafine details of material microstructure. There are two main types of electron microscopes: transmission electron microscopes (TEM) and scanning electron microscopes (SEM). The optics of the TEM is similar to the conventional transmission light microscope, while that of SEM is more like that of scanning confocal laser microscopes. This chapter introduces the basic working principles and features of transmission electron microscopes.

3.1 Instrumentation

A transmission electron microscope, similar to a transmission light microscope, has the following components along its optical path: light source, condenser lens, specimen stage, objective lens and projector lens as shown in Figure 3.1. The main differences are that, in a TEM, the visible light ray is replaced by an electron ray and glass lenses for visible light are replaced by electromagnetic lens for the electron beam. The TEM has more lenses (the intermediate lens) and more apertures (including the selected area aperture). The TEM contains further features arising from using electrons as illumination. For example, a vacuum environment is required in a TEM because collisions between high energy electrons and air molecules significantly absorb electron energy. Table 3.1 provides a comparison of a TEM with a light microscope. Structural features of the TEM are introduced in detail in the following sections.

Table 3.1 Comparison Between Light and Transmission Electron Microscopes

	Transmission light microscope	Transmission electron microscope
Specimen requirements	Polished surface and thin section ($<100\text{ }\mu\text{m}$)	Thin section ($<200\text{ nm}$)
Best resolution	$\sim 200\text{ nm}$	$\sim 0.2\text{ nm}$
Magnification range	2–2000 \times	500–1,000,000 \times
Source of illumination	Visible light	High-speed electrons
Lens	Glass	Electromagnetic
Operation environment	Air or liquid	High vacuum
Image formation for viewing	By eye	On phosphorescent plate

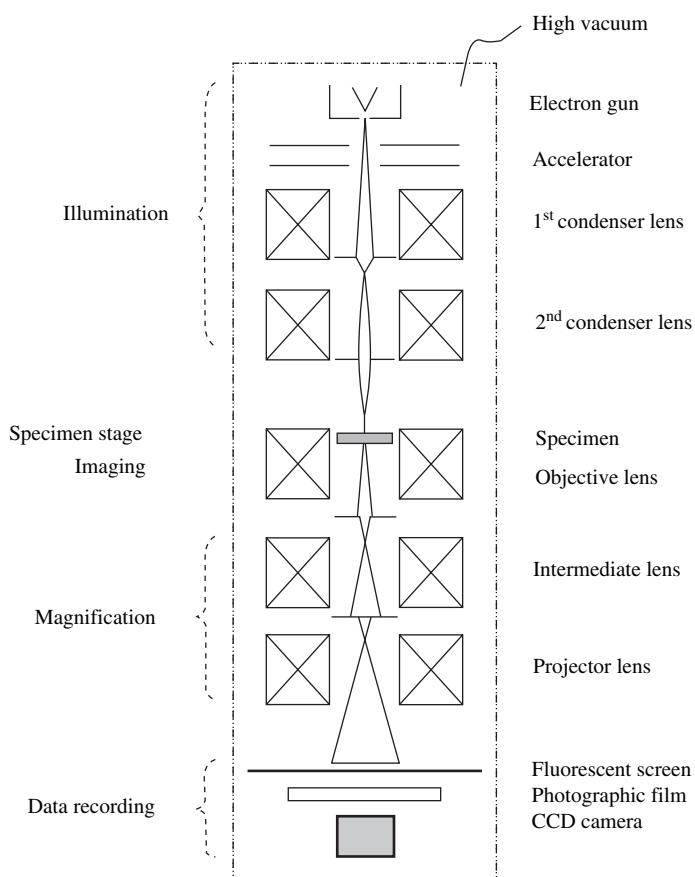
**Figure 3.1** Structure of a transmission electron microscope and the optical path.

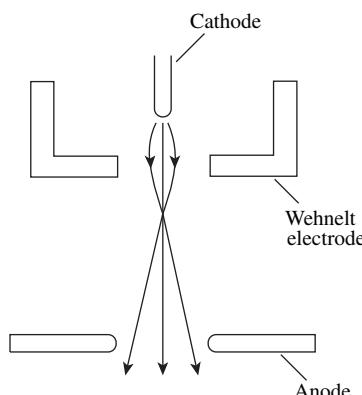
Table 3.2 Correlation between Acceleration Voltage and Resolution

Acceleration voltage (kV)	Electron wavelength (nm)	TEM resolution (nm)
40	0.00601	0.56
60	0.00487	0.46
80	0.00418	0.39
100	0.00370	0.35
200	0.00251	0.24
500	0.00142	0.13

3.1.1 Electron Sources

In a TEM system, an electron gun generates a high energy electron beam for illumination. In the electron gun, the electrons emitted from a cathode, a solid surface, are accelerated by high voltage (V_o) to form a high energy electron beam with energy $E = eV_o$. Because electron energy determines the wavelength of electrons and wavelength largely determines resolution of the microscope (Equation 1.3), the acceleration voltage determines the resolution to a large extent. Table 3.2 gives the relationship between the acceleration voltage and wavelength. It also provides the correlation between acceleration voltage and the TEM resolution, which is calculated with a realistic α angle of about 6.5×10^{-3} radians. To achieve a high resolution, the TEM is usually operated under an acceleration voltage of greater than 100 kV. In practice, 200 kV is commonly used and meets most resolution requirements. High-voltage electron microscopy (~ 1000 kV) requires extremely expensive instrumentation, and also it may damage a specimen by generating microstructural defects in it during observation.

The general structure of an electron gun is composed of three main parts: cathode or electron source, Wehnelt electrode and anode, as illustrated in Figure 3.2. Electrons are emitted from the surface of the cathode and accelerated by an electric field toward the cathode. The Wehnelt electrode is placed between the cathode and the anode. It is biased a few hundred volts negative with respect to the cathode in order to stabilize the electron beam against voltage fluctuation by

**Figure 3.2** General structure of an electron gun.

reducing the electron beam current whenever necessary. There are two basic types of electron guns: *thermionic emission* and *field emission*.

Thermionic Emission Gun

This is the most common type of electron gun and includes the *tungsten filament gun* and the *lanthanum hexaboride gun*. The tungsten gun uses a tungsten filament as the cathode. During operation, the filament is heated by electrical current (*filament current*) to high temperature (~ 2800 K). The high temperature provides kinetic energy for electrons to overcome the surface energy barrier, enabling the conduction electrons in the filament to leave the surface. The electrons leaving the surface are accelerated by a high electric voltage between the filament and anode to a high energy level. The intensity of the electron beam is determined by the filament temperature and acceleration voltage.

Lanthanum hexaboride (LaB_6) is a better cathode than tungsten and it is more widely used in modern electron microscopes. Electrons require less kinetic energy to escape from the LaB_6 surface because its surface work function is about 2 eV, smaller than the 4.5 eV of tungsten. Thus, a lower cathode temperature is required. A LaB_6 cathode can generate a higher intensity electron beam and has a longer life than a tungsten filament. One disadvantage is that a higher level of vacuum is required for a LaB_6 gun than for a tungsten filament gun.

Field Emission Gun

This type of electron gun does not need to provide thermal energy for electrons to overcome the surface potential barrier of electrons. Electrons are pulled out by applying a very high electric field to a metal surface. A tunneling effect, instead of a thermionic effect, makes metal electrons in the conducting band cross the surface barrier. During operation, electrons are physically drawn off from a very sharp tip of tungsten crystal (~ 100 nm tip radius) by an applied voltage field. The field emission gun generates the highest intensity electron beam, 10^4 times greater than that of the tungsten filament gun and 10^2 times greater than that of the LaB_6 gun. There are two kinds of field emission guns: thermal and cold guns. The thermal field emission gun works at an elevated temperature of about 1600–1800 K, lower than that of thermionic emission. It provides a stable emission current with less emission noise. The cold field emission gun works at room temperature. It provides a very narrow energy spread (about 0.3–0.5 eV), but it is very sensitive to ions in air. The bombardment of the tungsten tip by residual gaseous ions causes instability of the emission. Thus, regular maintenance (called flashing) of the tip is necessary. Generally, a vacuum of 10^{-10} torr is required to operate the cold field emission gun. Table 3.3 shows a comparison of various kinds of electron guns.

Table 3.3 Comparison of Electron Guns

	Tungsten filament	LaB_6	Field emission	
	Thermal		Cold	
Operation temperature (K)	~ 2800	~ 1800	$1600 \sim 1800$	~ 300
Brightness ^a At 200 kV (A cm ⁻² sr)	$\sim 5 \times 10^5$	$\sim 5 \times 10^6$	$\sim 5 \times 10^8$	$\sim 5 \times 10^8$
Requirement to vacuum (Torr ^b)	10^{-4}	$10^{-6} \sim 10^{-7}$	10^{-9}	$10^{-9} \sim 10^{-10}$

^a Intensity emitted per unit cathode surface in unit solid angle.

^b 1 torr = 133 Pa.

3.1.2 Electromagnetic Lenses

Lenses in light microscopes are made of glass; however, glass lenses cannot be used in electron microscopes because glass does not deflect or focus an electron beam. Noting that electrons have electric charges and that electric charges interact with magnetic fields, we can use a magnetic force to deflect or focus an electron beam. All lenses in an electron microscope are electromagnetic. Figure 3.3 illustrates the basic structure of an electromagnetic lens. The lens consists of a case and two poles. The case is made of soft-magnetic material, and encloses a solenoid. The poles are two pieces located at the narrow annular opening of the case; they are machined to very precise tolerances.

Such an arrangement creates a powerful magnetic field concentrated in the gap between the two pole pieces (N and S). Thus, when an electron beam passes through the lens, it is focused and deflected by the field's magnetic lines of force. Since the strength of a magnetic field can be easily manipulated by altering the electric current that generates magnetic field, an electromagnetic lens has a special advantage over a glass lens: the magnification power of an electromagnetic lens can be altered by simply changing the current passing through the solenoid of the lens.

The lens system of a TEM is more complicated than a light microscope, besides the electromagnetic lenses. There are two or more condenser lenses to demagnify the electron beam emitted from the electron gun. The condenser lens system controls the beam diameter and convergence angles of the beam incident on a specimen. During operation, the illumination area of specimen and illumination intensity can be adjusted by controlling the current in the electromagnetic lenses of condensers. The TEM has three lenses to ensure magnification capability of about 10^5 – 10^6 times: objective, intermediate and projector lenses. The intermediate lens is used to switch the TEM between an image mode and a diffraction mode that will be discussed in Section 3.4. For the image mode, the intermediate lens is focused on the image plane of the objective lens, and for the diffraction mode it is focused on the back-focal plane

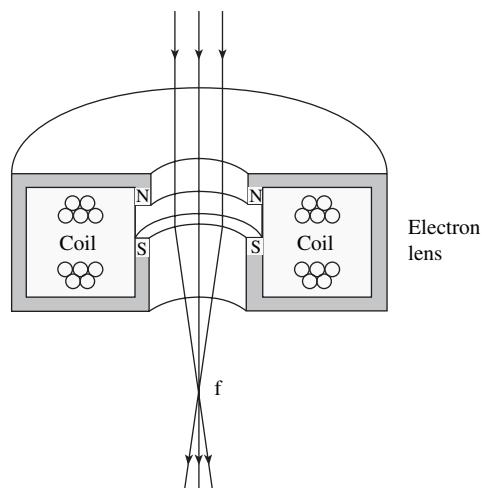


Figure 3.3 Structure of an electromagnetic lens. The magnetic field concentrated between the N and S poles deflects the electron beam.

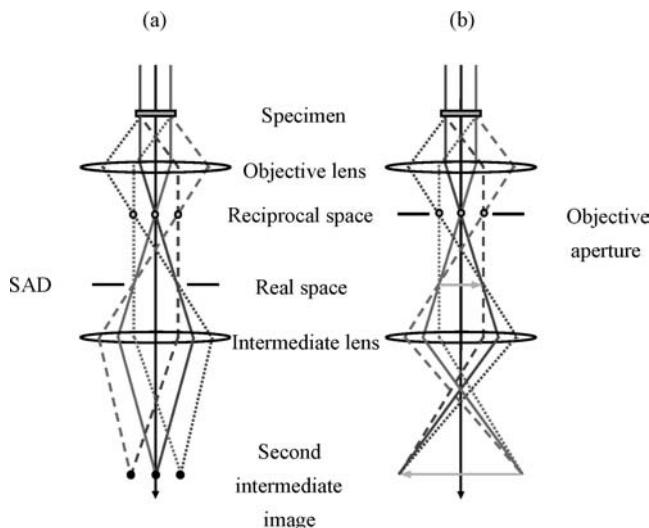


Figure 3.4 Optical paths of: (a) diffraction mode; and (b) image mode. SAD, selected area diffraction aperture.

of the objective lens where the diffraction pattern forms. Figure 3.4 schematically illustrates the switch between optical paths for image and diffraction modes by using the intermediate lens. The projector lens further magnifies the image or diffraction pattern and projects it onto the fluorescent screen for observation or the photographic plane.

The aperture of an electron microscope is a piece of metal with a small hole in the center, similar to that in a light microscope. The primary functions of the aperture are to limit light scattering and/or to select the non-diffracted or diffracted beams. The objective aperture in a TEM is located in the gap between the pole pieces of the electromagnetic lens, as illustrated in Figure 3.5. The hole diameter of the objective aperture in the TEM is only in the order of tens of micrometers.

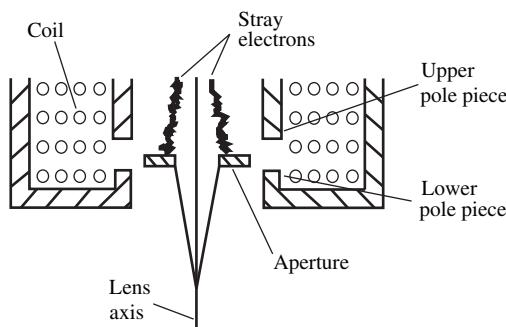


Figure 3.5 The objective aperture located in the electromagnetic lens.

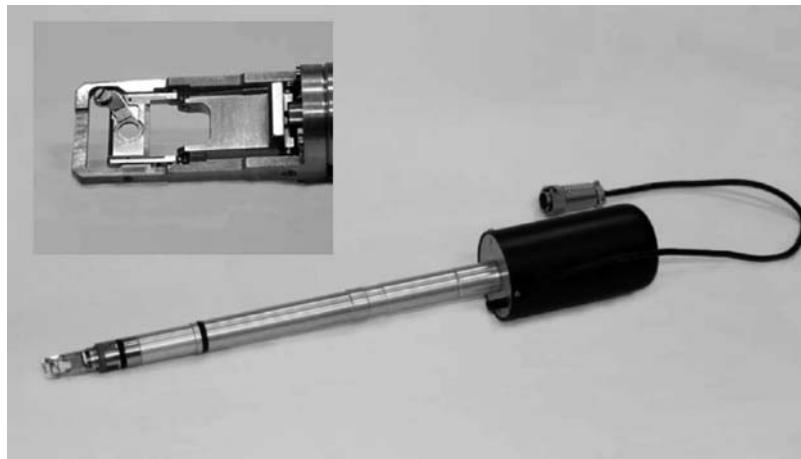


Figure 3.6 A specimen holder for a TEM (JEM-2010F) with a double-tilting function. The insert shows the enlarged part of the holder end where the 3-mm disc of the specimen is installed.

3.1.3 Specimen Stage

The TEM has special requirements for specimens to be examined; it does not have the same flexibility in this regard as light microscopy. TEM specimens must be a thin foil because they should be able to transmit electrons; that is, they are electronically transparent. A thin specimen is mounted in a specimen holder, as shown in Figure 3.6, in order to be inserted into the TEM column for observation. The holder requires that a specimen is a 3-mm disc. Smaller specimens can be mounted on a 3-mm mesh disc as illustrated in Figure 3.7. The meshes, usually made from copper, prevent the specimens from falling into the TEM vacuum column. Also, a copper mesh can be coated with a thin film of amorphous carbon in order to hold specimen pieces even smaller than the mesh size.

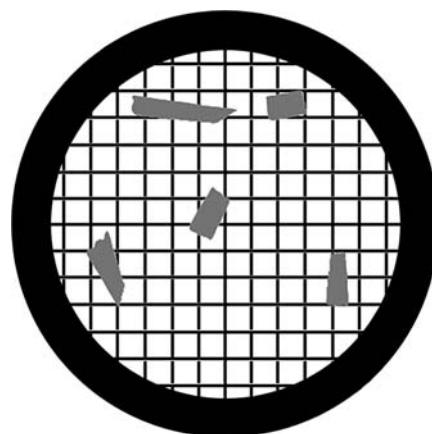


Figure 3.7 A metal mesh disc supporting small foil pieces of specimen.

The specimen holder is a sophisticated, rod-like device that not only holds the specimen but also is able to tilt it for better viewing inside the TEM column. Types of specimen holders used include single-tilt and double-tilt. The single-tilt holder allows a specimen to be tilted along one axis in the disc plane. The double-tilt holder allows the specimen be tilted independently along two axes in the disc plane. The double-tilt holder is necessary for analytical studies of crystalline structures.

3.2 Specimen Preparation

Preparation of specimens often is the most tedious step in TEM examination. To be electronically transparent, the material thickness is limited. We have to prepare a specimen with at least part of its thickness at about 100 nm, depending on the atomic weight of specimen materials. For higher atomic weight material, the specimen should be thinner. A common procedure for TEM specimen preparation is described as follows.

3.2.1 Pre-Thinning

Pre-thinning is the process of reducing the specimen thickness to about 0.1 mm before final thinning to 100 nm thickness. First, a specimen less than 1 mm thick is prepared. This is usually done by mechanical means, such as cutting with a diamond saw. Then, a 3-mm-diameter disc is cut with a specially designed punch before further reduction of thickness. Grinding is the most commonly used technique to reduce the thickness of metal and ceramic specimens. During grinding, we should reduce thickness by grinding both sides of a disc, ensuring the planes are parallel. This task would be difficult without special tool. Figure 3.8 shows a hand-grinding jig for pre-thinning. The disc (S) is glued to the central post and a guard ring (G) guides the grinding thickness. A dimple grinder is also used for pre-thinning, particularly for specimens that are finally thinned by ion milling, which is introduced in Section 3.2.2.



Figure 3.8 Hand-grinding jig for TEM specimen preparation. A disc (S) is glued to the center of the jig and the guide ring (G) controls the amount of grinding.

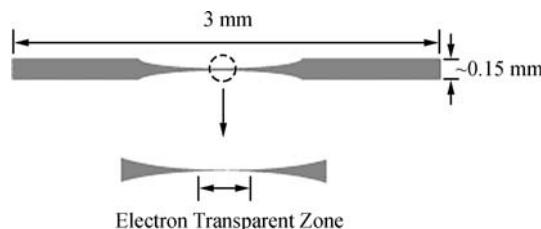


Figure 3.9 A dimple in the center of a specimen disc.

3.2.2 Final Thinning

Three methods of final thinning are described here: electrolytic thinning for metal specimens that are good electric conductors, ion milling for metal and ceramic specimens, and ultra-microtome cutting for polymeric and biological specimens.

Electrolytic Thinning

Electrolytic thinning and ion milling are methods for reducing specimen thickness to the scale of 100 nm. These methods create a dimpled area on pre-thinned specimens because it is almost impossible to reduce the thickness of specimens uniformly to the level of electron transparency. The dimpled area is likely to have regions of electron transparency as schematically shown in Figure 3.9. Electrolytic thinning is widely used for preparing specimens of conducting materials. A specimen is placed in an electrochemical cell with the specimen as anode. A suitable electrolyte (usually an acidic solution) is used to electrochemically reduce specimen thickness. A common technique is jet polishing, illustrated schematically in Figure 3.10. A specimen is placed in a holder with two sides facing guides of electrolyte jet streams, which serve as cathode. The electrolyte jet polishes both sides of the specimen until light transparency is detected by a light detector. Catching the precise moment that tiny holes appear is crucial, because only the edge of a tiny hole contains thin sections of electron transparency. Electrolytic thinning is very efficient and is completed in only 3–15 minutes.

Ion Milling

Ion milling uses a beam of energetic ions to bombard specimen surfaces in order to reduce the thickness by knocking atoms out of a specimen. The specimen does not need to be electrically conductive for ion milling. Thus, the technique is suitable for metal, ceramics and

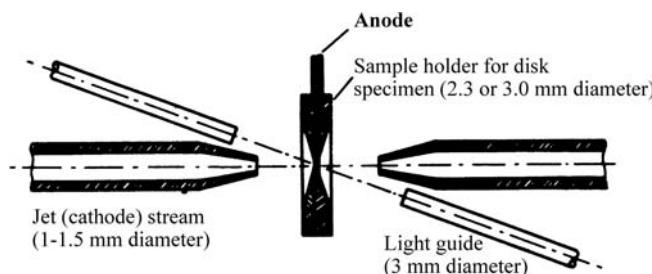


Figure 3.10 Electrochemical jet-polishing schematic diagram. (Reproduced with permission from M. von Heimendahl, *Electron Microscopy of Materials*, Academic Press, New York. © 1980 Elsevier B. V.)

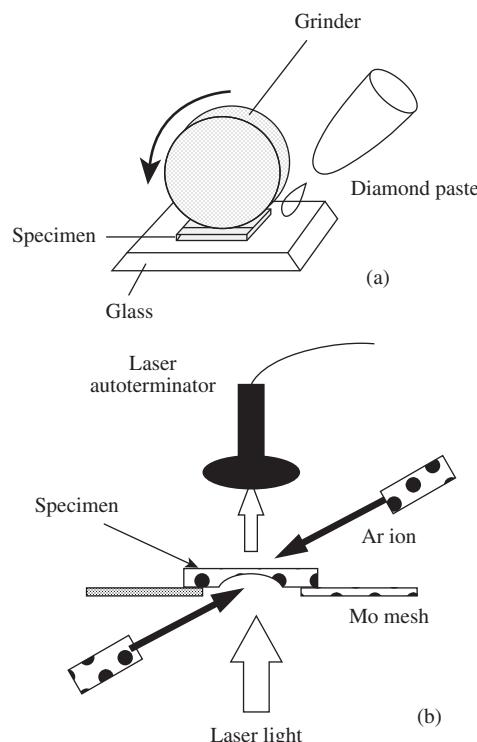


Figure 3.11 Ion thinning process: (a) dimple grinding; and (b) ion milling. (Reproduced with kind permission of Springer Science and Business Media from D. Shindo and T. Oikawa, *Analytical Electron Microscopy for Materials Science*, Springer-Verlag, Tokyo. © 2002 Springer-Verlag GmbH.)

other materials, as long as they are thermally stable. Figure 3.11 illustrates the general procedure of ion milling. Before ion milling, the specimen is often ground with a dimple grinding device in order to reduce the thickness in the central area of specimen (Figure 3.11a). Then, the ground specimen is cut as a 3-mm disc and placed in the ion milling chamber with a geometric arrangement as shown in Figure 3.11b. The edge of the specimen disc may need to be covered by a molybdenum ring in order to strengthen it during ion milling. An ion beam with energy of 1–10 keV bombards the specimen. The specimen is placed in the center at an angle of about 5–30° to the ion beam in order to have a high yield of sputtering. Light transparency is detected by a light detector aligned along the vertical direction. The ion beam can raise the temperature of the specimen rapidly and dramatically. Thus, cooling the chamber with liquid nitrogen is necessary in order to prevent specimen heating by the ion beam. Even with cooling, organic materials are not suitable for ion milling because the ion beam can cause severe damage to the microstructure of organic specimens such as polymers.

Ultramicrotomy

Ultramicrotomy is basically the same type of method as microtomy for preparing soft specimens for light microscopy. However, ultramicrotomy can be used to section a specimen to the 100 nm scale. It is commonly used to prepare polymeric or biological TEM specimens.

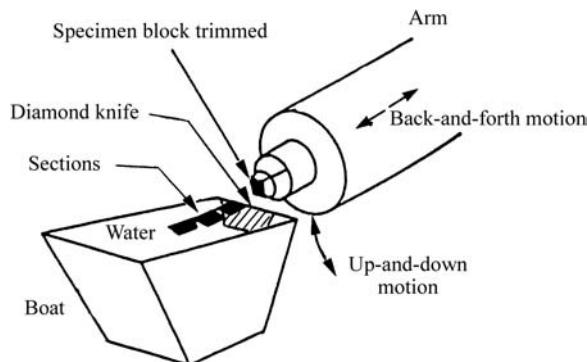


Figure 3.12 Schematic of ultramicrotomy. (Reproduced with kind permission of Springer Science and Business Media from D. Shindo and T. Oikawa, *Analytical Electron Microscopy for Materials Science*, Springer-Verlag, Tokyo. © 2002 Springer-Verlag GmbH.)

Figure 3.12 illustrates the working principles of the instrument. A specimen is mounted in a holder against the cutting tool (glass knife or diamond knife). The specimen should be trimmed to have a tip held against the knife. The cross-section of the trimmed tip usually is only about 1 mm^2 for diamond knife cutting. The holder gradually moves toward the knife while it repeatedly moves up and down. The firmly mounted specimen is sectioned as it passes the edge of the knife blade. To be sectioned by an ultramicrotome, materials must be softer than the knife. Even though diamond is the hardest material available, a specimen can still damage the tip of a diamond knife because the extremely sharp tip of a diamond knife makes it very fragile.

The aforementioned TEM specimen preparation techniques do not include all those that have been used. Various unique methods have been created for specific types of materials. For example, hammering has been used for very ductile metal such as gold, cleaving for materials such as mica and graphite flakes, and chemical thinning with suitable acids for nonmetallic materials. A ceramic specimen can be prepared by extracting its precipitates and separating them from a substrate with ultrasonic force in a liquid environment. This preparation technique avoids difficulties in thinning ceramic bulk materials. Success in TEM examination heavily relies on the quality of specimen preparation; thus, successful innovation in TEM specimen preparation can bring huge rewards.

3.3 Image Modes

As mentioned in Chapter 1, an observable image must display sufficient contrast (a brightness difference between two adjacent image points). Contrast in transmission light microscopy is generated by differences in light absorption in different regions of the specimen. In TEM, however, absorption of electrons plays a very minor role in image formation. TEM contrast relies instead on deflection of electrons from their primary transmission direction when they pass through the specimen. The contrast is generated when there is a difference in the number of electrons being scattered away from the transmitted beam. There are two mechanisms by which electron scattering creates images: *mass-density contrast* and *diffraction contrast*.

3.3.1 Mass-Density Contrast

The deflection of electrons can result from interaction between electrons and an atomic nucleus. Deflection of an electron by an atomic nucleus, which has much more mass than an electron, is like a collision between a particle and a wall. The particle (electron) changes its path after collision. The amount of electron scattering at any specific point in a specimen depends on the mass-density (product of density and thickness) at that point. Thus, difference in thickness and density in a specimen will generate variation in electron intensity received by an image screen in the TEM, if we only allow the transmitted beam to pass the objective aperture as shown in Figure 3.13. The deflected electron with scattering angle larger than the α angle (in the order of 0.01 radians) controlled by the objective aperture will be blocked by the aperture ring. Thus, the aperture reduces the intensity of the transmission beam as the beam passes through it. The brightness of the image is determined by the intensity of the electron beam leaving the lower surface of the specimen and passing through the objective aperture. The intensity of the transmitted beam (I_t) is the intensity of primary beam (I_o) less the intensity of beam deflected by object (I_d) in a specimen.

$$I_t = I_o - I_d \quad (3.1)$$

The contrast (C) in a TEM is defined by Equation 3.2.

$$C = (I_o - I_t)/I_o \quad (3.2)$$

This type of mass–density contrast exists in all types of materials. It is the main mechanism for image formation for non-crystalline materials such as amorphous polymers and biological

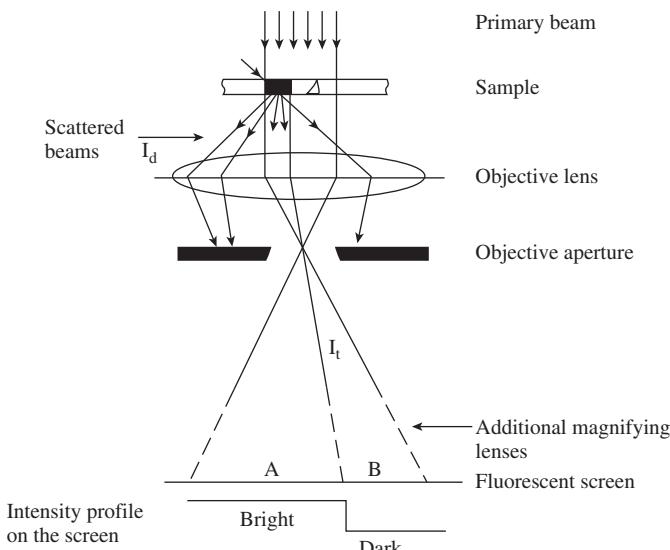


Figure 3.13 Optical arrangement of mass-density contrast in the TEM. (Reproduced with permission from M. von Heimendahl, *Electron Microscopy of Materials*, Academic Press, New York. © 1980 Elsevier B. V.)

specimens. The mass-density contrast is also described as ‘absorption contrast’, which is rather misleading because there is little electron absorption by the specimen.

In order to increase contrast, polymeric and biological specimens often are stained by solutions containing high density substances such as heavy metal oxides. The ions of metal oxides in solution selectively penetrate certain constituents of the microstructure, and thus increase its mass-density. Figure 3.14 is an example of a TEM image showing mass-density contrast in a specimen of polymer blend containing polycarbonate (PC) and polybutylene terephthalate (PBT). The PC phase appears darker because it is rich in staining substance, RuO₄. Besides staining, common effective techniques to increase contrast include using a smaller objective aperture and using a lower acceleration voltage. The negative side of using such techniques is reduction of total brightness.

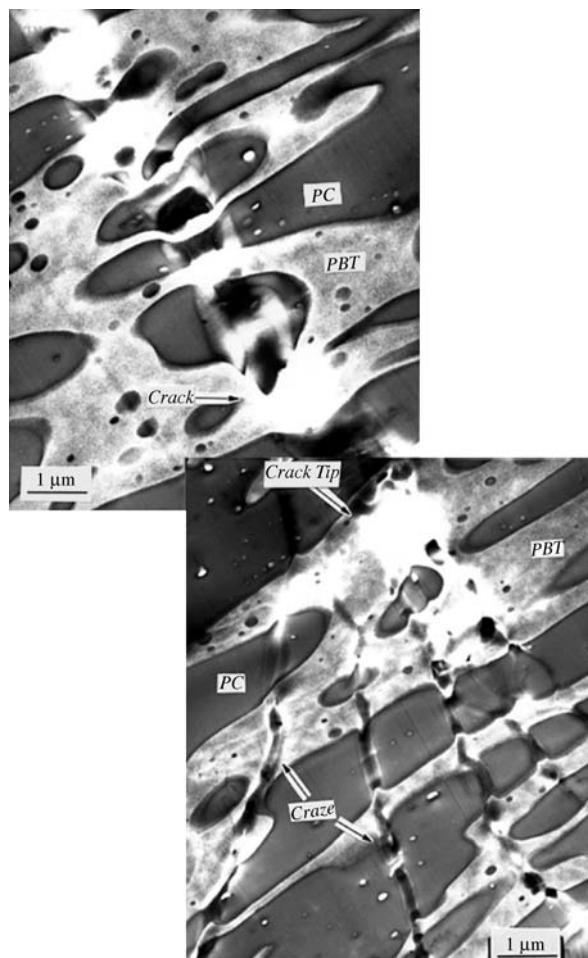


Figure 3.14 TEM image of PC–PBT polymer blends with mass-density contrast. The specimen is stained with RuO₄. (Reproduced with permission of Jingshen Wu.)

3.3.2 Diffraction Contrast

We can also generate contrast in the TEM by a diffraction method. Diffraction contrast is the primary mechanism of TEM image formation in crystalline specimens. Diffraction can be regarded as collective deflection of electrons. Electrons can be scattered collaboratively by parallel crystal planes similar to X-rays. Bragg's Law (Equation 2.1), which applies to X-ray diffraction, also applies to electron diffraction. When the Bragg conditions are satisfied at certain angles between electron beams and crystal orientation, constructive diffraction occurs and strong electron deflection in a specimen results, as schematically illustrated in Figure 3.15. Thus, the intensity of the transmitted beam is reduced when the objective aperture

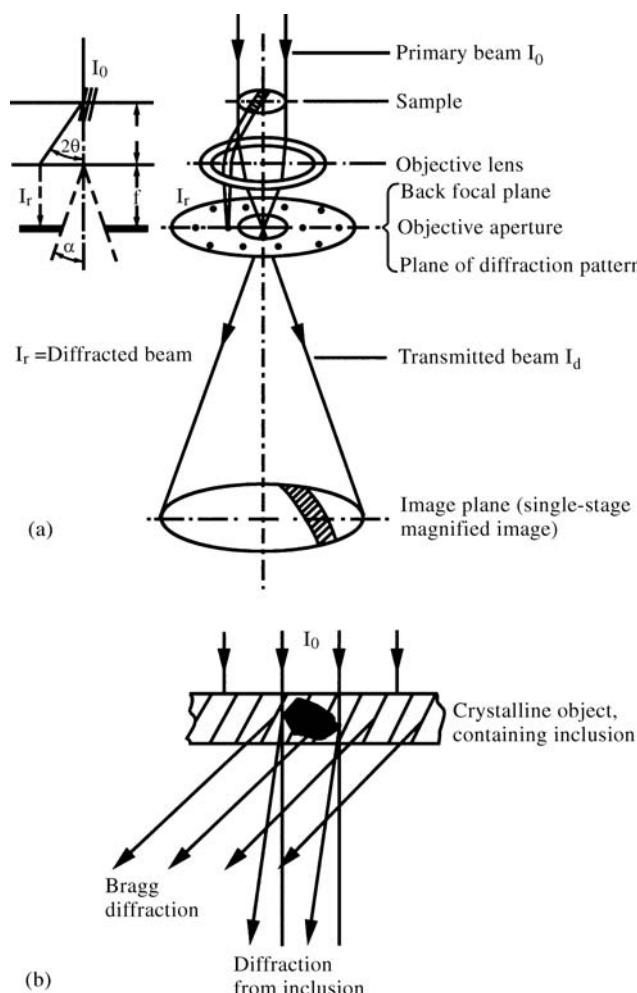


Figure 3.15 Electron deflection by Bragg diffraction of a crystalline specimen: (a) image formation in crystalline samples; and (b) diffraction at crystal lattice planes and at the contours of inclusions. (Reproduced with permission from M. von Heimendahl, *Electron Microscopy of Materials*, Academic Press, New York. © 1980 Elsevier B. V.)

blocks the diffraction beams, similar to the situation of mass-density contrast. Note that the main difference between the two contrasts is that the diffraction contrast is very sensitive to specimen tilting in the specimen holder but mass-density contrast is only sensitive to total mass in thickness per surface area.

The diffraction angle (2θ) in a TEM is very small ($\leq 1^\circ$) and the diffracted beam from a crystallographic plane (hkl) can be focused as a single spot on the back-focal plane of the objective lens. The Ewald sphere is particularly useful for interpreting electron diffraction in the TEM. When the transmitted beam is parallel to a crystallographic axis, all the diffraction points from the same crystal zone will form a diffraction pattern (a reciprocal lattice) on the back-focal plane as shown in Figures 2.9 and 2.11.

The diffraction contrast can generate bright-field and dark-field TEM images. In order to understand the formation of bright-field and dark-field images, the diffraction mode in a TEM must be mentioned. A TEM can be operated in two modes: the image mode and the diffraction mode. We can switch between the modes by changing the optical path in the TEM as illustrated in Figure 3.4. In the image mode, the image of the specimen is focused and projected onto a fluorescent screen. In the diffraction mode, the diffraction pattern formed on the back-focal plane is projected onto a fluorescent screen as shown in Figure 3.16. The TEM diffraction pattern shown in Figure 3.16 exhibits a plane of the reciprocal lattice of crystal (Section 2.1.1). Such a pattern is obtained by diffraction from a selected area in the specimen that contains a single crystal. The diffraction mode is designed for selected-area diffraction (SAD), which is discussed further in Section 3.4.

A bright-field image is obtained by allowing only the transmitted beam to pass the objective aperture, exactly the same as in mass-density contrast. A dark-field image, however, is obtained by allowing one diffraction beam to pass the objective aperture. Figure 3.17 illustrates the optical arrangement of bright- and dark-field contrast. To switch between bright- and dark-field modes, we need to operate in the diffraction mode first in order to select one diffraction spot in the diffraction pattern by moving the aperture to the center of that spot (Figure 3.18). However, what really happens in the TEM is that the direction of the incident beam is moved, rather than the aperture, to shift the selected diffraction spot to the central axis (Figure 3.17c).

Figure 3.19 shows an example of a bright-field image of a polycrystalline Al–Cu alloy with grain boundaries. The darker appearance of certain grains results from the strong diffraction from their crystallographic planes. When we tilt the specimen we change the diffraction angle, and the darker grains will become brighter and the brighter grains will become darker.

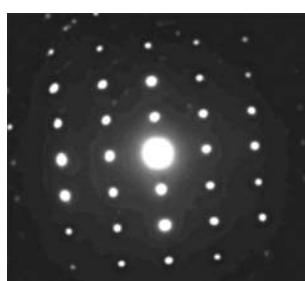


Figure 3.16 Photograph of a typical electron diffraction pattern of single crystal. The transmitted spot at the center is much brighter than the diffraction spots.

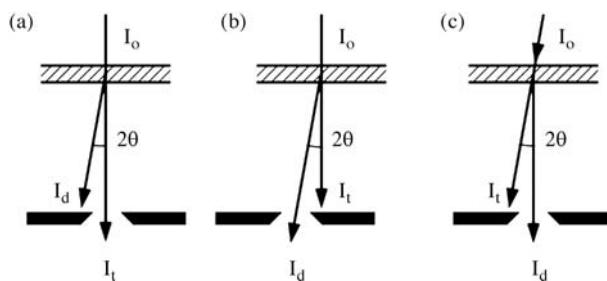


Figure 3.17 Arrangement for: (a) bright-field image; and (b, c) dark-field image.

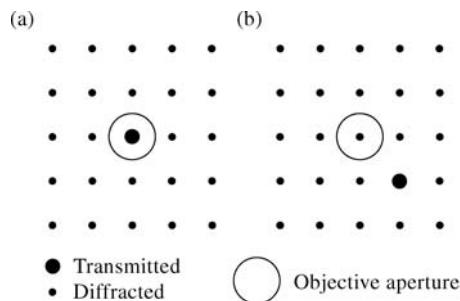


Figure 3.18 Selection of a diffraction spot with an objective aperture for dark-field imaging.

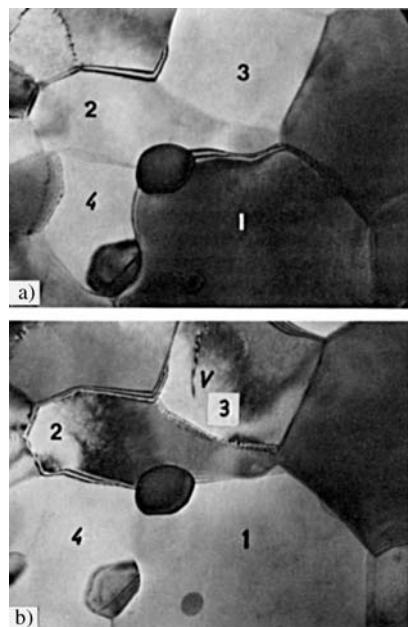


Figure 3.19 Bright-field images of aluminum alloy. The contrast difference between (a) and (b) is generated by tilting the specimen. Individual grains are marked with numbers. (Reproduced with permission from M. von Heimendahl, *Electron Microscopy of Materials*, Academic Press, New York. © 1980 Elsevier B. V.)

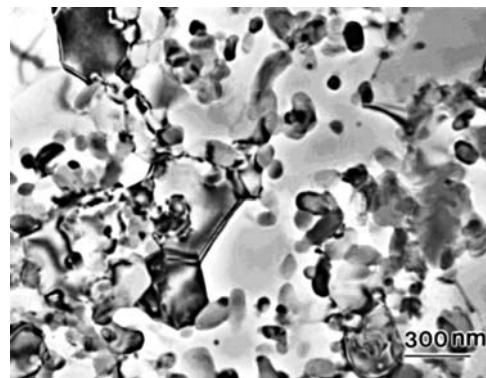


Figure 3.20 Bright-field image of fine grained Al–Fe–Si alloy. Only a few grains are visible at certain specimen orientations.

Diffraction contrast is very sensitive to specimen tilting. Even a tilt of less than 1° can cause the contrast to change from bright to dark. On the other hand, note that the grains are not always obvious in diffraction contrast. At certain diffraction angles, only a few grains might satisfy the Bragg conditions and be revealed (Figure 3.20).

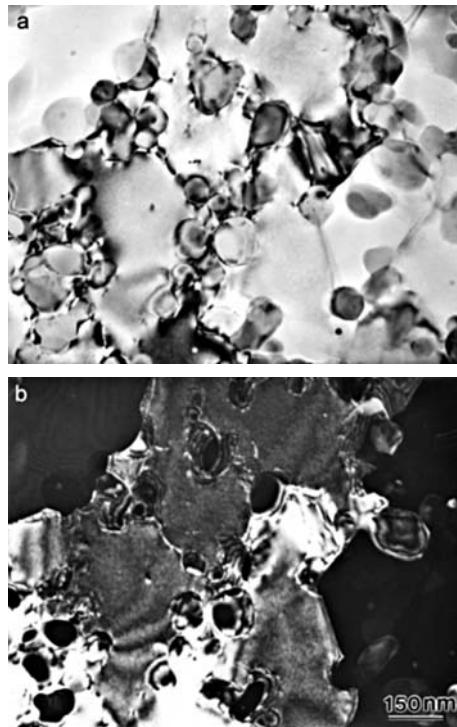


Figure 3.21 Comparison of bright- and dark-field image of Al–Fe–Si alloy: (a) bright-field; and (b) dark-field images.

Dark-field is not as commonly used as bright-field imaging. However, the dark-field image may reveal more fine features that are bright when the background becomes dark. Figure 3.21 shows a comparison between bright- and dark-field images for the same area in an Al–Fe–Si alloy. The dark-field image (Figure 3.21b) reveals finer details inside the grains. Interestingly, there are several dark-field images corresponding to a single bright-field image because we can select different diffraction spots (hkl) in Figure 3.18 to obtain a dark-field image. This is different from light microscopy in which there is a one-to-one correlation between bright- and dark-field images. When tilting a specimen to generate a high intensity diffraction spot (hkl), we may choose the spot of (hkl), which must exhibit low intensity in this circumstance, to form a dark-field image. This type of dark-field image is called the *weak-beam* dark-field image. Weak-beam imaging is useful for increasing the resolution of fine features such as dislocations in crystalline materials.

3.3.3 Phase Contrast

Both mass-density and diffraction contrasts are amplitude contrast because they use only the amplitude change of transmitted electron waves. Chapter 1 mentions that phase change generates contrast in light microscopy. Similarly, a TEM can also use phase difference in electron waves to generate contrast. The phase contrast mechanism, however, is much more complicated than that of light microscopy. The TEM phase contrast produces the highest resolution of lattice and structure images for crystalline materials. Thus, phase contrast is often referred as to *high resolution transmission electron microscopy* (HRTEM).

Phase contrast must involve at least two electron waves that are different in wave phase. Thus, we should allow at least two beams (the transmitted beam and a diffraction beam) to participate in image formation in a TEM. Formation of phase contrast in a crystalline specimen is illustrated simply in Figure 3.22. A crystalline specimen with a periodic lattice structure generates a phase difference between the transmitted and diffracted beams. The objective lens generates additional phase difference among the beams. Recombination of transmitted and diffracted beams will generate an interference pattern with periodic dark–bright changes on the image plane because of beam interferences, as shown in Figure 3.23. An interference pattern is a fringe type that reveals the periodic nature of a crystal.

Theoretical interpretation of phase contrast images is complicated, particularly when more than one diffraction spot participates in image formation. The schematic illustration in Figure 3.22 does not reflect the complexity of phase contrast. The electron wave theory of phase contrast formation is explained briefly in the following section.

Theoretical Aspects

The electron wave path of phase contrast in a TEM is illustrated by Figure 3.24. The properties of electron waves exiting the specimen are represented by an *object function*, $f(\mathbf{r})$, which includes the distributions of wave amplitude and phase. \mathbf{r} is the *position vector* from the central optical axis. The object function may be represented as a function in Equation 3.3.

$$f(\mathbf{r}) = [1 - s(\mathbf{r})] \exp [i\phi(\mathbf{r})] \quad (3.3)$$

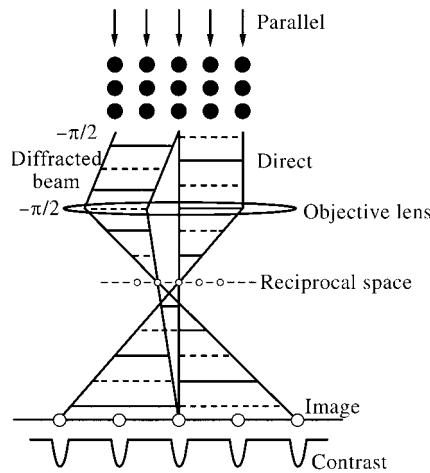


Figure 3.22 Phase contrast formation from a crystalline specimen. Phase difference is generated by crystal diffraction and an objective lens. This results in a phase difference of $-\pi$ between the direct and transmitted beams.

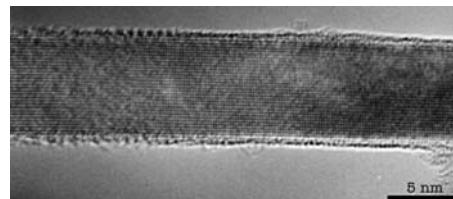


Figure 3.23 Phase contrast image of silicon nanowire. (Reproduced with permission of Ning Wang.)

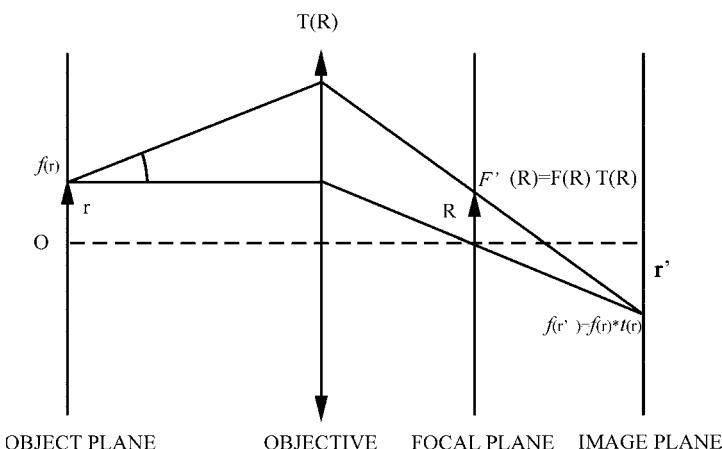


Figure 3.24 Two Fourier transformations during formation of a phase contrast image. (Reproduced with permission from J.P. Eberhart, *Structural and Chemical Analysis of Materials*, John Wiley & Sons Ltd, Chichester. © 1991 John Wiley & Sons Ltd.)

where $s(\mathbf{r})$ represents absorption by the specimen and $\phi(\mathbf{r})$ is a phase component of a wave. For a weakly scattering object (s and $\phi \ll 1$), $f(\mathbf{r})$ may be approximated.

$$f(\mathbf{r}) = 1 - s(\mathbf{r}) + i\phi(\mathbf{r}) \quad (3.4)$$

At the back-focal plane of objective lens, the diffraction pattern is considered as the *Fourier transform (FT)* of the object function.

$$F(\mathbf{R}) = FT f(\mathbf{r}) \quad (3.5)$$

where \mathbf{R} is the position vector of a diffraction spot from the central axis on the focal plane; a reciprocal vector. The Fourier transform $F(\mathbf{R})$ is defined by the following equation.

$$F(\mathbf{R}) = \int_{\infty} f(\mathbf{r}) \exp(2 \pi i \mathbf{r} \bullet \mathbf{R}) d\mathbf{r}^3 \quad (3.6)$$

in which $\mathbf{r} \bullet \mathbf{R}$ is defined.

$$\mathbf{r} \bullet \mathbf{R} = xX + yY + zZ \quad (3.7)$$

X , Y and Z are the magnitudes of reciprocal vector \mathbf{R} , corresponding to the magnitude (x , y and z) of vector \mathbf{r} in the real space. Since the electron waves pass the objective lens, the objective lens can affect the wave characteristics. The effect of the objective lens on the wave function can be represented by a *transfer function*, $T(\mathbf{R})$. Thus, the actual wave function on the back-focal plane becomes the following function.

$$F'(\mathbf{R}) = F(\mathbf{R})T(\mathbf{R}) \quad (3.8)$$

Wave theory tells us that the wave function of electrons on the image plane (fluorescent screen) should be the Fourier transform of $F'(\mathbf{R})$. For treatment simplicity, we may ignore the image magnification and rotation on the image plane. Then, the wave function on the image plane becomes $f'(\mathbf{r})$.

$$f'(\mathbf{r}) = FT\{F'(\mathbf{R})\} = FT\{F(\mathbf{R})T(\mathbf{R})\} = f(\mathbf{r})^* t(\mathbf{r}) \quad (3.9)$$

The results should be read as the *convolution product* of $f(\mathbf{r})$ by $t(\mathbf{r})$, where $t(\mathbf{r})$ is the Fourier transform of $T(\mathbf{R})$. The transfer function plays an important role in image formation. The transfer function can be written as the following equation.

$$T(\mathbf{R}) = A(\mathbf{R}) \exp[-i\gamma(\mathbf{R})] \quad (3.10)$$

where the amplitude factor (A) equals unity when the length of \mathbf{R} is less than the objective aperture radius on the focal plane; otherwise it equals zero. The *phase shift* (γ) that occurs as the waves pass objective lens should be considered. Phase difference occurs when the wave paths that reach a certain point are different. Spherical aberration causes phase shift because of the electron wave path difference between the waves near the central axis and the waves far

from the central axis of objective lens. The phase shift at R in the back-focal plane induced by spherical aberration can be expressed as the following.

$$\gamma_s(R) = \frac{\pi}{2} C_s \lambda^3 R^4 \quad (3.11)$$

C_s is the *spherical aberration coefficient*. This phase shift by spherical aberration can be significant for short wavelength waves such as electrons. Fortunately, this phase shift can be offset by phase shift caused by defocusing the image. We can defocus the image and make the focal point at the distance further from the focal plane (i.e. weaken the lens power). The phase shift from defocusing the image is a function of the *defocus distance* (D) from the focal plane.

$$\gamma_D(R) = -\pi \lambda D R^2 \quad (3.12)$$

The total phase shift in the transfer function should be the sum of two phase shifts.

$$\gamma = \gamma_s + \gamma_D = \pi \lambda \left(\frac{1}{2} C_s \lambda^2 R^4 - DR^2 \right) \quad (3.13)$$

This equation means that we can manipulate the amount of phase shift by defocusing the image. Formation of a phase contrast image requires that the wave function on the image plane satisfies the following condition.

$$f'(\mathbf{r}) = 1 + \phi(\mathbf{r}) \quad (3.14)$$

This is possible when the phase shift in the transfer function satisfies the following equation.

$$\sin \gamma \cong \pm 1 \quad (3.15)$$

This condition can be obtained in a defocused condition of the objective lens. As a result, an image of phase contrast may be interpreted in terms of periodic structures of a crystalline solid. Such an image of phase contrast is called the *interpretable structure image*.

In order to obtain an interpretable image, the objective aperture opening must be limited to a level corresponding to the optical resolution limit of the objective lens. In TEM operation, we can adjust the aperture opening on the image of the back-focal plane (Figure 3.25). Its dimension is the reciprocal of length (nm^{-1}). For example, the optical resolution limit is about 0.35 nm at an acceleration voltage of 100 kV, and about 0.25 nm at 200 kV. We should not open the aperture to include diffraction spots with length of \mathbf{R} larger than about $0.35^{-1} \text{ nm}^{-1}$ for 100 kV and about $0.25^{-1} \text{ nm}^{-1}$ for 200 kV. An aperture larger than the limit may generate artifacts in the phase contrast image.

Two-Beam and Multiple-Beam Imaging

We may use the transmitted beam and one diffraction beam from a crystallographic plane (hkl) to form a one-dimensional lattice image. This is the simplest way to obtain a set of fringes which is parallel to the crystallographic plane (hkl). Two-beam contrast can provide a lattice

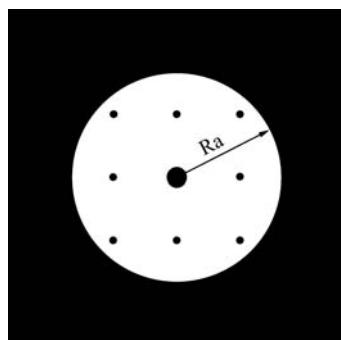


Figure 3.25 The objective aperture opening shown in the back-focal plane. Ra, diameter of the aperture.

image as long as the plane spacing is within the resolution of the TEM. Thus, we may directly measure plane spacing between two fringes. Figure 3.26 shows an example of lattice image of a carbon nanotube obtained by two-beam phase contrast.

Often, more than two beams are needed to obtain a structure image because better structure resolution can be obtained with multiple-beam imaging. The phase contrast, however, for a multiple-beam image is more complicated than for a two-beam case. Compensating the phase shift of spherical aberration with defocusing is difficult, unless the transmitted beam aligns with a highly symmetrical crystal axis, for example the [111] axis of cubic crystals. Although compensating spherical aberration with defocusing is not possible in some cases, we may still have interpretable structure resolution when the aperture opening is limited within the optical resolution.

When two or more diffraction beams participate in image formation, a two-dimensional lattice image that exhibits a lattice image of several sets of lattice planes will be observed, as shown in Figure 3.27. The bright parts of a phase contrast image generally correspond to small phase shifts in waves as they exit from a crystal. Thus, they may correspond to

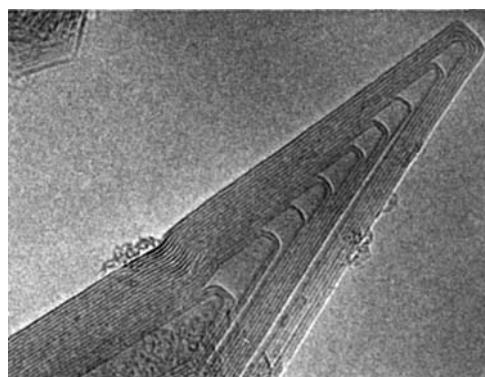


Figure 3.26 Two-beam phase contrast of a carbon nanotube at its end portion. (Reproduced with permission of Ning Wang.)

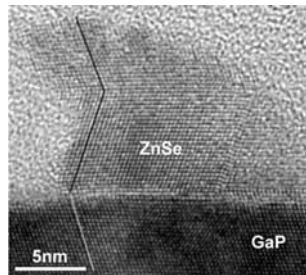


Figure 3.27 Multiple-beam lattice image of the interface between GaP and ZnSe crystals. (Reproduced with permission of Ning Wang.)

structure channels between atomic planes. Generally, neither a bright nor a dark dot in the image corresponds to an individual atom. Direct interpretation of a phase contrast image must be carefully handled. Often we have to rely on computer simulation based on phase contrast theory outlined in the previous section to interpret the structure images formed by multiple beams.

3.4 Selected Area Diffraction

A diffraction pattern is formed on the back-focal plane of the objective lens when an electron beam passes through a crystalline specimen in a TEM. In the diffraction mode, a pattern of selected area diffraction (SAD) can be further enlarged on the screen or recorded by a camera as illustrated in Figure 3.16. Electron diffraction is not only useful to generate images of diffraction contrast, but also for crystal structure analysis, similar to X-ray diffraction methods. SAD in a TEM, however, shows its special characteristics compared with X-ray diffraction, as summarized in Table 3.4. More detailed SAD characteristics are introduced in the following section.

3.4.1 Selected Area Diffraction Characteristics

Construction of an image of an SAD pattern in a TEM is illustrated in Figure 3.28. Constructive diffraction from a lattice plane (hkl) generates an intensity spot on the screen when

Table 3.4 Comparison between X-ray diffraction and SAD in TEM

	X-ray diffraction	SAD in TEM
Scattering Nature	Scattering by shell electrons	Scattering by atom nucleus
Wave paths	Reflection (XRD) Transmission (Laue)	Transmission
Diffraction angle θ	$0 \sim 180^\circ$	$0 \sim 2^\circ$
Intensity	low	$10^6 \sim 10^7$ times higher
Precision	high	Relatively low

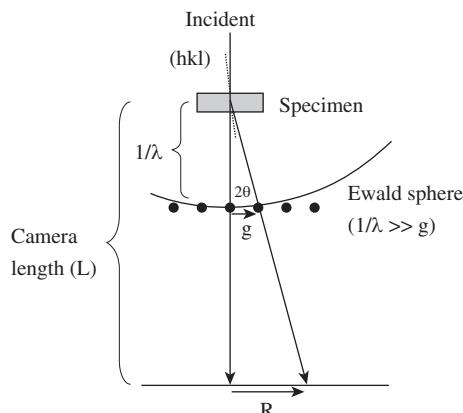


Figure 3.28 Formation of a diffraction pattern in a TEM.

the TEM is in the diffraction mode. The diffraction angle in the TEM is very small (commonly $\theta < 1^\circ$), because the reflecting lattice planes are nearly parallel to the primary beam. Using the approximation $\sin\theta \approx \theta$ for a small angle, we can rewrite Bragg's Law as the following.

$$\lambda = 2d\theta \quad (3.16)$$

Noting the geometric relation among θ , R and L demonstrated in Figure 3.28, we have the following equation.

$$\lambda L = Rd \quad (3.17)$$

where L is the *camera length* of a TEM, the distance between the crystal and photographic plate of camera. This is not a real physical distance in a TEM because the magnifying lens changes the geometry of the path between the crystal and photographic plate. In fact, L is an equivalent camera length for calculating the spacing of the lattice plane. λL is called as the *camera constant* of a TEM. Equation 3.17 is the *basic equation for electron diffraction* in a TEM. We can calculate the spacing of crystallographic planes by measuring R (the distance from the central spot of the transmitted beam to a diffraction spot) in photographic film using Equation 3.17. L can be changed in a TEM: for example, $L = 1.0$ m is often selected with 200-kV acceleration voltages. A proper L value allows us to have a reasonable R range in the image film.

It is important to note the features of diffraction pattern of a single crystal in TEM: the diffraction pattern represents a reciprocal lattice plane with reciprocal lattice points which are shown as the diffraction spots, and the reciprocal lattice plane contains the diffraction of lattice planes belonging to one crystal zone, of which the axis is parallel to the transmitted beam. These features are well illustrated in Figure 2.9 in which the zone axis is [001] of a cubic crystal and the reciprocal lattice plane consists of diffraction spots from lattice planes $(hk0)$.

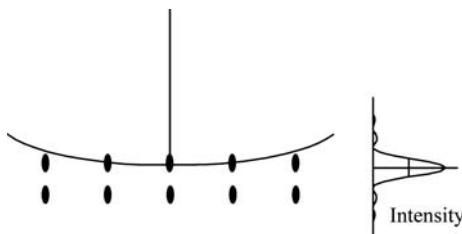


Figure 3.29 Diffraction pattern formation from a thin foil specimen in a TEM. The surface of Ewald sphere intersects the elongated diffraction spots. An elongated spot reflects the diffraction intensity distribution along the transmitted beam direction.

The Ewald sphere is very useful to explain the formation of diffraction patterns in a TEM (Figure 2.11). Recall that the short wavelength of high energy electrons makes the radius of the Ewald sphere (λ^{-1}) become so large that the spherical surface is nearly flat compared with the distances between reciprocal lattice points. On the other hand, a diffraction spot of a crystal with a small dimension in the vertical direction is not an exact reciprocal lattice point, but an elongated spot along that direction, as illustrated in Figure 3.29. The reason for spot elongation relates to diffraction intensity distribution, which can be fully explained by kinematic diffraction theory. Without more description of the theory, we may still understand it by recalling the discussion of crystal size effects on width of X-ray diffraction (XRD) diffraction peaks in Chapter 2. The width of XRD peaks increases with decreasing crystal size because a small crystal causes incomplete destructive interference. For the same reason, the thin crystal of a TEM specimen causes a widened intensity distribution and thus, the elongated diffraction spots along the vertical direction. Therefore, the flat surface of the large Ewald sphere can intersect even more diffraction spots on one reciprocal plane in a TEM as illustrated in Figure 3.29. Thus, we can see a diffraction pattern in the TEM as shown in Figure 3.16.

3.4.2 Single-Crystal Diffraction

To obtain a diffraction pattern as shown in Figure 3.16, we need to use the selected area aperture to isolate a single crystal in the image mode, and then switch to the diffraction mode. Commonly, tilting the specimen is necessary to obtain a diffraction pattern with high symmetry with respect to its central transmitted spot. A symmetrical pattern like Figure 3.16 must have a zone axis of the crystal aligned with the transmitted beam direction, as illustrated in Figure 2.9.

After obtaining a diffraction pattern, we may want to index the pattern in order to know the crystal structure or crystal orientation that the pattern represents. Indexing a pattern means to assign a set of h , k and l to each diffraction spot in the pattern. The information can be directly extracted from a pattern that includes the length of \mathbf{R} for each spot and angles between any two \mathbf{R} vectors. We can calculate the interplanar spacing (d_{hkl}) easily using Equation 3.17, with knowledge of the camera constant. Without additional information about the crystal such as chemical composition or lattice parameters, indexing a pattern could be a difficult task. However, indexing techniques for simple cases can be demonstrated as follows.

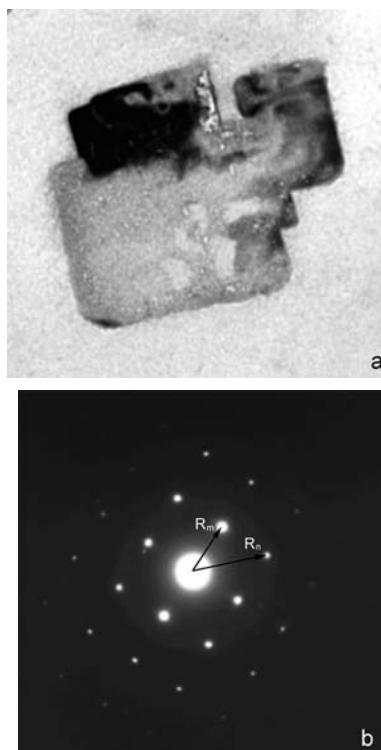


Figure 3.30 Single crystal of NaCl: (a) bright-field image; and (b) selected area diffraction pattern. R_m , R_n are the radii of spots m and n , respectively. The transmitted beam direction is parallel to [001].

Indexing a Cubic Crystal Pattern

The TEM bright-field image and an SAD pattern of sodium chloride (NaCl) are shown in Figure 3.30. In the pattern, the central spot of the transmitted beam is brightest and should be indexed as (0 0 0) of the reciprocal lattice plane. Then, we can select two spots, for example m and n in the pattern to measure their \mathbf{R} lengths and angle between them. Here, the \mathbf{R} lengths of m and n are measured as 8.92 mm and 12.6 mm, respectively; and the angle between \mathbf{R}_m and \mathbf{R}_n is measured as 45° . The indices of the spots can be determined by the following relationship for cubic crystals such as NaCl. The equation is obtained by combining Equation 3.17 and Equation 2.4.

$$R_{(hkl)} = \frac{\lambda L(h^2 + k^2 + l^2)^{1/2}}{a} \quad (3.18)$$

The wavelength of electrons (λ) for 200 kV, which is used to obtain the pattern, should be 0.00251 nm according to Table 3.2. The diffraction photograph is taken at the camera length (L) of 1.0 m. Since the lattice parameter of NaCl (a) is 0.563 nm, we can find out that the \mathbf{R}_m matches that of 200 and \mathbf{R}_n matches that of 220, according Equation 3.18. Then, we should check whether the angle ($\mathbf{R}_m \mathbf{R}_n$) matches that between specific planes. The plane angle in a

cubic crystal can be obtained using the following equation.

$$\cos \alpha = \frac{h_m h_n + k_m k_n + l_m l_n}{(h_m^2 + k_m^2 + l_m^2)(h_n^2 + k_n^2 + l_n^2)} \quad (3.19)$$

It is not difficult to find out that the measured angle α (45°) matches that of the angle between (200) and (220) or ($\bar{2}00$) and ($\bar{2}20$).

Actually, there is a 180° ambiguity in the pattern index: an identical pattern is obtained by rotating 180° around the transmitted beam. For convenience, we can index the spots of the right hand side as (200)(220). After indexing these two spots, the rest of the spots can be indexed using the rule of vector addition. The crystal zone axis can be identified by vector algebra.

$$[uvw] = \mathbf{R}_m \times \mathbf{R}_n \quad (3.20)$$

The diffraction pattern in Figure 3.30b should be indexed as illustrated in Figure 3.31. For the pattern in Figure 3.30, we can quickly determine the crystal zone axis as [001], according to Equation 3.20.

In certain cases, without knowing the lattice parameter of a crystal, we may still be able to index the pattern using the relationship between the R ratio and Miller indices. For example, if the crystal is cubic, the following holds.

$$\frac{R_m}{R_n} = \frac{(h_m^2 + k_m^2 + l_m^2)^{1/2}}{(h_n^2 + k_n^2 + l_n^2)^{1/2}} \quad (3.21)$$

In a cubic system, all the possible R ratios can be tabulated according to hkl pairs. We may find the match of R ratio for specific hkl pairs of a crystal structure. It should be remembered that diffraction must obey the structure extinction laws discussed in Chapter 2. Note that NaCl exhibits a face-centered cubic (FCC) structure for which the diffraction pattern should not include (hkl) indexes of mixed odd and even numbers. Thus, there are no diffraction spots for 100, 110... in Figure 3.30b. We can use the structure extinction laws to determine whether a

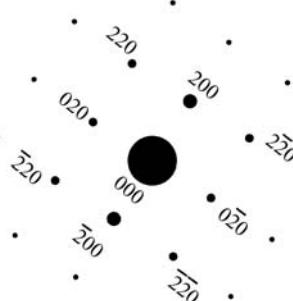


Figure 3.31 Indexing the diffraction pattern of NaCl shown in Figure 3.30b.

Table 3.5 Relationship between Real and Reciprocal Lattice Structure

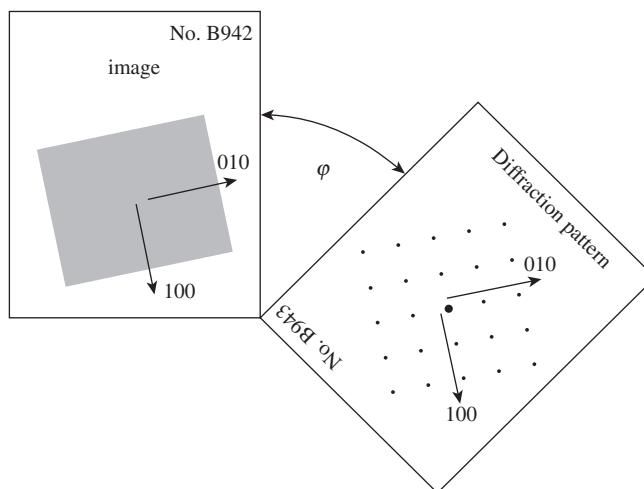
Real lattice	Real lattice parameter	Reciprocal lattice	Reciprocal lattice parameter
Simple cubic	a	Simple cubic	$a^* = \frac{1}{a}$
FCC	a	BCC	$a^* = \frac{2}{a}$
BCC	a	FCC	$a^* = \frac{2}{a}$

cubic crystal is FCC or body-centered cubic (BCC). According the structure extinction laws listed in Table 2.2, the reciprocal lattice of an FCC structure is a BCC type or vice-versa, as shown in Table 3.5.

The crystal orientation with respect to specimen geometry is important information. In Figure 3.30, the NaCl crystal orientation is already revealed by its bright field image, because its cubic shape implies its crystal structure and orientation. In most cases, crystalline materials do not show the geometrical shape representing their crystal structure and orientation. Thus, we need to determine the orientation from single-crystal diffraction patterns.

When paying attention to Figures 3.30a and 3.30b, we note that the crystal orientation of NaCl in the bright-field image does not match that of the diffraction pattern because the edge of the cube should be perpendicular to $\mathbf{R}_{(200)}$ or $\mathbf{R}_{(020)}$ in the diffraction pattern. Figure 3.30 tells us that there is image rotation of the diffraction pattern with respect to its crystal image. This rotation results from the change in magnetic field of the electromagnetic lens during switching from the image mode to the diffraction mode. The reason is that the image and the diffraction pattern are formed in different optical planes in the TEM. The image has undergone three-stage magnification, but the pattern has only undergone two-stage magnification.

The rotation angle is a function of TEM magnification, and has to be determined experimentally. For example, Figure 3.32 shows the rotation angle between a cubic single crystal and its

**Figure 3.32** The rotation angle between the image and diffraction pattern of a crystal.

diffraction pattern. Note that the angle is $\varphi + 180^\circ$, not φ , because each stage of magnification rotates the image by 180° . After calibrating the rotation angle at the given acceleration voltage, magnification and camera length, we should be able to determine the orientation of the crystal being examined. Some modern TEM systems have a function to correct the angle of image rotation automatically. Thus, manual correction is no longer needed when the images are recorded.

Identification of Crystal Phases

SAD is particularly useful for identifying the structure of tiny crystals such as crystalline precipitates. For example, three crystal structures may exist for calcium phosphates precipitated from supersaturated aqueous solutions: hydroxyapatite (HA), octacalcium phosphate (OCP) and dicalcium phosphate (DCPD). Figure 3.33 shows the precipitate images and their diffraction patterns of OCP and HA. Two diffraction patterns (Figures 3.33b and 33d) look similar, but careful examination reveals their differences. Both patterns exhibit a d -spacing of about 0.68 nm, which is calculated from the \mathbf{R}_1 length. The crystallographic data indicate that the (001) d -spacings of OCP and HA are about 0.68 nm, while DCPD does not have any d -spacing close to 0.68 nm. Thus, the precipitates should only be either OCP or HA, not DCPD. Note that the \mathbf{R}_2 length on Figure 3.33b is smaller than that on Figure 3.33d. The calculated d -spacing is 0.94 nm in Figure 3.32b and 0.82 nm in Figure 3.33d. The crystallographic data show that only OCP exhibits d -spacing of 0.94 nm in the $(\bar{1}10)$ plane. Thus, the precipitate in Figure 3.33a is likely to be OCP. In addition, the angle between $\mathbf{R}_1\mathbf{R}_2$ should be carefully measured and compared with crystallographic data. The angles between $\mathbf{R}_1\mathbf{R}_2$ in Figures 3.33b and 3.33d appear to be 90° , but careful measurement reveals the angle in Figure 3.33b is 90.3° , slightly larger 90° , while that in Figure 3.33d is exactly 90° . In fact, the angle in Figure 3.33b matches

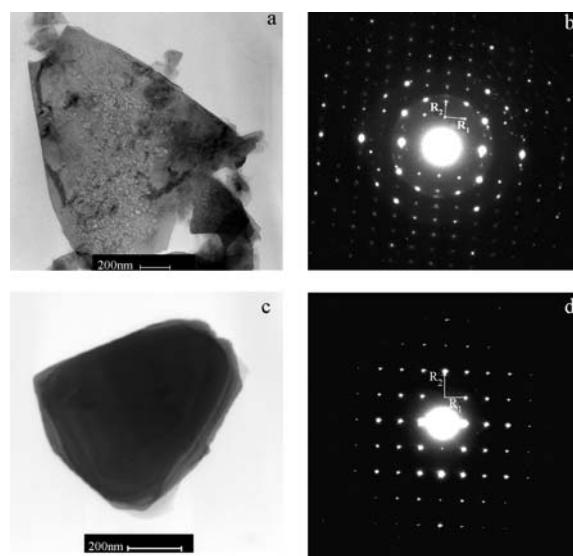


Figure 3.33 Calcium phosphate precipitates and their diffraction patterns: (a) image; and (b) diffraction pattern of octacalcium phosphate crystal; (c) image; and (d) diffraction pattern of hydroxyapatite crystal. \mathbf{R}_1 , \mathbf{R}_2 are vectors.

the crystallographic data of OCP, because the planar angle between OCP (001) and $(\bar{1}10)$ is 90.318° . Thus, we can confidently identify the crystal structure of the precipitate in Figure 3.33a as OCP, which has a triclinic structure, while the precipitate in Figure 3.33c is identified as HA with a hexagonal structure.

3.4.3 Multi-Crystal Diffraction

If an area selected for diffraction contains a large number of single crystals that are randomly oriented, a ring pattern will be obtained in SAD. Figure 3.34a shows the SAD ring pattern of a thallium chloride (TlCl) specimen composed of tiny grains. Formation of SAD ring patterns is similar to that of the X-ray Debye rings shown in Figure 2.12. The ring formation in the multi-crystalline specimen is equivalent to the rotation of the SAD diffraction of a single crystal around the axis of the transmitted electron beam, because the grains in a polycrystalline specimen may show all orientations similar to the pattern produced by single crystal rotation. However, rings in an SAD pattern are often broken as shown in Figure 3.34a, because it is not likely that a specimen contains grains in all possible orientations. The SAD ring pattern can also be formed when the selected area aperture includes many individual tiny crystals.

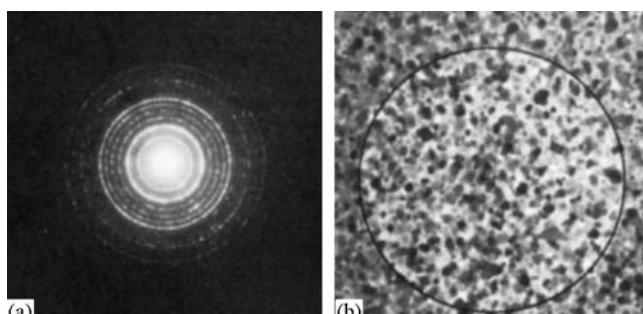


Figure 3.34 Images of polycrystalline specimen of thallium chloride: (a) SAD ring pattern; and (b) bright-field image in which the dark circle ($50 \mu\text{m}$ diameter) indicates the area selected for diffraction. (Reproduced with permission from M. von Heimendahl, *Electron Microscopy of Materials*, Academic Press, New York. © 1980 Elsevier B. V.)

3.4.4 Kikuchi Lines

Kikuchi lines are pairs of parallel lines consisting of one bright and one dark line in the diffraction mode as shown in Figure 3.35. Kikuchi lines are named after the Japanese scientist, Kikuchi, who discovered them in 1928. Kikuchi lines appear when the selected area for diffraction is moved to a thicker section in the specimen where the diffraction spots become weaker, or even disappear.

Kikuchi lines result from inelastic scattering of electrons in specimens. Generally, an electron scatters elastically when it interacts with an atomic nucleus. The mass of a nucleus is much larger than that of an electron. Thus, their interaction is similar to a ball hitting wall where the ball bounces without energy loss. However, when the electron interacts with an electron in an atomic shell, energy will transfer between the two electrons during collision, which is referred

Image Not Available In The Electronic Edition

as to inelastic scattering of electrons. A consequence of inelastic scattering is that the primary incident rays of electrons scatter from a location in a specimen as a cone of electron rays to all directions in a specimen as illustrated in Figure 3.36. The cone of electron rays becomes a secondary electron source with a maximum intensity in the direction of the primary beam. The intensity decreases with increasing angle between scattering and the primary ray direction. The formation of Kikuchi lines can be explained by the nature of this secondary electron source radiation in a specimen.

Figure 3.37 illustrates inelastic scattering at point P between two (hkl) planes inside a crystalline specimen. Inelastic scattering at P generates a cone of rays as a secondary source of incident electrons that form background intensity as shown in the lower part of Figure 3.37. Since the electron rays from this secondary source P are incident in all directions, there must be rays (represented by vectors \mathbf{k}_1 and \mathbf{k}_2) with the incident angle (θ) satisfying the Bragg condition for plane (hkl) and (hkl) . These rays \mathbf{k}_1 and \mathbf{k}_2 generate constructive diffraction on the back-focal planes at K_1 and K_2 . The intensity of \mathbf{k}_1 is larger than that of \mathbf{k}_2 because \mathbf{k}_2 is more inclined from the primary beam direction than \mathbf{k}_1 . As shown in Figure 3.37, the

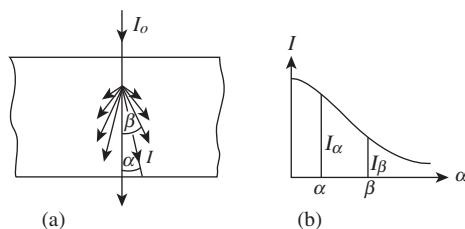


Figure 3.36 Inelastic scattering in a specimen: (a) cone of electron radiation is generated by inelastic scattering; and (b) intensity of inelastic scattering decreases with angle between the primary beam direction and the scattered ray. (Reproduced with permission from M. von Heimendahl, *Electron Microscopy of Materials*, Academic Press, New York. © 1980 Elsevier B. V.)

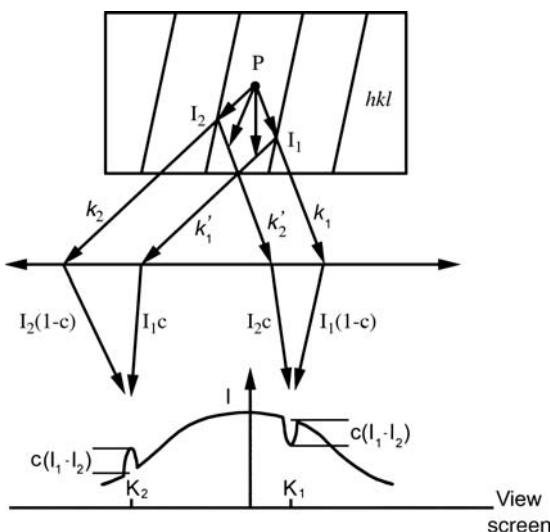


Figure 3.37 Kikuchi line formation by inelastic scattering of electrons at point P in a single crystal. The lower diagram illustrates the intensity of light on the view screen, which is affected by the inelastic scattering at P.

reflected ray of $\mathbf{k}_1(\mathbf{k}'_1)$ is parallel to \mathbf{k}_2 and the reflected ray of $\mathbf{k}_2(\mathbf{k}'_2)$ is parallel to \mathbf{k}_1 . We can determine the total intensity at K_1 and K_2 as follows.

$$\begin{aligned} I_{K_1} &= I_1(1 - c) + I_2c = I_1 - c(I_1 - I_2) \\ I_{K_2} &= I_2(1 - c) + I_1c = I_2 + c(I_1 - I_2) \end{aligned} \quad (3.22)$$

I_1 and I_2 are the background intensity at K_1 and K_2 , respectively, and c is ratio of the reflected intensity to incident intensity. Thus, contrast will be generated at K_1 and K_2 on the inelastic background; thus, a dark image will appear at K_1 and a bright image will appear at K_2 as illustrated in Figure 3.37.

Figure 3.36 and Figure 3.37 only illustrate the situation in the figure plane. In fact, inelastic scattering generates a cone of electron rays incident in front of and behind the figure plane. Consequently, those rays also generate diffraction spots in the back-focal plane. Thus, diffraction on the back-focal plane is no longer a single spot, but a locus of diffraction spots. It means that there should be a dark line and a bright line perpendicular to the view screen at positions K_1 and K_2 , respectively.

Kikuchi lines can be used to guide specimen tilting. For example, if the (hkl) plane is perfectly parallel to the radiation cone at P, then $I_1 = I_2$. The difference between the dark and bright lines will disappear. Figure 3.38 shows the Kikuchi pattern of silicon crystal with perfect alignment of the electron beam along the crystal [111] zone axis. Note that the area between two lines is darker than the background in the symmetric Kikuchi lines as shown in Figure 3.38. In such patterns, a lattice plane (hkl) parallel to the Kikuchi lines must be located at the center between the lines. We can use the symmetrical patterns of Kikuchi lines centered at a point to obtain perfect alignment of the incident beam with a crystal zone axis.

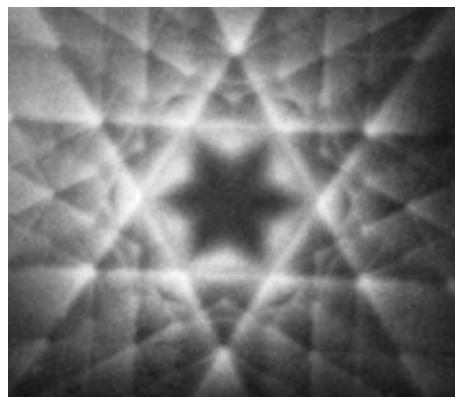


Figure 3.38 Kikuchi pattern of a silicon foil symmetrically oriented along the [111] zone. The pattern reveals six-fold symmetry. (Reproduced with permission from S. Amelinckx et al., *Handbook of Microscopy: Applications in Materials Science, Solid-State Physics and Chemistry*, Wiley-VCH Verlag GmbH, Weinheim. © 1997 Wiley-VCH.)

3.5 Images of Crystal Defects

Understanding TEM images is often not as straightforward a task as a light microscope image, except for images with mass-density contrast. For understanding images with diffraction contrast, we need more extensive knowledge of electron diffraction in crystalline solids. This section interprets several typical images of crystalline solids in diffraction contrast conditions.

3.5.1 Wedge Fringe

A fringe pattern (a set of parallel dark and bright lines) appears near the edge of thin foil specimens as shown in Figure 3.39, or near grain boundaries of a polycrystalline specimen as shown in Figure 3.40. The *wedge fringe* is generated in crystalline solid with a continuous

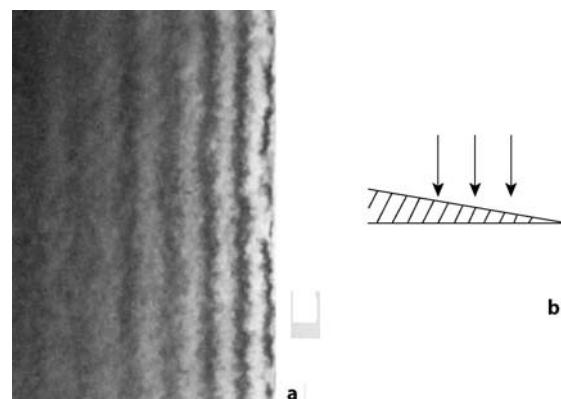


Figure 3.39 Wedge fringes at an edge of aluminum foil: (a) bright-field image ($20,000\times$ magnification); and (b) schematic illustration of the wedge shape of a foil edge. (Reproduced with permission from M. von Heimendahl, *Electron Microscopy of Materials*, Academic Press, New York. © 1980 Elsevier B. V.)

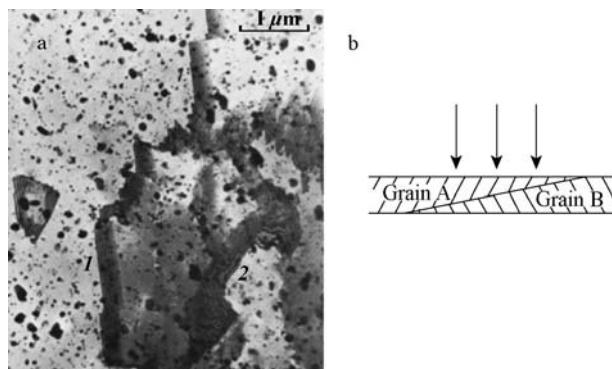


Figure 3.40 Wedge fringes at a grain boundary: (a) bright-field image; and (b) schematic illustration of grain boundary orientation in a specimen that generates the fringes in the bright-field image. (Reproduced with permission from M. von Heimendahl, *Electron Microscopy of Materials*, Academic Press, New York. © 1980 Elsevier B. V.)

change of thickness that forms a wedge shape. Full understanding of a wedge fringe requires knowledge of kinematic and dynamic theories of electron diffraction, which is beyond the scope of this book. Without details of these theories, a simple interpretation of the wedge fringe is provided as follows.

Electron diffraction in the TEM often does not satisfy the exact conditions of Bragg's Law. This relaxation of Bragg's Law in a TEM is illustrated in Figure 3.29, in which the reciprocal lattice points are elongated in the transmitted beam direction due to the small thickness of a TEM specimen. The deviation from the exact Bragg conditions can be represented by a *deviation*

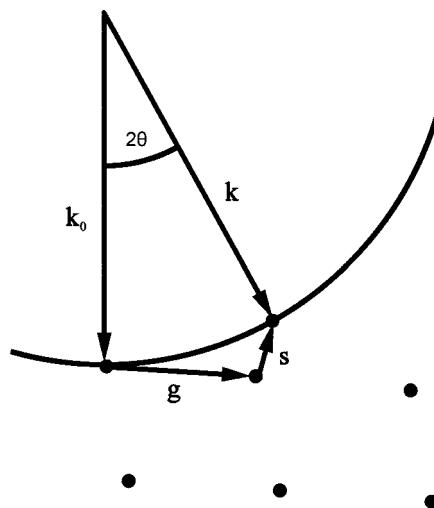


Figure 3.41 Deviation from Bragg's Law diffraction illustrated by the Ewald sphere method. The deviation from exact Bragg conditions is represented by vector \mathbf{s} in reciprocal space. \mathbf{k}_0 is the primary beam vector, \mathbf{k} is the scattering beam vector and \mathbf{g} is a reciprocal lattice vector.

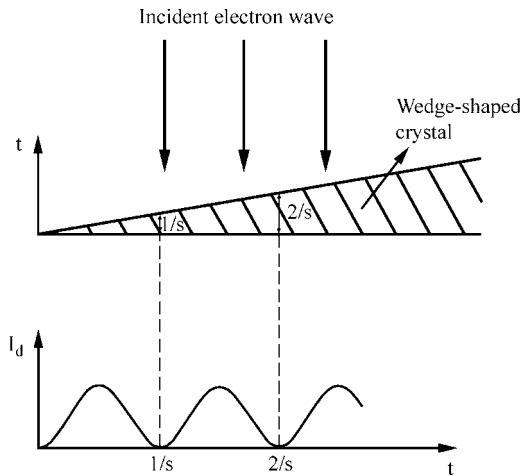


Figure 3.42 Electron intensity oscillation in the thickness direction of a specimen. A wedge-shaped crystal generates dark and bright fringes in the image because of the continuous change in thickness of the wedge.

vector (\mathbf{s}) in reciprocal space as shown in Figure 3.41. The kinematic theory of diffraction indicates that the diffraction intensity periodically changes with thickness of specimen (t).

$$I_d \propto \left(\frac{\sin \pi s t}{\pi s} \right)^2 \quad (3.23)$$

s is the component of vector \mathbf{s} parallel to the transmission vector \mathbf{k}_0 as illustrated in Figure 3.41. For a given crystal orientation, s is a constant in Equation 3.23, and thus, the intensity of the diffracted beam must oscillate between zero and a maximum with thickness increase, as illustrated in Figure 3.42. In a wedge-shaped crystal, we see a bright fringe where $t = ns^{-1}$ and a dark fringe where $t = n + \frac{s}{2}$, in which n is an integer.

The wedge fringe pattern is also seen at stacking faults. Stacking faults are a special type of planar defect in FCC crystals caused by disorder in the stacking sequence on the {111} planes. There is an interruption of the lattice structure at the fault plane. The stacking fault plane separates FCC crystal into two parts. Thus, the stacking fault plane is very similar to a grain boundary plane. Figure 3.43 shows a typical image of stacking faults that exhibit fringe patterns similar to a wedge fringe. The ends of fringes are the locations of partial dislocations separating a stacking fault from the rest of the crystal. Figure 3.44 schematically illustrates the formation of a wedge fringe when the stacking fault plane is inclined to the transmitted beam. The electron beam is intersected by the fault plane. The result will be generation of fringes in the diffraction contrast image.

3.5.2 Bending Contours

Dark bands in the background are often seen in images acquired in the diffraction contrast mode, as shown in Figure 3.45. The bands are called *bending contours* and commonly are diffuse and curved. The bending contours result from the distorting or bending of a specimen foil. In the

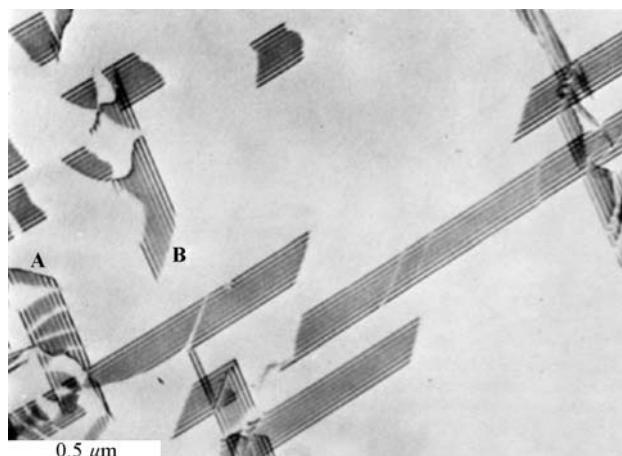


Figure 3.43 Bright-field image of stacking faults in a Co–Ni–Cr–Fe alloy. Partial dislocations are visible at A, but not at B. (Reproduced with permission from M. von Heimendahl, *Electron Microscopy of Materials*, Academic Press, New York. © 1980 Elsevier B. V.)

bent portion, the lattice plane is rotated with respect to the transmitted beam direction. At a certain location, such a rotated lattice plane may satisfy Bragg conditions exactly. Thus, a dark band appears in the bright-field image at such a location because the diffraction beams removed a significant portion of electrons from the transmitted beam. The bending contours often appear as pairs. The pairing of bending contours occurs as the specimen foil is bent symmetrically with respect to the transmitted beam, as shown in Figure 3.46. In a foil with symmetric bending, there are two locations where lattice planes (hkl) and (hkl) satisfy the exact Bragg conditions. Thus, we will see the two dark bands appear simultaneously in a bright-field image. Often, we can tilt a specimen to check whether the dark bands on the bright background are bending contours. Bending contours are sensitive to specimen tilting. They move quickly on the screen during tilting, but wedge fringes do not move quickly as they are less sensitive to tilting.

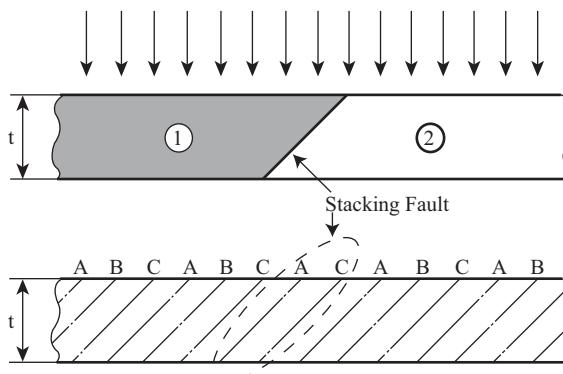


Figure 3.44 Wedge fringe formation at a stacking fault. The face of stacking fault divides a face-centered cubic crystal into parts 1 and 2. A transmitted beam column crossing the fault is similar to that crossing a grain boundary as shown in Figure 3.39b.



Figure 3.45 Bright-field image of bending contours. (Reproduced with permission of Ning Wang.)

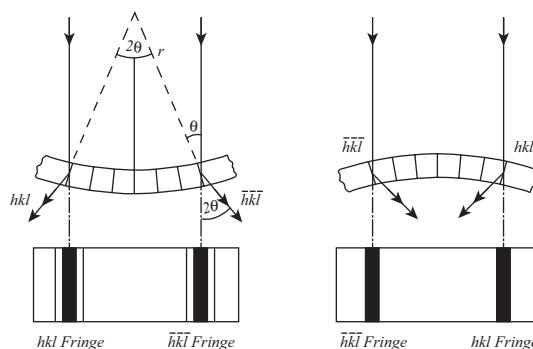


Figure 3.46 Formation of a pair of bending contours. (Reproduced with permission from M. von Heimendahl, *Electron Microscopy of Materials*, Academic Press, New York. © 1980 Elsevier B. V.)

3.5.3 Dislocations

One of the great advantages of a TEM is its ability to reveal dislocations, which are crystal line defects. Figure 3.47 shows typical images of dislocations in bright-field micrographs. The dislocations appear as fine, dark lines in the bright-field images and as bright lines in the dark-field images. The configurations of dislocations vary with specimens: they may be randomly distributed (Figure 3.47a), lined up (Figure 3.47b) and even form tangled cell walls (Figure 3.47c).

The formation of dislocation images can be graphically explained using an example of edge dislocations (Figure 3.48). An edge dislocation in a crystal generates local lattice distortion is schematically shown in Figure 3.48. A portion of distorted lattice planes is severely bent near the dislocation core. When the bent portion of the planes satisfies the Bragg conditions, a significant proportion of electrons in the transmitted beam will be deflected into the diffraction direction and the diffraction intensity immediately increases at such a portion of the lattice

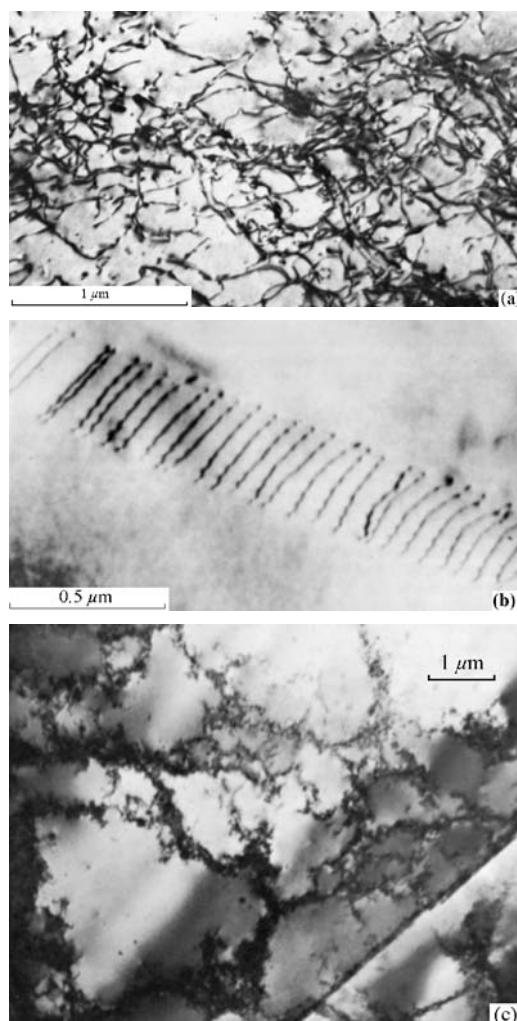


Figure 3.47 Bright-field images of dislocations: (a) in a Al-Zn-Mg alloys with 3% tension strain; (b) lining up in a Ni–Mo alloy; and (c) dislocation cell structure in pure Ni. (Reproduced with permission from M. von Heimendahl, *Electron Microscopy of Materials*, Academic Press, New York. © 1980 Elsevier B. V.)

plane. Thus, in the bright-field image, a dark image forms at point a near the dislocation core location b as illustrated in Figure 3.48. Since the dislocation is a line defect, we should see a dark line which is extended into the plane of Figure 3.48. A screw dislocation also has a dark line image for the same reason as an edge dislocation. A screw dislocation also bends lattice planes near its core to generate a plane orientation that satisfies the Bragg conditions locally. The difference between an edge and screw dislocation is the bending direction, which is determined by the Burger's vector of dislocations.

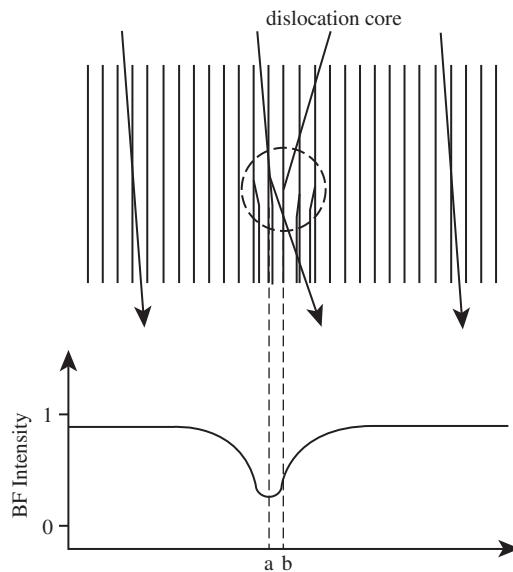


Figure 3.48 Formation of a dislocation image by deflection of transmitted electrons due to local diffraction near the core of an edge dislocation. BF, bright-field.

By understanding the formation of dislocation line images, the formation of the dislocation pattern in Figure 3.47b can be interpreted by a schematic illustration (Figure 3.49). Figure 3.47b shows a line-up of dislocations. The lined-up dislocations lie in a crystal slip plane that is inclined to the transmitted beam. The slip plane is cut off at the upper and lower surfaces of a

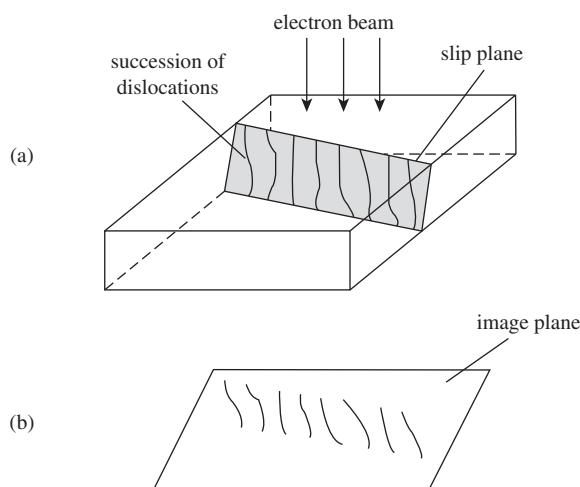


Figure 3.49 Configuration of dislocations in a thin foil when they line up in a single slip plane: (a) orientation relation between primary electron beam and dislocations in a crystal plane; and (b) the dislocation images projected on a TEM image plane.

thin foil specimen. Thus, only a short section of dislocations in the slip plane can be seen on the image screen of a TEM.

References

- [1] Flegler, S.L., Heckman, Jr., J.W. and Klomparens, K.L. (1993) *Scanning and Transmission Electron Microscopy, an Introduction*, W.H. Freeman and Co., New York.
- [2] von Heimendahl, M. (1980) *Electron Microscopy of Materials*, Academic Press, New York.
- [3] Goodhew, P.J. and Humphreys, F.J. (1988) *Electron Microscopy and Analysis*, 2nd edition, Taylor & Francis Ltd, London.
- [4] Eberhart, J.P. (1991) *Structural and Chemical Analysis of Materials*, John Wiley & Sons, Chichester.
- [5] Spence, J.C.H. (1988) *Experimental High-Resolution Electron Microscopy*, 2nd edition, Oxford University Express, Oxford.
- [6] Horiuchi, S. (1994) *Fundamentals of High-Resolution Transmission Electron Microscopy*, North-Holland, Amsterdam.
- [7] Fultz, B. and Howe, J.M. (2001) *Transmission Electron Microscopy and Diffractometry of Materials*, Springer.
- [8] Shindo, D. and Oikawa, T. (2002) *Analytical Electron Microscopy for Materials Science*, Springer-Verlag, Tokyo
- [9] Brandon, D. and Kaplan, W.D. (1999) *Microstructural Characterization of Materials*, John Wiley & Sons, Chichester.
- [10] Amelinckx, S., van Dyck, D., van Landuyt, J., van Tendeloo, G. (1997) *Handbook of Microscopy: Applications in Materials Science, Solid-State Physics and Chemistry*, Wiley-VCH Verlag GmbH, Weinheim.

Questions

- 3.1 Generally, the wavelength of electrons (λ) and the acceleration voltage (V_o) has a relationship given by: $\lambda \propto 1/\sqrt{V_o}$. Do the values in Table 3.2 satisfy the relationship exactly? Show the relationships between λ and V_o in Table 3.2 and compare them with values obtained from this relationship.
- 3.2 The resolution of a TEM is limited by diffraction (R_d) as discussed in Chapter 1, Equation 1.3, and by the spherical aberration which is expressed as: $R_{sph} = C_s \alpha^3$, where C_s is the spherical aberration coefficient and α is the half-angle of the cone of light entering the objective lens. The TEM resolution is the quadratic sum of the diffraction resolution and spherical aberration resolution, given by the expression: $(R = \sqrt{R_d^2 + R_{sph}^2})$. The minimum R is obtained when $R_d \cong R_{sph}$. Estimate the optimum value of α and resolution limit of TEM at 100 and 200 kV. $C_s \cong 1$ mm for an electromagnetic objective lens.
- 3.3 Explain how it is possible to increase the contrast of a mass-density image. Can we obtain a mass-density image of metal specimen? Can we obtain diffraction contrast in a polymer specimen?
- 3.4 Note that both the mass-density and the diffraction contrast require only a transmitted beam to pass through the objective aperture. How can you know you have diffraction contrast without checking selected area diffraction (SAD)?
- 3.5 Sketch diffraction patterns of single crystals in a TEM for (a) a face-centered cubic (FCC) crystal with transmitted beam direction (**B**) parallel to [001], and (b) a body-centered cubic (BCC) crystal with **B** = [001].
- 3.6 Is a zone axis of a single crystal diffraction pattern exactly parallel to the transmitted beam? Explain your answer graphically.
- 3.7 What kind of diffraction patterns you should expect if an SAD aperture includes a large number of grains in a polycrystalline specimen? Why?

- 3.8 You are asked to obtain maximum contrast between one grain and its neighbors. Which orientation of crystalline specimen do you wish to obtain? Illustrate your answer with a pattern of single crystal diffraction.
- 3.9 You have been told that thickness fringe and bending contours can be distinguished. Explain how.
- 3.10 The reason we see a dislocation is that it bends a crystal plane near its core region. Indicate a case where we may not be able to see the dislocation in diffraction contrast under a two-beam condition (a two-beam condition refers to a crystal orientation in which the intensity of one diffraction spot is much higher than those of the other diffraction spots). Indicate the answer either graphically or using the relation of vectors \mathbf{g} (the normal of planes which generate the diffraction beam) and \mathbf{b} (Burgers vector of dislocation).

4

Scanning Electron Microscopy

The scanning electron microscope (SEM) is the most widely used type of electron microscope. It examines microscopic structure by scanning the surface of materials, similar to scanning confocal microscopes but with much higher resolution and much greater depth of field. An SEM image is formed by a focused electron beam that scans over the surface area of a specimen; it is not formed by instantaneous illumination of a whole field as for a TEM. Perhaps the most important feature of an SEM is the three-dimensional appearance of its images because of its large depth of field. For example, the depth of field can reach the order of tens of micrometers at $10^3 \times$ magnification and the order of micrometers at $10^4 \times$ magnification. An SEM is relatively easily operated and maintained, compared with a TEM. In addition, an SEM system enables us to obtain chemical information from a specimen by using various techniques, including by equipping the X-ray energy dispersive spectrometer (EDS). EDS is separately introduced as a technique for chemical analysis in Chapter 6. This chapter introduces the basic working principles and features of an SEM for microstructural and surface morphology examination.

4.1 Instrumentation

4.1.1 Optical Arrangement

A scanning electron microscope consists of an electron gun and a series of electromagnetic lenses and apertures as shown in Figure 4.1, similar to TEM systems. In an SEM, however, the electron beam emitted from an electron gun is condensed to a fine probe for surface scanning. The electron gun for generating an electron beam is the same as in a TEM; either thermionic or field emission type guns that were introduced in Chapter 3 are used. Advanced SEM systems use a field emission gun because of its high beam brightness. Beam brightness plays an even more important role in imaging quality in an SEM than in a TEM. The relationship between the brightness and resolution will be discussed in Section 4.1.3. The acceleration voltage for generating an electron beam is in the range 1–40 kV, which is about one order of magnitude less than that for a TEM.

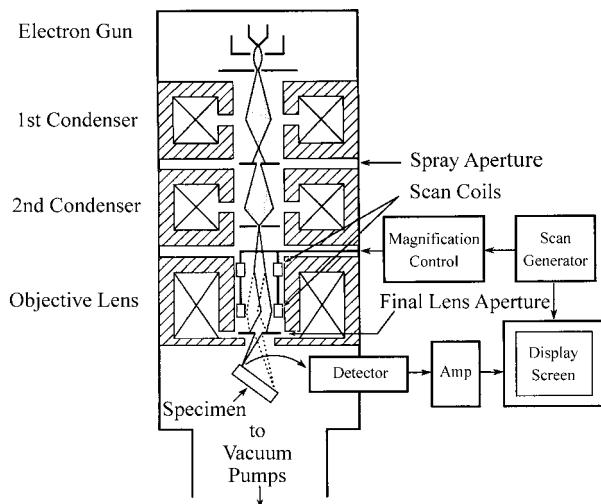


Figure 4.1 Structure of a scanning electron microscope (SEM). (Reproduced with kind permission of Springer Science and Business Media from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

An SEM optical path goes through several electromagnetic lenses, including condenser lenses and one objective lens, as shown in Figure 4.1. The electromagnetic lenses in an SEM are for electron probe formation, not for image formation directly as in a TEM. The two condenser lenses reduce the crossover diameter of the electron beam; then, the objective lens focuses the electron beam as a probe with a diameter on the nanometer scale. The objective lens should be considered as the third condenser lens in the SEM because it functions more like a condenser than an objective lens. The reason is that the objective lens in SEM demagnifies the cross-section of the electron beam; thus, it functions very differently from the objective lens in a TEM, which magnifies the electron beam. The SEM lens system demagnifies the electron beam by about $10,000\times$ for a thermionic source and $10\text{--}100\times$ for a field emission source.

Probe scanning is operated by a beam deflection system incorporated within the objective lens in an SEM. The deflection system moves the probe over the specimen surface along a line and then displaces the probe to a position on the next line for scanning, so that a rectangular raster is generated on the specimen surface. The signal electrons emitted from the specimen are collected by a detector, amplified, and used to reconstruct an image, according to one-to-one correlation between scanning points on the specimen and picture points on a screen of a *cathode ray tube* (CRT) or liquid crystal display. The deflection system of the electron probe is controlled by two pairs of electromagnetic coils (*scan coils*). The first pair of coils bends the beam off the optical axis of the microscope. The second pair of coils bends the beam back onto the axis at the pivot point of a scan. The apertures in an SEM, as shown in Figure 4.1, are mainly used for limiting the divergence of the electron beam in its optical path.

The magnification of an SEM is determined by the ratio of the linear size of the display screen to the linear size of the specimen area being scanned, not as in a TEM where magnification

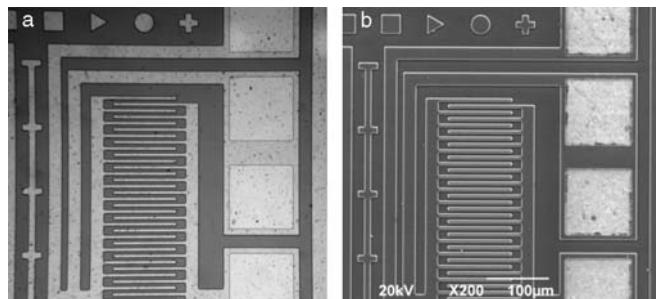


Figure 4.2 Comparison of images taken in: (a) a light microscope; and (b) an SEM. An SEM image (b) is able to provide the 3-D appearance of an integrated circuit while revealing the same in-plane details as the light microscopic image (a).

is determined by the power of the objective lens. The size of the scanned rectangular area (raster) can be varied over a tremendously wide range. Thus, an SEM is able to provide image magnification from about $20\times$ to greater than $100,000\times$. For low magnification imaging, an SEM is often more favorable than a light microscope (LM) because of the large depth of field in SEM. A comparison between LMs and SEMs is demonstrated by the images shown in Figure 4.2. Both the LM and SEM images in Figure 4.2 reveal the plane configuration of an integrated circuit, but the SEM image (Figure 4.2b) also reveals the out-of-plane details because of its three-dimensional appearances. The resolution of an SEM is controlled by the size of the electron probe scanning the specimen. Recalling that the effective magnification limit as discussed in Chapter 1 is determined by resolution of the microscope, for a probe size of 10 nm, SEM systems can generate effective magnification of $20,000\times$.

4.1.2 Signal Detection

The electron signals from each pixel of the raster are collected by a detector in order to generate a corresponding point-to-point image on a display screen. To understand signal detection, knowledge of electron signal types that are useful in an SEM is necessary: *backscattered electrons* and *secondary electrons*.

When high energy electrons strike a specimen, they produce either elastic or inelastic scattering. Elastic scattering produces the backscattered electrons (BSEs), which are incident electrons scattered by atoms in the specimen. Inelastic scattering produces secondary electrons (SEs), which are electrons ejected from atoms in the specimen. BSEs are typically deflected from the specimen at large angles and with little energy loss; they typically retain 60–80% of the energy of incident electrons. In contrast, SEs are typically deflected at small angles and show considerably low energy compared with incident electrons.

During inelastic scattering, an incident electron transfers kinetic energy to an electron in a specimen atom. Any electron in atoms in the specimen with sufficient kinetic energy will leave its orbital to become a secondary electron. The SE energy is usually in the range of about 3–5 eV. In terms of usefulness, SEs are the primary signals for achieving topographic contrast, while BSEs are useful for formation of elemental composition contrast.

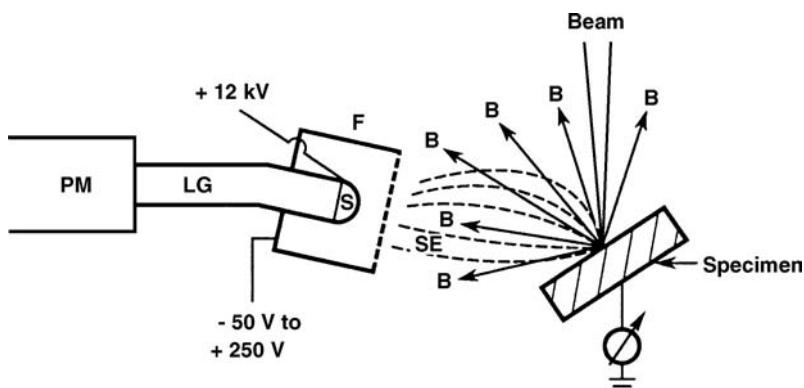


Figure 4.3 Signal collection by the Everhart–Thornley detector. B, backscattered electron trajectory; SE, secondary electron trajectory; F, Faraday cage; S, scintillator; LG, light guide; PM, photomultiplier tube. (Reproduced with kind permission of Springer Science and Business Media from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

Detector

A commonly used detector in an SEM is the *Everhart–Thornley (E–T) detector*, as illustrated in Figure 4.3. The SEs travel with large deflection angles toward the detector, while BSEs travel directly toward the detector. The *Faraday cage* in the front of the detector is either positively or negatively charged (250 V or –50 V), depending on signal selection. When given a positive charge, the detector attracts signal electrons, particular secondary electrons. When given a negative charge, it can screen out secondary electrons with energy less than 50 eV. The key element of the E–T detector is the *scintillator*, a disk of about 8–20 mm in diameter. The scintillator converts signal electrons into photons by accelerating the electrons with +12 kV and striking them onto a disk. The photons then travel through a *light guide* and enter the *photomultiplier tube* for signal gain (up to $\sim 10^6 \times$). The photomultiplier output is further amplified for display on a display screen.

Images in the modern SEM are recorded digitally. That is, the intensity of signal electrons for each pixel is digitized and saved on computer as a digital file for each scan. An advantage of digital imaging is the ability to generate an image by averaging multiple scans for the same area. Such a frame-averaging method can effectively reduce background noise in imaging.

4.1.3 Probe Size and Current

The resolution of SEM imaging is determined by the cross-sectional diameter of the scanning probe. Thus, the size of the probe limits the size of features on the specimen surface to be resolved. To obtain high resolution, we should know how to minimize probe size. The probe diameter is approximately expressed as d_p .

$$d_p = \left(\frac{4i_p}{\beta \pi^2 \alpha_f^2} \right)^{1/2} \quad (4.1)$$

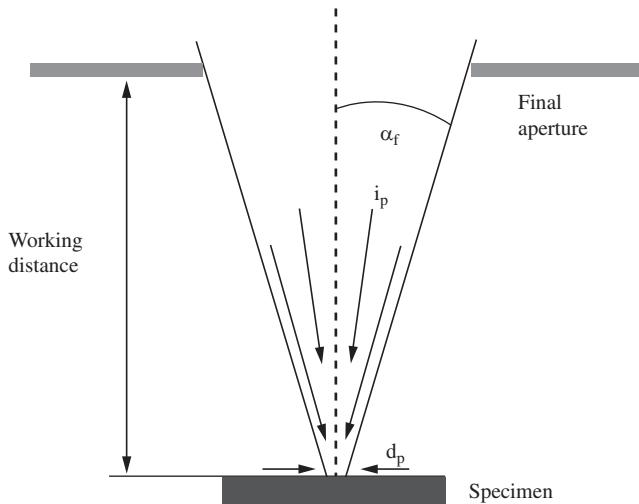


Figure 4.4 Relationship between the probe diameter, convergence angle and working distance.

i_p is the probe current, β is the beam *brightness*, which is controlled by the electron source, and α_f is the *convergence angle of the probe*, which is determined by the final aperture diameter and the distance between aperture and specimen surfaces (called the *working distance*) as illustrated in Figure 4.4.

The brightness of electron illumination or *beam brightness* (β) depends on the type of electron gun used. As indicated in Table 3.3, a field emission gun is $1000\times$ brighter than a tungsten thermionic gun and $100\times$ brighter than a LaB₆ thermionic gun. Thus, the field emission gun has been widely used in advanced SEM systems. In any case, regardless of gun type, the brightness is proportional to the acceleration voltage (V_o) of the gun.

$$\beta \propto e V_o \quad (4.2)$$

where e is the charge of an electron. To obtain a minimal probe size, we should increase the brightness as well as the convergence angle, as indicated in Equation 4.1. On the other hand, a large α_f likely introduces other optical problems, particularly spherical aberration. Thus, α_f must be optimized, not simply increased. With an optimized α_f (and neglecting chromatic aberration), the minimum probe size is approximated as d_{\min} .

$$d_{\min} = KC_s^{1/4} \left(\frac{i_p}{\beta} + \lambda^2 \right)^{3/8} \quad (4.3)$$

K is a constant close to unity, C_s is the spherical aberration coefficient and λ is the wavelength of the electrons. Equation 4.3 simply tells us that we should decrease wavelength and spherical aberration and increase brightness of electron illumination to obtain the minimal probe size. The importance of electron gun type in determining SEM resolution becomes obvious because there are significant differences in their brightnesses. Figure 4.5 shows the difference

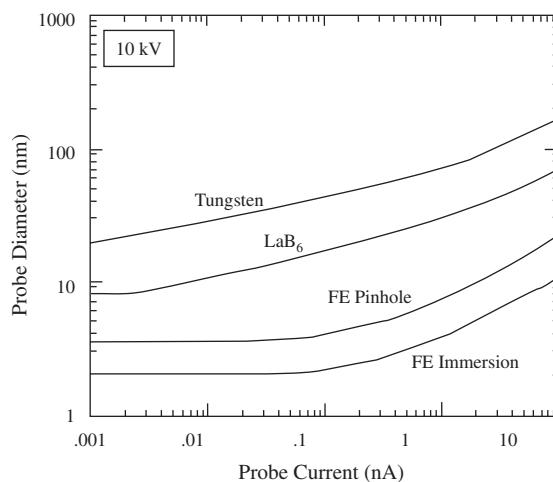


Figure 4.5 Probe size is determined by the type of the electron gun and probe current. (Reproduced with kind permission of Springer Science and Business Media from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

between probe sizes when different guns are used. The wavelength effects on probe size can be considered as the acceleration voltage effect, because of the relationship between the voltage and the wavelength (Table 3.2). Figure 4.6 shows the difference in probe size when different acceleration voltages are used.

Even though reducing the probe size is necessary, the reduction of probe size alone is not sufficient to obtain a high resolution image in an SEM. For obtaining a high resolution image,

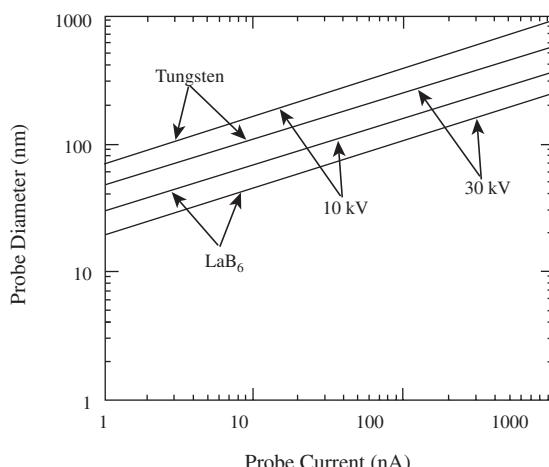


Figure 4.6 Probe size determined by the acceleration voltage of the electron gun. (Reproduced with kind permission of Springer Science and Business Media from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

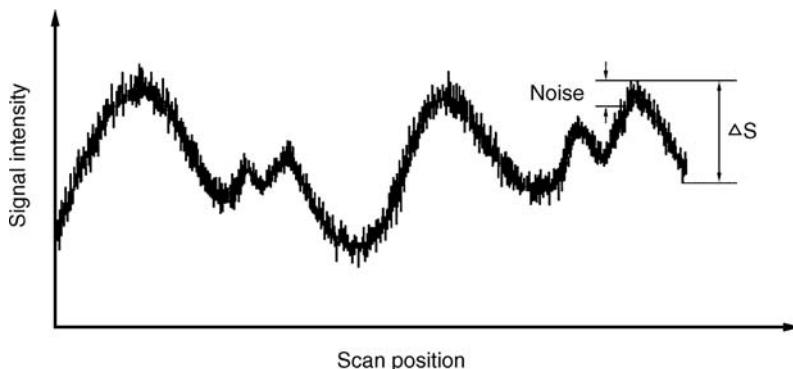


Figure 4.7 Signals and background noise produced when scanning a specimen.

there must be sufficient signal electrons generated by the probe for each pixel of an image in order to distinguish the signals from the background noise of electrons. The signal electrons are proportional to the amount of incident electrons in the probe. In other words, the probe current must be larger than a minimum value so that microscopic features of the specimen are visible in the SEM image.

The relationship between the probe current and image visibility can be understood by analyzing the basic requirements that an image will not be obscured by background noise. Background noise in an SEM system is generated by fluctuation of the electron beam current and signal amplification in the detector, and cannot be totally eliminated.

Figure 4.7 schematically shows the characteristics of signals and noise. The image signals from a specimen can be represented by a curved line varying with scan position. The background noise is represented by vertical fluctuation imposed on the curved line. The *Rose visibility criterion* states that, to distinguish two objects from the background noise, the change in signal (ΔS) due to the object contrast must be five times larger than the noise level (N).

$$\Delta S > 5N \quad (4.4)$$

The noise level can be expressed as the standard deviation of signals. Suppose that a signal S consists of n counts at the detector; its noise level is then considered to be the square root of its mean n value. Thus, we can rewrite the Rose visibility criterion.

$$\Delta S > 5\bar{n}^{1/2} \quad (4.5)$$

Recalling the definition of contrast (C) (Equation 1.6), we can express the criterion for signal counts.

$$C = \frac{\Delta S}{S} > \frac{5\bar{n}^{1/2}}{S} = \frac{5\bar{n}^{1/2}}{\bar{n}}$$

Thus,

$$\bar{n} > \left(\frac{5}{C}\right)^2 \quad (4.6)$$

In an SEM, the number of counts is the number of signal electrons being detected. The relationship between counts and current of signal electrons (i_s) for the dwell time of probe (τ) on an object is calculated by the following expression.

$$i_s = \frac{\bar{n}e}{\tau} \quad (4.7)$$

Here, i_s is not equal to the probe current (i_p) but proportional to it. By converting the criterion for electron counts to the probe current, we have the minimum requirement of probe current.

$$i_p > \frac{25e}{\varepsilon C^2 \tau} \quad (4.8)$$

ε is the proportionality factor between i_p and i_s . This minimum requirement, or threshold value, for probe current is necessary to observe a certain level of object contrast in an SEM. The dwell time of the probe (τ) is often expressed in terms of the time required for scanning one frame.

$$\tau = \frac{t_f}{n_{PE}} \quad (4.9)$$

t_f is the time required to scan one frame and n_{PE} is number of pixels in one frame. Thus, we should increase the frame time by reducing the scanning speed in order to satisfy the probe current criterion. Figure 4.8 shows the relationship among the contrast, frame time and the minimum probe current required.

In summary, the resolution of an SEM is determined by the probe size (Equation 4.3) and the probe current (Equation 4.8). The important points to note are that the minimum probe size and the maximum probe current cannot be obtained simultaneously, and a compromise between

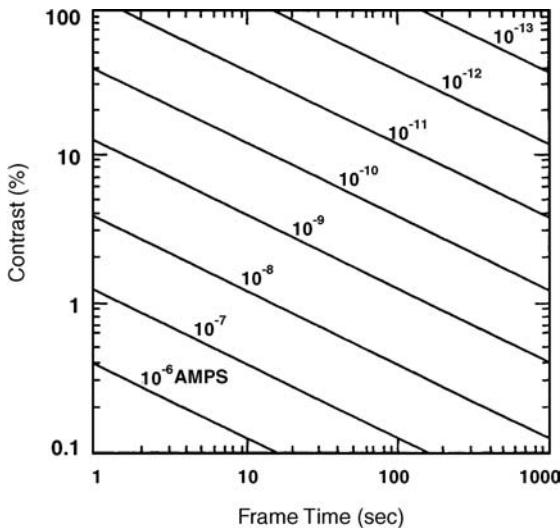


Figure 4.8 Image contrast variation with probe current and frame time. (Reproduced with kind permission of Springer Science and Business Media from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

the probe size and the probe current must be made when we try to obtain a high resolution image in SEM.

4.2 Contrast Formation

There are two types of contrast in SEM images: topographic and compositional. The secondary electrons are the primary source for surface topographic images, while backscattered electrons are mainly used for compositional images. Knowledge about the interaction between the high energy electron probe and the specimen is necessary for understanding these mechanisms of contrast formation.

4.2.1 Electron–Specimen Interactions

When high-energy electrons strike a specimen, the electrons are scattered by atoms of the specimen, as described briefly in Section 4.1. Electron scattering results in a change of direction of travel of electrons under the specimen surface. Electron trajectories during scattering in a specimen are schematically shown in Figure 4.9. As shown in Figure 4.9, the interaction between electrons and specimen atoms occurs within a certain volume under the specimen surface.

Both SEs and BSEs generated by scattering are used as signal sources for forming SEM images. However, SEs and BSEs, which are collected by a detector, escape from different locations in the specimen. Figure 4.10 illustrates the interaction zone where electrons scatter

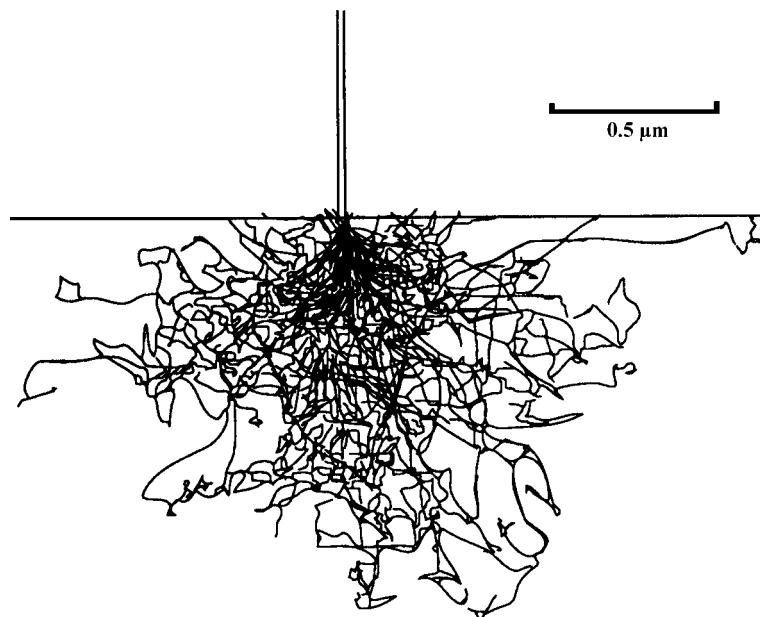


Figure 4.9 Monte Carlo electron trajectory simulation of an electron beam interaction with iron $E_0 = 20$ keV. (Reproduced with kind permission of Springer Science and Business Media from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

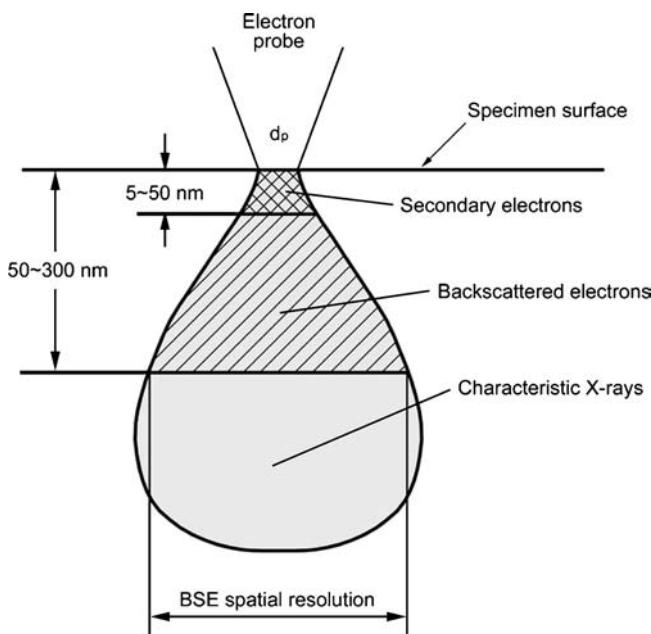


Figure 4.10 The interaction zone of electrons and specimen atoms below a specimen surface.

under the specimen surface. The zone is usually described as pear-shaped, and its size increases with the energy of incident electrons in the probe. Besides SEs and BSEs, characteristic X-rays are also produced in the interaction zone, and these are useful for chemical analysis, which is discussed in Chapter 6.

As mentioned previously, SEs are the products of inelastic scattering, and they have an energy level of several keV. In the interaction zone, SEs can escape only from a volume near the specimen surface with a depth of 5–50 nm, even though they are generated in the whole pear-shaped zone. In contrast, BSEs are the products of elastic scattering, and they have an energy level close to that of incident electrons. Their high energy enables them to escape from a much deeper level in the interaction zone, from depths of about 50–300 nm. The lateral spatial resolution of an SEM image is affected by the size of the volume from where the signal electrons escape. We can expect, as shown in Figure 4.10 that an image formed by SEs should have a better spatial resolution than that formed by BSEs. The characteristics of images formed by SEs and BSEs are discussed in the following sections.

4.2.2 Topographic Contrast

Topographic contrast in an SEM refers to variation in signal levels that corresponds to variation in geometric features on the specimen surface. Topographic contrast is the primary source of contrast in an SEM image. An SEM image with topographic contrast often has the stereoscopic appearance of a rough specimen surface. Topographic contrast occurs because signal electrons arise from two effects: the *trajectory effect* and the *electron number effect*. The trajectory effect arises from variations in how the specimen surface is oriented with respect to the detector. As

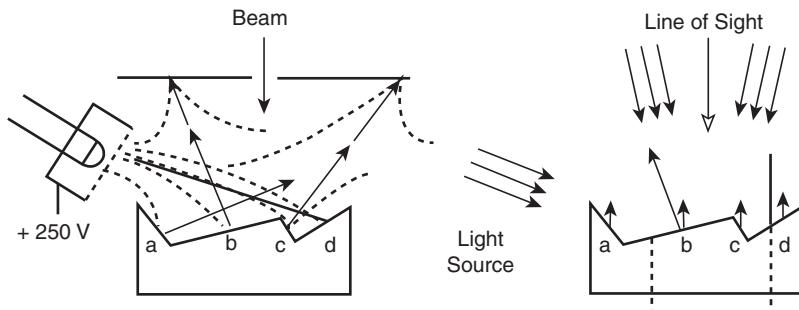


Figure 4.11 Generation of topographic contrast: (a) the trajectory effect, which arises from the orientation of surface with respect to the detector in an SEM, is similar to: (b) reflected light effects from the orientation of surface with respect to the light source in a light microscope. (Reproduced with kind permission of Springer Science and Business Media from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

schematically shown in Figure 4.11, the electrons emitted from the specimen surfaces facing the detector will be collected abundantly, and corresponding sites in the image will appear bright. The electrons emitted from the surfaces not facing the detector will reach the detector with difficulty, and thus corresponding sites in the image will appear dark. The contrast created by the trajectory effect is very much similar to the contrast we see with our eyes of a rough surface illuminated by light. Thus, there is little problem for us to interpret topographic contrast such as shown in Figure 4.12.

Nevertheless, topographic contrast in an SEM is not exactly the same as the contrast achieved by inclined light illumination. In an SEM image, the electron number effect will create bright areas in the image that do not correspond to surface contours on the specimen. Figure 4.13 illustrates the electron number effect. When the electron probe hits a surface at an angle, more

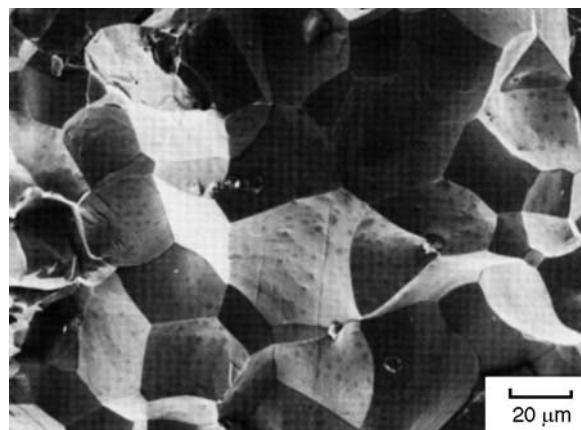


Figure 4.12 Secondary electron image of a fracture surface along grain boundaries. (Reproduced with kind permission of Springer Science and Business Media from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

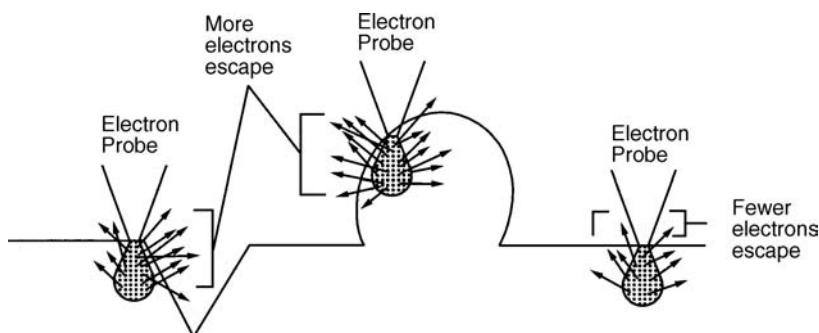


Figure 4.13 Electron number effects due to surface topography. More secondary electrons can escape from the edges of topographical features than from a flat surface.

electrons can escape from the specimen than when the probe hits a flat surface directly. Thus, certain areas of the specimen (such as edges of spherical particles, raised areas and cavities) will appear bright in an SEM image. Figure 4.14 shows the example of topographic contrast from the electron number effect.

Commonly, a topographic image is obtained by operating an SEM in the SE mode. However, both SEs and BSEs contribute to topographic contrast, but SEs are the primary signals for topographic contrast formation. Only those BSEs with trajectories on a direct path toward the scintillator of an E-T detector contribute to topographic contrast formation. It is important to note that in a topographic image the surface areas that do not face the detector are not totally dark. The SEs from such surface areas can be drawn to the detector when the Faraday cage is positively charged as shown in Figure 4.3, because SEs with low kinetic energy readily change their trajectories.

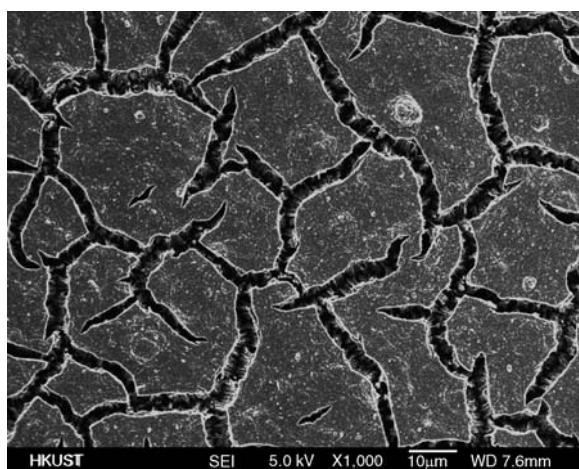


Figure 4.14 Secondary electron image of a cracked silicon surface. Electron number effects make the edges of surface cracks brighter than the rest of the surface.

4.2.3 Compositional Contrast

Compositional contrast refers to the variation in gray levels in an SEM image that correspond to variation in chemical composition in a specimen. An image formed by BSEs exhibits very useful compositional contrast if the specimen consists of more than one chemical element. The origin of compositional contrast arises because the capability of BSEs to escape from the specimen depends on the atomic numbers of the specimen atoms. The *backscatter coefficient* (η) characterizes such capability.

$$\eta = \frac{n_{BSE}}{n_i} \quad (4.10)$$

η is the ratio of the number of BSEs escaping from the specimen (n_{BSE}) to the number of incident electrons (n_i). It increases monotonically with the atomic number as shown in Figure 4.15. Thus, any area in a specimen containing chemical elements with higher atomic number will generate more BSEs. The difference in the number of BSEs collected by a detector will appear as differences in gray levels in a black and white image; that is, an area with atoms of higher atomic numbers will appear brighter. Thus, a BSE image shows the atomic number contrast or compositional contrast as demonstrated in Figure 4.16, in which there is a side-by-side comparison of an SE image and a BSE image for the same surface area of a nickel alloy. The SE signals also contributes to compositional contrast, but such contrast is small and unpredictable, mainly because the intensity of an SE signal does not have strong dependence on the atomic number.

When the E-T detector is used for compositional contrast, the Faraday cage should be negatively charged (-50 V) to exclude SEs with energies lower than the bias voltage of the cage. Some SEM systems are equipped with a dedicated detector for BSE collection, which is placed above the specimen. To obtain compositional contrast, we can switch detection from the SE mode to the BSE mode by simply pushing a button on the SEM control panel.

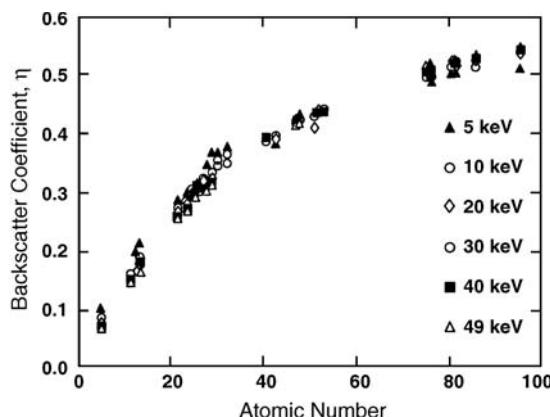


Figure 4.15 Backscatter coefficient as a function of atomic number of specimen atoms. (Reproduced with kind permission of Springer Science and Business Media from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

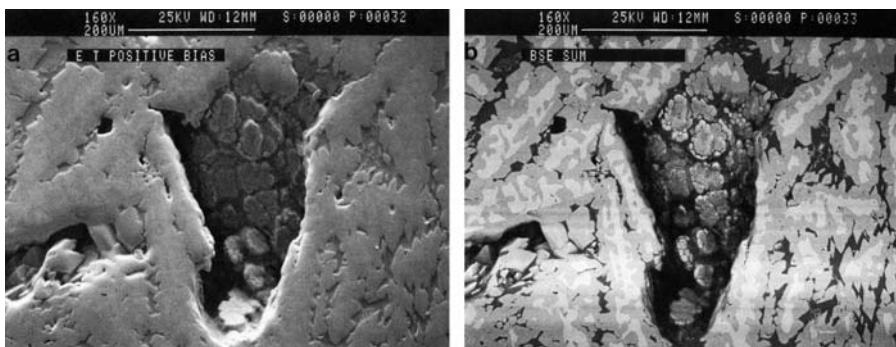


Figure 4.16 Comparison between: (a) a secondary electron image; and (b) a backscattered electron image for the same area of nickel alloy. Additional compositional information is obtained from the backscattered image. (Reproduced with kind permission of Springer Science and Business Media from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

In addition to topographic and compositional contrast, there are other types of contrast in the SEM, in particular, crystallographic contrast and magnetic contrast. Crystallographic contrast is also called the contrast of *electron channeling*. It arises because the backscattered electron coefficient is dependent on the orientation of crystallographic planes with respect to the direction of the electron beam. The plane orientation affects the penetration of the electron beam into the crystal because the atomic packing density varies with crystallographic plane. For higher penetration, the backscattering coefficient is lower. A large single crystal surface may exhibit an electron channeling pattern that reveals the crystal orientation of a specimen surface layer (~50 nm). The electron channeling image in an SEM is very similar to the Kikuchi pattern in a TEM. The magnetic domains of ferromagnetic materials may also exhibit special contrast in SEM. Detailed discussion of such contrast can be found in the references listed at the end of Chapter 4.

4.3 Operational Variables

It is necessary to understand the operational variables to obtain a good SEM image with desired depth of field and resolution. This section discusses how to control the depth of field and resolution by adjusting operation variables such as the working distance, final aperture size, acceleration voltage, probe current and astigmatism.

4.3.1 Working Distance and Aperture Size

The working distance and aperture size, illustrated in Figure 4.4, are the operation variables that strongly affect depth of field. We often need to make full use of the most important feature of an SEM: the three-dimensional appearance of topographic images. Recall that the depth of field is related to the resolution (R) and the convergence angle of objective aperture (α) as (Equation 1.7).

$$D_f = \frac{2R}{\tan \alpha}$$

Table 4.1 Depth of field of an SEM image (μm)

Magnification	α_f (radian)		
	5×10^{-3}	1×10^{-2}	3×10^{-2}
10×	4,000	2,000	670
50×	800	400	133
100×	400	200	67
500×	80	40	13
1,000×	40	20	6.7
10,000×	4	2	0.67
100,000×	0.4	0.2	0.067

(Data taken from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

The resolution of an SEM image on a display screen is limited by the pixel size of the screen. The product of the resolution and the SEM magnification (M) should be equal to the pixel size. Suppose that the pixel size is $100 \mu\text{m}$, the resolution of a SEM image on the display screen can be expressed as $100 M^{-1} \mu\text{m}$. The convergence angle of an SEM probe (α_f) is much less than 1 rad. Thus, the depth of field in SEM can be expressed as follows.

$$D_f \cong \frac{200}{\alpha_f M} (\mu\text{m}) \quad (4.11)$$

Table 4.1 lists the depth of field as a function of the convergence angle and magnification. α_f is controlled by the final aperture radius and the working distance between the aperture and specimen surface (Figure 4.4). Equation 4.11 can be rewritten to express the depth of field as a function of the aperture radius (R_{fap}) and the working distance (D_W) as follows.

$$D_f \cong \frac{200 D_W}{R_{fap} M} (\mu\text{m}) \quad (4.12)$$

Equation 4.12 simply tells us how to control the depth of field in an SEM by changing the aperture size and working distance. Figure 4.17 illustrates the relationship among depth of field, aperture size and working distance. A small aperture size and a long working distance results in a high depth of field. The SEM micrographs in Figure 4.18 show the differences in the depth of field when aperture radius changes from 25 to 85 μm and working distance changes from 25 to 48 mm. The aperture size and working distance can be easily changed during SEM operation.

Although selecting a combination of small aperture size and long working distance is favorable for high depth of field, such operation will result in negative effects on imaging resolution. A small aperture size reduces the probe current, and also may reduce the ratio of signal to noise. A long working distance is not desirable because it will decrease α_f if the aperture size is unchanged. A small α_f is not favorable for high resolution. Further, a long working distance increases the spherical aberration of a lens and consequently worsens the SEM resolution. Considering such tradeoffs between resolution and depth of field, it is wise to select an

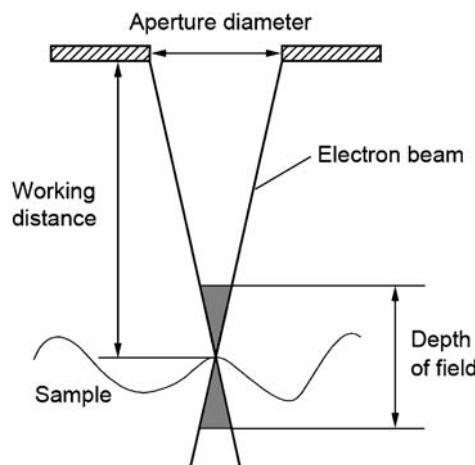


Figure 4.17 Relationship between the depth of field, aperture size and the working distance.

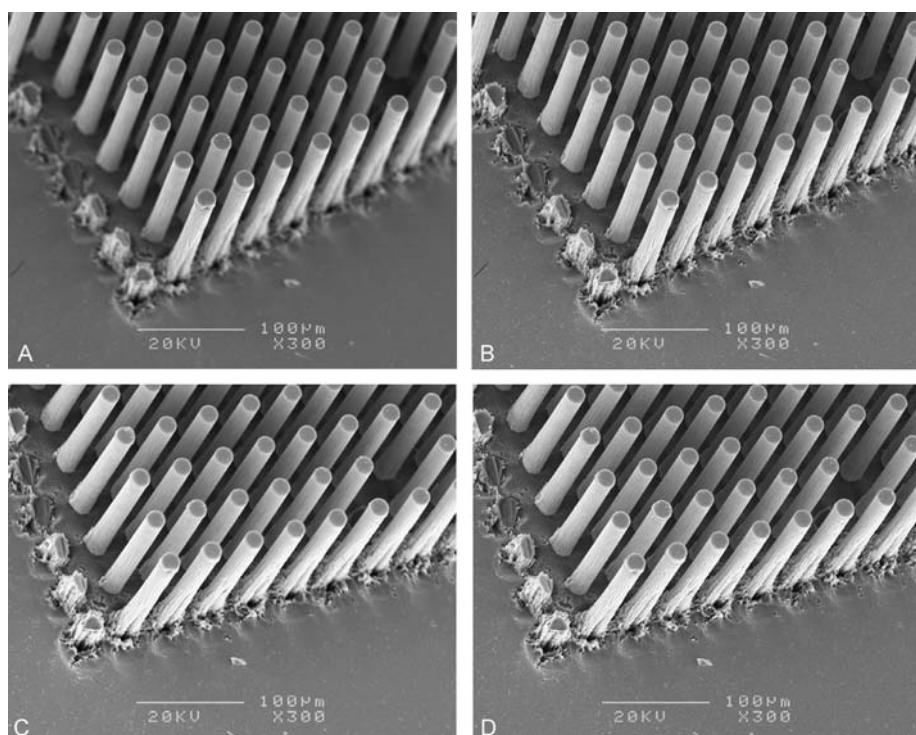


Figure 4.18 Effects of final aperture radius (R_{fap}) and working distance (D_W) on the depth of field: (a) 25-mm D_W , 85-μm R_{fap} ; (b) 25-mm D_W , 25-μm R_{fap} ; (c) 48-mm D_W , 85-μm R_{fap} ; and (d) 48-mm D_W , 25-μm R_{fap} .

intermediate aperture size and intermediate working distance, unless a very large depth of field or very high resolution is required.

4.3.2 Acceleration Voltage and Probe Current

As discussed in Section 4.1.3, the probe size and probe current are critical to achieving high resolution SEM images. The acceleration voltage of the electron gun, and probe current, are two primary operation variables used to adjust the resolution. Recall that acceleration voltage is the controlling factor of probe size, because the probe diameter decreases with decreasing wavelength of electrons, and the wavelength is, in turn, determined by acceleration voltage. Thus, increasing acceleration voltage, which is equivalent to decreasing wavelength, will reduce the probe size. The benefit of voltage increase is reflected in increases in electron beam brightness (Equation 4.2), which also results in reduction of probe size. Figure 4.19 shows a comparison of the resolution of SEM images when selecting 5 and 20 kV while maintaining other operation variables constant. The image at 20 kV reveals more surface details on a specimen of etched titanium than that at 5 kV because of higher resolution.

On the other hand, the negative effects of selecting higher voltage, which are due to the interaction between electron probe and specimen materials, should not be ignored. A higher voltage causes an increase in the interaction zone in the specimen as illustrated in Figure 4.10. An increase in the interaction zone in lateral directions will result in a decrease in the lateral spatial resolution of SEM images. Currently, more and more SEM systems are equipped with field emission guns, which provide high electron beam brightness without a high acceleration voltage. In a field-emission SEM system, a relatively low acceleration voltage (~ 5 kV) can also achieve high electron beam brightness and a small probe size.

Adjustment of the probe current during operation will help us to balance the requirements of probe size reduction by lowering probe current, and of signal-to-noise ratio. A low probe current produces weak signals collected by the detector. An excessive gain on the signal amplifier is needed for image formation when the probe current is too low. The excessive gain will generate a high level of electronic noise in an SEM image. Adjustment of probe current is necessary whenever the acceleration voltage and magnification are changed during operation.

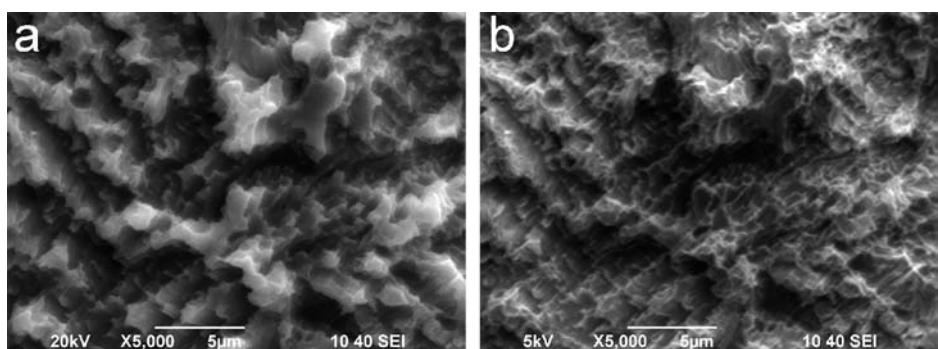


Figure 4.19 Comparison of the effect of acceleration voltage on topographic contrast of acid etched titanium specimen. (a) 20 kV, and (b) 5 kV.

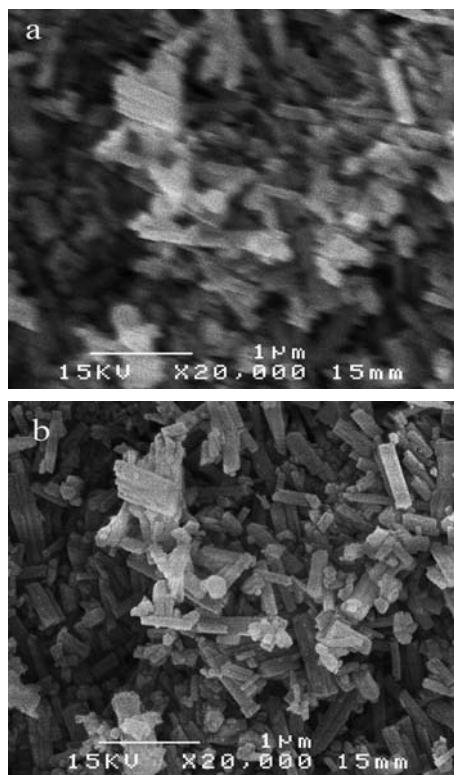


Figure 4.20 Effects of astigmatism: (a) blurred astigmatic image; and (b) sharp image after astigmatism has been corrected.

4.3.3 Astigmatism

Astigmatism is an operational variable that is often adjusted to obtain a good electron microscope image. Astigmatism is the lens aberration resulting from power differences of a lens in its lens plane perpendicular to the optical path (Figure 1.9), which is discussed in Chapter 1. For most SEM systems, astigmatism is not a serious problem when magnification of the image is less than 10,000 \times . At high magnification, astigmatism effects on an image are seen, as demonstrated by Figure 4.20. An image with astigmatism looks out of focus. In Figure 4.20a, the images of small particles are stretched by astigmatism. This can be conveniently corrected using the stigmator knobs on the SEM control panel. The stigmator adjusts the electromagnetic field of the objective lens in two perpendicular directions in order to eliminate the asymmetrical distortion of the lens. Figure 4.20b shows the same image as Figure 4.20a after correction by the stigmator.

4.4 Specimen Preparation

A specimen for SEM analysis can be in any form. This is one reason that SEMs are more widely used than TEMs, which requires a foil type of specimen. However, there are special

problems in SEM specimen preparation such as surface charging of specimens that have poor electric conductivity, and dehydration requirement for biological samples.

4.4.1 Preparation for Topographic Examination

For topographic examination, we should do minimal preparation of specimens in order to preserve features of their surfaces. The preparation should involve only sizing the specimens to fit a SEM specimen holder and removing surface contaminants. The specimen holder of an SEM varies with manufacturer. Some SEM systems can hold a specimen with dimensions of tens of centimeters. Common contaminants on the specimen surfaces are hydrocarbons from oil and grease, because an electron beam decomposes any hydrocarbon and leaves a deposit of carbon on the surface. The deposit generates an artifact of a dark rectangular mark in SEM images. Figure 4.21 shows an example of hydrocarbon contamination on an SEM image. A dark mark is readily seen on a hydrocarbon-containing specimen at lower magnification after examining the same location at higher magnification. The carbon deposition is formed quickly under a higher magnification because of the higher exposure rate of electron beam.

Surfaces with oil or grease contamination should be cleaned by an organic solvent such as methanol or acetone in an ultrasonic cleaner. It is also important to avoid touching the cleaned specimen with bare hands because fingerprints contain volatile hydrocarbon compounds. Sometimes a dark mark occurs on a well-cleaned specimen. This might result from hydrocarbon contamination from a dirty vacuum system. Hydrocarbons escaping from the diffusion pump oil into the SEM chamber can also contaminate a specimen under electron beam.

Charging and its Prevention

Surface charging causes SEM image distortions and should be carefully prevented. Surface charging is most likely encountered when examining electrically nonconductive surfaces.

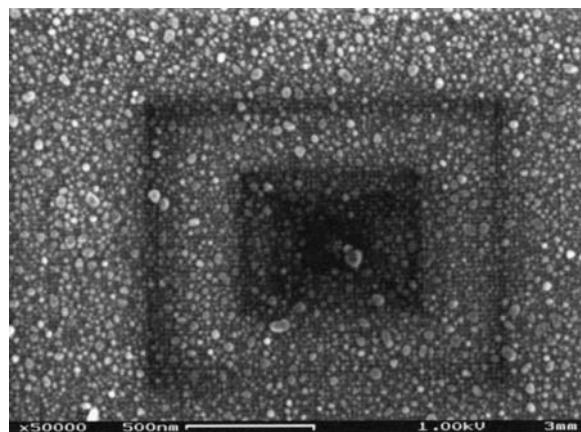


Figure 4.21 SEM micrograph of Al on TiN. The beam was first focused on the central region for about 30 seconds, and then the magnification was reduced. Contamination built up in the central region is visible at reduced magnification. (Reproduced with permission from D. Brandon and W.D. Kaplan, *Microstructural Characterization of Materials*, John Wiley & Sons Ltd, Chichester. © 1999 John Wiley & Sons Ltd.)

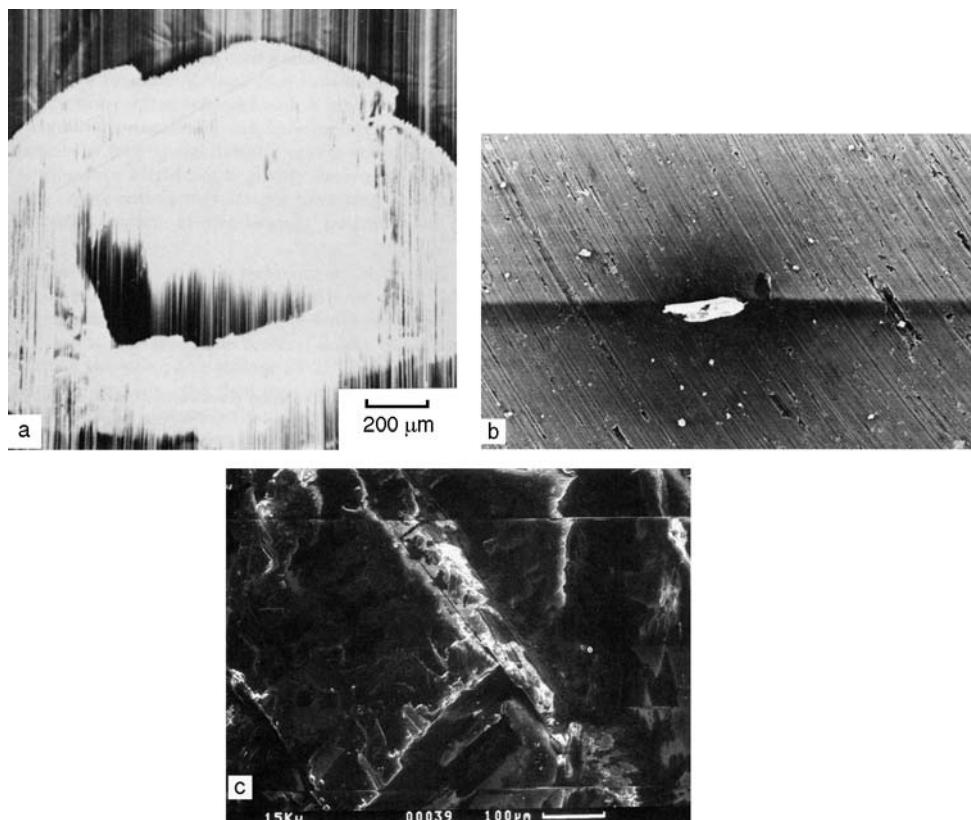


Figure 4.22 Image artifacts due to surface charging: (a) gross charging effects on TeflonTM; (b) negative charging of nonconductive particles on an Al substrate, which induces positive charge on the substrate as shown by dark contrast; (c) charging causes unstable signals as shown in discontinuities when scanning a calcite crystal. (Reproduced with kind permission of Springer Science and Business Media from: (a) and (c), J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science; (b), L. Reimer, *Scanning Electron Microscopy, Physics of Image Formation and Microanalysis*, 2nd ed., Springer-Verlag, Germany. © 1998 Springer-Verlag GmbH.)

It occurs when there are excessive electrons accumulated on the specimen surface where it is impinged by the electron beam. Accumulation of electrons on the surface builds up charged regions. In such regions, electric charges generate distortion and artifacts in SEM images. Figure 4.22 shows several examples of SEM image distortion due to surface charging.

Charging causes image distortion and artifacts because the charged regions deflect the incident electron probe in an irregular manner during scanning; charging alters secondary electron emission, and charging causes the instability of signal electrons. The third phenomenon can be understood as the sudden release of charge along a conducting path when charge has accumulated to a critical level in a particular region.

Surface charging should be a primary concern for nonconductive specimens such as ceramics, polymers and specimens containing biological substances. Surface charging is generally

not a problem for metallic specimens because their good electrical conduction ensures removal of excess electrons, which cause surface charging, to ground by electrical conduction between the specimen surface and ground. However, a metal oxide layer and nonconductive phases in metallic specimens may also generate surface charging.

The most common way to prevent charging is to coat a conductive film onto specimen surfaces to be examined. Conductive coating is ordinarily accomplished by means of either *vacuum evaporating* or *sputtering*. Vacuum evaporating deposits conductive atoms on a specimen in a vacuum chamber. The conductive substance has to be heated to a high temperature for it to evaporate. Sputtering strikes a conductive target with high energy ions which impart momentum to atoms of the target, enabling them to leave the target and deposit on the specimen. Sputtering is more widely used method than vacuum evaporating for conductive coating of SEM specimens.

Figure 4.23 shows a schematic illustration of sputtering mechanism. Argon gas is commonly used to produce ionized gas (plasma) in sputtering. Gold is often used as the target material. Sputter coating is more popular because of its short preparation time and coating uniformity on rough specimens. The coating thickness is normally controlled at about 5–20 nm depending on the magnification to be used in SEM examination. High magnification imaging requires a thin coating, 5 nm or less. Generally, we choose a thinner coating as long as it ensures conduction and surface coverage. A thick layer tends to be granular and cracked, which may distort the original topographic features of specimen surfaces.

A potential problem associated with sputtering is the possibility of thermal damage to sensitive specimens. Heating of the specimen during coating can cause cracking, pitting and melting, particularly for organic specimens. To avoid these problems, cool sputtering, known as *plasma-magnetron sputtering*, can be used. Plasma-magnetron sputtering can maintain the

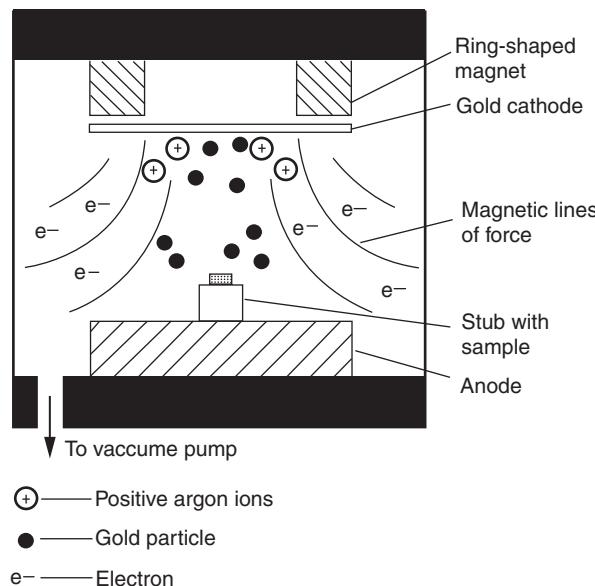


Figure 4.23 Sputtering gold onto a specimen surface.

specimen at ambient temperatures by introducing a magnetic field to manipulate the plasma flow and reduce heating of the specimen.

4.4.2 Preparation for Micro-Composition Examination

Any specimen prepared for topographic examination is generally suitable for composition examination using the BSE mode. Electrical conduction of specimens is still required for compositional examination. A specimen prepared for light microscopy examination may also be directly used for composition examination in an SEM. If the specimen is either not metallic or is a metallic specimen mounted in a nonconductive substrate, we should coat the specimen for composition examination as for topographic examination. It is common that an SEM specimen for composition examination is mounted in conductive epoxy resins or conductive thermoplastics before it is ground and polished, similar to the preparation of a light microscope specimen. The etching used for light microscopy is not necessary for compositional contrast using the BSE mode, while deeper etching is required for topographic contrast using the SE mode.

For polymer specimens, we can use a heavy metal element to stain the specimens to generate compositional contrast, similar to preparing polymer specimens for TEM. Compositional contrast resulting from staining can reveal a polymer specimen containing multiple phases because the staining level of the heavy metal element varies with molecular structures of polymer phases. Table 4.2 lists the staining agents for various polymer functional groups.

4.4.3 Dehydration

Dehydration is a special preparation technique required to examine a specimen containing water in an SEM. Any specimen derived from or containing biological substances may need processing to remove water. If water is not removed, the high-energy electron beam will heat the water in the specimen and cause it to burst through the specimen surface, thus destroying the surface morphology of the specimen. Dehydration may be accomplished either by *critical-point drying* or *freeze-drying*. These techniques remove water from a specimen without collapsing, flattening or shrinking the specimen. Figure 4.24 shows an example of using specimen dehydration before examining the behavior of bone cells on microgrooves of calcium

Table 4.2 Polymer Functional Groups and Staining Agents

Polymer groups	Staining agents
unsaturated hydrocarbons, alcohols, ethers, amine acids, esters	osmium tetroxide
unsaturated rubber	ebonite
saturated hydrocarbons: polyethylene (PE) and polypropylene (PP)	chlorosulfonic acid/uranyl acetate
amides, esters, PP	phosphotungstic acid
ethers, alcohols, aromatics, amines, rubber, bisphenol A, styrene	ruthenium tetroxide
esters, aromatic polyamides	silver sulfide
acids, esters	uranyl acetate

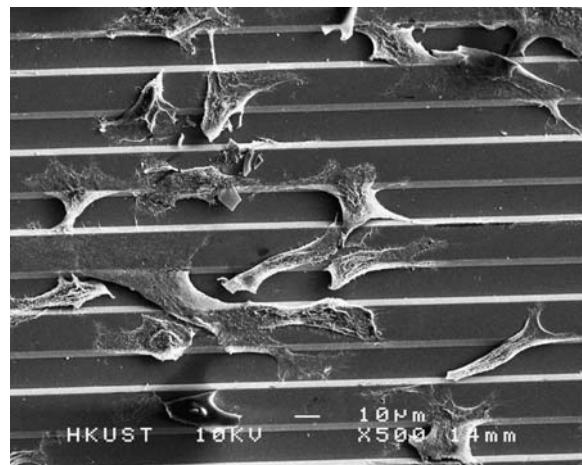


Figure 4.24 SEM micrographs of osteoblasts (a type of bone cell) on the microgrooved surface of calcium phosphate.

phosphate substrate. The cell morphology, which is critical for evaluating interactions between bone cells and the substrate, is preserved by critical point drying.

The critical point refers to the certain combination of temperature and pressure at which the liquid density is equal to the vapor density. At its critical point, liquid will become vapor and is easily removed. We cannot directly remove water using its critical point because the temperature and pressure of the critical point (374°C and 22 MPa) are too high and may damage the specimen. Alternatively, we can replace water with a transitional fluid that has a critical point with lower temperature and pressure. Liquid CO_2 or Freon is often used as the transitional fluid. The critical point for liquid CO_2 is 31.1°C and 7.4 MPa . The common procedure is described as follows. First, water content in a specimen is removed by dehydration with an ethanol series (30, 50, 75 and 100%). Then, the dehydrated specimen is transferred into an ethanol-filled and cooled chamber in the critical-point drying apparatus. The transitional fluid is introduced until it completely displaces ethanol in the chamber. The chamber is gradually heated and pressurized to reach the critical point of the transitional fluid. After reaching the critical point, the transitional fluid vaporizes, and this vapor is slowly released from the chamber until atmospheric pressure is reached. Then, we can retrieve the intact, dry specimen from the chamber.

The freeze-drying method uses sublimation (solid to vapor transformation) to dry specimens. This method requires freezing a specimen rapidly to below -80°C in a chamber. At that temperature, the chamber is degassed under a low pressure of below 0.1 Pa . The specimen can be dried in such a condition after several hours to several days.

References

- [1] Goldstein, J.I. and Yakowitz, H. (1977) *Practical Scanning Electron Microscopy*, Plenum Press, New York.
- [2] Goldstein, J.I., Romig, Jr., A.D., Newbury, D.E., Lyman, C.E., Echlin, P., Fiori, C., Joy, D.C. and Lifshin, E. (1992) *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd edition, Plenum Press, New York.

- [3] Flegler, S.L., Heckman, Jr., J.W. and Klomparens, K.L. (1993) *Scanning and Transmission Electron Microscopy, an Introduction*, W. H. Freeman and Co., New York.
- [4] Goodhew, P.J. and Humphreys, F.J. (1988) *Electron Microscopy and Analysis*, 2nd edition, Taylor & Francis Ltd, London.
- [5] Brandon, D. and Kaplan, W.D. (1999) *Microstructural Characterization of Materials*, John Wiley & Sons, Chichester.
- [6] Bozzola, J.J. and Russell, L.D. (1998) *Electron Microscopy*, 2nd edition, Jones and Bartlett Publishers, Sudbury.
- [7] Reimer, L. (1998) *Scanning Electron Microscopy, Physics of Image Formation and Microanalysis*, 2nd edition, Springer-Verlag, Germany.

Questions

- 4.1 Compare light microscopy, transmission electron and scanning electron microscopy in terms of optical arrangement, illumination source, working environment, imaging formation mechanism and specimen preparation; discuss their similarities and differences.
- 4.2 Given a specimen with surface fracture features with high differences up to $50 \mu\text{m}$, how do you select SEM operation parameters to ensure the whole field is in focus?
- 4.3 Estimate the effective magnifications of an SEM image showing on your computer screen when the probe size varies between 1 nm and $1 \mu\text{m}$.
- 4.4 How can you obtain SEM images with the best resolution by adjusting operational parameters?
- 4.5 When we find that the noise level is too high in an SEM image, how can we reduce noise by adjusting operational parameters?
- 4.6 Can backscattered electrons generate the same levels of resolution as secondary electrons? Why?
- 4.7 In an SEM image with composition contrast as shown in Figure 4.16b, which areas contain the heavy metallic elements: bright or dark? Explain.
- 4.8 Is it possible to eliminate surface charging of a polymer sample without conductive coating? Explain.

5

Scanning Probe Microscopy

Scanning probe microscopy (SPM) is a technique to examine materials with a solid probe scanning the surfaces. The SPM is relatively new for materials characterization compared with light and electron microscopy. It examines surface features whose dimensions range from atomic spacing to a tenth of a millimeter. The SPM is considered the most powerful tool for surface structure examination currently available because it lets us ‘see’ atoms. SPM started with the scanning tunneling microscope (STM) invented in 1982. An STM uses a tunneling current, a phenomenon of quantum mechanics, to examine material surfaces. The tunneling current flows through an atomic-scale gap between a sharp metallic tip and conducting surface atoms. The SPM added a popular member, the scanning force microscope (SFM), which was invented in the late 1980s. An SFM is commonly called an atomic force microscope (AFM). It uses near-field forces between atoms of the probe tip apex and the surface to generate signals of surface topography. The AFM is more widely used than the STM because it is not restricted to electrically conducting surfaces.

5.1 Instrumentation

The main characteristic of the SPM is a sharp probe tip that scans a sample surface. The tip must remain in very close proximity to the surface because the SPM uses *near-field interactions* between the tip and a sample surface for examination. This near-field characteristic eliminates the resolution limit associated with optical and electron microscopy as discussed in the previous chapters, because their resolution is limited by the *far-field interactions* between light or electron waves and specimens. Diffraction of light or electron waves associated with far-field interactions limit their resolution to wavelength scales. The near-field interactions in a SPM, however, enable us to obtain a true image of surface atoms. Images of atoms can be obtained by an SPM because it can accurately measure the surface atom profiles in the vertical and lateral directions. The lateral and vertical resolutions of an SPM can be better than 0.1 nm, particularly the vertical resolution. The lateral range of an SPM measurement is up to about 100 μm , and its vertical range is up to about 10 μm . However, the SPM must operate in a

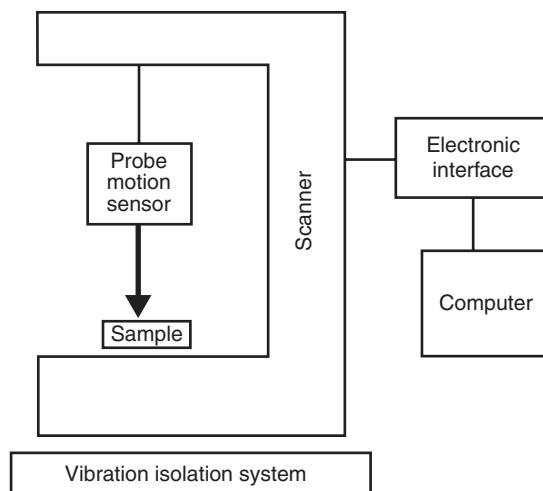
Table 5.1 Comparison of Surface Microscopy Techniques

	Optical Microscopy (reflected)	SEM	SPM
Operating environment	Ambient, liquid, vacuum	Vacuum	Ambient, liquid, vacuum
Depth of field	Small	Large	Medium
Depth of focus	Medium	Small	Small
Lateral Resolution	$1.0 \mu\text{m}$	$1\text{--}10 \text{ nm}$	$0.1\text{--}3.0 \text{ nm}$
Vertical Resolution	N/A ^a	N/A	0.01 nm
Magnification range	$1\times\text{--}2\times10^3\times$	$10\times\text{--}10^6\times$	$5\times10^2\times\text{--}10^8\times$
Sample preparation requirements	Polishing, etching	Dehydrated, often conductive coating	None
Sample requirements	Not completely transparent to light waves	No charge build-up on surface	No excessive height variations on surface

^a Not applicable.

vibration-free environment because the probe is only an atomic distance away from the examined surface. Table 5.1 provides a brief comparison between the SPM and reflected optical and scanning electron microscopes.

An SPM system consists of several basic components: probe and its motion sensor, scanner, electric controller, computer and vibration isolation system as schematically shown in Figure 5.1. These components are briefly introduced in the following sections.

**Figure 5.1** The main components of a scanning probe microscope (SPM).

5.1.1 Probe and Scanner

The probe is the basic SPM component that directly interacts with a sample surface to be examined or measured. Either the probe can scan over a stationary surface, or it can be stationary while the sample surface is moved by a scanner underneath.

The STM and AFM differ in probe material and in how the sensor works. In an STM, the probe is commonly made from tungsten wire. The tungsten wire conducts the tunneling current flowing between surface atoms and tip atoms. In the AFM, the probe tip is commonly made from SiO_2 or Si_3N_4 and mounted onto a cantilever spring. A probe motion sensor senses the spacing between the tip apex and the sample and provides a correction signal to control the piezoelectric scanner movement. In the STM, the tunneling current is used to sense the spacing between the probe and the sample surface. In the AFM, near-field forces between tip and sample detected by a beam deflection system are used to sense the spacing.

The scanner controls the probe that moves over the sample in three dimensions in a precise way. The scanning probe needs to be positioned with an accuracy of 1 pm (10^{-12} m) if atomic resolution is required. To achieve this level of precision, a scanner is made from piezoelectric materials.

Piezoelectric materials change shape in an electric field or generate an electric field due to shape change. The relative length of the change is proportional to applied voltage (E).

$$\frac{\Delta L}{L} = d_{ij}E$$

where d_{ij} is the piezoelectric coefficient; its typical value is about 0.3 nm mm V^{-1} . Thus, greater than 300 volts is required to achieve 1 μm movement for a 10 mm long piezoelectric tube.

Three-dimensional positioning of the scanner is obtained by placing three piezoelectric tubes in an orthogonal arrangement as shown in the illustration of the STM structure (Figure 5.2). The positioning accuracy of the scanner is determined by the characteristics of piezoelectric

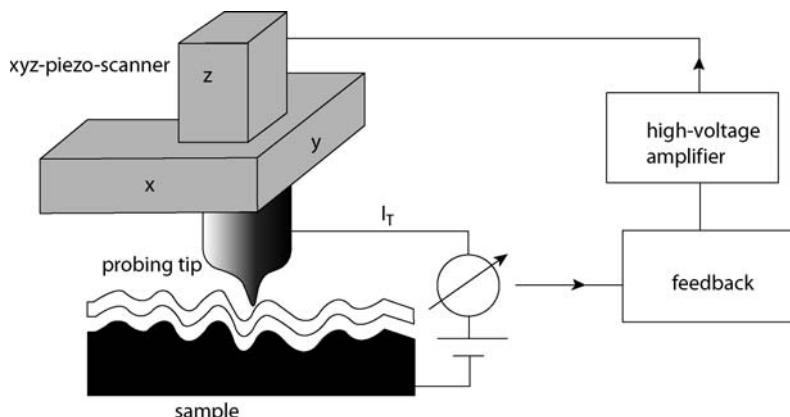


Figure 5.2 The scanning tunneling microscope (STM) system. The scanner moves the tip over the sample surface and the feedback loop keeps the tunneling current constant. (Reproduced with kind permission of Springer Science and Business Media from E. Meyer, J.H. Hug, and R. Bennewitz, *Scanning Probe Microscopy: the Lab on a Tip*, Springer-Verlag, Berlin. © 2004 Springer-Verlag GmbH).

effects. These effects include non-linear and hysteretic effects of piezoelectricity, as well as creep of piezoelectric materials. In order to obtain positioning accuracy, the scanner must be well-calibrated to eliminate or offset such piezoelectric effects.

5.1.2 Control and Vibration Isolation

A scanning probe microscope is a computer-based instrument, unlike light and electron microscopes that can operate without computers. The computer controls the scanner and probe through an electronic interface. The interface supplies the voltages that control the piezoelectric scanner, accepts the signal from the position sensing unit and contains the feedback control system for keeping the spacing between sample and tip constant.

The computer system processes the collected data; for some systems it even takes over feedback control. It can also correct possible data drift during instrumental instability as well as artifacts of tilt and curvature that can appear in topographic data due to piezoelectric effects. The computer generates SPM surface images using the collected data. The SPM images displayed by the computer usually are extensively processed by powerful software. For example, three-dimensional birds'-eye-view images and artificial color images are commonly used to enhance the contrast of images.

A scanning probe microscope must be isolated from any vibrations in its surroundings in order to achieve high resolution because its performance relies on precise control of the distance between the mechanical probe and sample surface. To maintain a constant near-atomic spacing during operation, any vibration from the environment must be eliminated. Common interference includes floor vibrations and acoustic vibration. Isolation from floor vibration is achieved by mounting the SPM on springs, rubber feet or air-damped feet. Isolation from acoustic vibration is achieved by operating the instrument in a solid box.

5.2 Scanning Tunneling Microscopy

5.2.1 Tunneling Current

Scanning tunneling microscopy (STM) relies on a tunneling current across a gap between the probe tip and a surface atom for surface examination. Electric current may penetrate an isolation gap between two conductors if high voltage is applied, but this is not the tunneling current. The tunneling current is a phenomenon of quantum mechanics. It results from electrons penetrating an energy barrier larger than the electron energy. To generate the tunneling current (I_t), a bias voltage is applied across a gap between the tip and the sample when the tip is kept near the sample. Without much knowledge of quantum mechanics, we may quantify tunneling current through the following equation.

$$I_t = V_b \exp(-Cd) \quad (5.1)$$

V_b is the bias voltage applied between the tip and the surface, C is a constant that depends on the nature of the sample material and d is the nearest distance between the tip and the sample surface. For generating a tunneling current, only a small bias voltage (typically between 1 mV and 4 V) is necessary when the gap is on the scale of interatomic distance. In an STM, the tunneling current (I_t) is between 10 pA and 10 nA. The tunneling current varies exponentially

when the distance changes as indicated in Equation 5.1. The tunneling current decays by an order of magnitude when the vacuum gap is changed by 0.1 nm. This is the reason that STMs provide resolution better than 0.1 nm, a truly atomic resolution.

According to quantum mechanics, the tunneling current is proportional to the local density of states (LDOS) of the sample at the Fermi level. The density of states is the number of electronic states per unit energy range in a solid. The Fermi level is the highest energy level occupied by electrons in metal at absolute zero temperature. The LDOS at the Fermi level is important with respect to many physical properties. In particular, it determines the numbers of electrons that can participate in electron conduction. The LDOS can be affected by surface chemistry. For example, a molecule adsorbed onto a metal surface may reduce the local density of states. Thus, the LDOS contour is not equal to the surface topographic contour. The surface topography is determined by the total charge density on the surface. It is important to note that a surface profile generated by an STM may differ from the surface topography of sample.

As shown in Figure 5.2, the tunneling current is used as the input signal for precisely controlling the motion of a piezoelectric scanner. The output signals of the feedback loop, determined by the tunneling current input, control the piezoelectric scanner in moving either up or down. A high-voltage amplifier provides the necessary electric potential to activate the piezoelectric scanner. Commonly, an STM is operated in a constant current mode. The spacing between tip and sample surface maintains the tunneling current as constant during scanning.

5.2.2 Probe Tips and Working Environments

A probe tip to detect a tunneling current is made from wire of tungsten or Pt–Ir alloys. An electrochemical etching method produces a sharp tip of tungsten wire. The surface oxide formed on the tip during etching must be removed to ensure good electric conductivity. Other methods of producing the tip include mechanical cutting or grinding. The size and sharpness determine the lateral resolution of an STM. Ideally, for atomic resolution, a smooth tip should have a radius of a single atom. This is impossible to achieve practically, and in fact is not necessary because only the outermost atoms on the tips dominate the measured tunneling current. Thus, a single-atom asperity on the tip can function like a single-atom probe for most purposes.

In principle, the STM can work in air and liquid environments. However, most STM work has been done in ultra-high vacuum. This is because most sample surfaces in an air environment quickly develop an oxide layer. An ultra-high vacuum environment can prevent possible surface oxide formation and maintain a conductive sample surface. Also, low temperature is preferred because it reduces thermal drift and diffusion of atoms and helps to obtain a static surface image of atoms. However, an elevated temperature provides an environment for observing dynamic processes of atoms and molecules.

5.2.3 Operational Modes

There are four operational modes in the STM: constant current, constant height, spectroscopic and manipulation modes. The most commonly used is the *constant current mode*. In the constant current mode, the feedback loop controls the scanner moving up and down to maintain a constant tunneling current. A corrugation contour of the LDOS during scanning can be obtained in the constant current mode. The corrugation contour resolves the locations of surface atoms of the sample because the LDOS is very sensitive to atom locations. Although the LDOS contour

is not identical to surface topography, an image generated in constant current mode often fairly accurately depicts the true topography when scanning is on a scale of a few nanometers.

The *constant height mode* can provide much higher scanning rates than that of constant current. It is useful for observing dynamic processes in which timing is important. The constant height mode can be obtained by turning off the feedback-loop control. This mode requires that the sample surface is well leveled without tilt. There is also a risk of crashing the tip onto the sample surface during scanning.

The *spectroscopic mode* refers to the operation of recording the tunneling current as a function of either tip–sample spacing or bias voltage. The spectroscopic mode is useful for studying surface properties such as superconductivity and molecular adsorption on metal. To record an STM spectrograph, the tip is stopped at a position above the sample, and the tunneling current is recorded while changing either the spacing or bias voltage between the tip and sample.

The *manipulation mode* refers to operations of relocating or removing atoms on a surface. Figure 5.3 illustrates relocating an adatom (an atom attached to the surface) by vertical and lateral manipulation. In vertical manipulation, an adatom is transferred from the surface to the probe tip, and then is deposited at another location. The attachment and detachment of the atom to and from the tip is controlled by voltage pulses. In lateral manipulation, an adatom remains adsorbed on the surface and is moved laterally by the tip when there is a weak bond between the adatom and the tip. This technique has been used for serious as well as fanciful work. For example, a scientist at the IBM Corporation made an ‘IBM’ symbol with a few atoms using the manipulation mode in an STM.

5.2.4 Typical Applications

The most exciting function of an STM is its ability to image atoms. The first atomically resolved STM image was of a silicon surface in 1982. A Si (111) surface was obtained by thermally

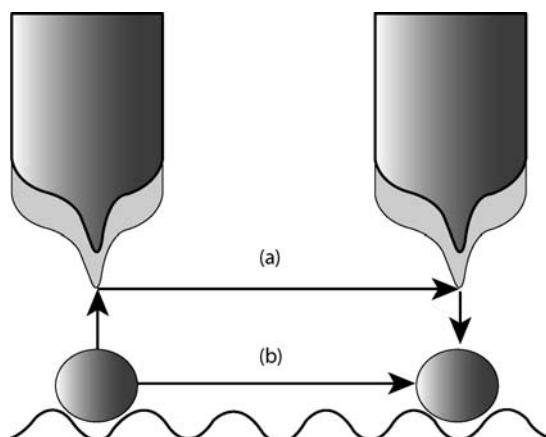


Figure 5.3 Manipulation mode in the STM: (a) vertical manipulation, where an adatom forms a strong bond with the tip and is detached from the surface, then it is transported by the tip and redeposited on the surface; and (b) lateral manipulation where the tip forms a weak bond with an adatom and moves it along the line of manipulation. (Reproduced with kind permission of Springer Science and Business Media from E. Meyer, J.H. Hug, and R. Bennewitz, *Scanning Probe Microscopy: the Lab on a Tip*, Springer-Verlag, Berlin. © 2004 Springer-Verlag GmbH.)

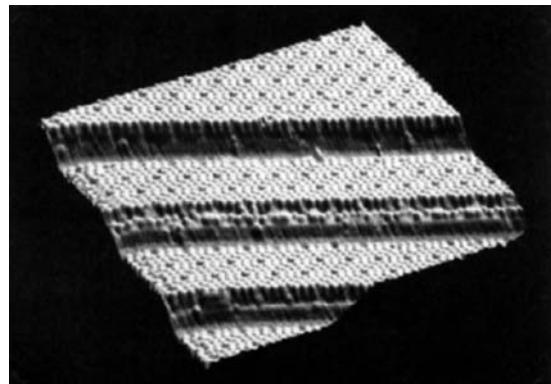


Figure 5.4 An STM image of a $32\text{ nm} \times 36\text{ nm}$ area on the Si (111) planes. Surface atoms have been reconstructed, showing 7×7 symmetry. Three atomic steps separate terraces. (Reproduced with permission from R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy*, Cambridge University Press, Cambridge, UK. © 1994 Cambridge University Press.)

annealing a silicon single crystal above 1000°C under ultra-high vacuum. The thermal treatment produced a sublimated surface SiO_2 layer. After slowly cooling, the surface atoms were reconstructed as a special lattice pattern. Figure 5.4 shows the reconstructed Si (111) image of a $32\text{ nm} \times 32\text{ nm}$ area, in which there are three surface steps separating reconstructed terraces. The STM has been successfully used to produce atomic images of various materials including metals (such as gold, copper, platinum and nickel) and layered crystals (such as graphite and MoS_2). The STM can even image adsorbed atoms and single molecules on surfaces.

While an STM image represents the topography of a surface, it must be interpreted carefully. The image may be altered by the bias voltage applied to a sample because the image reflects the surface of equal LDOS. For example, in an STM study of oxygen adsorption on an n-type GaAs (111) surface, an adsorbed oxygen atom may appear either as a protrusion or as a depression, depending on bias voltage directions (Figure 5.5). Correctly interpreting STM images requires an understanding of surface chemistry and physics, as well as the mechanism of the STM.

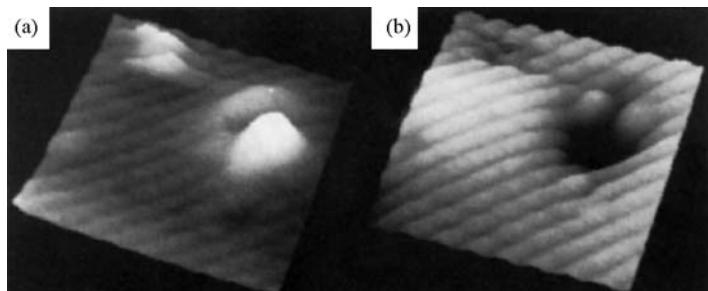


Figure 5.5 Oxygen adsorbed onto a GaAs surface: (a) the sample is under negative bias voltage (-2.6 V with respect to the tip); and (b) the sample is under positive bias voltage ($+1.5\text{ V}$ with respect to the tip). (Reproduced with permission from R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy*, Cambridge University Press, Cambridge, UK. © 1994 Cambridge University Press.)

In general, the STM's ability to obtain atom images is clearly outstanding. Unfortunately, there are some limitations, particularly the requirements that both the tip and sample surface are good conductors. This severely limits the materials that can be studied and has led to the development of scanning force microscopy (atomic force microscopy) which is described in Section 5.3.

5.3 Atomic Force Microscopy

5.3.1 Near-Field Forces

Like the STM, the atomic force microscope (AFM) uses a very sharp tip to probe and map sample topography. The AFM detects near-field forces between the tip and sample, instead of detecting the tunneling current. Knowledge of the near-field forces acting between tip and sample is necessary for us to understand the AFM working principles. There are several types of the near-field forces which are briefly described as follows.

Short-Range Forces

The short-range forces refer to atomic forces between atoms when their distance is close to atomic spacing. Overlapping of their electron wave functions, which can be considered as overlapping of their electron clouds, causes either attractive or repulsive forces. The forces will be attractive when the overlap reduces their total energy, and the forces will be repulsive when the Pauli exclusion principle is in play. The attractive short-range forces are in the range of 0.5–1 nN per pair of interactive atoms between a typical tip and a sample. The decay length of the forces is in the range of 0.05–0.2 nm, larger than that of a tunneling current. In any case, the variation of the short-range forces on the atomic scale makes the AFM capable of obtaining topographic images of atoms.

Van der Waals Forces

The van der Waals forces are the interactive forces between dipoles of molecules. The dispersion force, a type of van der Waals force between dipoles that arises from thermal fluctuation or electric field induction, is always present between molecules. The van der Waals forces (F_{vdW}) between tip and sample can be estimated by the following equation, if we assume the tip is a sphere with radius, R .

$$F_{vdW} = \frac{HR}{6d^2} \quad (5.2)$$

H is the *Hamaker constant* and is in the order of 10^{-19} J, and d is the spacing between tip and sample. For a tip radius of 30 nm and vacuum gap of 0.5 nm, the van der Waals forces are around 2 nN.

The van der Waals forces are significantly affected by the medium of the gap between tip and sample. For example, when the medium is water rather than a vacuum, the forces will be greatly reduced because the dielectric constant and refractive index of water are closer to those of a solid than of a vacuum.

Electrostatic Forces

Electrostatic forces refer to the interactive forces between the electric charges of tip and sample. The charges can be easily trapped at a sample surface and at a tip if they are insulated.

Electrostatic forces can also act between a conductive sample and a tip if there is an electric potential difference. The electrostatic forces can be estimated by the follow equation if the spacing between a tip and a sample is small.

$$F_{els} = \pi \varepsilon_0 \frac{R}{d} (V_{bias} - V_{cpd})^2 \quad (5.3)$$

ε_0 is permittivity of vacuum, V_{bias} is the bias voltage applied between tip and sample and V_{cpd} represents the contact potential difference due to the difference in work functions of the tip and the sample. For a tip radius of 20 nm and a distance between sample and tip (d) of 0.5 nm, the electrostatic force is about 0.5 nN when the difference between V_{bias} and V_{cpd} equals 1 V.

We, however, should know that the electrostatic force is considered as long-range compared with the short-range atomic force. The force is proportional to d^{-2} as described by Coulomb's law, not proportional to d^{-1} as shown in Equation 5.3.

Capillary Forces

Capillary forces are forces resulting from water vapor condensation between tip and sample. In an air environment, water vapor is likely to form a condensed nucleus there. The water nucleus condensed between tip and sample will form a meniscus as schematically shown in Figure 5.6, similar to the water surface in a capillary tube. The capillary force on the tip can be estimated.

$$F_{cap} = \frac{4 \pi R \gamma \cos\theta}{1 + d/[R(1 - \cos\phi)]} \quad (5.4)$$

γ is the surface tension of water nucleus. The meanings of angles θ and ϕ , and d and R are indicated in Figure 5.6. For a tip radius of 100 nm, the maximum capillary force reaches about 90 nN. This force is significantly larger than other forces being discussed. If an attractive force in the order of 10^{-8} – 10^{-7} N is recorded, it most likely arises from the capillary. Capillary forces play an important role in AFM measurements in air. We can reduce these forces by covering the sample with hydrophobic molecules. AFM measurements in liquid are not affected by capillary forces.

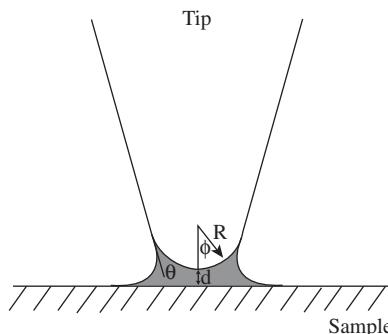


Figure 5.6 Capillary force between the tip and sample due to water formation between them.

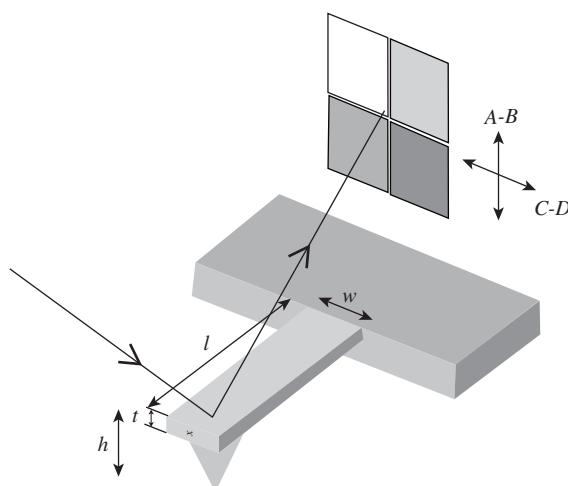


Figure 5.7 Detection of cantilever deflection by a laser beam reflected from the rear side of the cantilever. The reflection is measured by four-quadrant photodiodes. Cantilever deflection is converted to the change in beam position by photodiodes. t , l and w , cantilever dimensions; h , distance from back of cantilever to probe tip. (Reproduced with kind permission of Springer Science and Business Media from E. Meyer, J.H. Hug, and R. Bennewitz, *Scanning Probe Microscopy: the Lab on a Tip*, Springer-Verlag, Berlin. © 2004 Springer-Verlag GmbH.)

5.3.2 Force Sensors

The key element of the AFM, or the main difference in structure from the STM, is its microscopic force sensor. The most common force sensor is a cantilever. The tip is mounted at the end of a cantilever as shown in Figure 5.7. The force detection is based on a beam deflection method. The force between the tip and sample generates elastic bending of the cantilever. The amount of bending is monitored and recorded by position-sensitive photodiodes which are arranged in four quadrants. The photodiodes receive the laser beam reflected from the rear side of the cantilever, which often has a thin metal coating, making it a mirror. Any small deflection of the cantilever will tilt the laser beam and change its striking position on the photodiodes. The difference between the two photodiode signals indicates the amount of cantilever deflection. The amount of deflection, in turn, can be mathematically converted into the force on the tip, according to the elastic properties of the cantilever. The distance between the cantilever and the photodiodes is about three orders of magnitude greater than the length of the cantilever. Such an optical arrangement can greatly magnify the tip motion and provide extremely high sensitivity. Other methods have been used for detecting cantilever deflection, such as interferometry and piezoelectricity. The photodiode arrangement, however, is most commonly used because it is inexpensive, easy to use and acceptably accurate.

There are two types of cantilever: single beam (Figure 5.8), and triangular beam (Figure 5.9). A single-beam cantilever beam is usually made from silicon with dimensions of about $100\text{--}500\ \mu\text{m}$ long and $0.5\text{--}5\ \mu\text{m}$ wide. The triangular beam is commonly made from silicon nitride. Most cantilevers are produced with integrated tips in a micro-fabrication process, similar to the process for manufacturing integrated circuits in the electronics industry. Such a tip can be envisaged as a cone with an opening angle and a finite radius at the apex. The apex

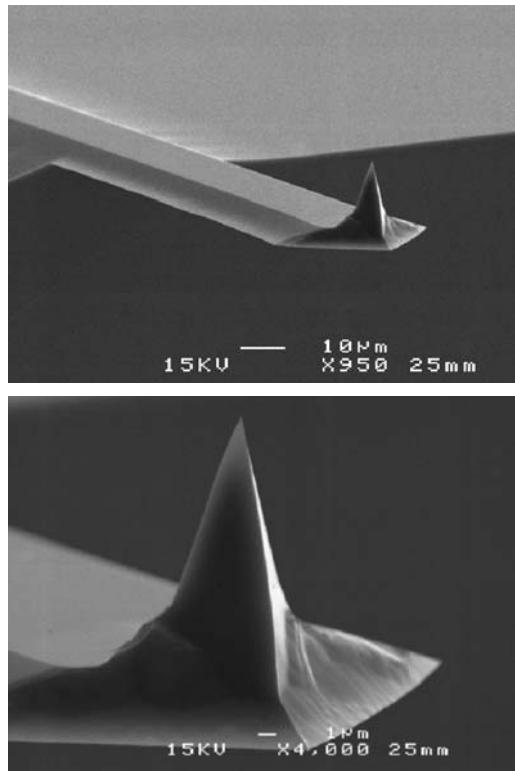


Figure 5.8 SEM micrographs of a single beam with integrated tip: (a) 950 \times ; and (b) 4000 \times .

radius and open-angled tip shape are crucial parameters for controlling the AFM lateral resolution. A tip radius of about 20 nm has been achieved. The open angle, however, is often limited by etching characteristics of crystalline materials. Etched surfaces of crystalline materials are often comprised of certain crystallographic planes. For example, the silicon tip pointing to the $\langle 100 \rangle$ direction has a cone angle of about 54.7° , which is the angle between $\{100\}$ and $\{111\}$ planes, because $\{111\}$ planes are the etched surface of crystalline silicon.

5.3.3 Operational Modes

Generally, the operational modes in an AFM can be divided into *static* or *dynamic modes*. In static modes, the cantilever statically deflects to a certain degree and the feedback loop maintains this set value of deflection during scanning. In the dynamic modes, the cantilever oscillates at a designated frequency, and the feedback loop tries to maintain a set value of amplitude of oscillation during scanning. In static modes, the tip on the cantilever physically ‘touches’ the surface to be examined. Thus, static modes must be of a *contact type*. The dynamic modes include both contact and non-contact types. The most widely used dynamic mode is the *intermittent contact mode*, also called the *tapping mode*. The tapping mode reduces possible damage to a sample during scanning and also can provide information on chemical and physical properties of the surface, in addition to topography.

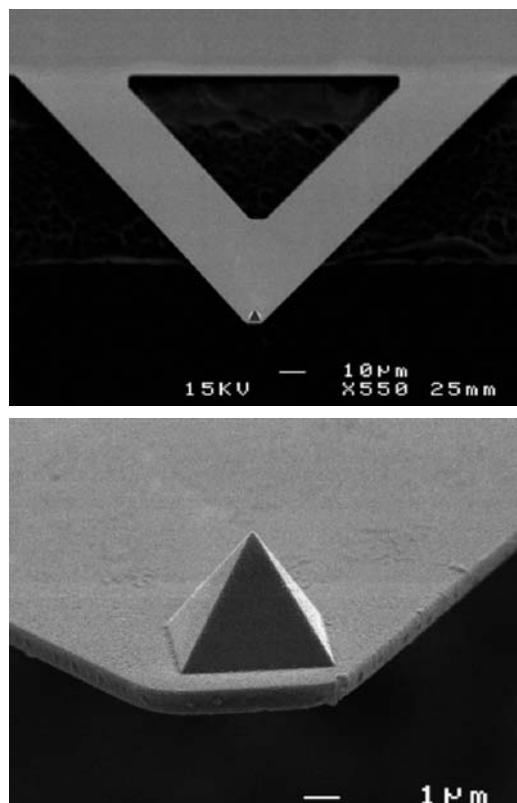


Figure 5.9 SEM micrographs of a triangular silicon nitride cantilever with integrated tip: (a) 550 \times ; and (b) higher magnification view.

Figures 5.10a and 5.10b illustrate the difference between interaction forces in contact and non-contact modes. In the non-contact mode (Figure 5.10a), all forces between tip and sample are attractive, and the short-range force between the outermost atoms of tip and sample atoms makes a significant contribution to imaging. In the contact mode (Figure 5.10b), a repulsive force occurs between tip apex atoms and the surface atoms, even though the total force is still attractive.

The difference between the static and dynamic modes can be illustrated using a graph of force versus distance between the tip and a sample. In Figures 5.10c and 5.10d, solid lines represent the total force and the dashed lines represent the short-range force. The dynamic mode uses stiff cantilevers while static mode uses soft cantilevers. Figure 5.10c illustrates a case of dynamic mode. The closest distance between the tip and the sample is marked as z , where the situation is comparable to that illustrated in Figure 5.10a. The arrow marked A indicates the typical amplitude of tip oscillation. The closest distance is the turning point of force curve in order to prevent the tip from being trapped at the sample surface.

Figure 5.10d illustrates a case of contact mode. The contact position (z) is on the repulsive side of the short-range force curve. The arrow pointed to the left on Figure 5.10d represents the phenomenon of tip *jump-to-contact* with the sample caused by cantilever elastic bending when

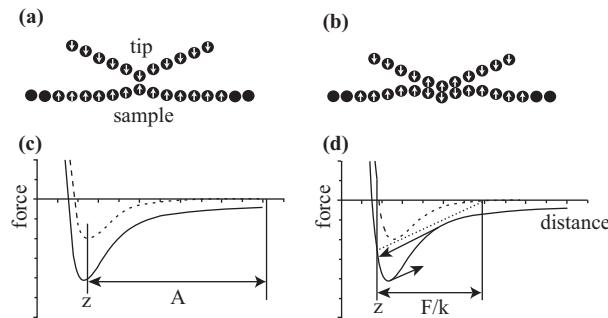


Figure 5.10 Interaction forces between a tip and sample atoms in: (a) non-contact mode; and (b) contact mode. Force versus distance between the tip and sample in: (c) non-contact mode; and (d) contact mode. The solid line marks total force between tip and sample; dashed line marks short-range force; A, oscillation amplitude of the tip; arrows mark the jump-in and spring-off contact in the static mode; and the straight broken line marks the value of cantilever spring constant (k). (Reproduced with kind permission of Springer Science and Business Media from E. Meyer, J.H. Hug, and R. Bennewitz, *Scanning Probe Microscopy: the Lab on a Tip*, Springer-Verlag, Berlin. © 2004 Springer-Verlag GmbH.)

the tip is brought near to the surface. The arrow pointed to the right represents the phenomenon of tip *jump-off-contact* caused by elastic recovery of the cantilever when the tip is retracted from the surface. The phenomena of jump-to and jump-off are characteristics of cantilever behavior in the contact mode. The ‘jump positions’ on the force curve depend on the spring constant of the cantilever (k).

Static Contact Modes

The static contact mode is often simply called the contact mode. In this case, the tip is brought in contact with a sample surface and the cantilever is constantly deflected during scanning, as schematically shown in Figure 5.11. The tip apex is brought into the region of repulsive force, while the rest of the tip experiences attraction to the sample. The cantilever force, proportional to the vertical deflection, will balance with forces on the tip. The short-range repulsive force experienced on the tip apex is most sensitive to topographic features of a sample. Thus, this

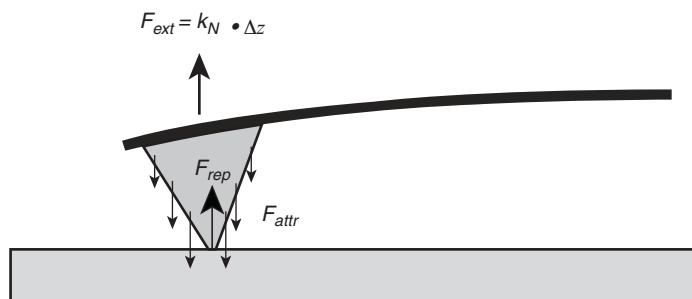


Figure 5.11 Force balance in the contact mode. The attractive long-range forces (F_{attr}) are balanced by the short-range repulsive force (F_{rep}) at the contact area and the spring force of the cantilever. (Reproduced with kind permission of Springer Science and Business Media from E. Meyer, J.H. Hug, and R. Bennewitz, *Scanning Probe Microscopy: the Lab on a Tip*, Springer-Verlag, Berlin. © 2004 Springer-Verlag GmbH.)

short-range force will provide high resolution images of topography. A three-dimensional topographical map of the surface can be constructed by plotting the local sample height versus the horizontal position of the tip.

The resolution of the contact mode depends on a contact area at the tip apex. The contact diameter (a) can be estimated with the Hertz model in which the contact area increases with applied force (F) on the tip due to elastic deformation between tip and sample.

$$a = 2(DRF)^{1/3} \quad (5.5)$$

where R is tip radius and D represents combined elastic properties of the tip and the sample.

$$D = \frac{1 - v_t^2}{E_t} + \frac{1 - v_s^2}{E_s} \quad (5.6)$$

With Young's modulus of the tip (E_t) and surface (E_s) in the order of $1.5 \times 10^{11} \text{ Nm}^{-2}$, R of 90 nm and applied F in the range 1–100 nN, a will be several nanometers. v is Poisson's ratio. This is not small enough to resolve an atomic image. The contact area diameter can be reduced further in ultra-high vacuum and in liquid because long-range forces between tip and sample are reduced. Nevertheless, in general, it is difficult to obtain atomic resolution with the contact mode.

Lateral Force Microscopy

Lateral force microscopy (also called *friction force microscopy*) is a variation of the static contact mode. When the cantilever moves in the lateral direction across the sample surface during scanning, the friction force between tip and sample will generate lateral force on the cantilever and will distort it as illustrated in Figure 5.12. The degree of cantilever torsion will be detected by the four-quadrant photodiodes and converted to image signals for lateral force contrast. The measurement of torsion provides friction characteristics of microscopic surface areas with a high resolution. Lateral force microscopy also characterizes the friction as a function of tip velocity moving on the surface. The lateral force contrast cannot only be generated by friction variation but also by geometric features, such as steps and holes. The

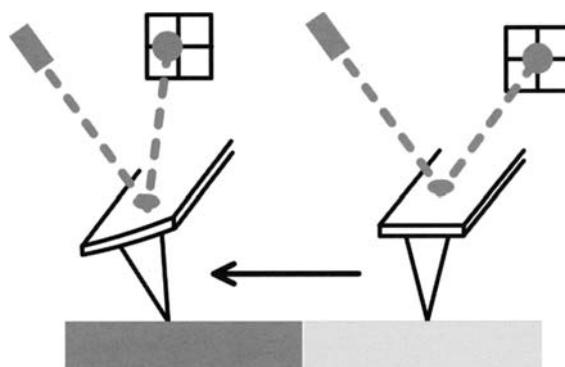


Figure 5.12 Detection of distortion of the cantilever due to lateral force between tip and sample in the lateral force microscope. (Reproduced with permission from P.C. Braga and D. Ricci (eds), *Atomic Force Microscopy*, Humana Press. © 2004 Humana Press.)

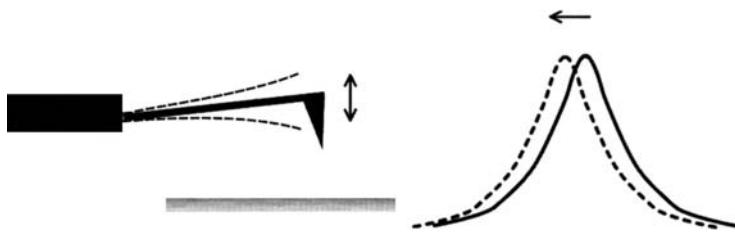


Figure 5.13 The dynamic non-contact mode. The frequency shift of the cantilever oscillation is used to operate the feedback loop. (Reproduced with permission from P.C. Braga and D. Ricci (eds), *Atomic Force Microscopy*, Humana Press. © 2004 Humana Press.)

contrast may even be generated by the local variation of normal force due to the chemical inhomogeneity of materials.

Dynamic Operational Modes

Several dynamic modes have been developed to overcome the limits of static contact modes. The dynamic modes include non-contact, intermittent contact (*tapping mode*) and contact types. The dynamic contact mode is also called *force modulation* and is described in this section.

Dynamic Non-Contact Mode

This mode is often simply called the non-contact mode. This mode can provide true atomic resolution and image quality comparable to an STM. The cantilever is excited by the piezo-actuator to oscillate at or near its resonant frequency. The vibrating tip is brought near a sample surface, but not touching the surface, to sense the weak attractive force between tip and sample instead of strong repulsive force in the contact mode. As illustrated in Figure 5.13, the interactions between the tip and the sample shift the oscillation frequency. The frequency shift signals are used to control the tip–sample distance. The interaction forces detected in the non-contact mode provide excellent vertical resolution. However, this mode cannot be operated in a liquid environment. Even in air, liquid attached on the tip can cause failure in operation.

Tapping Mode

This mode is an intermittent contact mode in which the cantilever oscillates and brings the tip to lightly touch the surface to provide high resolution and then lifts the tip off the surface to avoid surface damage as illustrated in Figure 5.14. As the oscillating cantilever begins



Figure 5.14 The tapping mode. The oscillation of the cantilever is damped when the tip touches the sample surface at each cycle. The image is formed by keeping the oscillation amplitude constant. (Reproduced with permission from P.C. Braga and D. Ricci (eds), *Atomic Force Microscopy*, Humana Press. © 2004 Humana Press.)

to intermittently contact the surface, the cantilever oscillation amplitude is reduced due to energy loss caused by the tip contacting the surface. The reduction of oscillation amplitude from the amplitude in the free oscillation state can be used to identify and measure surface features. During operation, the oscillation frequency is maintained constant by a feedback loop. Selection of the optimal oscillation frequency of the cantilever is done by computer software. The tip force on the sample is automatically set and maintained at the lowest possible level. When the tip passes over a bump in the surface, the cantilever has less room to oscillate and the amplitude of oscillation decreases. Conversely, when the tip passes over a depression, the cantilever has more room to oscillate and the amplitude increases (to approach the maximum free-air amplitude). The oscillation amplitude of the tip is measured by the detector and input to the controller electronics. The digital feedback loop then adjusts the tip–sample distance to maintain constant amplitude and force on the sample.

The tapping mode inherently prevents the tip from sticking to the surface and from damaging the sample surface during scanning. Unlike the contact mode, in this mode when the tip contacts the surface, it has sufficient oscillation amplitude to overcome the tip–sample adhesion forces. Further, the surface material is not pulled sideways by shear forces since the applied force is always vertical. This is particularly important for soft samples such as polymers and biological materials.

Phase imaging is a powerful extension of the tapping mode that provides nanometer-scale information about surface structures often not revealed by other SPM techniques. By mapping the phase of the cantilever oscillation during the tapping-mode scan, phase imaging goes beyond simple topographical mapping. It can detect variations in properties including composition, adhesion, friction and viscoelasticity.

In phase imaging, the phase lag of the cantilever oscillation relative to the signal sent to the cantilever's piezoelectric actuator is simultaneously monitored and recorded as illustrated in Figure 5.15. The phase lag is very sensitive to variations in material properties such as adhesion and viscoelasticity. Once a force microscope is engaged in the tapping mode, phase imaging is enabled simply by displaying a second image and selecting the phase data type in software. Both the topography and phase images are viewed side-by-side in real time. Phase imaging can also act as a technique to enhance real-time contrast of topographic images, because phase imaging highlights edges and is not affected by large-scale height differences. Thus, it provides clear observation of fine features such as grain edges, which can be obscured by rough topography.

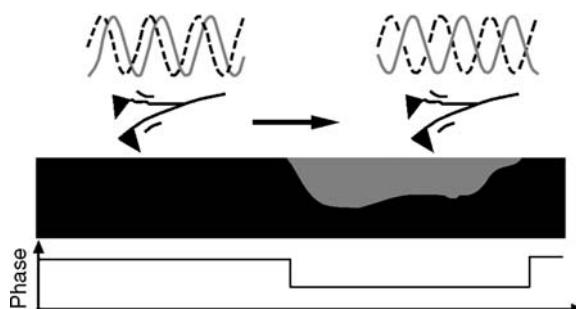


Figure 5.15 Phase image in the tapping mode. The phase lag of cantilever oscillation is detected. The phase lag is very sensitive to variations in material properties such as adhesion and viscoelasticity.

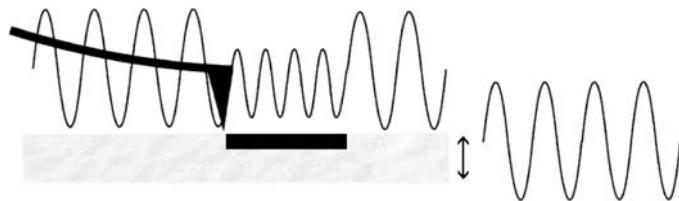


Figure 5.16 Force modulation. The tip is kept in contact with the sample. Differences in local properties of the sample will be reflected in the amplitude change of the cantilever. (Reproduced with permission from P.C. Braga and D. Ricci (eds), *Atomic Force Microscopy*, Humana Press. © 2004 Humana Press.)

Force Modulation

Force modulation imaging is the dynamic contact mode that identifies and maps differences in surface stiffness or elasticity. These techniques use a variety of surface properties to differentiate among materials where topographical differences are small or not measurable.

In the force modulation operation, a low frequency oscillation is induced (usually large enough to bring the tip in contact with the sample); when the tip contacts the sample, a constant force on the sample is applied. The cantilever deflection during scanning is recorded. Under the same applied force, a stiffer area on the sample will undergo less elastic deformation than a soft area. Thus, a stiffer area generates greater resistance to the vertical oscillation and less bending of the cantilever as illustrated in Figure 5.16. The variation in cantilever deflection amplitude is a measure of the relative stiffness of the surface. Topographical information is collected simultaneously with the force modulation data.

5.3.4 Typical Applications

Static Mode

The static contact mode is the simplest mode to use for obtaining basic topographic information of solid surfaces. Figure 5.17 is an image of surface crystalline polyethylene obtained with the contact mode. The contact mode was able to detect the nano-scale roughness of crystalline polyethylene, which is believed to be chain folds of polyethylene. The scale of the topography in Figure 5.17 implies that the primary function of an AFM is not the same as a surface profiler

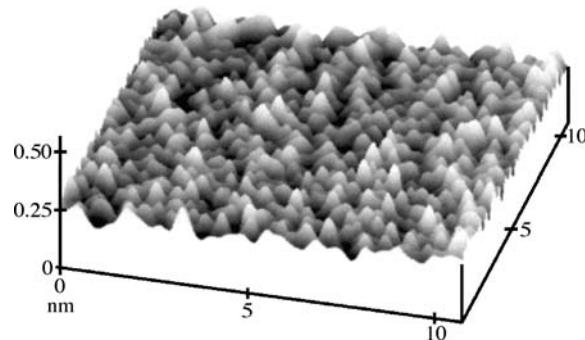


Figure 5.17 Topography of crystalline polyethylene. The chain folds of the crystal structure generate the rough appearance. (Reproduced with permission from B.D. Ratner and V.V. Tsukruk (eds), *Scanning Probe Microscopy of Polymers*, American Chemical Society, Washington D.C. © 1998 American Chemical Society.)

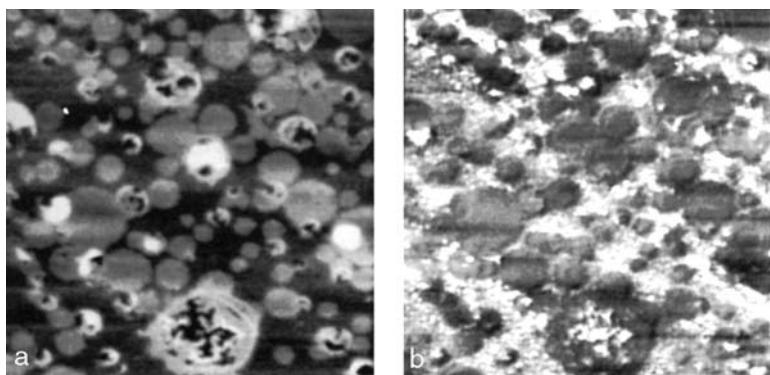


Figure 5.18 A scanning force microscope (SFM) image ($5 \mu\text{m} \times 5 \mu\text{m}$) of a mixed film of fluorocarbon and hydrocarbon carboxylates. The circular domains were assigned to the hydrocarbon component and the surrounding flat film to the partially fluorinated component: (a) topographic image; and (b) lateral force image obtained simultaneously with the topography. The measured lateral force generates higher contrast due to the difference in friction. (Reproduced with permission from R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy*, Cambridge University Press, Cambridge, UK. © 1994 Cambridge University Press.)

for surface roughness measurement. It means that AFM samples should have limited height variations on their surfaces, usually less than $1 \mu\text{m}$, because vertical movement of the scanner is limited. Interpretation of AFM images such as Figure 5.17 also relies on knowledge of the tip being used because the tip apex size determines the resolution of the topography. The distortion of topography due to a blunt tip might have hindered revealing the polymer chain folds protruding from the surface.

Lateral force microscopy, as one type of contact mode, can generate contrast in an image from geometrical features such as surface steps and holes because such features change local friction. More importantly, contrast can be generated by materials whose surface properties vary. For example, Figure 5.18 shows lateral force contrast on a film consisting of two types of polymers. Comparing the two images in Figure 5.18, we realize the difference between material-specific contrast and topographic contrast.

There has been a general trend since the mid-1990s for the static contact mode for surface topography to be gradually replaced by the dynamic non-contact or tapping modes, except for materials examined in a liquid environment. The dynamic modes provide better image resolution and are less likely to damage the sample surface.

Dynamic Non-Contact Mode

The dynamic non-contact mode can provide atomic resolution in an ultra-high vacuum, even though it is more difficult to generate an atomic image with an AFM than with an STM. Figure 5.19 shows an example of atomic resolution imaging with a non-contact mode. The hydrogen atoms absorbed on the Si (100) surface in an ultra-high vacuum of $2 \times 10^{-8} \text{ Pa}$ are revealed as the bright spots in Figure 5.19. The interaction between the hydrogen atom on the topmost layer and the silicon tip apex generates a frequency shift of cantilever oscillation. The frequency shift signals were converted to image contrast that reveals the locations of hydrogen atoms as bright spots.

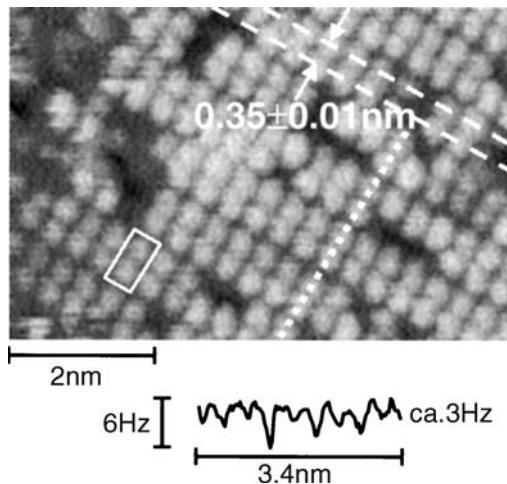


Figure 5.19 A non-contact mode image of a hydrogen-terminated silicon monohydride surface. Line profile of frequency shift along the white dotted line indicates the location of hydrogen atoms. The initial frequency of the cantilever is about 3 Hz. (Reproduced with kind permission of Springer Science and Business Media from S. Morita, R. Wiesendanger, and E. Meyer (eds), *Noncontact Atomic Force Microscopy*, Springer-Verlag GmbH, Berlin. © 2002 Springer-Verlag GmbH.)

Tapping Mode

The tapping mode is becoming popular for topographic measurements. As an example, Figure 5.20 shows an image, created in the AFM by tapping mode, of the topography of a high-speed transistor. The image clearly reveals the multiple layers of $\text{Al}_x\text{Ga}_{(1-x)}\text{N}$ deposited

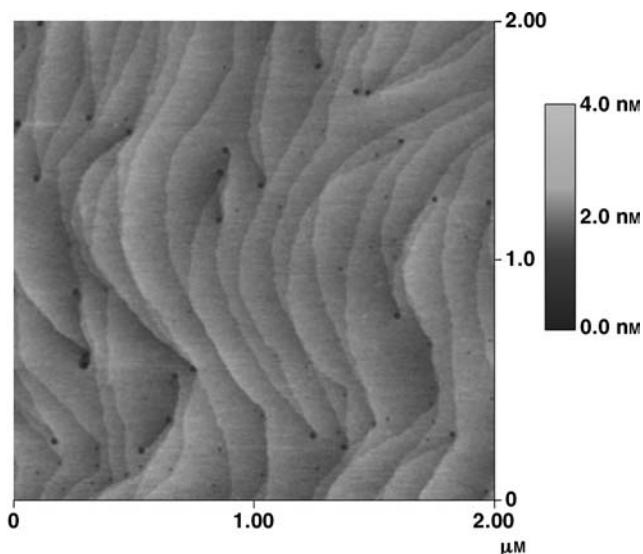


Figure 5.20 Tapping-mode image of multiple layers of $\text{Al}_x\text{Ga}_{(1-x)}\text{N}$ deposited on a sapphire substrate. (Reproduced with permission of Kei May Lau.)

on a sapphire substrate by metal–organic-chemical vapor deposition (MOCVD) with very precise control.

Force Modulation

Force modulation as a dynamic contact mode has the advantage of being able to distinguish materials with difference elastic properties. For example, Figure 5.21 shows the comparison between topographic and force modulation images of carbon fiber-reinforced epoxy. The topographic image reveals the difference in height of a composite cross section (Figure 5.21a), but it does not provide clear distinction between the fibers and matrix. The force modulation image (Figure 5.21b), resulting from the elastic stiffness difference between the fibers and matrix, reveals the location of individual carbon fibers in the cross-section of the composite sample. High stiffness of carbon fibers compared to epoxy generates high contrast in the image.

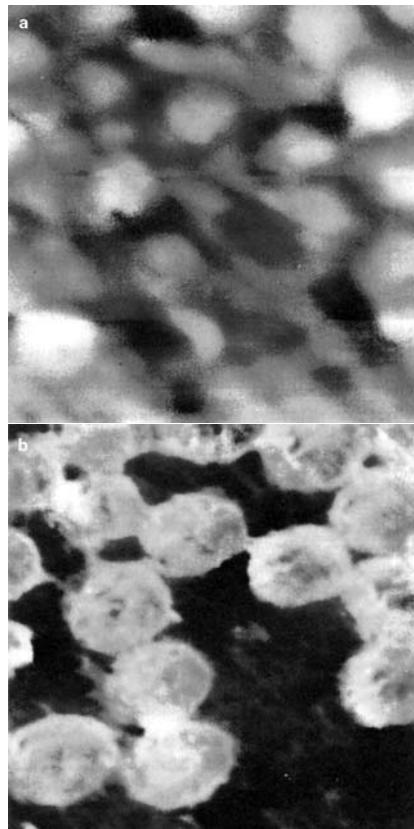


Figure 5.21 Cross-section of carbon fibers in an epoxy matrix: (a) topographic image; and (b) force modulation image. The image width is $32\ \mu\text{m}$. (Reproduced with permission from R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy*, Cambridge University Press, Cambridge, UK. © 1994 Cambridge University Press.)

5.4 Image Artifacts

While SPM offers unique advantages for surface measurement in terms of simplicity and high resolution, it also has its unique problems of image artifacts caused by near-field interactions between tip and sample. The artifacts in imaging may originate from the tip, the scanner, vibrations and/or image processing. Typical artifacts as introduced from several sources are briefly described in this section.

5.4.1 Tip

The geometry of the tip can generate artifacts. In SPM operation, the tip either contacts or nearly contacts the sample surface. Sharpness of the tip affects the mapping of the sample surfaces unless the tip is sharper than the surface features being mapped. Two geometrical parameters of the shape of the tip that are relevant here are tip apex and sidewall angle (also called the opening angle) of the tip.

The tip geometry may alter the lateral dimension of surface features in an image. For example, if a surface feature has a sharper edge than the tip's sidewall angle, the image of the surface feature will be distorted in its lateral direction because the imaged edge cannot be sharper than the sidewall angle. For a small surface protrusion, the image size could be larger than the real feature as illustrated in Figure 5.22a. When the tip maps a surface pit or hole, the sidewall angle and size of tip apex may prevent the tip from reaching the bottom of the pit or hole.

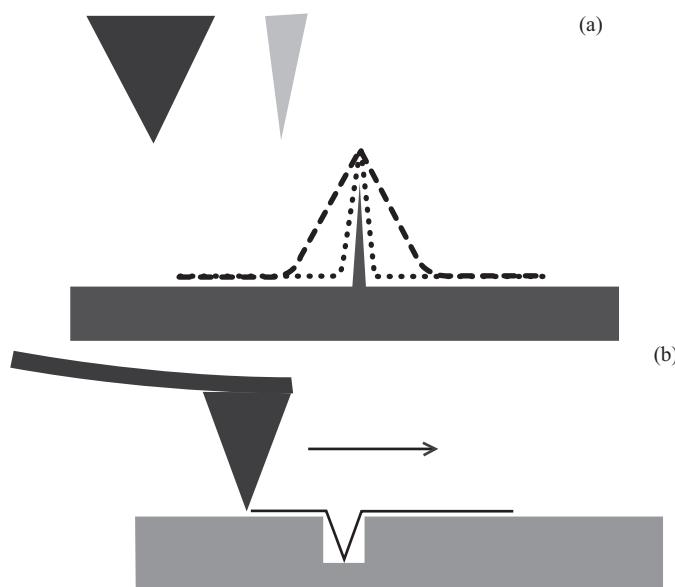


Figure 5.22 Artifacts generated by tip: (a) the side of tip causes broadening of the image of the protruding object; and (b) limits in the width the of tip limits cause distortion of the hole geometry in the image. (Reproduced with permission from P.C. Braga and D. Ricci (eds), *Atomic Force Microscopy*, Humana Press. © 2004 Humana Press.)

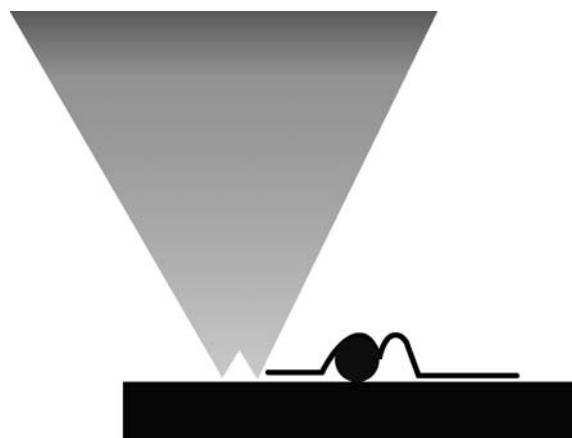


Figure 5.23 Image distortion by a tip with a double apex. (Reproduced with permission from P.C. Braga and D. Ricci (eds), *Atomic Force Microscopy*, Humana Press. © 2004 Humana Press.)

Consequently, the image of the pit or hole would be shallower and smaller than the real feature as illustrated in Figure 5.22b.

The tip geometry may also cause double imaging when the tip does not have a single apex. When the size of a surface feature is significantly smaller than the tip radius, such as a nano-scale particle, two protrusions on one tip can physically map the feature twice and generate two images of one feature as illustrated in Figure 5.23. Ideally, the tip apex for atomic-scale imaging should be a single atom. A tip with two or more atomic apices is more practically possible.

Tip artifacts may also result from a damaged tip or a tip contaminated by debris. For example, when a damaged tip maps a surface feature with lateral symmetry (such as an etched pit), the tip will generate an asymmetric image profile as illustrated in Figure 5.24. A tip with contaminating debris can cause image contrast changing and blurring during surface mapping.

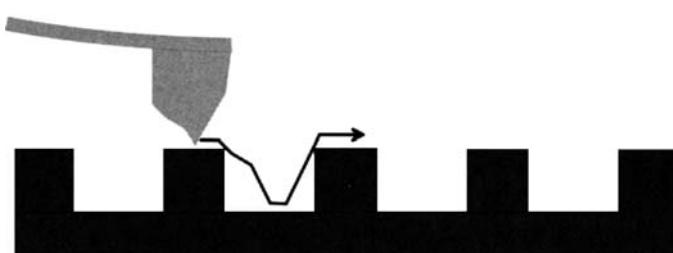


Figure 5.24 Image distortion by a damaged tip. (Reproduced with permission from P.C. Braga and D. Ricci (eds), *Atomic Force Microscopy*, Humana Press. © 2004 Humana Press.)

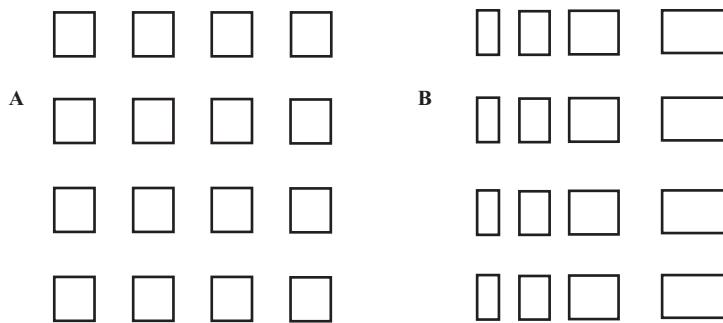


Figure 5.25 Image distortion caused by nonlinearity of the piezoelectric scanner: (a) regular grid; and (b) distorted image of grid. (Reproduced with permission from P.C. Braga and D. Ricci (eds), *Atomic Force Microscopy*, Humana Press. © 2004 Humana Press.)

5.4.2 Scanner

The scanner controlling the tip position and tip scanning is made from piezoelectric materials. The intrinsic physical and mechanical properties of piezoelectric materials, such as nonlinear piezoelectricity, hysteretic piezoelectricity and creep deformation affect the performance of the scanner. Such properties can generate distortion of SPM imaging during scanning.

The relationship between the electric voltages applied to the scanner and the scanner extension is not perfectly linear (the piezoelectric coefficient is not truly constant). Consequently, an image may be distorted in its lateral dimensions. The nonlinearity effects should be of concern when the lateral scanning range is greater than $100\text{ }\mu\text{m}$. Nonlinear distortion can be revealed by imaging a surface grid with periodic structure as illustrated in Figure 5.25. Such a grid pattern can be used to calibrate the linearity of a scanner. The nonlinearity of the scanner has less effect on measurement in the vertical direction because the vertical movement of the scanner is about one order of magnitude smaller than its lateral movements. The SPM is commonly equipped with linearity corrections made by either software or hardware.

The hysteretic behavior of piezoelectric materials is similar to the hysteretic behavior of magnetic materials. It occurs because there is an electric voltage difference for scanner elongation in the forward and backward directions. Thus, the tip may not return to its original position after one voltage loop applied to the scanner. The tip position can deviate by as much as 15% between forward and backward scanning. Image distortion caused by the hysteretic properties of the scanner can be illustrated by comparing surface profiles obtained by forward and backward scanning as shown in Figure 5.26. The hysteresis of the scanner generates a lateral shift between two profiles.

Creep deformation causes the scanner to show time-dependent extension. Creep deformation of the scanner will extend its length with time at a constant voltage. The profile of surfaces can be different between a first scan and a second scan, after a certain period of applying a voltage. Creep in the vertical direction leads to ‘overshoot’ of the scanner position at the leading and trailing edges of surface features in slow scanning (Figure 5.27). Consequently, the images will have a shaded edge in a lateral direction.

Thermal drift also can cause instability of mapping surface. Thermal drift refers to the effect of temperature change on the scanner. Thermal drift may come from applied electric power and

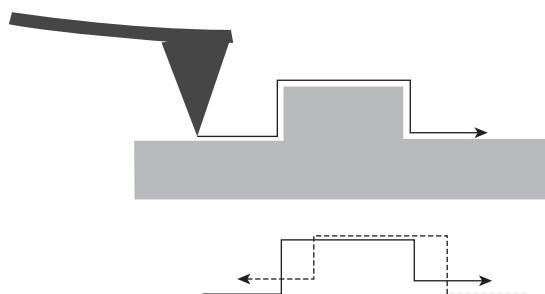


Figure 5.26 Hysteresis effects of a piezoelectric scanner on imaging. The bottom figure illustrates the profile difference between forward and backward scanning. (Reproduced with permission from P.C. Braga and D. Ricci (eds), *Atomic Force Microscopy*, Humana Press. © 2004 Humana Press.)

the SPM working environment. High resolution imaging is likely to be affected by a fraction of a degree change in temperature. Thus, it is advisable to have a waiting period for SPM operation and for control of environments.

5.4.3 Vibration and Operation

As aforementioned, the SPM requires excellent vibration isolation. Artifacts in images can be generated by floor vibration as well as acoustic vibration. Artifacts due to vibration appear as oscillation patterns in images. The floor can vibrate in the vertical direction with a magnitude of several micrometers and at a frequency less than 5 Hz. SPM operation can be affected even by people walking in a hallway. While a good vibration insulation system should be installed, even better insulation is achieved if the SPM system is placed on the ground floor of a building, and ideally at a corner of a room. The SPM can be insulated from acoustic vibration caused by talking and music by the installation of an acoustic hood.

Artifacts can also be generated by inappropriately selected operational parameters including the feedback-loop control, oscillation parameters of the AFM cantilever and scan speed. Setting appropriate parameters for operation relies primarily on an operator's experience; however,

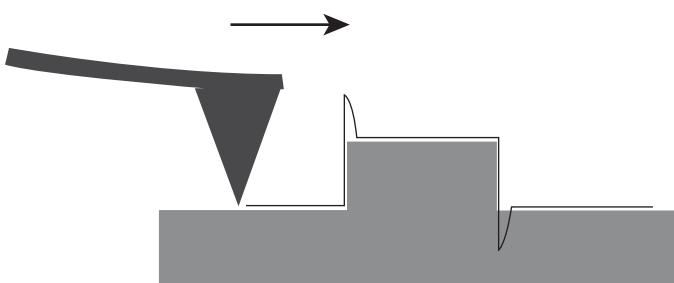


Figure 5.27 Creep effects of a piezoelectric scanner on image topography. In slow scanning, the “overshoot” of the scanner occurs at two edges of a surface feature. (Reproduced with permission from P.C. Braga and D. Ricci (eds), *Atomic Force Microscopy*, Humana Press. © 2004 Humana Press.)

there are several general guidelines we can follow. The gain parameter of the feedback loop determines the sensitivity of control over the surface features. A high-gain parameter may cause scanner system oscillation and may generate high frequency noise in an image. A low-gain parameter can ensure that the tip tracks the surface topography, and thus, the surface features are not smeared out. Soft samples should be scanned at slow scanning speed and with low interaction forces in the AFM. Rough samples also need a slow scanning speed.

There are several ways to detect possible artifacts in images:

- Repeat the scan for the same area;
- Change the scanning direction;
- Change the scanned area size;
- Rotate the sample and conduct a second scan; and
- Change the scanning speed.

Artifacts are most likely to be revealed if the image changes after changing such scanning methods.

References

- [1] Meyer, E., Hug, H.J. and Bennewitz, R. (2004) *Scanning Probe Microscopy: the Lab on a Tip*, Springer-Verlag, Berlin.
- [2] Birdi, K.S. (2003) *Scanning Probe Microscopes*, CRC Press, Boca Raton.
- [3] Wiesendanger, R. (ed.) (1998) *Scanning Probe Microscopy*, Springer, Berlin.
- [4] Wiesendanger, R. (1994) *Scanning Probe Microscopy and Spectroscopy*, Cambridge University Press, Cambridge, UK.
- [5] Braga, P.C. and Ricci, D. (eds) (2004) *Atomic Force Microscopy*, Humana Press.
- [6] Morita, S., Wiesendanger, R. and Meyer, E. (eds) (2002) *Noncontact Atomic Force Microscopy*, Springer-Verlag GmbH, Berlin.
- [7] Ratner, B.D. and Tsukruk, V.V. (eds) (1998) *Scanning Probe Microscopy of Polymers*, American Chemical Society, Washington DC.

Questions

- 5.1 Is it practical to measure the topography of a 1-mm x 10-mm area with an SPM? Give your reasons.
- 5.2 Is it practical to measure the topography of a sample with roughness $\Delta z > 10 \mu\text{m}$? Why?
- 5.3 What characteristics of a tunneling current enable it to be used for imaging atoms?
- 5.4 Why is an ultra-high vacuum environment often necessary for operation of an STM?
- 5.5 A hole appears in an STM image with atomic resolution. Can we assume it is an atomic vacancy? Why?
- 5.6 Why, in your opinion, are people more likely to use an AFM than an STM for topographic examinations?
- 5.7 To obtain atomic resolution, the force between the sample and tip of an AFM should decay significantly at a distance of atomic spacing. Plot the van der Waals force as the function of distance between the tip and sample and the tip radius using Equation 5.2. Can we obtain atom images with the van der Waals force?

- 5.8 Estimate the capillary force between the tip and sample using Equation 5.4, and compare the level of maximum capillary force with the van der Waals force estimated in Problem 5.8. Assume the surface tension of water $\gamma = 0.074 \text{ Nm}^{-1}$ at 20°C .
- 5.9 Compare the contact areas in contact mode using Equation 5.5: (1) a silicon tip with an iron sample, and (2) a silicon tip with a polymer sample. Assume that the contact force is 10 nN and radius of the tip is 80 nm ; the Young's modulus of tip, iron and polymer are about 0.17×10^{12} , 0.2×10^{12} and $0.1 \times 10^{12} \text{ Nm}^{-2}$, respectively; and their Poisson ratio is about 0.3.
- 5.10 Is there any way, not using a standard grid, to check whether there is non-linearity in the scanner? Explain.
- 5.11 Suggest methods to check possible damage on a tip.

6

X-ray Spectroscopy for Elemental Analysis

X-ray spectroscopy is a technique of using characteristic X-rays to identify chemical elements; it is to be distinguished from X-ray diffraction for crystal structure analysis. X-ray spectroscopy determines presence and quantities of chemical elements by detecting characteristic X-rays that are emitted from atoms irradiated by a high-energy beam. From the characteristic X-rays emitted from sample atoms, chemical elements can be identified either from the X-ray wavelength, as in X-ray wavelength dispersive spectroscopy (WDS), or from the X-ray energy, as in X-ray energy dispersive spectroscopy (EDS). The most commonly used spectrometers for X-ray spectroscopy include X-ray fluorescence spectrometers and microanalyzers in electron microscopes. The X-ray fluorescence spectrometer is stand-alone equipment for elemental analysis. It uses X-ray radiation to excite the emission of characteristic X-rays from sample atoms. The word ‘fluorescence’, is used to distinguish the secondary X-ray emission of sample atoms from the primary X-rays irradiating the sample. A microanalyzer is an EDS type of X-ray spectrometer that is commonly found in a scanning or a transmission electron microscope. The EDS microanalyzer in an electron microscope that uses a primary electron beam to excite the emission of characteristic X-rays from the sample atoms. Since the electron beam can be readily focused on a microscopic area on a sample, the EDS microanalyzer can examine chemical compositions in a microscopic area; while the X-ray fluorescence spectrometer is mainly used to examine overall chemical compositions in a sample. This chapter introduces both X-ray fluorescence spectrometers and microanalyzers in electron microscopes.

6.1 Features of Characteristic X-rays

Characteristic X-rays are introduced in Chapter 2 when discussing monochromatic X-ray radiation for the X-ray diffractometer. It is necessary, however, to further discuss the features of characteristic X-rays in this chapter because they are the essential signals used in X-ray spectroscopy for identifying chemical elements.

Figure 6.1 schematically illustrates the generation of characteristic X-rays when an atom is irradiated by high energy particles. When a high energy particle, such as an X-ray photon, an

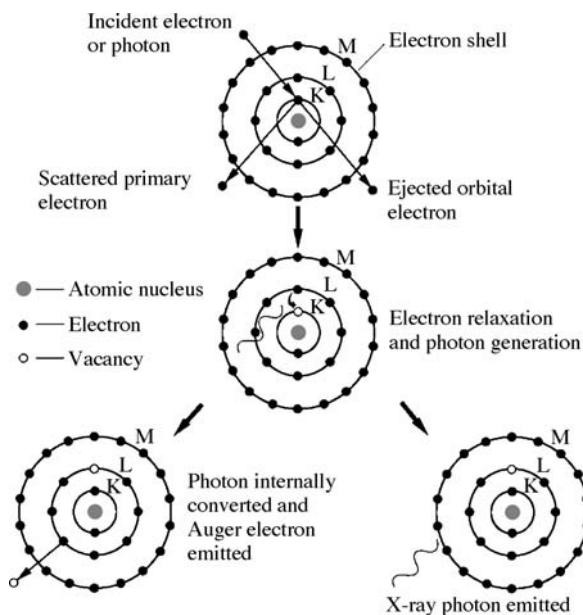


Figure 6.1 Excitation of a characteristic X-ray photon or an Auger electron by a high energy X-ray photon or electron.

electron or a neutron, strikes an electron in the inner shell of an atom, the energy of the particle can be high enough to knock an electron out of its original position in an atom. The knocked-out electron leaves the atom as a ‘free’ electron and the atom becomes ionized. Because ionization is an excited state, the atom will quickly return to its normal state after refilling the inner electron vacancy by an outer shell electron. In the meantime, the energy difference between an outer shell electron and an inner shell electron will generate either an X-ray photon (characteristic X-ray) or another characteristic free electron that is emitted from the atom. The latter is called as the *Auger electron*, which also can be used for elemental analysis; this is discussed in Chapter 7.

The energy of characteristic X-rays is the energy difference between two electrons in different shells. It is well defined and dependent on the atomic number of the atom. For example, the energy of the X-ray $K\alpha$ line is the energy difference between a K shell electron and L shell electron. Thus, we can identify a chemical element from the characteristic X-rays that it emits. Moseley’s Law defines the relationship between wavelength of characteristic X-rays (λ) and atomic number (Z).

$$\lambda = \frac{B}{(Z - \sigma)^2} \quad (6.1)$$

B and σ are constants that depend upon the specific shells. This relationship can also be expressed as the energy of characteristic X-rays and atomic number using the energy-wavelength relationship (Equation 9.1 and Equation 9.4).

6.1.1 Types of Characteristic X-rays

A number of characteristic X-rays can be emitted from an atom if a number of inner shell electrons have been knocked out by high energy particles. The individual characteristic X-rays are marked $K\alpha$, $K\beta \dots$ as mentioned in Chapter 2. It seems that there are many possible ways in which outer shell electrons can fill inner shell vacancies; however, the possibilities are limited and such electron transitions in an atom are controlled by the *selection rules*.

Selection Rules

Each electron in an atom is defined by four quantum numbers: n , l , m and s . The principal quantum number (n) defines the shell; for example, the K shell as 1, L shell as 2 and the M shell as 3. The angular quantum number (l) defines the number of subshells, taking all values from 0 to $(n - 1)$. The magnetic quantum number (m) defines the number of energy states in each subshell, taking values $-l$, 0 and $+l$. The spin quantum number (s) defines two spin moments of electrons in the same energy state as $+\frac{1}{2}$ and $-\frac{1}{2}$. The quantum numbers of electrons in K , L and M shells are listed in Table 6.1. Table 6.1 also gives the total momentum (J), which is the sum of $(l + s)$. No two electrons in an atom can have same set quantum numbers (n, l, m, s).

Selection rules for electron transitions between two shells are as follows:

1. Change in n must be at least 1 ($\Delta n \geq 1$);
2. Change in l must be 1 ($\Delta l = \pm 1$); and
3. Change in J must be either 0 or 1 ($\Delta J = \pm 1$, or 0).

Characteristic X-rays are divided into K , L and M series according to the shells of electron vacancies being filled. For example, ‘ K series’ means the characteristic X-rays emitted when an outer shell electron fills an electron vacancy in the K shell. According to the selection rules, the

Table 6.1 Electron Structure and Quantum Numbers in K , L and M Shells

Shell (Electrons)	n	l	m	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	1	1	$\pm \frac{1}{2}$		
	2	1	0	$\pm \frac{1}{2}$	2p	$\frac{1}{2}; \frac{3}{2}$
	2	1	-1	$\pm \frac{1}{2}$		
M (18)	3	0	0	$\pm \frac{1}{2}$	3s	$\frac{1}{2}$
	3	1	1	$\pm \frac{1}{2}$		
	3	1	1	$\pm \frac{1}{2}$	3p	$\frac{1}{2}; \frac{3}{2}$
	3	1	-1	$\pm \frac{1}{2}$		
	3	2	2	$\pm \frac{1}{2}$		
	3	2	1	$\pm \frac{1}{2}$		
	3	2	0	$\pm \frac{1}{2}$	3d	$\frac{3}{2}; \frac{5}{2}$
	3	2	-1	$\pm \frac{1}{2}$		
	3	2	-2	$\pm \frac{1}{2}$		

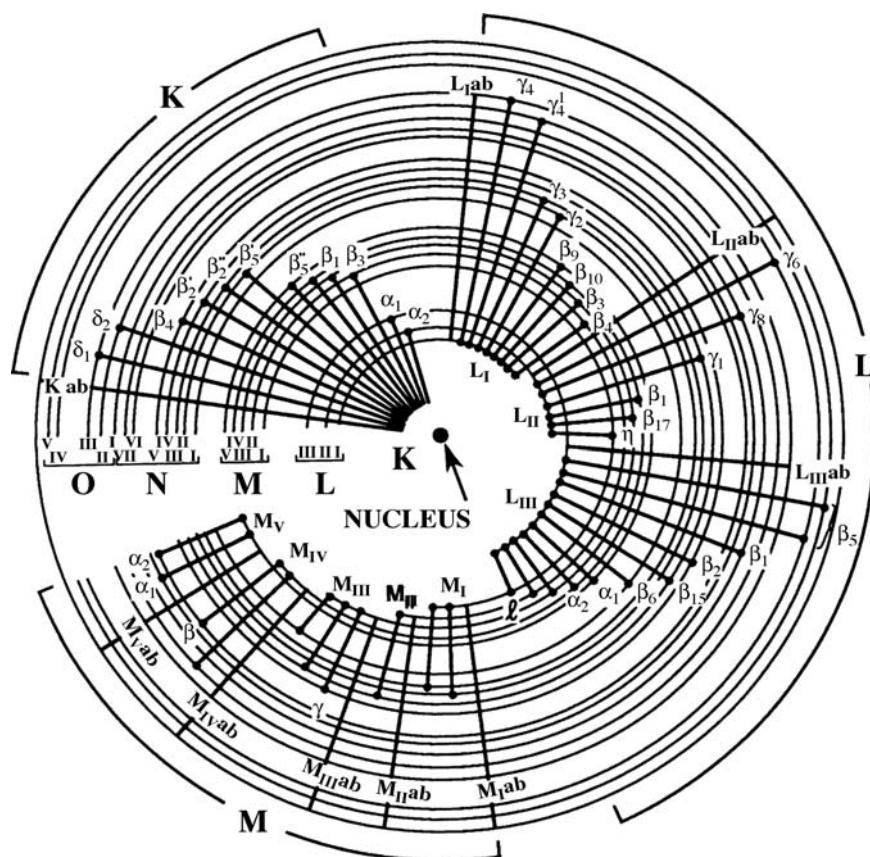


Figure 6.2 Allowable electron transitions in K , L and M shells. (Reproduced with kind permission of Springer Science and Business Media from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

electron transition from the L_1 to the K shell is forbidden because the l change would be zero, which would violate the second selection rule. Thus, the only possible transitions from the L to K shell are L_3 to K and L_2 to K , which are known as $K\alpha_1$ and $K\alpha_2$, respectively, as mentioned in Chapter 2. Figure 6.2 shows the shell structures of electrons and the possible characteristic X-rays emitted when an electron transitions to a K , L or M shell. Figure 6.2 also shows that the characteristic X-ray can be classified as K , L or M series, according the ground state of the electron transition. The *Siegbahn Notation* is commonly used to specify a characteristic X-ray line. The Siegbahn Notation comprises a capital English letter, a lower-case Greek letter and a subscript number.

$K\alpha_1$	$L_3 \rightarrow K$
$K\beta_1$	$M_3 \rightarrow K$
$L\alpha_1$	$M_5 \rightarrow L_3$
$L\beta_1$	$N_3 \rightarrow L_3$

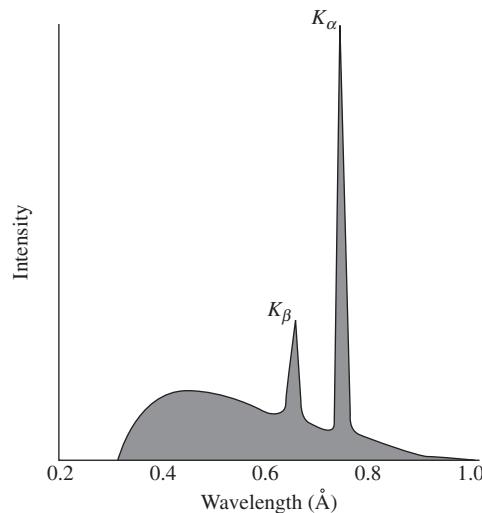


Figure 6.3 Primary X-ray radiation generated by the X-ray tube, including continuous and characteristic radiations.

$L\gamma_1$	$N_4 \rightarrow L_2$
$M\alpha_1$	$N_7 \rightarrow M_5$
$M\beta$	$N_6 \rightarrow M_4$

Generally, the Greek letters indicate the general order of the X-ray intensity, where the X-ray intensity of $\alpha > \beta > \gamma$. Figure 6.3 illustrates the relative intensities of $K\alpha$ and $K\beta$ on top of continuous X-ray radiation. The subscript number also indicates the relative intensity, with number 1 as the highest.

6.1.2 Comparison of K , L and M Series

An atom's ability to generate characteristic X-ray photons varies when it is irradiated by high energy X-ray photons or electrons. First, there is an emission competition between the characteristic X-ray photon and the Auger electron when an electron refills an inner shell vacancy. Second, there are competitions among generations of K , L and M series of the X-rays. A parameter, *fluorescent yield*, has been used to measure the relative effectiveness of X-ray generation. Figure 6.4 shows the fluorescent yield (ω) variation in K , L , and M series for a range of atomic numbers. Table 6.2 lists the values of fluorescent yields of some elements in the K , L and M series.

For an atom with an atomic number lower than 4 (Be), the fluorescent yield is zero, and for an atomic number lower than 8 (O), the yield is less than 0.5%. Generally, the fluorescent yield increases with atomic number as shown in Figure 6.4. Thus, there is an intrinsic disadvantage for detection of light elements using X-ray spectroscopy. Figure 6.4 also indicates that the generation of the K series X-rays is more effective than the L and M series X-rays. It means that the intensity of the K series X-rays is always higher than the L series, and in turn, the L series is always higher than the M series. The fluorescent yield is zero for the L series when the atomic number is lower than 20 (Ca), and for the M series when the atomic number is less than 57 (La).

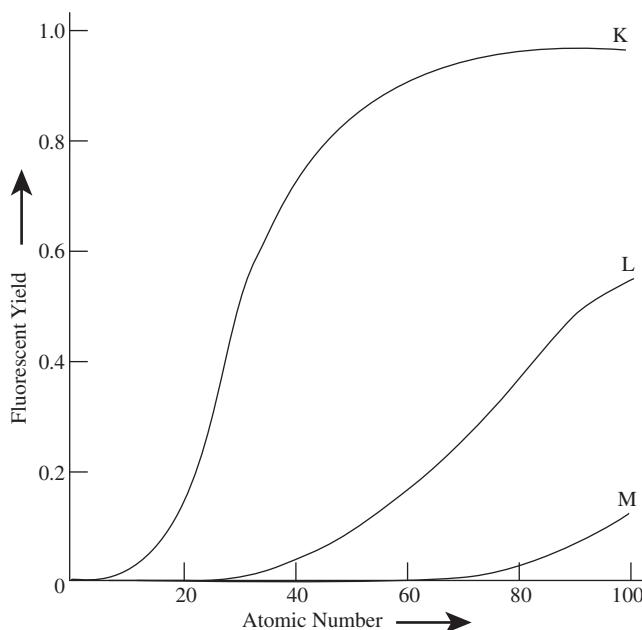


Figure 6.4 Fluorescent yield variation with atomic number and with the *K*, *L* and *M* characteristic lines.

Another difference among the *K*, *L* and *M* lines is their energy levels (or wavelengths). The energy of *K* series is the highest, followed by *L* and *M* series as shown in Figure 6.5. Table 6.3 lists *K*, *M* and *L* wavelengths and energies for selected elements. For example, the energy of $\text{AuK}\alpha_1$ reaches 66.99 keV, but the energy of $\text{AuL}\alpha_1$ and $\text{AuM}\alpha_1$ are 9.713 keV and 2.123 keV, respectively. A typical X-ray energy spectrum is in a range from 0.2 to 20 keV. Thus, the $\text{AuK}\alpha_1$ will be absent in such a spectrum. In general, the *L* and the *M* lines of heavy elements are much more likely than their *K* lines to be detected by an X-ray spectrometer because the energies of the *K* lines are likely to be higher than 20 keV. An X-ray spectrum of a high atomic number element is complicated because of the existence of X-rays from different series. On the other hand, the spectra of atomic numbers lower than 20 (Ca) contain only the *K* lines. The energies of *L* and *M* lines for low atomic numbers, as well as the fluorescent yield, can be too low to be detected in the X-ray spectrometer.

6.2 X-ray Fluorescence Spectrometry

X-ray fluorescence spectrometry (XRF) analyzes the chemical elements of specimens by detecting the characteristic X-rays emitted from the specimens after radiation by high-energy primary X-rays. The characteristic X-rays can be analyzed from either their wavelengths or energies. Thus, there are two types of XRF: wavelength dispersive spectroscopy (WDS) and energy dispersive spectroscopy (EDS). Figure 6.6 schematically illustrates the structural similarities of, and differences between, WDS and EDS instruments. An XRF instrument consists of three main parts: the X-ray source, detection system, and data collection and processing system.

Table 6.2 Fluorescent Yield of Chemical Elements

Atomic number	Element	ω_k	ω_L	ω_M
4	Be	0.00045	—	—
5	B	0.00101	—	—
6	C	0.00198	—	—
7	N	0.00351	—	—
8	O	0.00579	—	—
9	F	0.00901	—	—
10	Ne	0.0134	—	—
11	Na	0.0192	—	—
12	Mg	0.0265	—	—
...
19	K	0.138	—	—
20	Ca	0.163	0.00067	—
21	Sc	0.190	0.00092	—
22	Ti	0.219	0.00124	—
23	V	0.249	0.00163	—
24	Cr	0.281	0.0021	—
...
56	Ba	0.900	0.126	—
57	La	0.906	0.135	0.00111
58	Ce	0.911	0.144	0.00115
59	Pr	0.915	0.153	0.00120
...

(Data taken from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

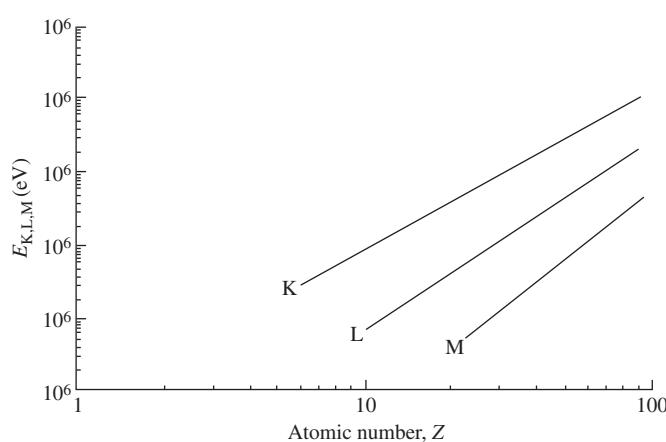
**Figure 6.5** Energies of characteristic X-ray photons in the *K*, *L* and *M* lines.

Table 6.3 Characteristic X-ray Wavelengths and Energies of Selected Elements

Atomic number	Element	$K\alpha_1$		$L\alpha_1$		$M\alpha_1$	
		$\lambda(\text{\AA})$	$E(\text{keV})$	$\lambda(\text{\AA})$	$E(\text{keV})$	$\lambda(\text{\AA})$	$E(\text{keV})$
4	Be	114	0.109	—	—	—	—
5	B	67.6	0.183	—	—	—	—
6	C	44.7	0.277	—	—	—	—
7	N	31.6	0.392	—	—	—	—
8	O	23.62	0.525	—	—	—	—
9	F	18.32	0.677	—	—	—	—
10	Ne	14.61	0.849	—	—	—	—
11	Na	11.91	1.041	—	—	—	—
12	Mg	9.89	1.254	—	—	—	—
13	Al	8.339	1.487	—	—	—	—
14	Si	7.125	1.740	—	—	—	—
15	P	6.157	2.014	—	—	—	—
16	S	5.372	2.308	—	—	—	—
17	Cl	4.728	2.622	—	—	—	—
...
19	K	3.741	3.314	—	—	—	—
20	Ca	3.358	3.692	36.33	0.341	—	—
21	Sc	3.031	4.091	31.35	0.395	—	—
22	Ti	2.749	4.511	27.42	0.452	—	—
23	V	2.504	4.952	24.25	0.511	—	—
24	Cr	2.290	5.415	21.64	0.573	—	—
25	Mn	2.102	5.899	19.45	0.637	—	—
26	Fe	1.936	6.404	17.59	0.705	—	—
27	Co	1.789	6.930	15.97	0.776	—	—
28	Ni	1.658	7.478	14.56	0.852	—	—
29	Cu	1.541	8.048	13.34	0.930	—	—
30	Zn	1.435	8.639	12.25	1.012	—	—
...
45	Rh	0.6133	20.13	4.597	2.697	—	—
46	Pd	0.5854	21.18	4.368	2.839	—	—
47	Ag	0.5594	22.16	4.154	2.984	—	—
...	—	—
56	Ba	0.3851	32.19	2.776	4.466	—	—
57	La	0.3707	33.44	2.666	4.651	14.88	0.833
58	Ce	0.3571	34.72	2.562	4.840	14.04	0.883
59	Pr	0.3441	36.03	2.463	5.034	13.34	0.929
...
78	Pt	0.1855	66.83	1.313	9.442	6.047	2.051
79	Au	0.1851	66.99	1.276	9.713	5.840	2.123
80	Hg	0.1751	70.82	1.241	9.989	5.645	2.196
...

(Data taken from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

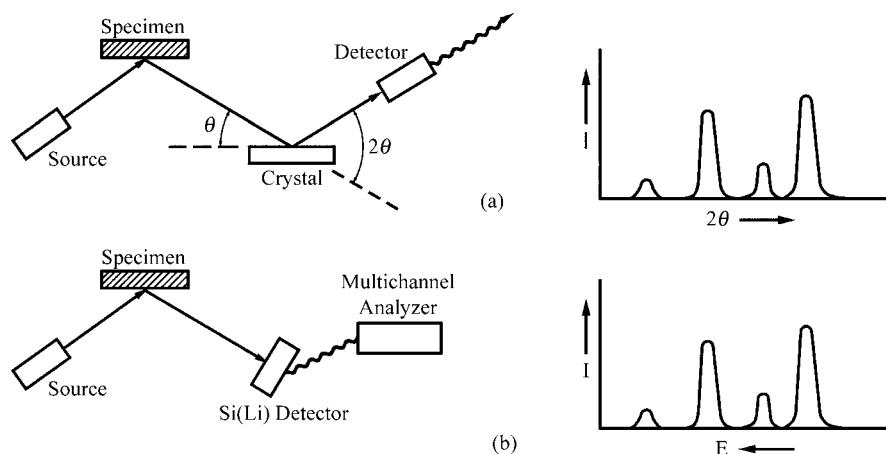


Figure 6.6 Main components and dispersive spectra of: (a) WDS; and (b) EDS.

The X-ray source generates the primary X-rays to excite atoms of specimens. The most commonly used source is the X-ray tube, similar to that used in the X-ray diffractometer introduced in Chapter 2. The X-ray tube is operated at a power of 0.5–3 kW and with high voltage of 30–50 kV. The reason for using high voltage is to ensure that the X-rays impinging the specimen exceed the critical potential of exciting characteristic X-rays from the elements in specimens. The optimum ratio of tube voltage to the critical potential is about 3–5. The X-ray radiation is produced by electrons striking a metal target in the tube. The target metals include Cr, Rh, W, Ag, Au and Mo. The emitted X-rays from the source consist of both continuous and characteristic radiations (Figure 6.3).

The main difference between the WDS and EDS instrumentation is in the X-ray detection systems. The WDS type uses single crystal diffraction to detect characteristic wavelengths emitted from the specimen, because, according to Bragg's Law, the single crystal is able to diffract a specific wavelength at a specific angle between the incident X-ray beam and a crystallographic plane. The EDS instrument uses a photon detector, typically a Si(Li) diode, to separate the characteristic X-ray photons according to their energies. The features and properties of the WDS and EDS systems are introduced in the following subsections.

6.2.1 Wavelength Dispersive Spectroscopy

X-ray fluorescence spectrometry was introduced as wavelength dispersive spectrometry (WDS) in the early 1950s; the EDS type came along years later. WDS provides better resolution and a wider range of elemental analysis than EDS, although its instrumentation is more complicated. The WDS systems can resolve relative changes in wavelength ($\frac{\Delta\lambda}{\lambda}$) in the range 0.002–0.02. This range corresponds to the energy range 0.01–0.1 keV, which is about one order of magnitude better than that of EDS. Modern WDS systems can detect elements from upward of C ($Z = 6$). Figure 6.7 illustrates the structure of a WDS system, which is similar to an X-ray diffractometer. There is a rotating X-ray detector system (an analyzing crystal and X-ray photon counter in Figure 6.7) to collect the diffraction beam, and collimators to align the characteristic X-ray

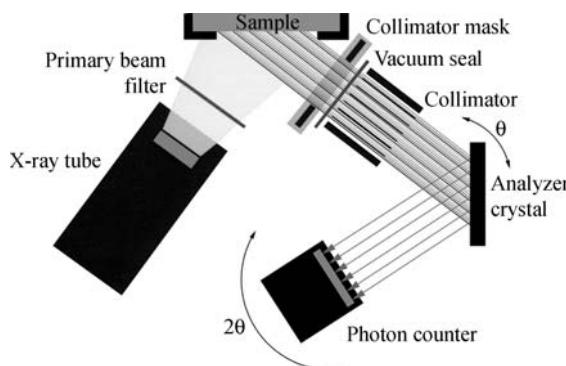


Figure 6.7 WDS apparatus, which includes the X-ray tube, specimen, primary collimator, analyzing crystal, flow counter, auxiliary collimator and scintillation counter.

beam from the specimen and the beam diffracted from the analyzing crystal. The rotating X-ray photon counter scans a range of 2θ to detect specific wavelengths of characteristic X-rays from the specimen. We can understand that the signal detection in WDS is sequential in nature because of 2θ scanning. A WDS system may have one crystal/detector set (single channel) or a number of crystal/detector sets (multi-channel). The multi-channel system can partially overcome the drawbacks of sequential detection, and increase analysis speed.

Analyzing Crystal

In WDS, the analyzing crystal should be carefully selected because it determines the range of detectable atomic numbers. The wavelength that can be detected by a crystal is defined by Bragg's Law.

$$\lambda = \frac{2ds\sin\theta}{n} \quad (6.2)$$

Normally, the maximum achievable θ angle in a WDS system is about 73° . Thus, the maximum λ of characteristic X-rays being diffracted is about $1.9d$ of the analyzing crystal. Table 6.4 lists the commonly used analyzing crystals for WDS, which include LiF, pentaerythrital (PE), thallium acid phthalate (TAP) and layered synthetic microstructure (LSM). For detecting light

Table 6.4 Analyzing Crystals used in Wavelength Dispersive Spectrometers (WDS)

Crystal	Plane	Plane spacing, $2d$ (\AA)	Atomic number range	
			K lines	L lines
Lithium fluoride (LiF)	(220)	2.848	>Ti(22)	>La(57)
Lithium fluoride (LiF)	(200)	4.028	>K(19)	>Cd(48)
Pentaerythrital (PE)	(002)	8.742	Al(13)–K(19)	–
Thallium acid phthalate(TAP)	(001)	26.4	Fe(9)–Na(11)	–
Layered synthetic microstructure (LSM)	–	50–120	Be(4)–F(9)	–

(Reproduced with permission from R. Jenkins, *X-ray Fluorescence Spectrometry*, 2nd ed., John Wiley & Sons Inc., New York. © 1999 John Wiley & Sons Inc.)

elements, which have long wavelengths of characteristic X-rays, an analyzing crystal with a large atomic spacing must be used. For detecting elements with atomic numbers less than 9, LSM materials with extremely large *d*-spacings should be used. Although an analyzing crystal with small *d*-spacing limits the range of detectable atomic numbers, the small *d*-spacing can increase angular dispersion. This can be seen by differentiating the Bragg's Law equation.

$$\frac{d\theta}{d\lambda} = \frac{n}{2dcos\theta} \quad (6.3)$$

Equation 6.3 implies that we will obtain the wave spectrum of higher resolution when using a small *d*-spacing crystal. However, using the analyzing crystal with a small *d* will reduce the range of wavelength to be detected because of the $1.9d$ limit.

Wavelength Dispersive Spectra

A WDS spectrum is presented as a diagram in which the characteristic X-ray lines are dispersed in a range of the X-ray wavelengths as shown in Figure 6.8. The relative intensities of the individual X-ray lines are represented by their heights in the spectrum; however, there is no scale to indicate the real intensities of the X-rays. Similar to the spectrum of X-ray diffraction, only relative intensities among the lines, and with respect to the background, are important. An individual line is marked with the corresponding element that generates it and its specific

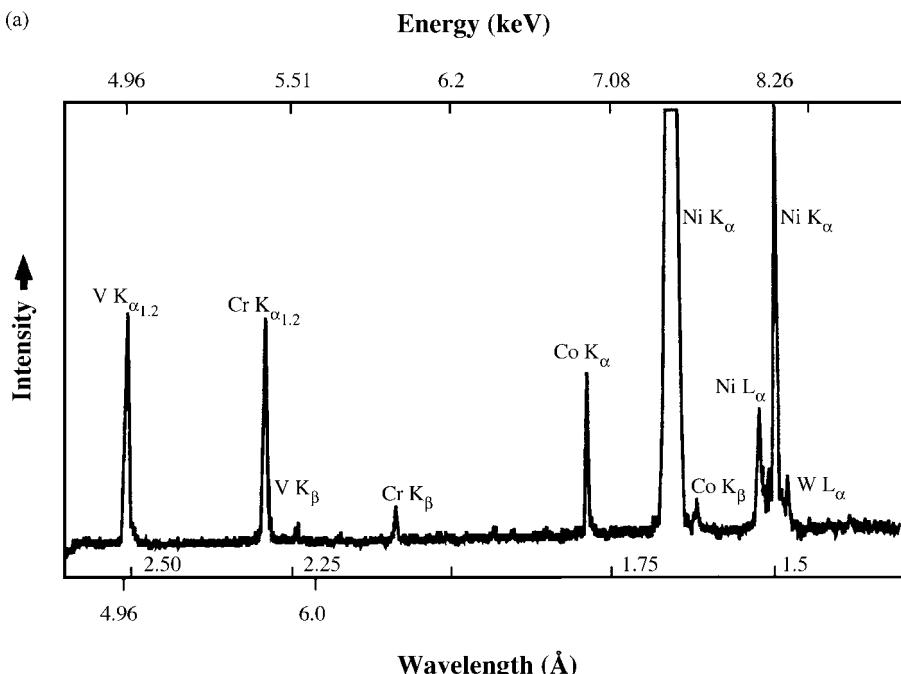


Figure 6.8 WDS spectra of a nickel-based superalloy: (a) spectrum obtained with a LiF analyzing crystal; and (b) spectrum obtained with a TAP crystal. (Reproduced with kind permission of Springer Science and Business Media from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Micro-analysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

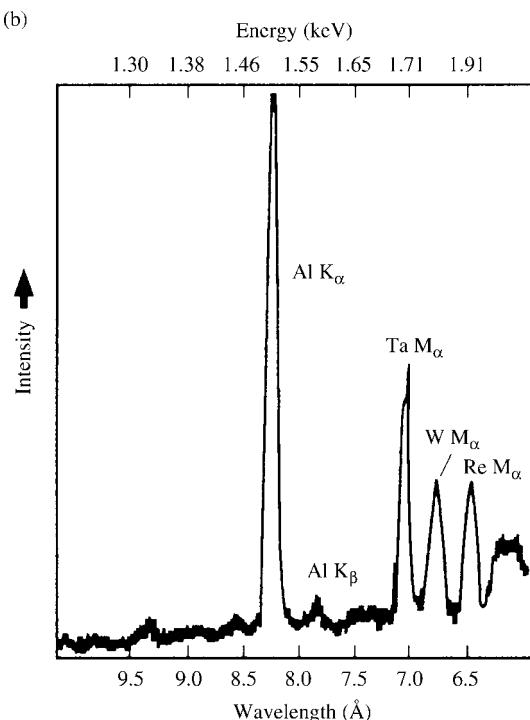


Figure 6.8 (continued)

line code, such as $\text{Cr}K\alpha$ or $\text{Cr}K\alpha_{1,2}$. The reason is that the wavelength resolution is unable to separate $\text{Cr}K\alpha_2$ from $\text{Cr}K\alpha_1$. Often, we need more than one type of analyzing crystal to obtain the whole range of wavelength dispersion from a specimen, because the limitation of wavelength ranges by the crystal's *d*-spacing. For example, Figure 6.8 shows the spectra of a nickel-based alloy obtained by LiF and TAP analyzing crystals. The spectrum obtained with LiF covers a short wavelength range and that obtain with TAP covers a long wavelength range. The two spectra provide a complete chemical analysis of the nickel-based alloy that contains Cr, Co, W, Ta, V, Al and Re. The relative intensities of spectrum lines representing individual elements provide a rough idea of the relative concentrations of those elements in the alloy. Precise measurement of element concentrations requires more complicated quantitative analysis, which will be introduced in Section 6.4.

While WDS provides high resolution and high ratio of signal to background, analysis of the spectra can be complicated. Equation 6.2 (Bragg's Law) indicates that the orders of diffraction (*n*) may generate multiple wavelength peaks in a spectrum for the same characteristic line. For example, given a specific characteristic X-ray line with the wavelength (λ) emitted from a specimen, there may be three peaks corresponding to $n = 1, 2$ and 3 . In addition, if a specific wavelength is nearly equal to a multiple of another wavelength, then their peaks will appear superimposed on each other in the spectrum. For example, $S\text{K}\alpha$ ($n = 1$) has a wavelength of 5.372 \AA , which is very close to three times the $\text{Co}K\alpha$ wavelength ($1.789 \text{ \AA} \times 3 = 5.367 \text{ \AA}$). Thus, the $S\text{K}\alpha$ line will likely be superimposed on the third order ($n = 3$) $\text{Co}K\alpha$ line. We should be aware of all of these special problems in interpreting a WDS spectrum.

6.2.2 Energy Dispersive Spectroscopy

Energy dispersive spectroscopy became a commercial product in the early 1970s, and rapidly overtook WDS in popularity. An EDS system is structurally simple because it does not have moving parts such as the rotation detector with WDS. EDS systems are relatively faster because the detector collects the signals of characteristic X-rays energies from a whole range of elements in a specimen at the same time rather than collecting signals from X-ray wavelength individually. For EDS, the typical resolution of energy dispersion is about 150–200 eV, worse than the corresponding resolution of WDS, and the lightest element that can be detected is O ($Z=8$), not C ($Z=6$). But these disadvantages are not as important as the advantages of an EDS system, which are low cost and fast analysis.

Detector

The Si(Li) is the most commonly used detector in an EDS system. The Si(Li) detector consists of a small cylinder of p-type silicon and lithium in the form of a Si(Li) diode, as schematically shown in Figure 6.9. X-ray photons collected by the detector generate a specific number of electron–hole pairs. The average energy of photons needed to generate as electron–hole pair (e_i) is about 3.8 eV in the Si(Li) diode. The higher the photon energy, the more pairs are generated. Characteristic X-ray photons can be separated by their energy levels according to the numbers of electron–hole pairs they generate.

The electron–hole pairs, as the electric charge, are swept from the detector diode. A preamplifier collects the charge to produce an output of electrical pulse whose voltage amplitude is proportional to the X-ray photon energy. The energy resolution of the detector (R) in eV can be estimated.

$$R = \sqrt{\sigma_{\text{noise}}^2 + [2.35(e_iFE)]^2} \quad (6.4)$$

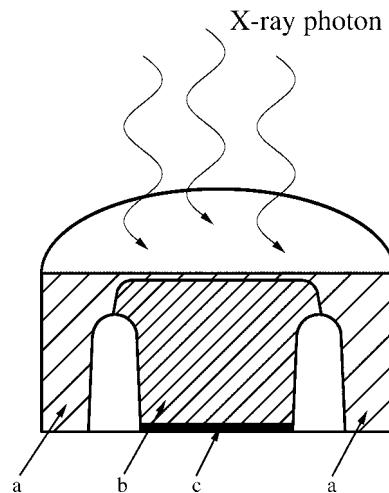


Figure 6.9 A Si(Li) cylinder with annular groove construction: (a) p-type silicon; (b) lithium compensated region; and (c) n-type silicon.

E is the energy of characteristic X-ray line and F is a constant called the *Fano factor*, which has a value of 0.12 for Si(Li). σ_{noise} , the electronic noise factor, plays an important role in the resolution. Reduction of the electronic noise will improve the resolution of the EDS detector. Thus, the Si(Li) diode and the preamplifier are mounted in a cylindrical column (the cryostat) so that they can operate at the temperature of liquid nitrogen (-196°C) in order to reduce the electrical noise and increase the signal-to-noise ratio.

X-ray photons must pass through a window to reach the Si(Li) diode. This window is typically made of a light element, beryllium. Beryllium, like any material, will absorb X-ray photons, even though its absorption is weak. Thus, the sensitivity of the detector will be affected by the beryllium window. In order to reduce the photon absorption, the beryllium has to be as thin as possible, especially for detecting light elements of which the characteristic energies are on the order of 100 eV. The typical thickness of the beryllium window is about 10 μm .

Energy Dispersive Spectra

An EDS spectrum is presented as the intensity of characteristic X-ray lines across the X-ray energy range. A spectrum in a range from 0.1 to about 10–20 keV can show both light and heavy elements because both K lines of light elements and M or L lines of heavy elements can be shown in this range. For example, Figure 6.10 shows the EDS spectrum of a glass specimen containing multiple elements including Si, O, Ca, Fe, Al and Ba in an energy range up to 10 keV. EDS spectra are similar to WDS spectra, but identification of individual elements from EDS spectra is more straightforward than from WDS spectra because each characteristic line generated by a specific element exhibits a unique X-ray energy. However, the signal-to-noise ratio is lower than that of WDS, and the resolution, in terms of energy, is about 10 times lower than that of WDS. Even so, EDS is an attractive technique for qualitative analysis of elements.

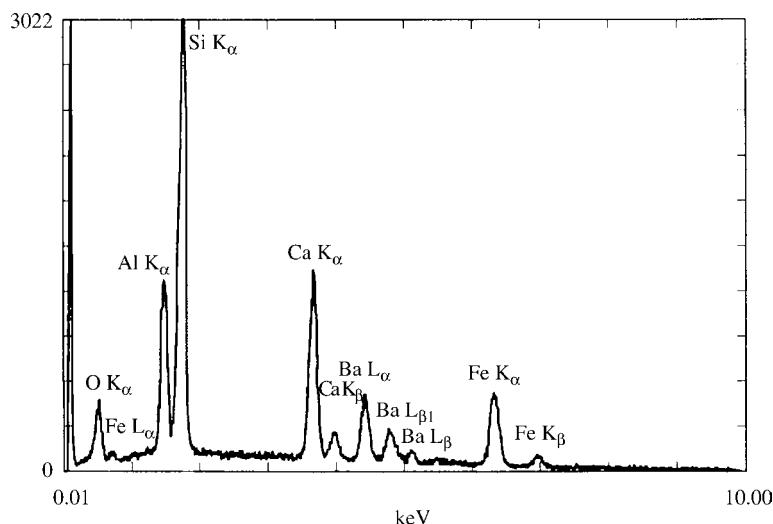


Figure 6.10 EDS spectrum of glass which includes Si, O, Ca, Al, Fe and Ba. (Reproduced with kind permission of Springer Science and Business Media from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

Advances in Energy Dispersive Spectroscopy

All the early EDS systems were operated in a primary mode in which X-rays emitted from the X-ray tube irradiate the specimen directly. This had drawbacks because the X-rays could excite more photons than could be counted by the detector, because there is a maximum counting rate of photons for an energy detector. If the number of characteristic photons excited in the specimen is too great, a portion of photons may not be counted. Thus, the measured intensities may be lower than the actual emitted intensity of X-ray photons. The period during which the detector cannot respond to the number of photons correctly is called the *dead time* of the detector. The dead time is in the order of 0.5×10^{-7} seconds.

Later, a secondary mode of EDS was developed to overcome this intrinsic limitation of the detector when handling a large influx of photons. In the secondary mode, a selected pure element standard with absorption filters is placed in the optical path between the primary X-ray source and the specimen. The element standard only allows the X-ray photons in a selected range of energy (secondary photons) to strike the sample. Thus, in secondary photon radiation, the amount of photons emitted from the sample is controlled by preventing unwanted X-ray excitation of the specimen.

Another problem of early EDS systems was the difficulty in examining small samples, of which the signals of characteristic X-rays are weak compared with background noise in a spectrum. This problem can be solved by use of the *total reflection spectrometer* (TRXRF). The TRXRF technique is based on the phenomenon that the reflectivity of a flat target strongly increases below a critical glancing angle. This instrumentation is illustrated in Figure 6.11. The Si(Li) detector is placed close to (~ 5 mm) and directly above the specimen. The primary X-ray radiation enters the specimen at a glancing angle of less than 0.1° , rather than at 40° as in a conventional EDS. The sample is presented as thin film on the optically-flat quartz plate. Reflectors and apertures between the X-ray tube and sample ensure that the angle at which radiation enters the sample is slightly less than the *critical angle of total reflection*. The critical angle refers to an incident angle below which no penetration of radiation into the quartz substrate occurs. TRXRF eliminates primary X-rays radiating substrate materials to a large extent. Thus, the X-ray signals in the EDS spectrum from sources other than the sample can

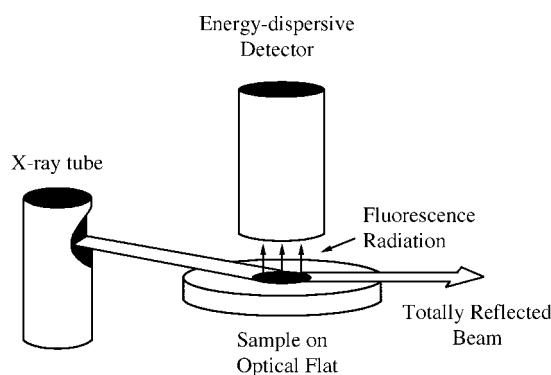


Figure 6.11 Geometric construction of a total reflection fluorescence spectrometer. The incident angle is less than 0.1° so that the primary X-ray beam is totally reflected from the flat substrate. (Reproduced with permission from R. Klockenkämper, *Total-Reflection X-ray Fluorescence Analysis*, John Wiley & Sons Inc., New York. © 1997 John Wiley & Sons Inc.)

be significantly reduced. This technique enables us to measure elements from solid samples with low mass down to 10^{-12} g.

6.3 Energy Dispersive Spectroscopy in Electron Microscopes

The EDS type of X-ray spectrometer is commonly included as a part of SEMs and TEMs. The reason for using EDS rather than WDS is simply its compactness. With EDS in an electron microscope, we can obtain elemental analysis while examining the microstructure of materials. The main difference between EDS in an electron microscope and in a stand-alone XRF is the source to excite characteristic X-rays from a specimen. Instead of using the primary X-ray beam, a high energy electron beam (the same beam for image formation) is used by the X-ray spectrometer in the microscopes. EDS in an electron microscope is suitable for analyzing the chemical elements in microscopic volume in the specimen because the electron probe can be focused on a very small area. Thus, the technique is often referred to as microanalysis.

6.3.1 Special Features

The structure of EDS in an electron microscope is illustrated as in Figure 6.12, using an SEM system as an example. EDS in the SEM is fundamentally similar to a stand-alone EDS except for the primary beam source. In the SEM, the electron beam aligns with the vertical axis of the microscope so that the Si(Li) detector has to be placed at a certain angle from vertical. The angle between the surface plane of the specimen and detector is called the *take-off angle* and is often referred to as the angular position of the detector. The take-off angle can be changed by rotating the specimen surface with respect to the detector. For a low take-off angle, a rough surface may interfere with collection of X-ray photons emitted from a valley area as illustrated in Figure 6.13. Such problems do not occur if the specimen has a microscopically flat surface.

Although the ability to focus the electrons as a probe enables us to analyze chemical elements in a microscopic area, we should know that the minimum specimen area that EDS can resolve is not equal to the diameter of the electron probe. The characteristic X-rays are excited from

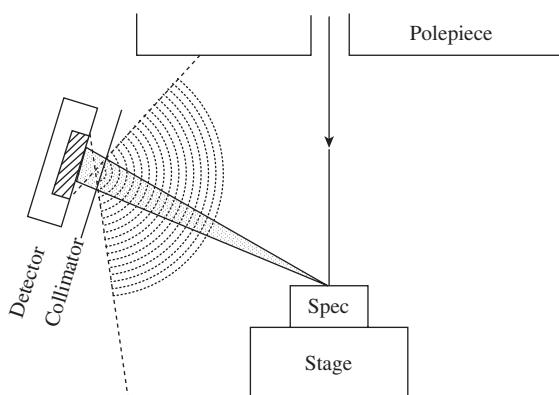


Figure 6.12 Geometrical arrangement of EDS in a scanning electron microscope (SEM). (Reproduced with kind permission of Springer Science and Business Media from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

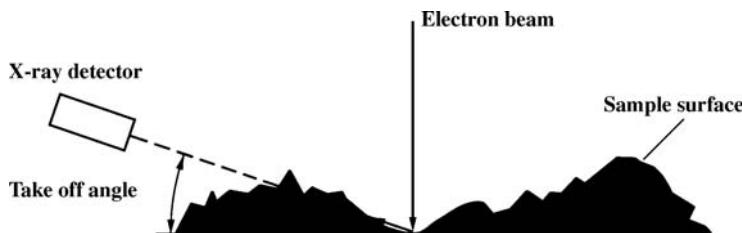


Figure 6.13 Potential interference of X-ray detection due to low take-off angle in the SEM.

a volume under the surface of the specimen, not from an area on the surface. As shown in Figure 6.14, the volume is pear-shaped and its lateral size can be significantly larger than the size of the probe. It is important for us to know, when analyzing a microscopic area of a bulk specimen in an SEM, that EDS signals are emitted from a lateral area that is much larger than the probe diameter. Thus, the spatial resolution of EDS should be considered when we are interested in elemental analysis of a microscopic feature in the SEM.

The volume emitting X-rays (emission volume) is that in which primary electron energy is still greater than the energy required to excite the specific line of characteristic X-rays (E_{ex}). The emission volume is determined by the acceleration voltage of electrons (E_o) and the characteristics of the specimen's component elements, such as mass density (ρ) and E_{ex} . The emission volume has been estimated with the depth of characteristic X-ray generation (R).

$$R = \frac{0.064(E_o^{1.7} - E_{\text{ex}}^{1.7})}{\rho} \quad (6.5)$$

It is reasonable to assume the lateral size of the emission volume is about the same as the depth of the emission volume. Thus, R can represent the spatial resolution of EDS microanalysis. Figure 6.14a illustrates the depth of X-ray generation (R) with the same electron beam energy in two microscopic areas with different densities. Figure 6.14b shows that the resolution will be improved by reducing the acceleration voltage of electron beam (E_o) for characteristic lines of different elements. To obtain a better spatial resolution, we should choose an acceleration voltage that is not much higher than that necessary to excite characteristic X-rays in the specimen.

For thin specimens, as used in a TEM, the spatial resolution is of less concern. The lateral size of the emission volume is not significantly larger than the probe size because the thinness of the specimen limits the spread of the emission volume. For example, with a TEM acceleration voltage of 100 kV and probe diameter of 20 nm, the lateral size will be about 23 nm for aluminum. The electron beam size can be reduced to about 0.5 nm in advanced TEM systems. Thus, chemical analysis of nanometer-sized particles can be achieved in TEM systems.

6.3.2 Scanning Modes

EDS offers two modes of examination: stationary and scanning. In the stationary mode, the probe stays at one location until the collection of X-ray photons is complete. The dwell time of the probe for EDS analysis is determined by the number of X-ray photon counts received by the detector. Elemental detection relies on the signal-to-noise ratio. As discussed in Chapter 4, the

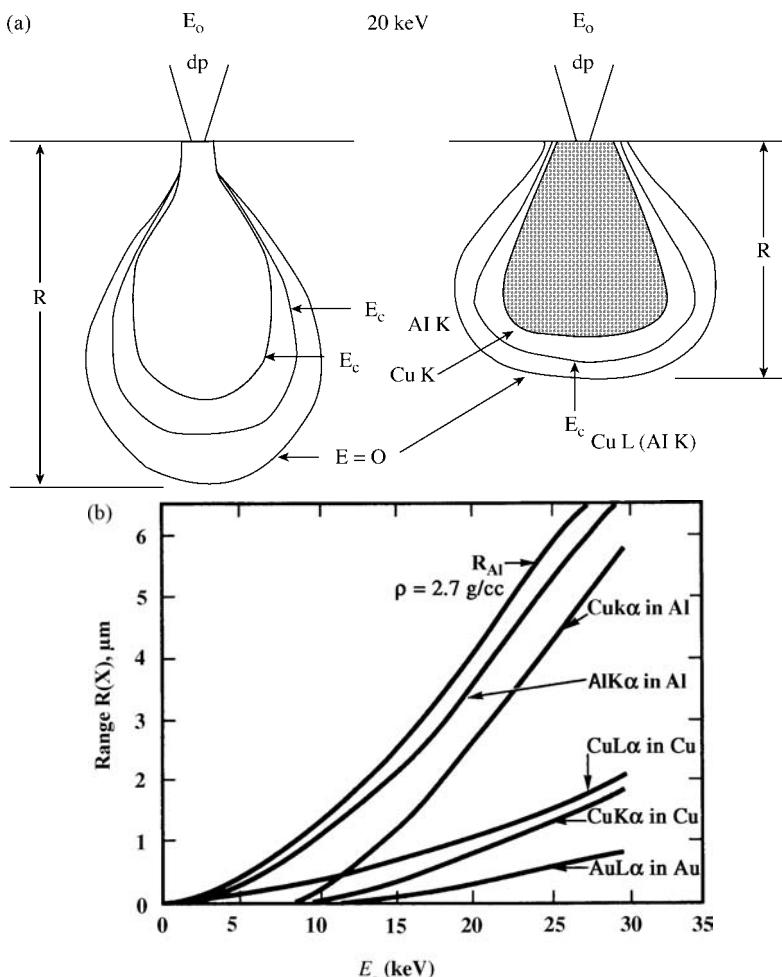


Figure 6.14 (a) Comparison of X-ray production regions with specimens of different mass densities ($\rho = 3$ and 10 g cm^{-3} for left and right examples, respectively); and (b) spatial resolution of EDS as a function of acceleration voltage of electrons and specific lines of characteristic X-rays emitted from bulk specimens. (Reproduced with kind permission of Springer Science and Business Media from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

longer the dwell time, the higher the counts of characteristic X-ray photons. Low concentration of trace elements (<1 wt%) requires longer dwell times for their detection.

In the scanning mode, the electron probe moves over the specimen surface, similar to the way in which the probe moves for obtaining an electron image in the SEM. The intensity of specific characteristic X-rays can be recorded and superimposed on the corresponding electron image as shown in Figure 6.15. The line scans of elements help us to identify the element distribution shown in the SEM image. Also, we can use the scanning mode to make composition maps in which the distributions of elements are clearly revealed. Figure 6.16 shows element maps of an alloy sample in which distributions of alloy elements are revealed by gray scales.

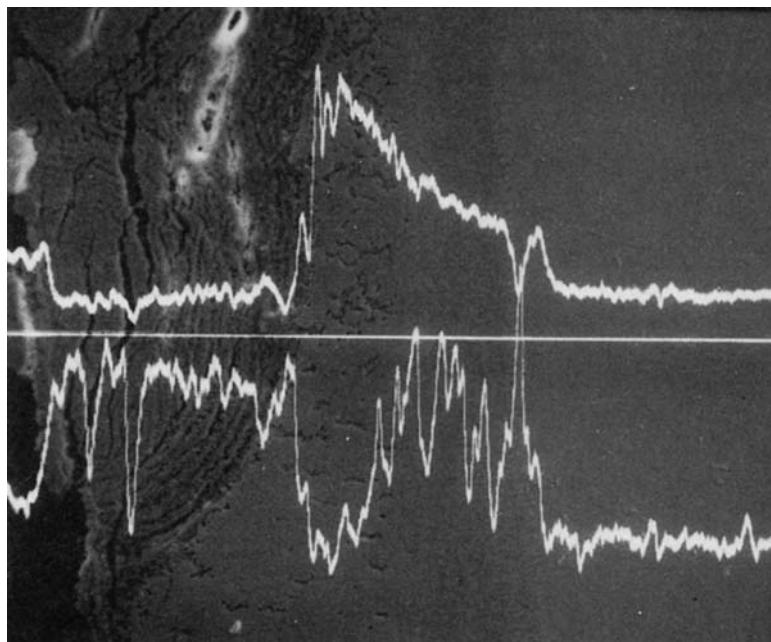


Figure 6.15 Co and Cr line scans across an oxidized high-temperature alloy. The top curve is for cobalt, and the bottom curve is for chromium. (Reproduced with kind permission of Springer Science and Business Media from J.I. Goldstein et al, *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd ed., Plenum Press, New York. © 1992 Springer Science.)

We may select the stationary mode to analyze the elements in microscopic features such as impurities, precipitates and grain boundaries. We may select the scanning mode to examine the composition change in a certain dimension of the specimen or composition variations in a selected area of a specimen.

6.4 Qualitative and Quantitative Analysis

6.4.1 Qualitative Analysis

Identifying the chemical elements in specimens by X-ray spectrometers is in many ways similar to identifying chemical compounds by XRD. However, identifying elements is relatively easy because there are only about 100 elements, while there are several million chemical compounds. We may only find a major peak with a few minor peaks of a certain element in WDS and EDS, not several tens of peaks for a crystalline compound as in an XRD spectrum. For example, Figure 6.17 shows an XRF spectrum (EDS type) for a Ni-Ti alloy. The spectrum shows one major $K\alpha$ peak and one minor $K\beta$ peak for each of Ni and Ti. The $K\alpha$ includes $K\alpha_1$ and $K\alpha_2$, which are often overlapped in the spectrum due to the small difference in their energy levels. The relative positions and intensities of the α and β lines also can help us to identify chemical elements if there is possibility that the $K\alpha$ or $K\beta$ line of one element overlaps another element's L lines.

We should note that the primary X-ray source of XRF includes both continuous and characteristic radiation. The continuous X-rays generate the background of the spectrum. The primary

Image Not Available In The Electronic Edition

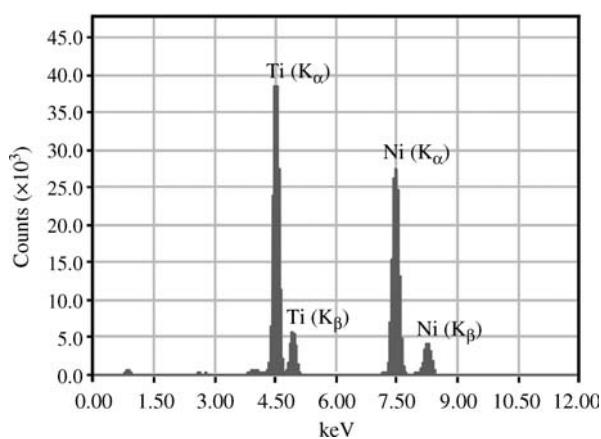


Figure 6.17 EDS spectrum of a Ni–Ti alloy obtained in an XRF spectrometer.

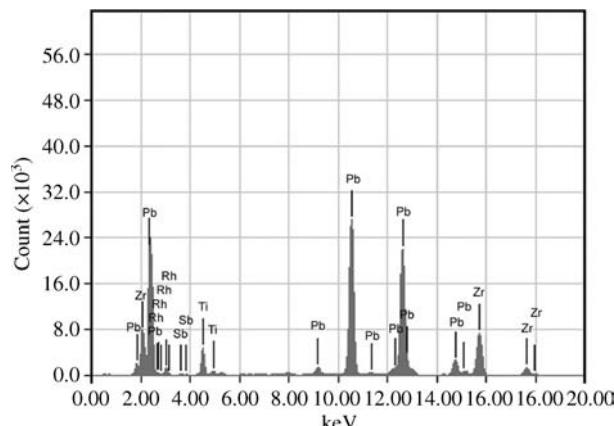


Figure 6.18 XRF spectrum of a Pb alloy with primary X-ray radiation from the X-ray tube with a Rh target. The Rh characteristic lines are not overlapped with those of elements in the alloy specimen.

characteristic X-rays will generate additional peaks in a spectrum. Thus, such background and additional peaks of a target material of the X-ray tube should be deducted from a spectrum. For XRF, an uncommon element such as rhodium (Rh) is often selected as the target material of X-ray tube in order to reduce confusion in identifying elements in samples. For example, Figure 6.18 shows the EDS spectrum of a lead alloy in which the Rh peaks are not overlapped with peaks of sample elements.

Modern X-ray spectrometers are equipped with computer software that is fully capable of identifying the possible elements from a spectrum. We can input the elements that possibly exist in the specimen into the software for qualitative analysis. The computer software will mark peak positions of the input elements in the spectrum. Also, the software will generate such peak lines with correct relative intensities in a spectrum, for example, a correct intensity ratio of $K\alpha$ to $K\beta$. With computer assistance, the errors in element identification can be reduced to minimum.

6.4.2 Quantitative Analysis

We often want to know not only the identity of chemical elements but also their concentrations in a specimen. Thus, quantitative elemental analysis is often required. The concentrations of elements must relate to their peak intensities in the spectrum, similar to the relationship between weight fractions of crystalline phases and their peak intensities in the XRD spectrum. Chemical compositions should be calculated by comparing the ratios of integrated peak intensities among elements in the specimen. In general, the weight fraction (C) of an element in relation to the relative intensities of its peaks (I_R) is mainly affected by the instrument factor (K), and the matrix factor of specimen (M).

$$C = M \times K \times I_R \quad (6.6)$$

The instrument factor K includes conditions of primary sources, the geometrical arrangement of specimen respect to radiation and detection, and detector characteristics. The matrix factor refers to the interactions among the elements in the specimen.

There are three main matrix effects in XRF: *primary absorption*, *secondary absorption* and *secondary fluorescence*. Primary absorption refers to the radiation that is absorbed on the beam's path to reach the atoms to be excited. Secondary absorption refers to absorption of fluorescent radiation from atoms that occur along its path inside the specimen to the detector. Secondary fluorescence refers to the fluorescent radiation from the atoms which are excited by the fluorescent radiation of atoms with a higher atomic number in the same specimen. This phenomenon is possible when energy of the primary fluorescent radiation from heavier atoms is sufficient to excite secondary fluorescence from lighter atoms in the specimen. The absorption effects reduce the intensity of characteristic X-ray lines in spectrum, while secondary fluorescence increases the intensity of lighter elements. The matrix factors of EDS analysis in an electron microscope (EM) are described later in Section 6.8.

Specimen preparation conditions may also affect the relation between intensity and concentration, such as chemical homogeneity. For accurate quantitative analysis, either a standard sample of pure elements to be analyzed or a sample with similar composition to that of the specimen is needed to determine factors M and K in Equation 6.6 precisely.

The following sections will briefly introduce the methods of quantitative analysis by XRF and EDS in EM, separately. There are differences in the factors affecting the analysis because EDS in EM uses an electron beam rather than X-rays as the primary source.

Quantitative Analysis by X-ray Fluorescence

Internal Standard Method

There are several methods of quantitative analysis in XRF. For analyzing a single element in a specimen, we may use an *internal standard method*, one of the most useful techniques for determining a single element weight fraction in a known or unknown matrix. The internal standard refers to adding a standard element of a known concentration into the specimen. This method is based on the assumption that the instrumental and matrix effects are the same for the standard and the element to be analyzed. Thus, we can obtain the weight fraction of element (C_x) from its measured X-ray intensity (I_x) with knowledge of the weight fraction of standard element (C_s) and its corresponding intensity (I_s) by the following relationship.

$$\frac{C_x}{C_s} = \frac{I_x}{I_s} \quad (6.7)$$

The instrument factor for the standard and the element to be analyzed should be the same when their intensities are obtained under identical instrumental conditions. However, the matrix factor of the standard element and the element to be analyzed could be different. The differences may arise from difference in the chemical element or concentrations. To reduce the difference in the matrix factor, the standard element should have an atomic number close to the element to be analyzed and should have the same concentration level as the element to be analyzed. Also, the internal standard method should be used for analyzing an element with low concentration (<10%) to avoid excessive amounts of the standard element being added to the specimen. A significant amount of the standard element would change the characteristics of the matrix. In addition, good mixing of the standard elements with the specimen and similar particle size between the standard and specimen are necessary.

Fundamental Parameter Method

For a modern XRF equipped with a powerful computer system, the *fundamental parameter method* (FP method) is most widely used for quantitative analysis. The method determines the concentration of an element when its theoretical intensity matches its measured intensity. The fluorescence X-ray intensity of a given composition can be calculated using theoretical formulas with given specimen physical and instrumental parameters. The physical parameters include specimen density, thickness, X-ray absorption coefficients and fluorescence yield. The instrumental parameters include excitation voltage of the X-ray tube, optical geometry and detector characteristics.

During quantitative analysis, the computer estimates the composition of the specimen from a measured intensity spectrum. Then, its theoretical intensity is calculated using the FP method. The theoretical intensity, which is on a relative scale, should be converted to the same scale as measured intensity by multiplying with an instrumental sensitivity. The intensity difference between the measured intensity and calculated intensity will trigger a new estimation of composition and start a new cycle of calculation. The final composition is obtained when the difference between the calculated intensity and measured intensity is less than a threshold level. This instrumental sensitivity for each element should be obtained by either measuring the intensity of a standard sample containing the known elemental concentration under the same instrumental conditions or by consulting a sensitivity database stored in the computer. The FP method requires extensive calculation, which is only practical since personal computers have become powerful. The great advantage of the FP method is that it can analyze the specimen when relevant standard samples are not available. This feature is extremely useful in many research situations.

Quantitative Analysis in Electron Microscopy

In general, quantitative analysis by EDS in EM is similar to that of XRF. The analytical methods, however, are different for two main reasons. First, the interactions between the electron beam and specimen are different from those of primary X-ray radiation. Second, an EM specimen for chemical analysis cannot be modified as in the internal standard method. For accurate quantitative analysis of EDS in EMs, separate standard samples containing the elements in the specimen to be analyzed are necessary. The standards should be measured at identical instrumental conditions to the specimen. It means that the spectra of specimen and standard should be collected under the same conditions with regard to the following parameters:

- Acceleration voltage of the electron beam;
- X-ray take-off angle from specimen to detector; and
- Detector coincidence loss (detector dead time).

Even so, the differences in matrix effects between the specimen and standard cannot be ignored. The general formula of quantitative analysis with a standard is expressed by an equation that includes matrix effects.

$$\frac{C_i}{C_{(i)}} = (Z_i A_i F_i) \frac{I_i}{I_{(i)}} = (Z_i A_i F_i) k_i \quad (6.8)$$

The expression $(Z_i A_i F_i)$ represents the matrix factors affecting the X-ray intensity of element i in the specimen (I_i) and in standard ($I_{(i)}$), which can be divided into three major types: *atomic*

number effect (Z_i), *X-ray absorption* (A_i) and *X-ray fluorescence* (F_i). Equation 6.8 represents the quantitative method, called the ZAF method. It is the most commonly used technique that determines composition by calculating the $(Z_i A_i F_i)$ for each element in the specimen. The ZAF method is suitable for quantitative analysis of heavy and medium-weight elements. To understand the method, we need to know about Z , A and F factors.

The atomic number factor (Z) counts the difference in characteristic X-rays generation due to atomic number change in the matrix. It arises from two phenomena: electron backscattering (R) and electron retardation (S). A fraction of primary electrons ($R < 1$) is backscattered and does not generate X-rays. Recalling Chapter 4, backscattering effects are a function of atomic number. An element in a matrix with elements heavier than it will generate a lower intensity than it does in pure form. Electron retardation represents the continuous electron energy loss that occurs when electrons travel in a specimen due to inelastic scattering. This effect is expressed as the *stopping factor* (S).

$$S = \frac{1}{\rho} \frac{dE_o}{ds} \quad (6.9)$$

S is a function of electron travel distance in the specimen (s), acceleration voltage of electrons (E_o), and specimen density (ρ). S decreases while R increases with increase in atomic number.

The X-ray absorption factor (A) counts the decrease in density of characteristic X-rays from emitting location to detector. The absorption follows an exponential relationship as shown in Equation 2.2.

$$I_x = I_o \exp [-(\mu/\rho) \rho t] \quad (6.10)$$

The absorption is strongly affected by the absorption coefficient (μ), density (ρ) and path length (t) that X-rays travel within a specimen before reaching its surface. The X-ray absorption is the greatest factor affecting X-ray intensity. For a certain path length, the X-ray intensity is significantly affected by the mass absorption coefficient ($\mu\rho^{-1}$). For example, the $\mu\rho^{-1}$ difference between Fe and Al in a Ni matrix is more than a factor of 5, comparing 90 with $4837 \text{ cm}^2 \text{ g}^{-1}$.

The fluorescence factor (F) arises from the excitation of characteristic X-rays of the element to be analyzed by the characteristic X-rays emitted from matrix atoms. This phenomenon occurs when the characteristic X-rays from matrix elements have energies greater than that required for exciting characteristic X-rays of the element to be analyzed. Thus, the resulting X-ray intensities of the element to be analyzed will be enhanced by the fluorescence effect. The fluorescence factor is usually least important in the matrix factor, because fluorescence X-rays may not exist, or the concentration of elements emitting fluorescence X-rays may be small.

The ZAF method relies extensively on a computer, similar to the fundamental parameter method. First, the weight fraction of elements in the specimen is estimated from their relative intensities ($k_i / \sum_i k_i$). Then, the $(Z_i A_i F_i)$ factor for each element is estimated and is used for calculating k_i . The calculated k_i will compare with the measured k_i . The iteration of computation continues until the two values converge.

The possible errors for ZAF correction in the analysis will be amplified for light elements ($Z < 10$), because in theoretical calculations of $(Z_i A_i F_i)$ factors, approximations that are valid

for medium-weight and heavy elements fail to apply for light elements. Thus, other methods specially designed for light elements are needed.

The ZAF method is based on assumptions that the specimen is in bulk and its surface is microscopically flat. Thus, we should be cautious when a specimen does not meet these assumptions, such as a thin film specimen or a specimen with a rough surface. For example, with a thin film specimen, the matrix effect will be reduced and the intensity–weight fraction relationship will be simplified.

Standardless quantitative analysis is often available in the EDS system of an EM. The standardless technique is desirable when standards are not available. The standardless technique is based on prediction of the X-ray intensity that would be obtained from a pure element standard under the same instrumental conditions. This method works well for a specimen consisting of transition elements of which the K lines span the range 5–10 keV, such as stainless steel. The method should be used cautiously because it potentially generates large errors that we may not recognize. Standardless analysis may generate large errors for specimens in which the elements generate low energy lines (<3 keV). The standardless method may generate considerable errors when multiple X-ray families, K , L and M , are involved in analysis. There are several ways we can double-check the result of standardless analysis; for example, we can use a different family of X-ray lines (K and M), change electron beam energy, or change the count rate.

References

- [1] Brandon, D. and Kaplan, W.D. (1999) *Microstructural Characterization of Materials*, John Wiley & Sons, Chichester.
- [2] Jenkins, R. and Snyder, R.L. (1996) *Introduction to X-Ray Powder Diffractometry*, John Wiley & Sons, New York.
- [3] Jenkins, R. (1999) *X-ray Fluorescence Spectrometry*, 2nd edition, John Wiley & Sons, New York.
- [4] Lifshin, E. (1999) *X-ray Characterization of Materials*, Wiley-VCH Verlag GmbH, Heinheim.
- [5] Goldstein, J.I., Romig, Jr., A.D., Newbury, D.E., Lyman, C.E., Echlin, P., Fiori, C., Joy, D.C. and Lifshin, E. (1992) *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd edition, Plenum Press, New York.
- [6] Jenkins, R., Gould, R.W. and Gedcke, D. (1995) *Quantitative X-ray Spectrometry*, 2nd edition, Marcel Dekker, New York.
- [7] Klockenkamp, R. (1997) *Total-Reflection X-ray Fluorescence Analysis*, John Wiley & Sons, New York.
- [8] Goldstein, J.I. and Yakowitz, H. (1977) *Practical Scanning Electron Microscopy*, Plenum Press, New York.
- [9] Goodhew, P.J. and Humphreys, F.J. (1988) *Electron Microscopy and Analysis*, 2nd edition, Taylor & Francis Ltd, London.

Questions

- 6.1 Compare the similarities of, and differences between, X-ray diffractometry and X-ray fluorescence spectrometry.
- 6.2 For an industrial application that requires fast identification of elements, what XRF system (WDS or EDS) would you choose? What are the benefits and disadvantages of your choice?
- 6.3 Count the total number of K lines for electron transitions from L and M shells to the K shell.
- 6.4 How would you adjust the X-ray tube to ensure the excitation of fluorescence X-rays with energy of 15 keV? Why should we choose appropriate target materials in the X-ray tube for XRF?

- 6.5 Select analyzing crystals of WDS to detect the *K* lines of elements Ca, Si, Na and N.
- 6.6 Commonly, an EDS spectrum is recorded in the range 0–20 keV. What are the possible *K*, *L* and *M* lines to be detected in a specimen containing Mg, Zn, Ag and Au?
- 6.7 Why should the Si(Li) detector of an EDS instrument be cooled by liquid nitrogen? Why may we detect light elements such as C and O without a beryllium window on the EDS detector?
- 6.8 Why are we likely to overestimate the weight fraction of Ni from the WDS spectrum of a specimen containing Ni and Ag? Is there any way to detect this error by examining the spectrum?
- 6.9 What change in the EDS spectrum do you expect when a specimen is tilted to face the detector in an SEM?
- 6.10 Can you use the ZAF method for quantitative analysis for EDS in a TEM? Why or why not?

7

Electron Spectroscopy for Surface Analysis

Electron spectroscopy is a technique that uses characteristic electrons emitted from a solid for elemental analysis, not for imaging as in electron microscopy. The characteristic electrons (either Auger electrons or photoelectrons) exhibit characteristic energy levels, revealing the nature of chemical elements in specimens being examined. Auger or photoelectrons can only escape from the uppermost atomic layers of solid (a depth of 10 nm or less) because their energies are relatively low (generally 20–2000 eV), while the characteristic X-rays can escape from a much greater depth (several micrometers from the surface). Thus, electron spectroscopy is a technique for surface chemical analysis. There are two types of electron spectroscopy: Auger electron spectroscopy (AES) and X-ray photoelectron spectroscopy (XPS). Auger electrons and photoelectrons are different in their physical origins, but both types of electrons carry similar information about chemical elements in material surfaces.

7.1 Basic Principles

7.1.1 X-ray Photoelectron Spectroscopy

The *X-ray photoelectron* is an electron ejected from an electron shell of an atom when the atom absorbs an X-ray photon. Figure 7.1a schematically illustrates the emission of a photoelectron from an atom when it is excited by an X-ray photon. As mentioned in Chapter 6, an incident X-ray photon can have sufficient energy (a value of $h\nu$) to knock out an inner shell electron, for example from the atom's *K* shell. In such a case, the *K*-shell electron would be ejected from the surface as a photoelectron with kinetic energy E_K . Knowing the kinetic energy E_K , we can calculate the binding energy of the atom's photoelectron (E_B) based on the following relationship.

$$E_B = h\nu - E_K - \Phi \quad (7.1)$$

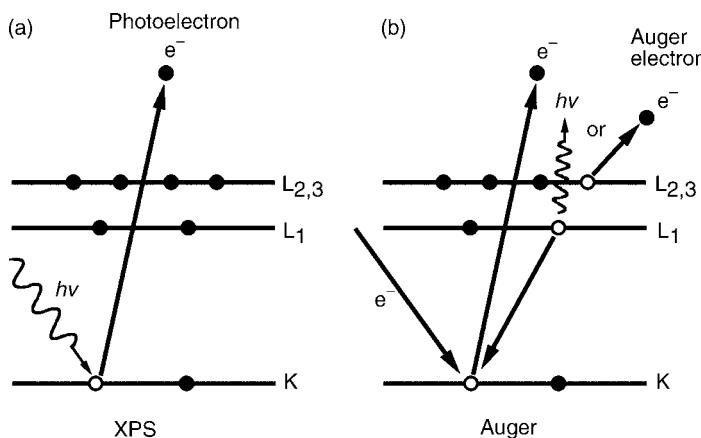


Figure 7.1 Emission processes of characteristic electrons: (a) a 1s photoelectron; and (b) a $KL_1L_{2,3}$ Auger electron

Φ is the parameter representing the energy required for an electron to escape from a material's surface, h is Planck's constant and ν is the frequency. The value of Φ depends on both the sample material and the spectrometer. The binding energies (E_B) of atomic electrons have characteristic values, and these values are used to identify elements, similar to the way that characteristic X-ray energy is used in X-ray spectroscopy. X-ray photoelectron spectroscopy (XPS) identifies chemical elements from the binding energy spectra of X-ray photoelectrons. A typical XPS spectrum is a plot of intensity versus binding energy. Photoelectrons are ejected from different electronic shells and subshells. Each binding energy peak is marked as an element symbol plus a shell symbol from where the photoelectron was emitted, for example Al 2p, O 1s... as shown in Figure 7.2. The photoelectrons emitted by subshells p, d and f are commonly marked with an additional fraction number; for example, Cu 2p $\frac{1}{2}$. There are $\frac{1}{2}$ and $\frac{3}{2}$ for the p subshell, $\frac{3}{2}$ and $\frac{5}{2}$ for the d subshell, and $\frac{5}{2}$ and $\frac{7}{2}$ for the f subshell. These fractions represent the quantum number of total angular momentum (J) for an individual shell electron, as mentioned in section 6.1.1. An XPS spectrum may also contain peaks from Auger electrons. For example, the spectrum shown in Figure 7.2 includes Auger electron peaks marked as OKLL and CKLL. The generation of Auger electrons is discussed in the following section.

7.1.2 Auger Electron Spectroscopy

Auger electrons were named after Pierre Auger who, together with Lise Meitner, discovered Auger electron emission in 1920s. The Auger electrons were mentioned in the discussion of the interaction between incident electrons and atoms in Chapter 6. Figure 7.1b illustrates the emission process of an Auger electron. When a high-energy electron (or X-ray photon) strikes an inner shell electron of an atom, the energy of the incident particle can be high enough to knock out the inner shell (K shell in Figure 7.1b) electron. Thus, the atom becomes ionized and in an excited state. The atom will quickly return to its normal state after refilling the inner electron vacancy with an outer shell electron. In the meantime, the energy difference between the outer shell electron and the inner shell may cause emission of either an Auger

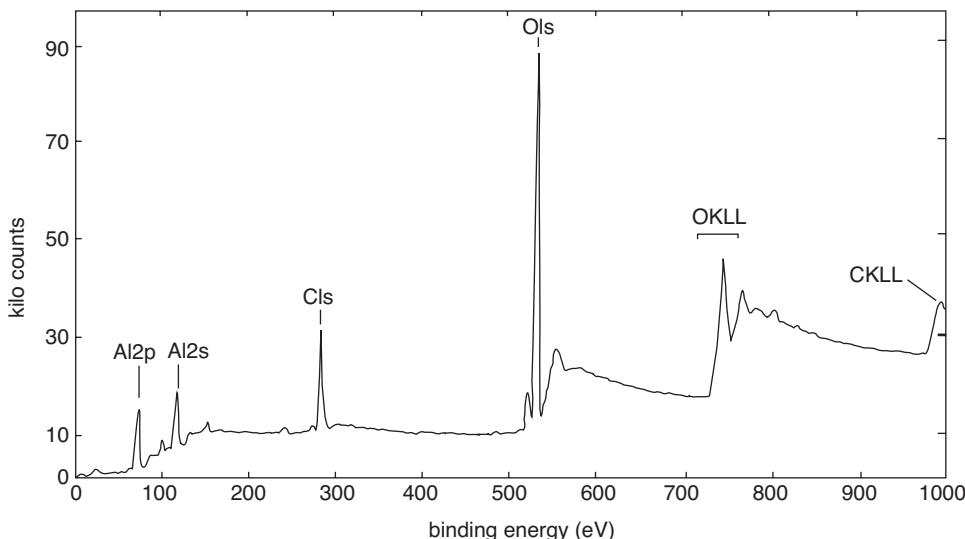


Figure 7.2 XPS spectrum of an oxidized aluminum surface. (Reproduced with permission from J.F. Watts, *An Introduction to Surface Analysis by Electron Spectroscopy*, Oxford University Press, Oxford. © 1990 Royal Microscopy Society.)

electron from an electron shell or a characteristic X-ray photon. The kinetic energy of an Auger electron is approximately equal to the energy difference between binding energies in the electron shells involved in the Auger process. For example, the kinetic energy of an Auger electron in Figure 7.1b is approximated by the following equation.

$$E_{KL_1L_{2,3}} \approx E_{BK} - E_{BL_1} - E_{BL_{2,3}} \quad (7.2)$$

The notation for kinetic energy of Auger electron describes its origin, but the nomenclature is rather complicated. For example, the kinetic energy of an Auger electron in Equation 7.2 is from a process illustrated in Figure 7.1b; that is, an incident electron knocks out a K shell electron, a L_1 shell electron refills the K shell vacancy, and a $L_{2,3}$ shell electron is ejected as the Auger electron. The subscript of E_{BX} indicates the binding energy of electron shell X; for example, E_{BK} is the binding energy of the K shell. Auger electron spectroscopy (AES) identifies chemical elements by measuring the kinetic energies of Auger electrons. In an AES spectrum, an individual kinetic energy peak from an Auger electron is marked with an elemental symbol and subscripts indicating the electron shells or subshells involved, for example, Al_{KLL} , O_{KLL} . . . A typical AES spectrum is a plot of intensity versus kinetic energy; most commonly, it is a plot of the first derivative of intensity versus the kinetic energy as shown in Figure 7.3.

The Auger peaks appear small against the background of the direct mode spectrum. This occurs because the signals from Auger electrons are relatively weak compared with those of secondary electrons escaped from a solid surface. The interactions between incident electrons with material atoms are discussed in Chapter 4. Recalling Figure 4.10, electrons are either elastically or inelastically scattered when a primary electron beam strikes a solid. Figure 7.4 illustrates numbers of Auger electrons compared with numbers of other electrons that escape

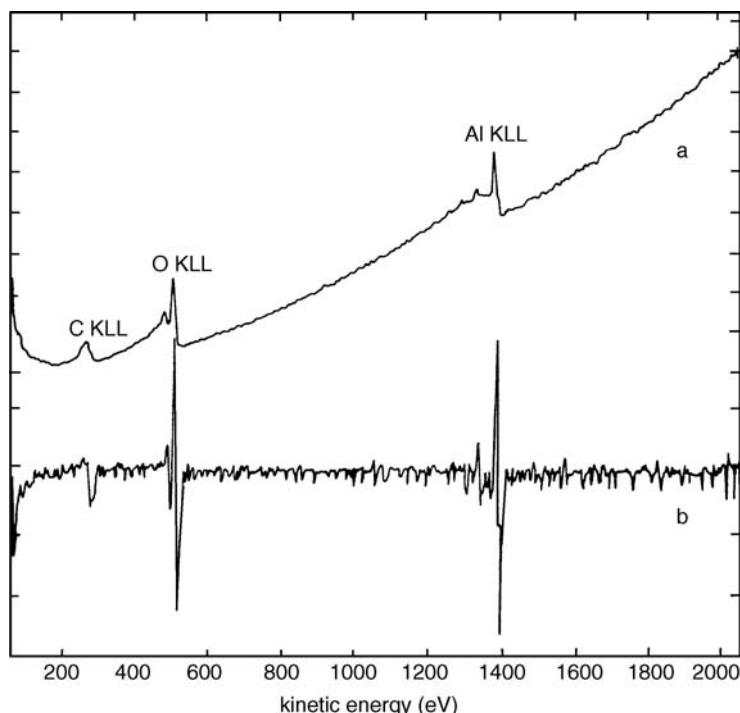


Figure 7.3 AES spectra of an oxidized aluminum surface: (a) direct spectrum of intensity versus kinetic energy of Auger electrons; and (b) differential spectrum of intensity versus kinetic energy of Auger electrons. (Reproduced with permission from J.F. Watts, *An Introduction to Surface Analysis by Electron Spectroscopy*, Oxford University Press, Oxford. © 1990 Royal Microscopy Society.)

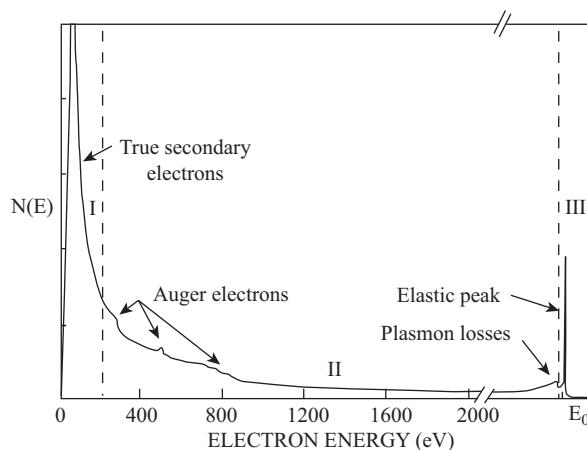


Figure 7.4 Schematic comparison of Auger peak intensity with other electrons escaped from a solid surface. E_0 indicates energy of incident electrons. The kinetic energy of electrons can be divided into three regions I, II and III from low to high. (Reproduced with permission from M. Prutton and M.M. El Gomati, *Scanning Auger Electron Microscopy*, John Wiley & Sons Ltd, Chichester. © 2006 John Wiley & Sons.)

from a solid surface. The primary electrons ejected from a solid surface by inelastic scattering comprise the background of an AES spectrum in the region of high kinetic energies while the secondary electrons comprise the background in the region of low kinetic energies. The number of Auger electrons ejected from a sample is much less than that of scattered primary electrons and secondary electrons. Thus, the direct-mode spectrum is not favored except for quantitative analysis.

7.2 Instrumentation

A modern instrument for electron spectrometry contains both XPS and AES in a single chamber as a multifunctional surface analysis system. A scanning electron microscope (SEM) system may also be included in order to image the microscopic area to be examined by electron spectroscopy. Figure 7.5 schematically illustrates a system combining AES and XPS, which includes an electron gun, an X-ray gun and a shared analyzer of electron energy. The electron beam for generating the Auger electron emission can be easily focused and scanned over a sample surface to obtain two-dimensional mapping for AES analysis. Thus, AES analysis is commonly the scanning type, called *scanning Auger microscopy*.

7.2.1 Ultra-High Vacuum System

Instruments for electron spectrometry require an ultra-high vacuum environment with a vacuum pressure in the range 10^{-8} – 10^{-10} mbar ($1 \text{ mbar} = 100 \text{ Nm}^{-2} = 1.33 \text{ torr}$). Reasons for such a high vacuum are to reduce the chances of low-energy electrons being scattered by gas molecules on their way to reach the detector, and to keep the sample surface free from gas molecule contamination. Low-energy photoelectrons and Auger electrons are easily scattered by gas molecules. Scattering will reduce signal intensity and increase background noise in spectra. Surface contamination by gas molecules is a major concern for surface chemical analysis. It is possible for a surface to adsorb a monolayer of gas molecules within one second in a vacuum environment of 10^{-6} mbar. However, the typical time to collect XPS and AES spectra is more than several hundred seconds. Only an ultra-high vacuum environment ensures satisfactory analysis, particularly for surfaces containing elements that exist in gaseous environments such as C, O and N.

The ultra-high vacuum chamber is commonly made from stainless steel, and joints of chamber parts are made from crushed copper gaskets. Ultra-high vacuum pressure can be achieved using diffusion pumps, sputter ion pumps or turbomolecular pumps. Currently, sputter ion pumps and turbomolecular pumps are more commonly used than diffusion pumps that may contaminate the chamber by leaking oil gas molecules.

Vacuum pumping alone cannot achieve the necessary ultra-high vacuum because gas molecules will come from the chamber in such a low-pressure environment. Thus, the vacuum chamber must be baked at an elevated temperature (250–350 °C) and pumped. This baking process allows the gas molecules adsorbed onto the chamber walls to be pumped out. An additional requirement for the chamber is magnetic shielding, because the trajectory of signal electrons is strongly affected by any magnetic field, even the Earth's magnetic field.

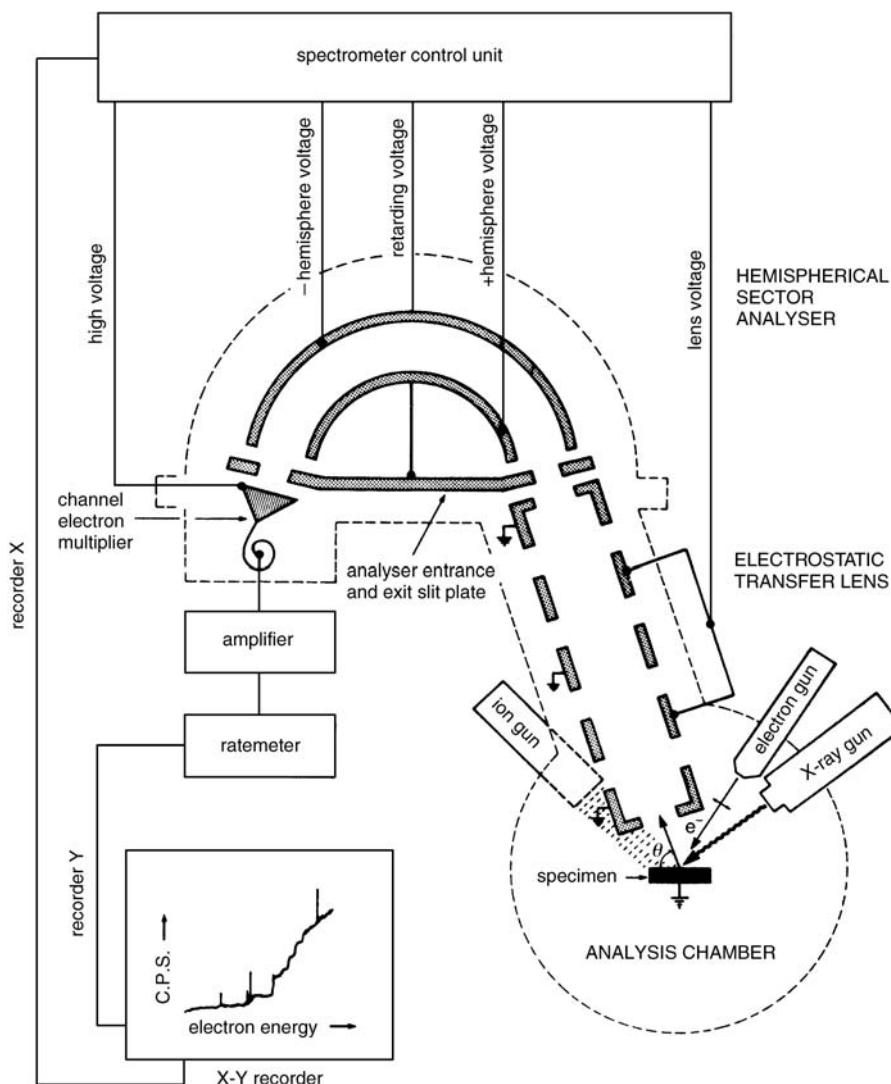


Figure 7.5 Structure of an electron spectrometer. (Reproduced with permission from J.F. Watts, *An Introduction to Surface Analysis by Electron Spectroscopy*, Oxford University Press, Oxford. © 1990 Royal Microscopy Society.)

7.2.2 Source Guns

X-ray Gun

An electron spectrometer system contains an **X-ray gun** for XPS analysis. The working principles of the X-ray gun are similar to the X-ray tube used for X-ray diffractometry introduced in Chapter 2. X-ray photons are generated by high-energy electrons striking a metal anode, commonly Al or Mg for XPS spectrometry. The X-ray gun produces a characteristic X-ray

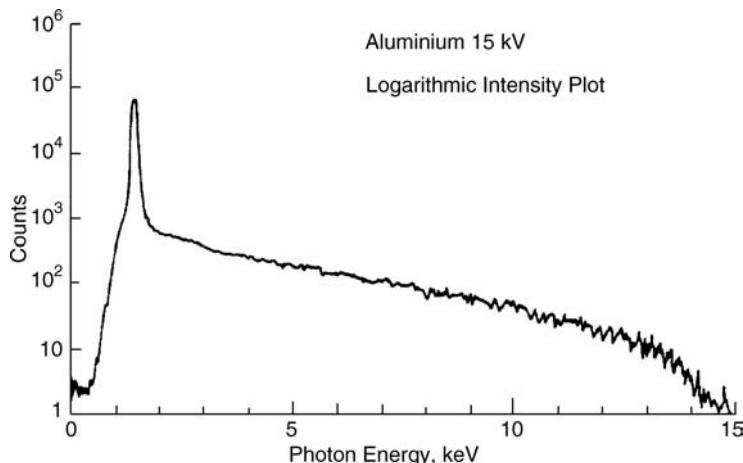


Figure 7.6 Non-monochromatic X-ray radiation from an X-ray gun with an Al target. The characteristic AlK α line is at about 1.5 keV. (Reproduced with permission from D. Briggs and J.T. Grant, *Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, IM Publications and Surface Spectra Ltd, Chichester. © 2003 IM Publications.)

line to excite atoms of the surface to be analyzed. XPS uses both non-monochromatic and monochromatic X-ray sources. The output from a non-monochromatic X-ray source consists of a continuous energy distribution with high intensity of K α characteristic lines as shown in Figure 7.6. The output of the monochromatic source is produced by removing continuous X-rays from a radiation spectrum. The monochromatic source is useful for obtaining XPS spectra with reduced background intensity.

Energies of the characteristic X-ray lines used in XPS are lower than those used in X-ray diffractometry. For example, the energies of AlK α and MgK α are 1.4866 and 1.2536 keV, respectively; however, the energies of CuK α and MoK α , which are commonly used in X-ray diffractometry, are 8.04 keV and 17.44 keV, respectively. The reason to choose lower energy X-rays, also called ‘soft’ characteristic X-rays, is their narrow line width. The line width of characteristic X-rays refers to their range of energy. XPS requires a line width less than 1.0 eV to ensure good energy resolution. Both AlK α and MgK α exhibit line widths less than 1.0 eV and also have sufficient energies (>1000 eV) for photoelectron excitation. Different from the X-ray tube used in an X-ray diffractometer, a single X-ray gun in XPS has both Al and Mg anodes; its structure is schematically shown in Figure 7.7. Such a gun structure makes switching between MgK α and AlK α simple.

Electron Gun

The electron guns used in AES analysis is similar to those used in electron microscopy. LaB₆ and field emission guns are commonly used in electron spectrometers. The LaB₆ gun can provide an electron beam of high brightness with spatial resolution of 200 nm after being focused by an electromagnetic lens. Field emission guns are increasingly used in modern instruments. These guns provide superior brightness and higher spatial resolution than LaB₆; in addition, their emitting surface remains clean during operation without adsorption of gas molecules.

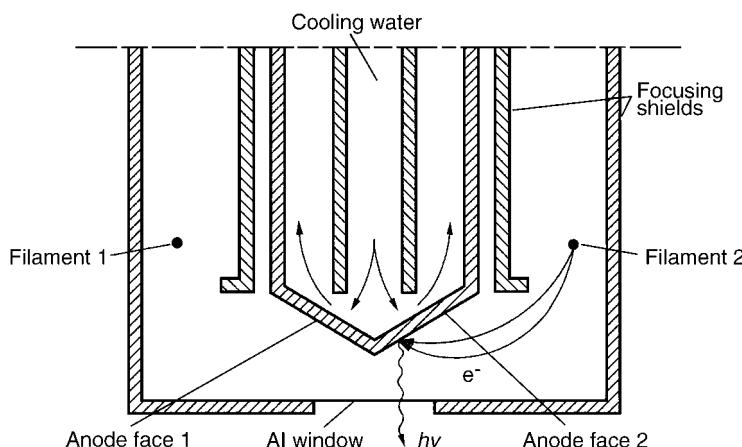


Figure 7.7 An X-ray gun with two-anodes. Two tapered anode faces (one is Al and the other is Mg) have semi-circular filaments, which are near ground potential. An acceleration voltage of about 15 kV between a filament and anode generates X-rays that exit through an Al window. (Reproduced with permission from D. Briggs and M.P. Seah, *Practical Surface Analysis*, Volume 1, John Wiley & Sons Ltd, Chichester. © 1990 John Wiley & Sons Ltd.)

Ion Gun

An electron spectrometer is also equipped with an ion gun as shown in a system combining AES and XPS (Figure 7.5). The functions of an ion gun are twofold. First, it provides a high-energy ion flux to clean sample surfaces before examination. This is necessary because the signal electrons come from the surface atom layers of a sample. Sample surfaces are commonly contaminated with adsorbed hydrocarbons, water vapor and oxides that need to be removed before surface analysis. The second function of the ion gun is to sputter out sample atoms layer by layer so that an elemental depth profile can be revealed. The ion gun produces an argon ion beam by either electron impact or gaseous discharge. This beam has an energy level of 0.5 to 5.0 keV, and can be focused to a diameter down to several tens of micrometers. The ion beam can scan a surface area as large as 10×10 mm. The ion gun will be discussed in more detail in Chapter 8 with secondary ion mass spectroscopy.

7.2.3 Electron Energy Analyzers

An electron spectrometer relies on an electron energy analyzer to obtain XPS and Auger spectra. The most commonly used analyzer is the *concentric hemispherical analyzer* (CHA), also called the *hemispherical sector analyzer* (HSA) as shown in Figure 7.5. The analyzer is composed of two concentric hemispheres with radii R_1 and R_2 and its working principles can be understood from the schematic illustration in Figure 7.8. Negative potentials V_1 and V_2 are applied to the inner and outer hemispheres, respectively. The applied potential generates a median equipotential surface with radius of R_o . The potential along the median surface (V_o) is called the *pass energy* of the CHA. A slit at one end of the CHA allows electrons from the sample to enter, and a slit at the other end lets electrons pass through to an electron detector. The CHA only allows the electrons with energy $E = eV_o$, which are injected tangentially to the median surface, to pass through its channel and reach the detector.

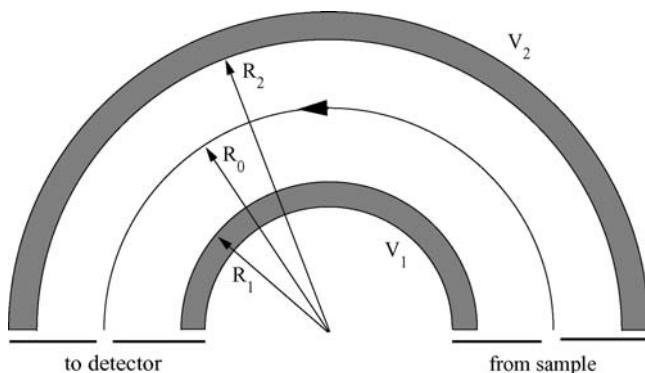


Figure 7.8 Working principles of a concentric hemispherical analyzer.

Before being injected into the CHA, electrons are focused by an electrostatic transfer lens, and their energies are reduced to a certain level. Electron energy reduction by electrostatic processing is called *electron retardation*. The electrons can be analyzed by one of two modes: the *constant analyzer energy* (CAE) mode or *constant retarding ratio* (CRR) mode. Commonly, the spectrometer is run in the CAE mode for XPS and in the CRR mode for Auger spectroscopy. The CAE mode maintains the pass energy of the CHA as a constant. Electron energies are analyzed by recording the change of electron retardation. The CRR mode maintains the ratio of electron retardation as a constant and the electron energies are analyzed by recording the change of pass energy.

It is necessary to have knowledge of electron energy resolution and CHA characteristics to understand the reasons for using different modes. Electron energy resolution can be expressed either in an absolute form (ΔE) or as a relative form ($\Delta E/E$). XPS requires high absolute resolution of about 0.5 eV in the whole range of a spectrum. However, the CHA has a relative resolution limit, which is determined by the geometric configuration of the CHA, including such factors as hemisphere radii and slit opening. A resolution of 0.5 eV is readily achievable when the photoelectron energy is low because the relative resolution is also low. For example, for $E = 200$ eV, a CHA requires a relative resolution of 0.025 to satisfy the XPS absolute resolution of 0.5 eV. However, for $E = 1500$ eV, a CHA requires a relative resolution of 0.003 to do so, which is not practical. Thus, to ensure absolute resolution in a whole spectrum range, XPS analysis requires retarding the electron energy to achieve a low CHA pass energy, commonly 10–100 eV. For example, a photoelectron with energy of 1000 eV will be retarded to a pass energy of 50 eV. An XPS spectrum is recorded by changing the retarding electric potential in the CAE mode.

Auger analysis requires suppressing the electron signal at the low energy end of its spectrum. The CRR mode meets the Auger analysis requirement because the CHA mode exhibits a low transmission rate with low pass energy. When a constant retardation ratio is applied, a low Auger electron energy generates low CHA pass energy. For example, with a retardation ratio of 10, the pass energy is only 10 eV for $E = 100$ eV, while the pass energy is 100 eV for $E = 1000$ eV. In other words, electron transmission through a CHA is lower at a pass energy of 10 eV than at 100 eV. An Auger spectrum is commonly recorded by changing the CHA pass energy in the CRR mode.

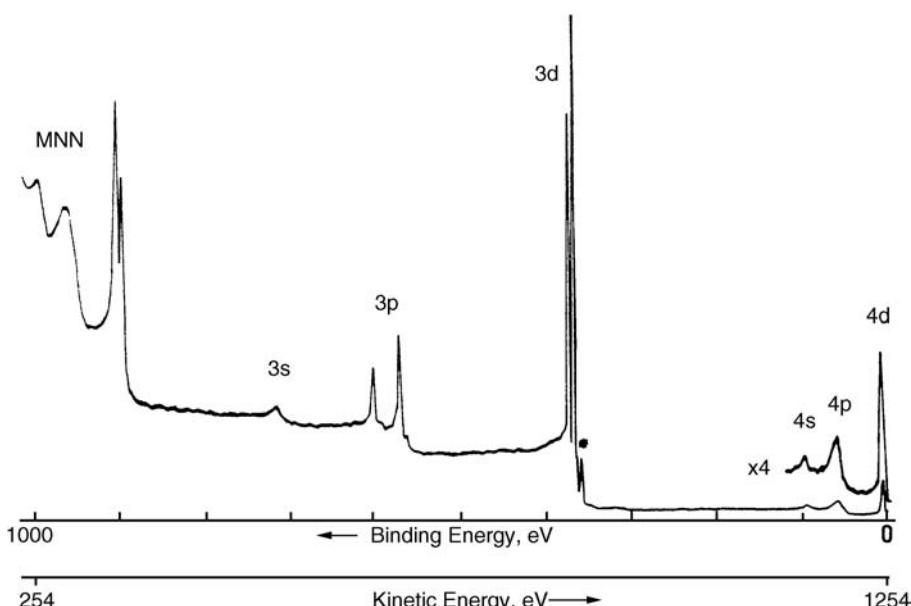


Figure 7.9 An XPS spectrum of silver excited $MgK\alpha$ with pass energy of 100 eV. (Reproduced with permission from D. Briggs and M.P. Seah, *Practical Surface Analysis*, Volume 1, John Wiley & Sons Ltd, Chichester. © 1990 John Wiley & Sons Ltd.)

7.3 Characteristics of Electron Spectra

7.3.1 Photoelectron Spectra

XPS spectra are expressed as photoelectron intensity versus binding energy (E_B) as shown in Figure 7.9. An XPS spectrum can have three types of peaks on a background: photo-emission from core electron levels, photo-emission from valence levels and Auger emission excited by X-rays. Figure 7.9 shows an XPS spectrum of a clean silver surface in which the core-level peaks are marked as 3s, 3p and 3d; the valence-level peak is marked as 4d and the Auger peaks are marked as *MNN*. The core-level photoelectron peaks are the primary peaks for elemental analysis. The valence-level peaks are those at low binding energy (0–20 eV) and are primarily useful in studies of the electronic structure of materials. The Auger peaks, arising from X-rays excited Auger process, are also useful for chemical analysis.

XPS spectra have a step-like background, increasing with binding energy. They result from inelastic scattering of photoelectrons in a solid. It should be noted that the X-ray radiation is commonly not exactly monochromatic. The photoelectron emission can also be excited by continuous X-rays, which are commonly associated with characteristic X-ray radiation. Such photoelectron emission generates a background in the low binding energy region of spectrum. The XPS spectrum can also include extra, or satellite, peaks associated with a main core-level peak as shown in Figure 7.10.

Shake-up satellites are the extra peaks which result from interaction between a photoelectron and a valence electron. A photoelectron can excite (shake-up) a valence electron to a higher energy level and thereby lose a few electron volts of kinetic energy. This will create a satellite

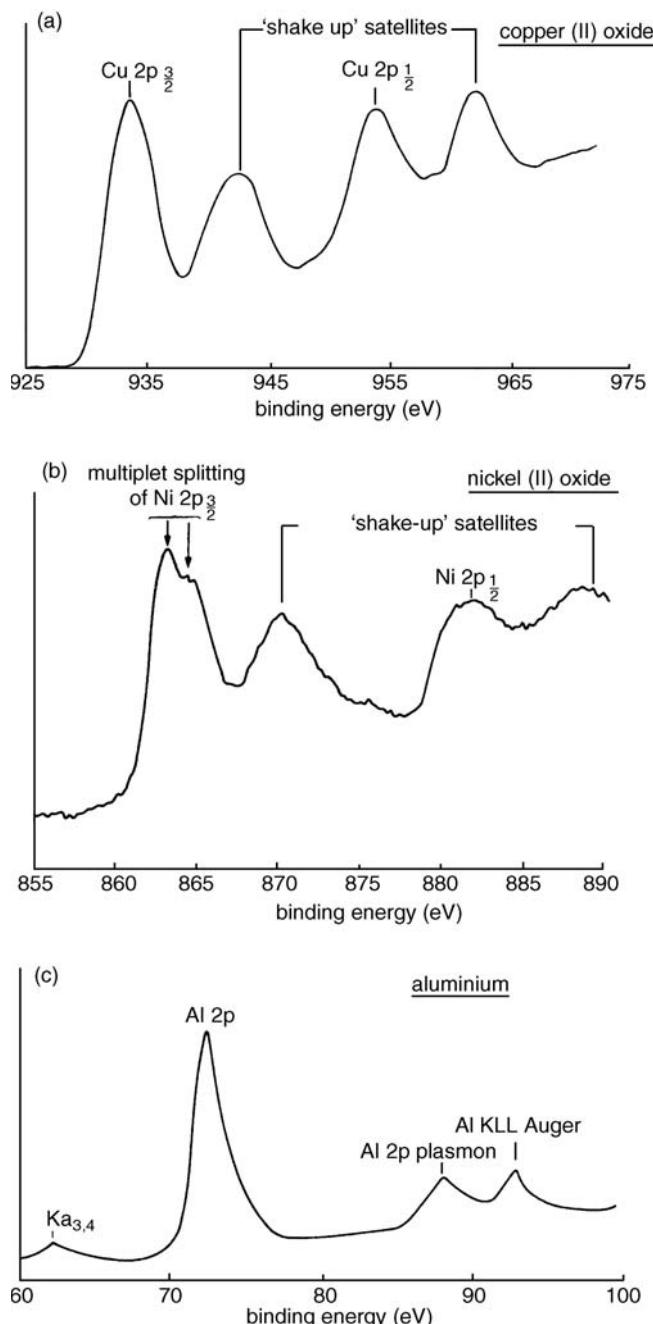


Figure 7.10 Examples of several types of satellite peaks in XPS spectra: (a) shake-up peaks in a CuO spectrum; (b) shake-up peaks and multiplet splitting in a NiO spectrum; and (c) plasmon loss peak in a clean Al spectrum. (Reproduced with permission from J.F. Watts, *An Introduction to Surface Analysis by Electron Spectroscopy*, Oxford University Press, Oxford. © 1990 Royal Microscopy Society.)

peak associated with a core-level peak of photoelectrons as shown in Figure 7.10a. For certain transition and rare earth metals with unpaired electrons in 3d and 4f shells, the shake-up satellites produce strong peaks. Organic materials with aromatic systems also show shake-up satellites with intensity up to approximately 10% of core-level peaks. Shake-up satellites are useful for chemical analysis. For example, a shake-up satellite is associated with the Cu 2p of CuO, but it is absent around the Cu 2p peak of Cu₂O.

Multiplet splitting of a core-level peak may occur in a compound that has unpaired electrons in its valence level. For example, the core-level peak of Ni 2p_{3/2} of NiO in Figure 7.10b shows multiplet splitting. Detailed explanation of multiplet splitting involves knowledge of quantum theory of electron structure, which is beyond the scope of this book. Multiplet splitting is also useful in chemical analysis. For example, Ni(OH)₂ can be distinguished from NiO because the 2p_{3/2} of Ni(OH)₂ does not have multiplet splitting.

Plasmon loss generates another type of satellite peaks; these peaks do not provide useful information but they complicate a spectrum. Plasmon loss refers to the energy loss of a photoelectron because it excites collective vibrations in conduction electrons in a metal. The vibrations require a characteristic amount of energy. Such characteristic amounts of photo-electron energy loss in photoelectrons will generate satellite peaks as shown in Figure 7.10c. Plasmon loss peaks are shown in XPS spectra of clean metal surfaces and also in Auger spectra.

7.3.2 Auger Electron Spectra

Auger spectra can be expressed in two modes: a direct mode and differential mode as shown in Figure 7.3. The direct mode presents the intensity distribution in a range of electron kinetic energies. The differential mode presents the derivative of intensity versus the kinetic energy. The differential mode is more widely used because the Auger peaks are more obvious than in the direct mode. Figure 7.11 shows the Auger spectra of contaminated tungsten foil as an example. The CRR mode of the CHA generates a constant relative resolution ($\Delta E/E$) over the whole range of the spectrum. The direct mode spectrum required with the CRR mode expresses the energy distribution as the number of electrons multiplied by its kinetic energy, $EN(E)$, as shown in Figure 7.11a.

The differential mode spectrum is produced by taking the first derivative of the curve in the direct mode using computer software. The ordinate should be $EdN(E)$ in the differential mode spectrum with the CRR acquisition as shown in Figure 7.11b. The differential mode effectively reduces the background and enhances the Auger peaks. The Auger energy in the differential mode is represented by a sharp minimum position corresponding to an Auger signal. For example, Figure 7.12 shows the energy of carbon KLL line is 272 eV in the differential mode; however, this energy is not exactly identical to the carbon KLL peak energy in the direct mode. It is slightly higher because the exact peak energy in the direct mode should equal to that of the differentiated curve passing zero. In practice, such difference is often ignored for element identification.

A light element is often identified from its KLL Auger lines, which dominate in the Auger spectrum range. However, for an element with atomic number higher than 15, either LMM or MNN Auger lines are dominant. Figure 7.13 is a chart of the principal Auger electron energies for different elements. The LMM lines for an element are divided into three, as triplets. The LMM triplet feature results from the difference in subshells involved in the Auger process. Figure 7.14 shows triplet spectra of three transition metals. In iron, for example, triplet lines at 703, 651 and 598 eV correspond to the Auger processes of $L_3M_{45}M_{45}$, $L_3M_{23}M_{45}$

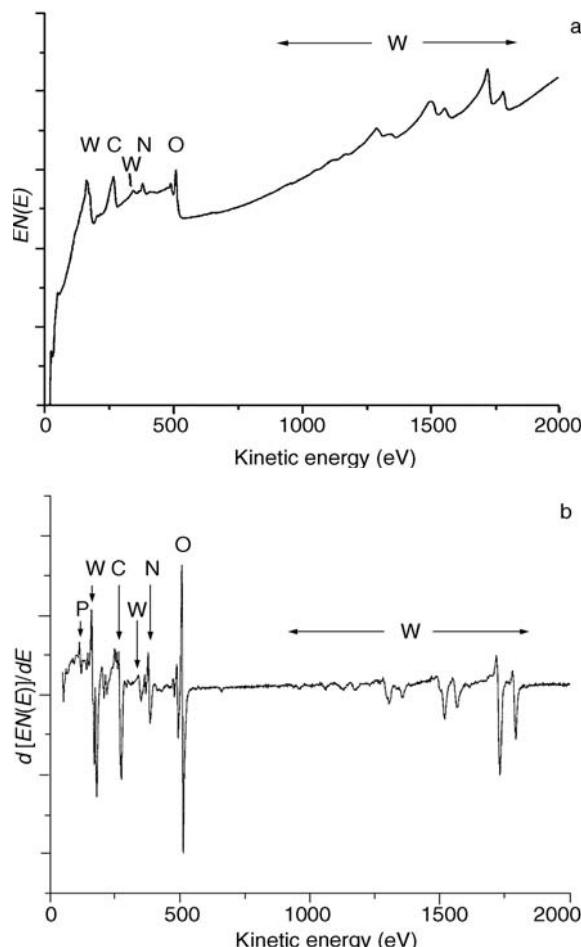


Figure 7.11 Auger spectra of a contaminated tungsten foil acquired in a fixed retarding ratio mode with 0.6% relative resolution: (a) direct spectrum; and (b) differential spectrum. Elements P, N, O, W, C are indicated. (Reproduced with permission from D. Briggs and J.T. Grant, *Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, IM Publications and Surface Spectra Ltd, Chichester. © 2003 IM Publications.)

and $L_3M_{23}M_{23}$, respectively. The KLL auger electrons can also result from more than one Auger process and produce several KLL lines including KL_1L_1 , KL_1L_{23} and $KL_{23}L_{23}$; however, $KL_{23}L_{23}$ is dominant as shown in Figure 7.12. Understanding such characteristics of Auger peaks is necessary in identifying chemical elements from Auger spectra.

7.4 Qualitative and Quantitative Analysis

7.4.1 Qualitative Analysis

XPS and AES are powerful tools for surface chemical analysis. They can identify chemical elements in the layer within several nanometers from the surface. Importantly, peak positions

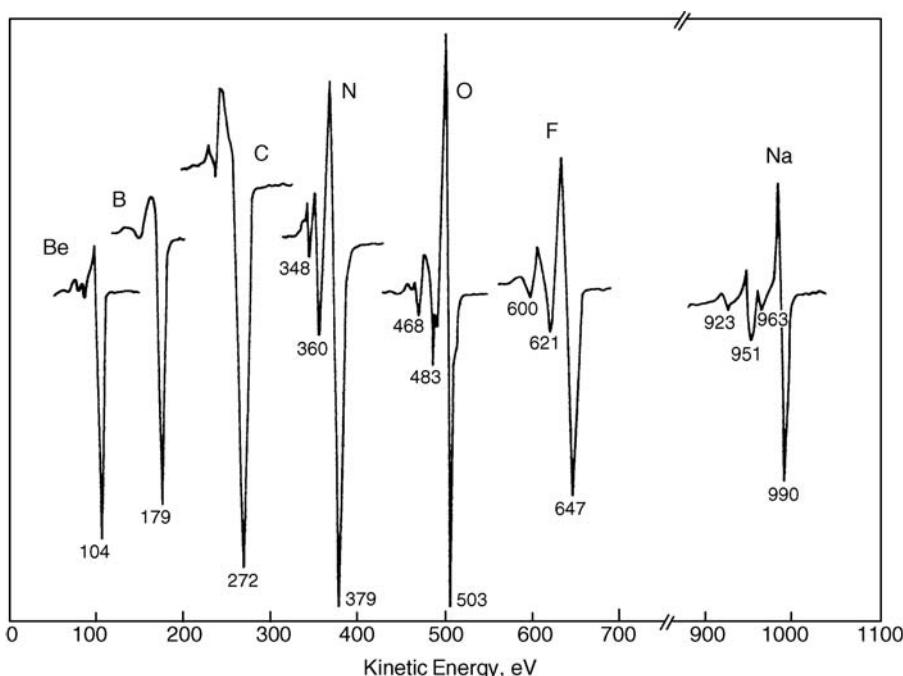


Figure 7.12 Principal Auger KLL peaks of light elements, Be, B, C, N, O, F and Na. $KL_{23}L_{23}$ is the most visible KLL peak for each element; for example, OKL_1L_1 (468 eV), OKL_1L_{23} (483 eV) and $OKL_{23}L_{23}$ (503 eV). (Reproduced with permission from D. Briggs and M.P. Seah, *Practical Surface Analysis*, Volume 1, John Wiley & Sons Ltd, Chichester. © 1990 John Wiley & Sons Ltd.)

of elements in XPS and AES spectra are sensitive to their chemical status; for example, carbon peak positions in CO_2 are different from those in saturated hydrocarbons. This phenomenon, called the *chemical shift* in XPS and AES spectra, provides extra information for chemical analysis. Also, XPS and AES instruments are able to generate depth profiles of the chemistry of the surface layer of bulk materials and reveal the spatial distributions of elements on the surface of bulk materials.

Peak identification in spectra is the primary task in qualitative analysis. For example, data in Figure 7.13 can be used to identify peaks in AES spectra. Table 7.1 lists the binding energies of elements in core levels. Such energies are the primary source for peak identification of XPS spectra. Peak identification is a non-trivial task in electron spectroscopy such as XPS. In addition to the spectrum features discussed in the previous section, other factors such as chemical shift and surface charge can complicate peak identification. The following section briefly reviews the important factors that, particularly for XPS, affect peak identification.

Peak Identification

Peak identification in AES spectra is straightforward. The peaks in an AES spectrum can be identified by comparing the experimental peaks with standard peaks found in reference books or computer databases. Peak identifications in XPS spectra, however, are more complicated because Auger peaks may be present. We use two different X-ray sources to distinguish the

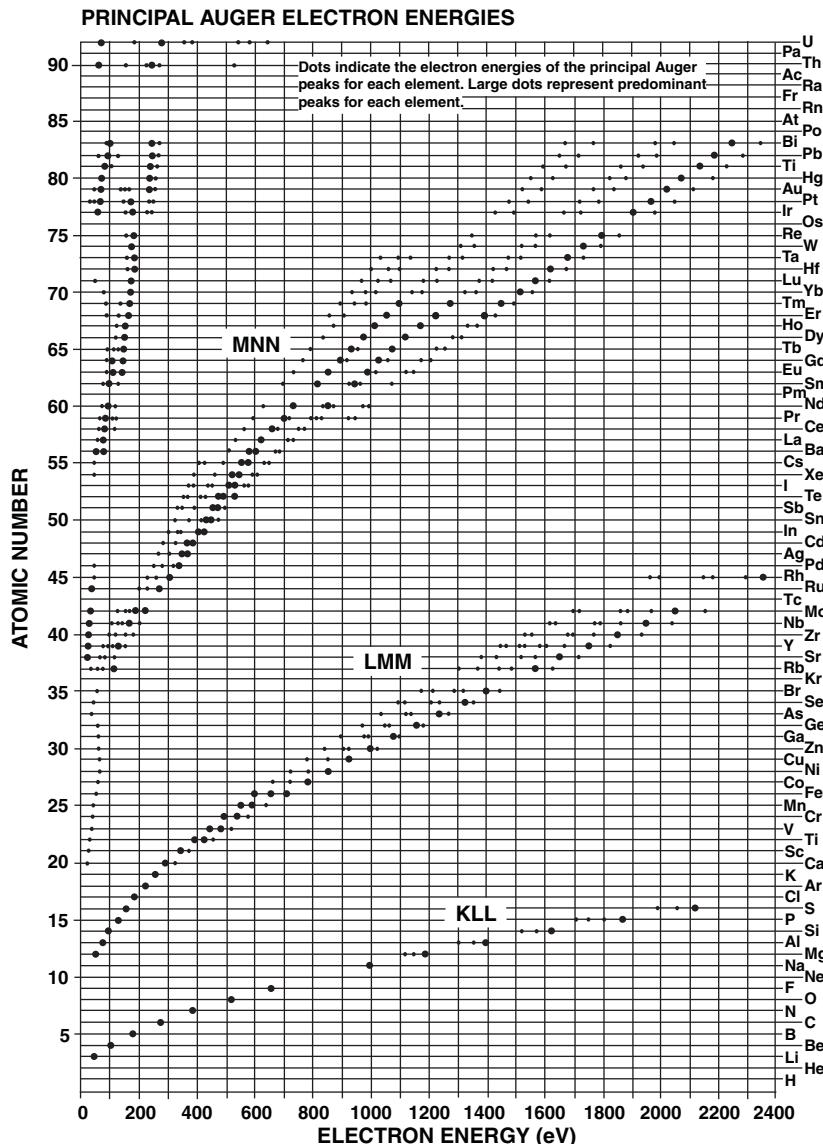


Figure 7.13 Chart of principal Auger electron energies of *KLL*, *LMM* and *MNN* lines (Reproduced with permission from J.F. Watts, *An Introduction to Surface Analysis by Electron Spectroscopy*, Oxford University Press, Oxford. © 1990 Royal Microscopy Society.)

Auger peaks from photoelectron peaks. An Auger peak represents the kinetic energy of an Auger electron which changes with energy of primary X-rays. Thus, an Auger peak will shift in apparent binding energy in an XPS spectrum when we change the X-ray source. For example, an Auger peak shifts by 233 eV in the XPS spectrum when we change the radiation from $MgK\alpha$ (1253.6 eV) to $AlK\alpha$ (1486.6 eV).

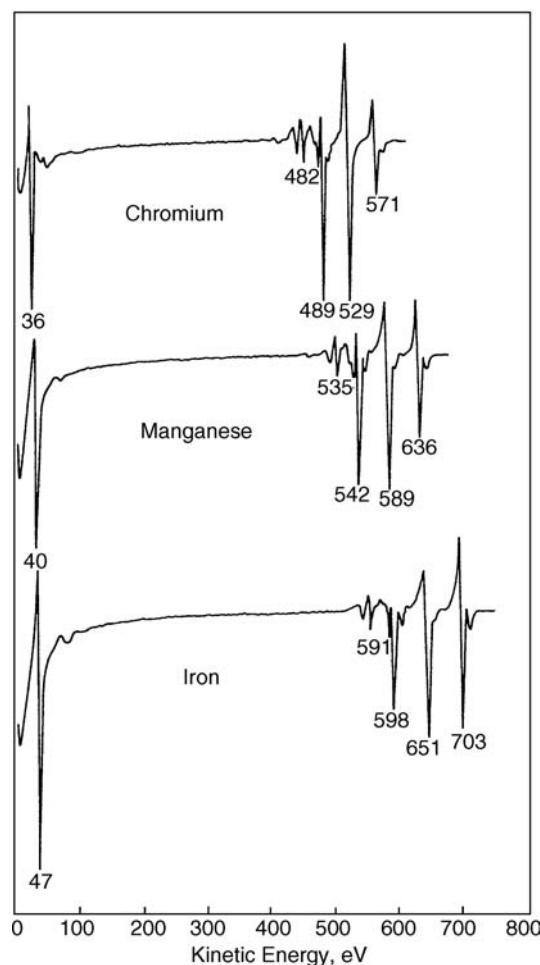


Figure 7.14 Triplet peaks of Auger spectra for Cr, Mn and Fe. The LMM triplets occur in transition metals. The low kinetic energy peaks are of $L_{2,3}VV$ where V represents the level of valence electrons. (Reproduced with permission from D. Briggs and M.P. Seah, *Practical Surface Analysis*, Volume 1, John Wiley & Sons Ltd, Chichester. © 1990 John Wiley & Sons Ltd.)

Peak positions in an XPS spectrum are likely to be affected by spectrometer conditions and the sample surface. Before XPS peak identification, we often need to calibrate the binding energy. Calibration is particularly important for samples with poor electrical conductivity. Calibration can be done with an internal standard that has a peak that shows little or no chemical shift, for example, elemental Si. The most common method is use of the C 1s peak at 285 eV from carbon adsorbed on the sample surface. The carbon from organic debris (as C–H or C–C) in air is found on all samples exposed to the environment. The peaks of core level binding energies as listed in Table 7.1 are sufficiently unique for element identification.

Table 7.1 XPS Peak Positions of Elements Irradiated With Al X-rays

Z	Element	$1s_{\frac{1}{2}}$ K	$2s_{\frac{1}{2}}$ L ₁	$2p_{\frac{1}{2}}$ L ₂	$2p_{\frac{3}{2}}$ L ₃	$3s_{\frac{1}{2}}$ M ₁	$3p_{\frac{1}{2}}$ M ₂	$3p_{\frac{3}{2}}$ M ₃	$3d_{\frac{3}{2}}$ M ₄	$3d_{\frac{5}{2}}$ M ₅
1	H	14								
2	He	25								
3	Li	55								
4	Be	111								
5	B	188			5					
6	C	284			6					
7	N	399			9					
8	O	532	24		7					
9	F	686	31		9					
10	Ne	867	45		18					
11	Na	1072	63		31		1			
12	Mg	1305	89		52		2			
13	Al	1560	118	74	73	1				
14	Si	1839	149	100	99	8				
15	P	2149	189	136	135	16	10			
16	S	2472	229	165	164	16	8			
17	Cl	2823	270	202	200	18	7			
18	Ar	3202	320	247	245	25	12			
19	K	3608	377	297	294	34	18			
20	Ca	4038	438	350	347	44	26		5	
21	Sc	4493	500	407	402	54	32		7	
22	Ti	4965	564	461	455	59	34		3	
23	V	5465	628	520	513	66	38		2	
24	Cr	5989	695	584	757	74	43		2	
25	Mn	6539	769	652	641	84	49		4	
26	Fe	7114	846	723	710	95	56		6	
27	Co	7709	926	794	779	101	60		3	
28	Ni	8333	1008	872	855	112	68		4	
29	Cu	8979	1096	951	932	120	74		2	
30	Zn	9659	1194	1044	1021	137	90		9	
31	Ga	10367	1299	1144	1117	160	106		20	
42	Mo	20,000	2866	2625	2520	505	410	393	208	205
46	Pd	24,350	3604	3330	3173	670	559	531	340	335
48	Ag	25,514	3806	3523	3351	718	602	571	373	367
73	Ta ^a	67,416	11,681	11,136	11,544	566 ^a	464 ^a	403 ^a	24 ^a	22 ^a
79	Au ^a	80,724	14,352	13,733	14,208	763 ^a	643 ^a	547 ^a	88 ^a	84 ^a

^a 4s, 4p and 4f levels indicated, respectively.

Z, atomic number.

(Data reproduced with permission from J.C. Vickerman, *Surface Analysis: The Principal Techniques*, John Wiley & Sons Ltd, Chichester. © 1997 John Wiley & Sons Ltd.)

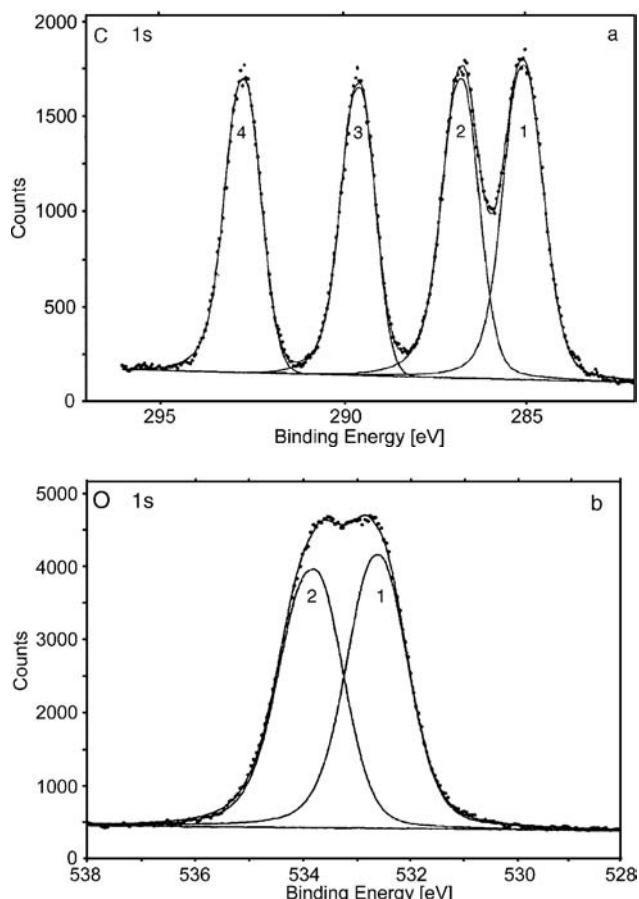


Figure 7.15 XPS spectrum of poly(vinyl trifluoroacetate): (a) C 1s; and (b) O 1s with monochromatic AlK α excitation. (Reproduced with permission from D. Briggs and J.T. Grant, *Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, IM Publications and Surface Spectra Ltd, Chichester. © 2003 IM Publications.)

Chemical Shifts

Chemical shifts of binding energy peaks for an element are caused by the surrounding chemical state of the element. Knowing possible chemical shifts is necessary to identify the peaks correctly. We can use the features of chemical shifts to identify not only elements but also chemical compounds. For example, Figure 7.15 shows the XPS spectrum of poly(vinyl trifluoroacetate) (PVTFA). The carbon atoms in different environments generate distinctive peaks as shown in Figure 7.15a. Similarly, oxygen atoms in two different environments also generate two peaks as shown in Figure 7.15b. The XPS spectra shown in Figure 7.15 with defined peak positions and relative peak intensities can also be used as a fingerprint to identify PVTFA. Small shifts in binding energy may cause peak overlap as shown in Figure 7.15b. Thus, we often need to carefully resolve the overlapped peaks with assistance of computer software.

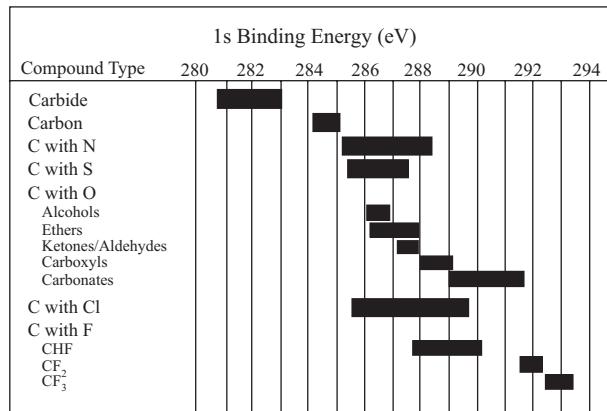


Figure 7.16 Chart of carbon chemical shift in XPS spectra. (Reproduced with permission from D. Briggs and J.T. Grant, *Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, IM Publications and Surface Spectra Ltd, Chichester. © 2003 IM Publications.)

The chemical shifts of the C 1s peak are summarized in Figure 7.16. The range of the C 1s shift extends over 12 eV. We can identify the different carbon bonds according to the amounts of chemical shifts. The binding energy difference between C–C and C=O is about 3 eV. This is extremely useful for identifying polymers by revealing various carbon bonds. Chemical shifts can also reveal the degree of oxidation in molecular solids. The larger the number of electrons transferred, the higher the chemical shift. Chemical shifts are less obvious in metallic and semiconductive elements because the shifts decrease with increasing atomic number. For example, the shift range of Si 2p is less than 6 eV.

Chemical shifts also occur in AES spectra, and the chemical shifts can be significantly larger than the shifts in XPS. For example, the shift between metallic and oxide Al peaks of $\text{Al}_{KL_2L_3}$ is more than 5 eV while the corresponding shift of Al 2p binding energy is only about 1 eV. In addition, Auger peak shape may be affected by the chemical state. For example, Figure 7.17 shows the position and shape of oxygen KLL Auger peaks in different types of oxides. Separation between the $\text{O}_{KL_1L_{23}}$ and $\text{O}_{KL_{23}L_{23}}$ peaks varies by up to 5 eV and the shape of the $\text{O}_{KL_{23}L_{23}}$ peak also varies between a single and a double type.

Problems with Insulating Materials

We may find that peak positions in the spectra of insulating materials do not match well with standards provided in handbooks. To understand this phenomenon in XPS, we have to know more about the basic equation for binding energy calculation (Equation 7.1). For a conducting sample that electrically contacts with a spectrometer, Φ in Equation 7.1 will be a work function of the spectrometer and its value is a constant (~ 4.5 eV). For an insulating sample, however, Φ is not certain and it is related to the surface potential of the sample. Photoemission from an insulating sample results in build-up of positive charge on the sample surface. This build-up will change the surface potential of the sample during XPS data collection. Also, the surface of an insulating sample tends to charge negatively in AES because incident electrons cannot be removed by electrical conduction. Change of surface potential can make the measured kinetic energy or binding energy tens of eV different from a standard value in a published table.

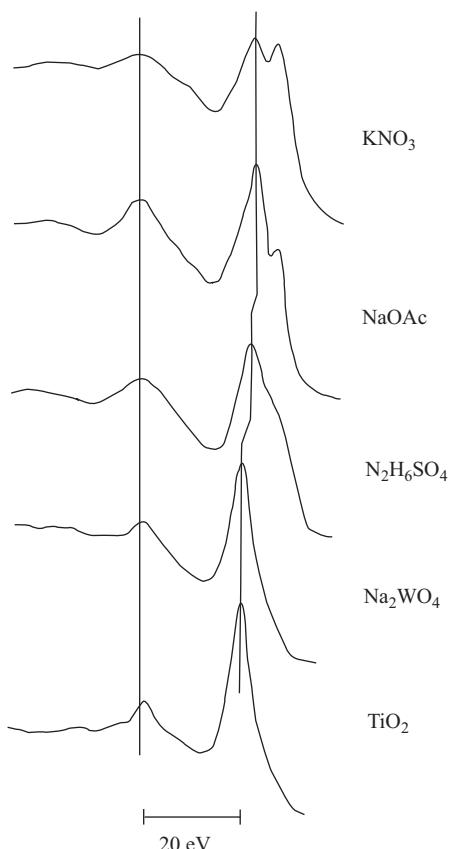


Figure 7.17 Comparison of positions and shapes of OKLL Auger peaks in several solid oxides. (Reproduced with permission from D. Briggs and M.P. Seah, *Practical Surface Analysis*, Volume 1, John Wiley & Sons Ltd, Chichester. © 1990 John Wiley & Sons Ltd.)

The method used to control the surface potential problem is *charge neutralization*. This is performed in order to provide a stable and uniform surface potential on the solid. In XPS, the sample surface is provided with flux of low-energy electrons using an *electron flood gun* as shown in Figure 7.18. The electron flux compensates for charge loss due to photoelectrons escaping from the surface. The flux helps maintain a stable surface potential. Figure 7.19 shows the differences in XPS spectra of a polymer sample with and without charge neutralization. In performing XPS examination of insulating samples, the surface charge problem is much more severe with monochromatic than with non-monochromatic X-ray radiation. The reason is that non-monochromatic X-rays illuminate the window of the X-ray gun as well as other surfaces inside the vacuum chamber, producing sufficient secondary electron flux to neutralize the positively charged sample surface. Hence, using non-monochromatic X-rays has much the same effect as using an electron flood gun.

For AES, this surface charge problem with insulating samples is more difficult to overcome because the electrons have to be removed from the insulating surface, instead of compensating for electron loss. AES does not work well with totally insulating materials. Commonly used

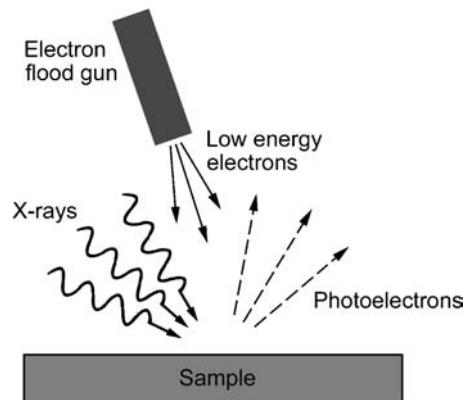


Figure 7.18 Compensating the electron loss due to photoelectron escape using low-energy electron flux from a flood gun.

methods in AES for reducing surface charge are reduction of primary electron beam energy and application of a conductive coating. The surface potential can be stabilized when the energy of the primary electron beam is reduced, for example, from 10 keV to about 2–5 keV. We also can coat the surface near the area (but not in the area) to be examined with a thin layer of metal such as silver or gold to achieve the same effect.

Composition Imaging

AES and XPS can also be used to analyze the distributions of chemical elements on surfaces by producing composition maps, similar to the EDS mapping discussed in Chapter 6. Composition maps are commonly obtained by scanning the sample surface with a focused source beam.

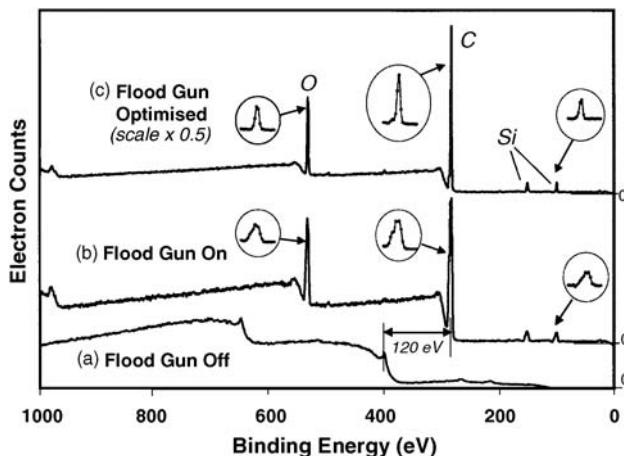


Figure 7.19 Comparison of XPS spectra of plastic tape with a monochromatic X-ray source: (a) spectrum with electron flood gun off; (b) spectrum with the flood gun on, but with incorrect electron compensation; and (c) spectrum with the flood gun on and proper electron compensation. (Reproduced with permission from D. Briggs and J.T. Grant, *Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, IM Publications and Surface Spectra Ltd, Chichester. © 2003 IM Publications.)

The spatial resolution of the composition maps mainly depend on the diameter of the beam. Auger imaging is a well developed technique because focusing the electron beam is readily achieved by the use of an electromagnetic lens. The diameter of the electron beam can be as small as about 10 nm when a field emission gun is used. Focusing the X-ray beam in XPS is difficult because the X-rays are electrically neutral and cannot be focused with an electromagnetic field. For modern XPS instruments, a monochromatic X-ray beam can be focused down to a diameter of about 10 μm using a special X-ray gun and a special crystal monochromator.

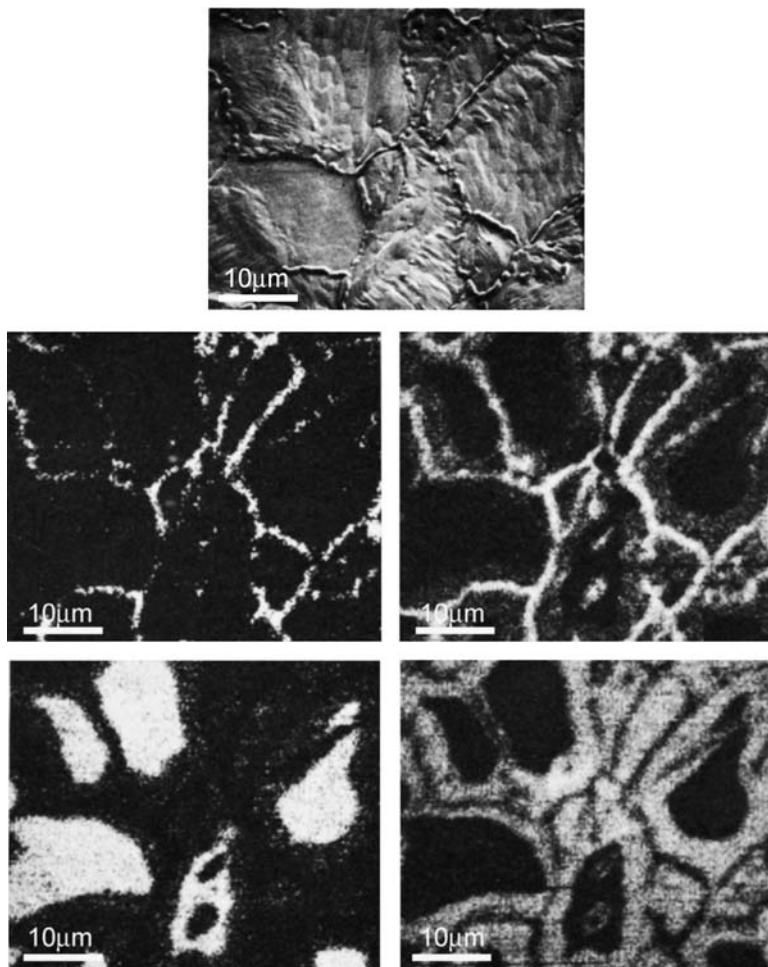


Figure 7.20 Comparison between images of gold-coated stainless steel: (a) a scanning electron microscope (SEM) secondary electron image; (b) iron Auger image; (c) oxygen Auger image; (d) gold Auger image; and (e) nickel Auger image. (Reproduced with permission from D. Briggs and M.P. Seah, *Practical Surface Analysis*, Volume 1, John Wiley & Sons Ltd, Chichester. © 1990 John Wiley & Sons Ltd.)

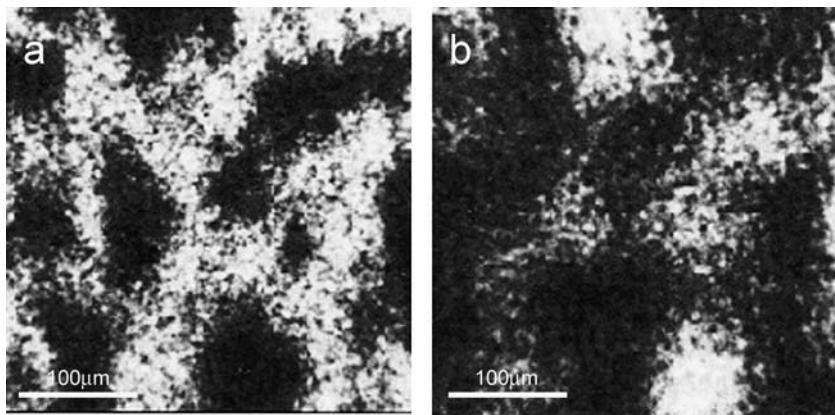


Figure 7.21 XPS images of a TiAlN thin film on a stainless steel substrate: (a) Ti 2p photoelectron image; and (b) Fe 2p photoelectron image. (Reproduced with permission from D. Briggs and J.T. Grant, *Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, IM Publications and Surface Spectra Ltd, Chichester. © 2003 IM Publications.)

Figure 7.20 shows examples of AES images obtained from an area of a gold-coated stainless steel lead frame. The grain structure is revealed by secondary electron imaging (Figure 7.20a). The Auger images disclose the presence of iron, oxygen, gold and nickel on the steel surface. Iron and oxygen on the surface are concentrated along grain boundaries. Nickel is diffused around grain boundaries. Figure 7.21 shows XPS images obtained from a surface area of an oxidized TiAlN film on a stainless steel substrate. The oxidized film contains iron that has migrated from the substrate. The spatial resolution of XPS images is obviously poorer than that of AES images as the comparison of Figures 7.20 and 7.21 shows.

7.4.2 Quantitative Analysis

Peaks and Sensitivity Factors

Quantitative elemental analysis in electron spectroscopy is similar to that in X-ray spectroscopy. Analysis quantifies the concentrations of chemical elements on a sample surface from the peak intensities of the spectra. In theory, the quantitative relationship between the intensities of electron signals and atomic fractions of elements can be calculated. In practice, for quantification in both XPS and AES, most parameters for calculations are not available. Thus, the following empirical equation is commonly used.

$$X_i = \frac{I_i/S_i}{\sum I_j/S_j} \quad (7.3)$$

An atomic fraction of a surface element (X_i) can be calculated from known its peak intensity (I_i) and *sensitivity factor* (S_i). The sensitivity factor for each element is experimentally determined and varies with conditions of the instrument and sample surface. It is desirable to measure the sensitivity factor for a given spectrometer and sample type. Common practice is to use the sensitivity factors from published handbooks with certain corrections according to the

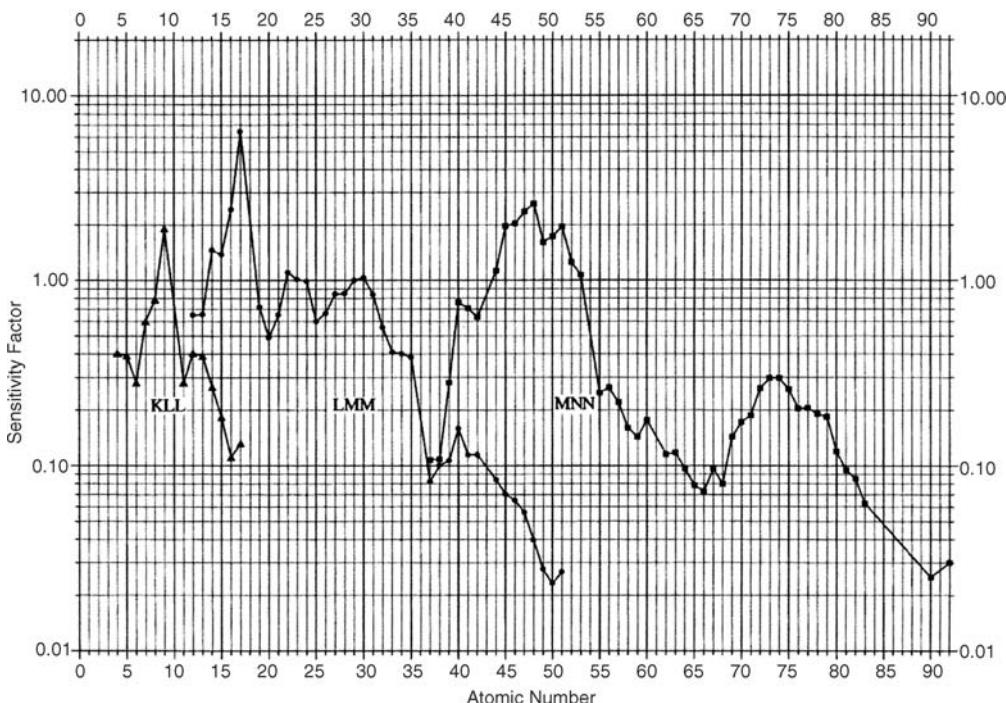


Figure 7.22 AES sensitivity factors normalized to the Cu_{LMM} line for 10 keV electron radiation. Sensitivity factors are calculated from the peak-to-peak heights in differential spectra. (Reproduced with permission from D. Briggs and J.T. Grant, *Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, IM Publications and Surface Spectra Ltd, Chichester. © 2003 IM Publications.)

instrument characteristics, because it is not feasible to compile a set of in-house sensitivity factors. For example, Figure 7.22 shows the Auger sensitivity factors of elements published in a handbook of Auger electron spectroscopy. The sensitivity factors are values relative to that of the Cu_{LMM} Auger line, which is set as unity. Also, note the sensitivity factors in Figure 7.22 are only useful for a primary electron energy of 10 keV.

The quantitative calculations based on Equation 7.3 include assumptions and approximations. The calculations can readily lead to erroneous results unless caution is exercised. The intensities of XPS peaks should be calculated from the peak areas after subtracting the background. The intensity of AES peaks should be the peak-to-peak height in differential spectra. Unfortunately, the peak-to-peak heights may vary with instrument resolution and the differentiation (modulation) process. This problem may be resolved by using the AES direct spectrum. However, the relative intensities of peaks to background are low. Small errors in background subtraction can generate large variation in peak height measurement.

The published sensitivity factors are obtained with specific instruments and standard samples of pure elements. Sensitivity factors are subject to change with different samples and developments in instruments. For example, chemical contamination on a sample surface affects the accuracy of using Equation 7.3. For an AES spectrum, the matrix composition affects the efficiency of Auger emissions because the backscattered electrons in the matrix

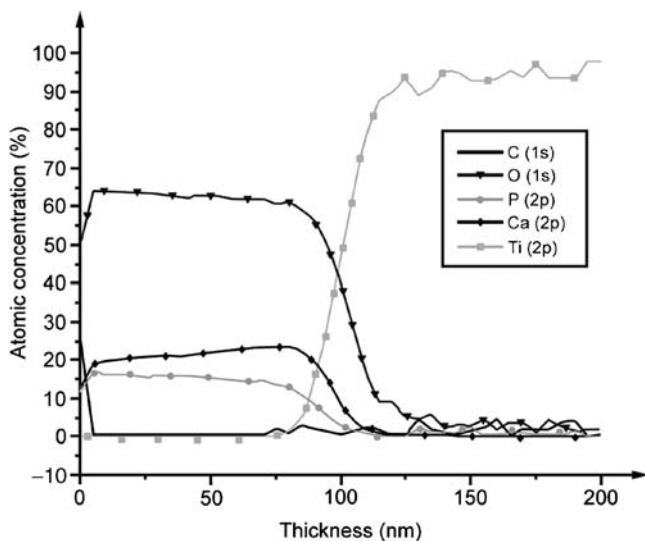


Figure 7.23 XPS Depth profile of a titanium surface coated with calcium phosphate which contains Ca, P, and O. C in the profile is from surface contamination.

also excite Auger electron emission. Thus, we expect that the intensity of element *i* in a matrix with multiple elements differs from that of the same amount of *i* element in its pure solid state.

7.4.3 Composition Depth Profiling

We often need to characterize the changes in composition with distance from a surface plane. XPS and AES can analyze such composition changes. The most commonly used method to obtain a depth profile of composition is sputter depth profiling using an ion gun. The ion gun provides a flux of positively charged argon ions with current density of $1\text{--}50 \mu\text{A mm}^{-2}$ and energy of $0.5\text{--}5 \text{ keV}$. The ion beam bombards a sample, causing it to eject atoms from the surface. A rastering ion beam directed at a surface produces a crater. We can obtain a profile of the near-surface compositions by examining the AES and XPS spectra of the crater at its different depths. For example, Figure 7.23 shows an XPS depth profile of a titanium surface coated with calcium phosphate. The depth scale in the profile is not from a direct thickness measurement; instead, it is calculated from the etching rate of the sample caused by ion beam bombardment. The etching rates depend on the *sputter yield*, which is the number of atoms ejected by an ion. Sputter yield data, unfortunately, are not commonly available and should be experimentally determined. It is strongly suggested to calibrate the depth scale when the etching rate is obtained from published data of sputter yields.

Quantitative analysis of a depth profile, in addition to quantifying concentrations, needs to measure depth accurately. There are several factors we should be aware of for resolution of depth measurement.

First, the etching rate for each type of atom is not constant even though the ion beam energy and ion density are exactly the same. The consequence of this phenomenon is that

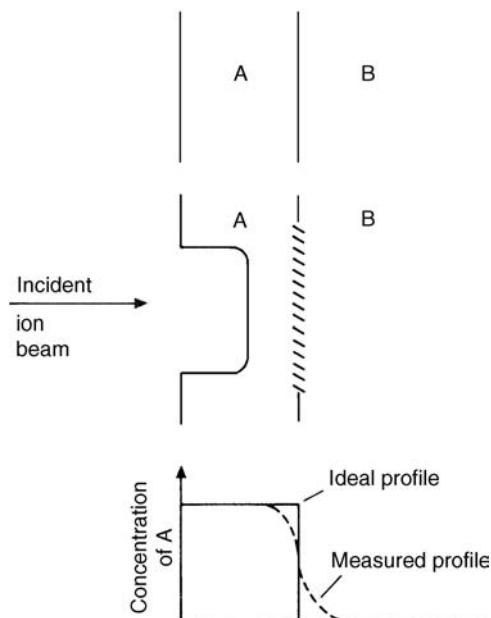


Figure 7.24 Smearing of the interface in a depth profile due to ion bombardment. (Reproduced with permission from G.C. Smith, *Quantitative Surface Analysis for Materials Science*, Institute of Metals, London. © 1991 Institute of Metals.)

preferential etching occurs in samples containing multiple elements. In samples comprising several elements which are not uniformly distributed, certain areas will be etched faster than others because of preferential etching. Thus, a rough crater surface is likely generated with increasing depth of etching. The depth scale of the profile is based on the average etching rates. Thus, the measured concentration of elements with higher etching rates (higher sputter yield) will be higher than the real concentration.

Second, the incident ions can smear the interface in the depth profiles. Figure 7.24 schematically illustrates interface smearing during depth profiling. When an ion beam bombards a sample of thin film that has a smooth interface with a substrate, the depth profile of element A in the thin film often exhibits a diffuse rather than sharp interface profile as should be the real case, as shown on the lower part of Figure 7.24. This is also the case shown in Figure 7.23 in which the smooth interface between the coating and titanium substrate becomes diffuse in the profile. The main reason for interface smearing is that the incident ions cause some coating atoms and/or substrate atoms to cross the interface. The atom displacements caused by ion bombardment is not difficult to understand, considering the energy brought into the sample by the ions.

References

- [1] Watts, J.F. (1990) *An Introduction to Surface Analysis by Electron Spectroscopy*, Oxford University Press, Oxford.
- [2] Smith, G.C. (1991) *Quantitative Surface Analysis for Materials Science*, Institute of Metals, London.
- [3] Briggs, D. and Seah, M.P. (1990) *Practical Surface Analysis*, Volume 1, John Wiley & Sons, Chichester.

- [4] Briggs, D. and Grant, J.T. (2003) *Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, IM Publications and Surface Spectra Ltd, Chichester.
- [5] Brandon, D. and Kaplan, W.D. (1999) *Microstructural Characterization of Materials*, John Wiley & Sons, Chichester.
- [6] Eberhart, J.P. (1991) *Structural and Chemical Analysis of Materials*, John Wiley & Sons, Chichester.
- [7] Prutton, M. and El Gomati, M.M. (2006) *Scanning Auger Electron Microscopy*, John Wiley & Sons, Chichester.
- [8] Vickerman, J.C. (1997) *Surface Analysis: The Principal Techniques*, John Wiley & Sons, Chichester.
- [9] Beamson, G. and Griggs, D., (1992) *High Resolution XPS of Organic Polymers*, John Wiley & Sons, Chichester.

Questions

- 7.1 Explain why H and He cannot be detected by Auger electron spectroscopy.
- 7.2 How many types of KLL Auger electrons are there? Rank them according to their kinetic energies.
- 7.3 Why are the electron energies of O_{KLL} and C_{KLL} in Figure 7.3 different from those in Figure 7.2? Estimate the work function of an aluminum oxide surface from Figures 7.2 and 7.3.
- 7.4 Why are differential spectra preferred in AES analysis? Why are direct spectra often used for quantitative analysis?
- 7.5 Why are $MgK\alpha$ and $AlK\alpha$ used as X-ray sources for XPS? Do they excite 1s photoelectron emission of metals such as Ti and Fe? If not, how do we detect such elements?
- 7.6 What are the differences between electron energy analyzers for AES and XPS? Why?
- 7.7 How can you tell Auger peaks from photoelectron peaks in an XPS spectrum?
- 7.8 The chemical shifts of binding energy are related to electronegativity (charge on an atom). Using Figure 7.25, find out the chemical shifts of C 1s for a polymer by identifying the binding energy peaks for different carbon atoms in its chemical structure.

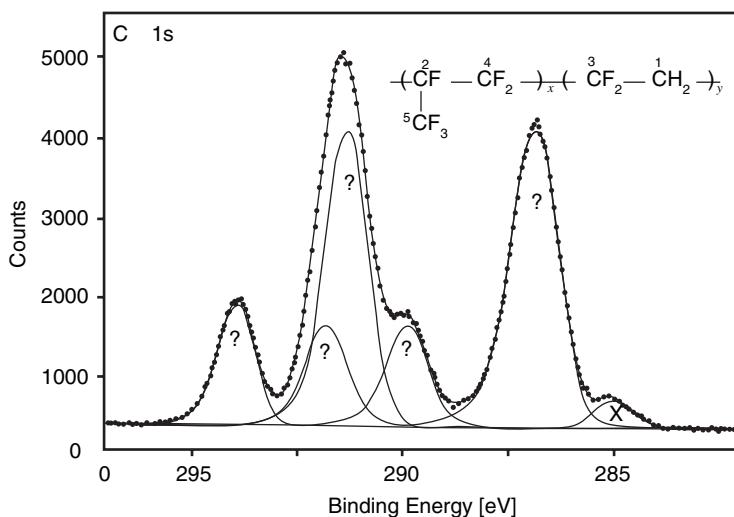


Figure 7.25 An XPS spectrum showing carbon (1s) binding energies of a polymer sample. (Reproduced with permission from G. Beamson and D. Griggs, *High Resolution XPS of Organic Polymers*, John Wiley & Sons Ltd, Chichester. © 1992 John Wiley & Sons Ltd.)

- 7.9 Composition maps may be obtained by EDS, AES and XPS techniques. Please indicate the right technique(s) for each of the following types of sample: thin film of oxides, polycrystalline metals, and nanometer-thick coatings of metal on a polymer.
- 7.10 Compare AES and XPS with respect to the following aspects: energy source, signals from the sample, detectable elements, spatial resolution and applicability to organic samples.

8

Secondary Ion Mass Spectrometry for Surface Analysis

Secondary ion mass spectrometry (SIMS) is a relatively new technique for surface chemical analysis compared with Auger electron spectroscopy (AES) and X-ray photoelectron spectroscopy (XPS). SIMS examines the mass of ions, instead of energy of electrons, escaped from a solid surface to obtain information on surface chemistry. The term ‘secondary ion’ is used to distinguish ‘primary ion’ that is the energy source for knocking out ions from a solid surface. The advantages of SIMS over electron spectroscopy are:

- Detection of all the chemical elements in the periodic table, including hydrogen which cannot be detected by the AES or XPS;
- Detection of elements in concentrations as low as 10^{-6} , while AES or XPS detection limits are concentration levels of 0.1 atom%;
- Limitation of the detection to the top one or two atomic layers of a solid surface ($<1\text{ nm}$); and
- Distinguish between different isotopes of elements.

Such features make SIMS a powerful technique for surface analysis. However, SIMS as a surface analysis technique has not yet reached a mature stage because it is still under development in both theoretical and experimental aspects. This lack of maturity is attributed to the complicated nature of secondary ion yield from a solid surface. Complexity of ion yield means that SIMS is less likely to be used for quantitative analysis because the intensity of secondary ions is not a simple function of chemical concentrations at the solid surface. SIMS can be either destructive or non-destructive to the surface being analyzed. The destructive type is called *dynamic SIMS*; it is particularly useful for depth profiling of chemistry. The nondestructive type is called *static SIMS*. Both types of SIMS instruments are widely used for surface chemical examination.

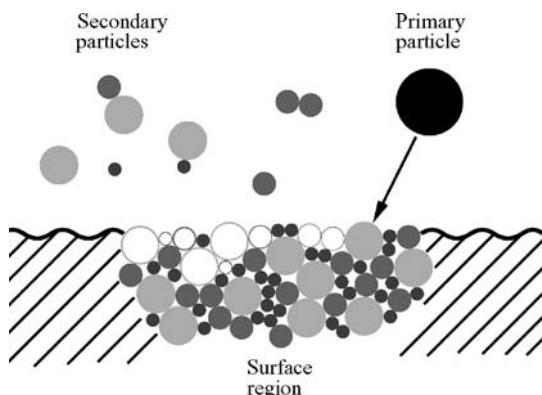


Figure 8.1 Secondary particle generation by an energetic primary particle.

8.1 Basic Principles

SIMS uses energized primary particles, usually ions such as Ar^+ , Ga^+ and Cs^+ , to bombard a solid surface in order to induce sputtering of secondary particles from an area as illustrated in Figure 8.1. The interactions between primary ions and the solid are rather complicated. First, the secondary particles include electrons, neutral species of atoms or molecules, and ions. The majority of the secondary particles are neutral and not useful in SIMS. Only the secondary ions generated by the bombarding process carry chemical information. Second, the interactions are often more than a simply one-to-one knock-out of a surface ion by a primary ion. Commonly, the primary ions induce a series of collisions (*collision cascade*) in a solid because the energy of a primary ion is transferred by collisions between atoms in a solid before secondary ions on the surface are emitted. The generation of secondary ions involves sputtering and ionization, which are introduced in the following section.

8.1.1 Secondary Ion Generation

Emission of secondary particles can result from *collision sputtering* or from other processes such as *thermal sputtering*, which are illustrated in Figure 8.2. Collision sputtering includes *direct collision sputtering* and *slow collision sputtering*. The former can be considered as a direct impact between a primary ion and a surface atom as illustrated in Figure 8.2a. Direct collision sputtering is extremely fast, and occurs in the range of 10^{-15} to 10^{-14} seconds after the primary ion strikes a surface. Slow collision sputtering, illustrated in Figure 8.2b, represents the case that a primary ion never has chance to collide with a surface atom; instead, atoms in the solid transfer the impact energy to surface atoms after a series of collisions. The time scale of slow collision sputtering is in the range of 10^{-14} to 10^{-12} seconds. Thermal sputtering occurs when the density of primary ions is high. It is a process of transient vaporization of the surface area impacted by primary ions as illustrated in Figure 8.2c. Thermal sputtering is slower than collision sputtering, and occurs in the range of 10^{-13} to 10^{-10} seconds. Thermal sputtering not only includes thermal but also electronic excitation when the solid is poor in electronic conduction.

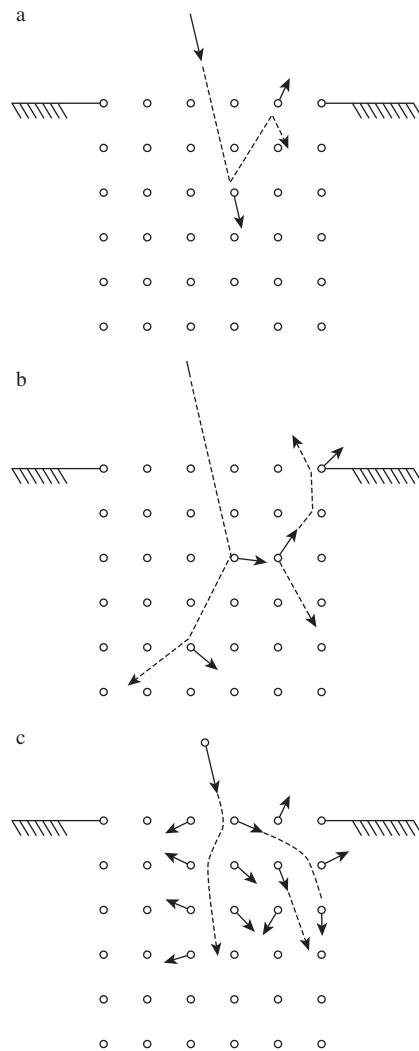


Figure 8.2 The sputtering process in SIMS: (a) direct collision sputtering; (b) collision cascade; and (c) thermal sputtering. (Reproduced with kind permission of Springer Science and Business Media from R. Behrisch, Ed., *Sputtering by Particle Bombardment I*, Springer-Verlag GmbH, Berlin. © 1981 Springer Science.)

Only a small portion of secondary particles ($\sim 1\%$ of total secondary particles) are ionized and become the secondary ions that are analyzed in SIMS. A sputtered particle faces competition between ionization and neutralization processes when it escapes a sample surface. *Ionization probability* represents the chance of a sputtered particle being an ion. The ionization probability is strongly affected by the electronic properties of the sample matrix. Ionization directly affects the signal intensity of secondary ions as shown in the basic equation of secondary ion yield.

$$I_m = I_p Y_m \alpha^+ \theta_m \eta \quad (8.1)$$

The secondary ion current (I_m) of chemical species m is collectively determined by the primary ion flux (I_p), the sputter yield (Y_m) and the ionization probability. α^+ represents the probability for positive ions, θ_m is the fractional concentration of species m in the surface layer and η is the *transmission* of the detection system. The transmission is defined as the ratio of the ions detected to ions emitted, and it varies from 0 to 1 depending on the analyzer. The sensitivity of SIMS to element m is controlled by the factors Y_m , α^+ and η . Y_m generally increases with beam energy and with the primary ion mass. It also varies with the types of atomic bonding in target samples.

Under the same experimental conditions, the yield of elemental secondary ions can vary by several orders of magnitude across the periodic table as shown in Figure 8.3. Also, the secondary ion yield for a particular element dramatically varies with chemical state, as shown in Table 8.1. For example, the positive Al ion yield in its oxide state is 100 times higher than in its pure state. Such effects are mainly attributed to the ionization process because the sputter yield only varies by a factor of about 3–5 across the periodic table at a given bombardment energy.

There is no simple method to determine the ionization probability because the understanding of the ionization process is still limited. Ionization mechanisms differ among materials. For metals, a departing atom may transfer its electrons to the surface, while for ionic solids, secondary ions may form by breaking atomic bonds. It is also likely that rapid electronic transition in the surface region will neutralize any ions before they can escape from the surface. Thus, ionization may occur by fragmentation of secondary particles within a certain distance from the surface. The fragmentation of a neutral particle may result from its high vibrational motion because its internal energy increases in collision sputtering.

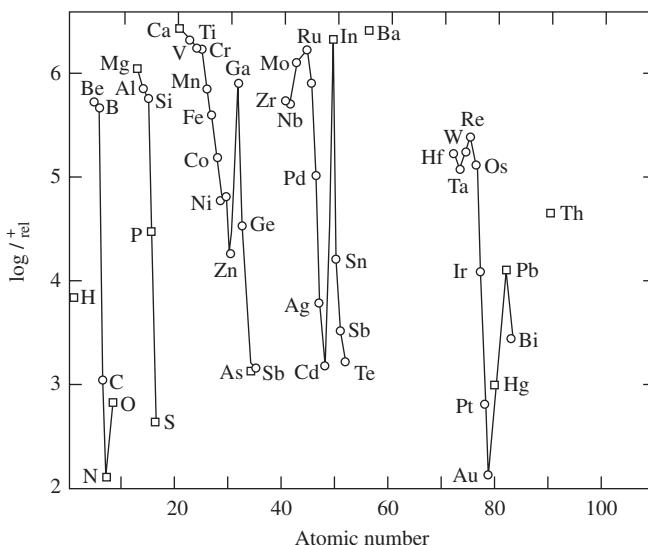


Figure 8.3 Variation of positive ion yield among chemical elements under bombardment of 13.5 keV O^{-1} . (o) indicates a pure element, (□) indicates a compound. (Reproduced with permission from H.A. Storms, K.F. Brown, and J.D. Stein, “Analytical Chemistry,” 49, no. 13, 2025, figure 3. © 1977 American Chemical Society.)

Table 8.1 Comparison of Secondary Ion Yield in Pure Metals and Their Oxide

Metal	Clean elements M^+ yield	Oxide M^+ yield
Al	0.007	0.7
Ti	0.0013	0.4
Cr	0.0012	1.2
Fe	0.0015	0.35
Cu	0.0003	0.007
Mo	0.00065	0.4
Ba	0.0002	0.03
W	0.00009	0.035
Si	0.0084	0.58
Ge	0.0044	0.02

(Data reproduced with permission from D. Briggs, A. Brown, and J.C. Vickerman, *Handbook of Static Secondary Ion Mass Spectrometry*, John Wiley & Sons Ltd, Chichester. © 1989 John Wiley & Sons Ltd.)

The important features of the sputter ionization process can be summarized as follows. First, the point at which the secondary particles escape from the surface is not the point of the initial impact by the primary ion. Instead, there is a surface damage zone that includes the points of primary particle impact and secondary particle escape. Second, the cascade collision results in generation of secondary ions with much lower energy than that of primary ions. Third, there is significant variation in secondary ion yield with chemical elements and chemical states of the surface, which makes quantitative analysis difficult.

8.1.2 Dynamic and Static SIMS

Sputtering is, fundamentally, a process of damaging a surface. The energetic primary ions remove atoms, atom clusters, molecules or molecular fragments from the surface region of a sample. Thus, chemical structures of the surface may be destroyed during SIMS examination. This is especially true when the high flux of primary ions bombards a surface. SIMS with high flux of primary ion bombardment is referred to as *dynamic SIMS*. Dynamic SIMS is able to remove many layers of atoms in the surface region and provide elemental distributions in a depth profile, similar to AES and XPS.

In order to analyze surface chemical structures, *static SIMS* has been developed in which a low flux of primary ions is used in order to avoid the possibility that any surface area is bombarded by primary ions twice. To do so, the dose of primary ions never exceeds 10^{13} ions cm^{-2} of surface area. We estimate the density of surface atoms as about 10^{15} atoms cm^{-2} by assuming the area of each atom as about 0.1 nm^2 . The low dose of primary ions used in static SIMS means that less than 10% of the surface atoms or molecules are bombarded. Figure 8.4 schematically illustrates static SIMS. A primary ion bombardment only generates damage to a zone with a cross-section (σ) less than 10 nm^2 ($= 10^{-13} \text{ cm}^2$) on the surface. A damaged surface area is defined as a region that includes the impact point of a primary ion and escape point of a secondary ion. Thus, static SIMS is considered as a non-destructive method for surface analysis. For static analysis, the integrity of the surface should be maintained within the timescale of the SIMS process.

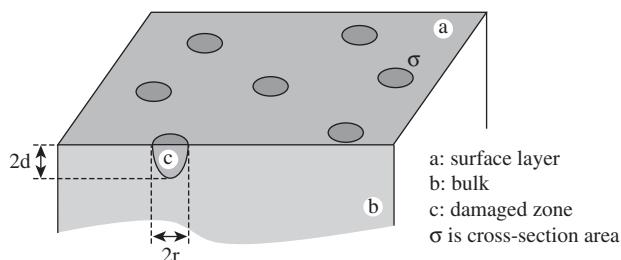


Figure 8.4 Static SIMS. Primary ion bombardment results in surface damage and emission of ions and neutral atoms. The damaged area is quantified by the damage cross-section (σ). In static SIMS the total primary ion dose should be limited so that only small portion of the surface is damaged. a, surface layer; b, bulk; c, damaged zone; d, depth of damage zone; r, radius of damage zone.

We can estimate the time scale in which the whole surface layer is affected by the primary ions. The *lifetime of a surface* may be simply estimated from the primary ion flux (I_p) and damage cross-section (σ) generated by each impact. I_p is commonly measured in A cm^{-2} ($1 \text{ A} = 6.2 \times 10^{18}$ charged particles per second). Assume that each primary ion generates $\sigma = 10^{-13} \text{ cm}^2$. Then, 10^{13} primary ions cm^{-2} will affect the whole surface area of 1 cm^2 . It means that the lifetime of a surface with the flux density $I_p = 1 \mu\text{A cm}^{-2}$ ($= 6.2 \times 10^{12} \text{ ions cm}^{-2}$) is less than 1 second. Apparently, $1 \mu\text{A cm}^{-2}$ of flux density for primary ions is too high for static SIMS. Since it is commonly accepted for the static SIMS condition to limit the total amount of primary ions up to $10^{13} \text{ ions cm}^{-2}$, for a 10-min duration of static SIMS examination a primary flux density of about 2.7 nA cm^{-2} is required to preserve the chemical structure of the surface top layer where the secondary ions are emitted. This flux is extremely low compared with that of dynamic SIMS, which requires a flux density of greater than $1 \mu\text{A cm}^{-2}$ to ensure a reasonable erosion rate of surface for depth profiling.

8.2 Instrumentation

SIMS requires an ultra-high vacuum environment, similar to AES and XPS. The ultra-high vacuum environment ensures that trajectories of ions remain undisturbed during SIMS surface analysis. The SIMS vacuum chamber and pumping system is not much different from that for AES and XPS. Figure 8.5 illustrates common SIMS structure in a vacuum chamber in which there are two main components: a primary ion system and a mass analyzer system. The primary ion system includes a primary ion source, a *Wien filter* and ion beam deflector. The mass analysis system includes a secondary ion extractor filter, mass analyzer and ion detector.

8.2.1 Primary Ion System

A system generating a primary ion beam consists of three main parts: ion source, ion filter and deflector as schematically illustrated in Figure 8.5. The ion source produces primary ions with a certain kinetic energy. The ion filter purifies the primary ions and rejects unwanted ions in the primary ion beam. The purified ions are condensed to a focused beam using an electromagnetic lens. The deflector makes the focused ion beam raster on the sample surface in two orthogonal directions; its function is similar to the electron beam deflector in the scanning electron microscope (SEM).

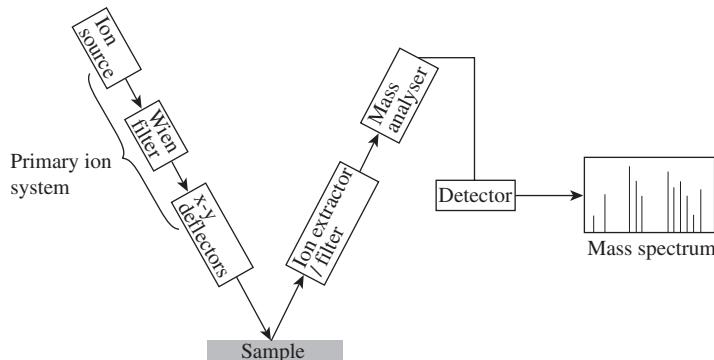


Figure 8.5 SIMS instrumentation.

Ion Sources

The commonly used primary ions include argon (Ar^+), xenon (Xe^+), oxygen (O_2^+), gallium (Ga^+) and cesium (Cs^+) ions. Heavier metal ions such as bismuth (Bi^+) are also available in modern static SIMS instruments. The ions of elements normally occurring in gaseous phases, such as oxygen and argon, are produced by *electron bombardment sources* or *plasma ion sources*. Electron bombardment sources use a circular filament cathode surrounding a cylindrical grid anode as illustrated in Figure 8.6. The gas to be ionized is injected into the open space of the grid, and the gas molecules are bombarded by the electrons emitted from the

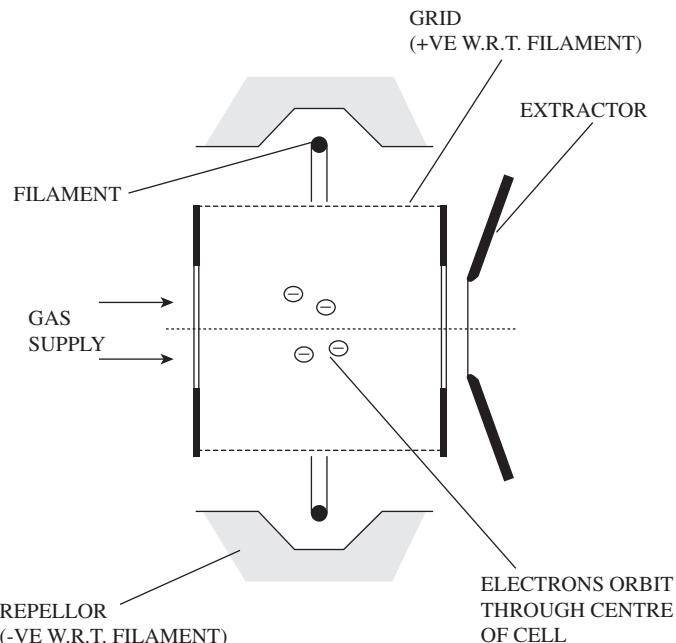


Figure 8.6 Electron bombardment source. (Reproduced with permission from J.C. Vickerman and D. Briggs, *ToF-SIMS Surface Analysis by Mass Spectrometry*, IM Publications and SurfaceSpectra, Chichester and Manchester. © 2001 IM Publications.)

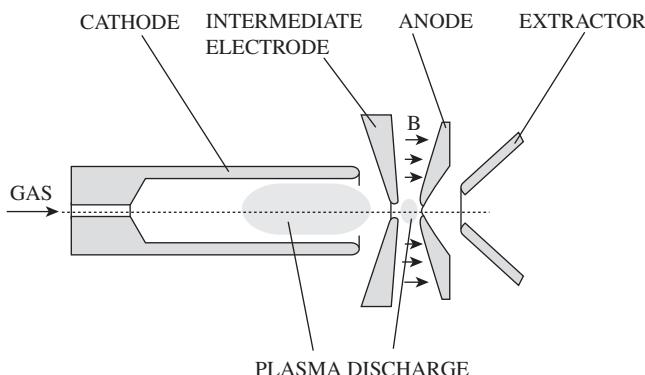


Figure 8.7 Duoplasmatron ion source. The magnetic field marked **B** intensifies the plasma by confining electrons close to the axis. (Reproduced with permission from J.C. Vickerman and D. Briggs, *ToF-SIMS Surface Analysis by Mass Spectrometry*, IM Publications and SurfaceSpectra, Chichester and Manchester. © 2001 IM Publications.)

cathode. The electrons travel in orbit inside the grid in order to increase their chances of striking gas molecules. An extraction field inducted in the grid will draw the gas ion beam through an opening in the center of the extractor. The electron bombardment sources provide moderate brightness of primary ions ($\sim 10^5 \text{ A m}^{-2}$ per solid angle).

Plasma ion sources can produce much higher ion beam brightness than the electron bombardment source, up to 10^7 A m^{-2} per solid angle. Plasma sources use a source gas at high pressure. The *duoplasmatron source* is a commonly used plasma source, illustrated in Figure 8.7. It has a hollow cathode containing a high pressure source gas, an intermediate electrode and an anode mounted along the axis of the cathode hollow. The current between cathode and anode generate arcs between the cathode and intermediate electrode, as well as between the intermediate electrode and anode. The plasma beam will flow through a small aperture in the center of the anode. A beam with high brightness is produced and is especially suited for dynamic SIMS, which needs a high yield of secondary particles and high erosion rates. The O_2^+ ion beam is commonly used because it can increase positive secondary ion yield.

Liquid metal ion sources are used for producing a Ga^+ ion beam. The working principles of a liquid metal ion source are similar to a field emission gun for electrons. A strong electrical field strips off electrons from the source liquid metal to generate a metal ion flow from the fine tip of a needle. Gallium is a metal with a low melting temperature of 29.8°C , and hence it is suitable for this purpose. In a liquid metal ion source, liquid gallium is coated on a metal reservoir and a needle is welded to a heater filament as illustrated in Figure 8.8. The liquid metal can flow over the needle tip and extend to form a cone (the *Taylor cone*) under a high electric field between the tip and extractor. During operation, a dynamic equilibrium between extraction force and surface tension is reached in the liquid metal. A constant ion current will flow out of the tip area as small as 10 nm and generate an ion beam with high current density. This type of source exhibits very high brightness ($\sim 10^{10} \text{ A m}^{-2}$ per steradian solid angle) and high spatial resolution. Liquid metal ion sources are also used to produce other ion beams of heavy metals with relatively low melting temperature, such as bismuth (271.3°C).

Surface ionization sources are commonly used for producing a Cs^+ ion beam. Surface ionization is produced in a vacuum by heating a source metal which is adsorbed onto the

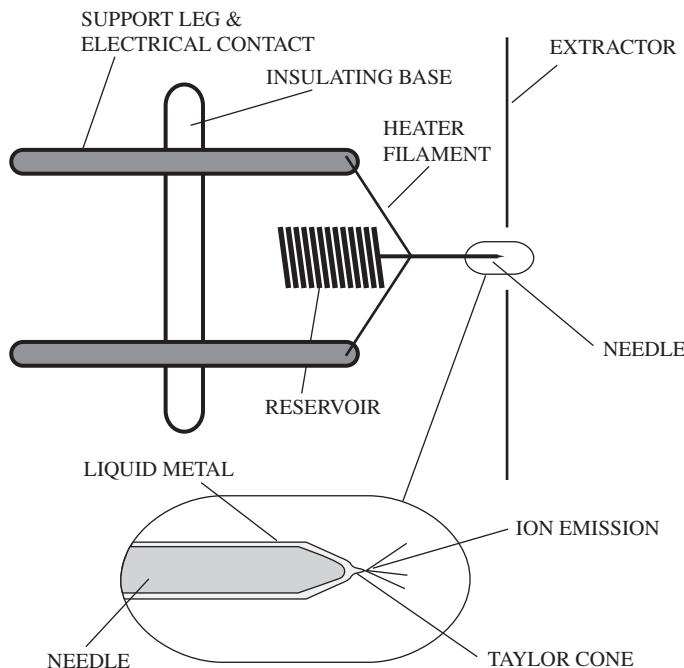


Figure 8.8 Liquid metal ion source. The primary ion beam is extracted from the Taylor cone of liquid metal. (Reproduced with permission from J.C. Vickerman and D. Briggs, *ToF-SIMS Surface Analysis by Mass Spectrometry*, IM Publications and SurfaceSpectra, Chichester and Manchester. © 2001 IM Publications.)

surface of a metal with a high work function. The ionization of source metal atoms is realized by transferring electrons from the adsorbed metal to the substrate of the metal because the work function of the substrate is higher than the ionization potential. Then, positive Cs^+ ions can be extracted from the metal substrate to form an ion beam. The Cs^+ ion beam is particularly useful for enhancing negative secondary ions. The Cs^+ ion beam exhibits uniform kinetic energy and reasonable spatial resolution.

Wien Filter

A primary ion beam for SIMS analysis should ideally contain ions with identical mass. Under the same electric field for accelerating the primary ions, the time for two ions to reach the sample will be different if their mass is different. Time accuracy is extremely important in the most widely used form of static SIMS, *time-of-flight SIMS*, in which primary ions are emitted in a pulse length of about 10^{-9} seconds. The primary ion beam likely contains ions with different mass such as isotopic ions and impurity ions. Thus, the primary ion beam should be purified before bombarding the sample using mass filters. The Wien filter is a simple device to filter ions according to their mass. Figure 8.9 illustrates the working principle of the Wien filter. The magnetic and electric fields are mutually orthogonal. Also, these two fields generate forces lateral to an ion current, which is perpendicular to both fields. The lateral forces generated from magnetic and electric fields will cancel out when the velocity of ions reaches a specific value (v_s). The velocity of ions is determined by their mass, because all the primary ions from

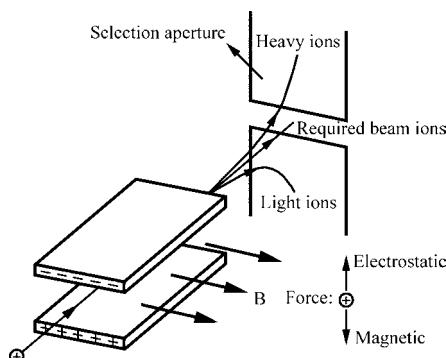


Figure 8.9 The Wien filter for ion mass selection.

a source have same kinetic energy ($E_k = \frac{1}{2}mv^2$). Thus, only ions that have a specific mass (m_s) do not experience a lateral force and pass the filter.

8.2.2 Mass Analysis System

The mass analysis system collects and analyzes the ion masses to produce mass spectra with the assistance of a computer. The extractor filter extracts secondary ions from the surface, selects a mass range of ions to analyze, and eliminates scattered primary ions from a mass spectrum. The mass analyzer separates secondary ions according to their ratios of mass to electric charge, and these ratios are the basic signals for SIMS analysis. The detector counts the numbers of ions each with specific ratios of mass to electric charge. The mass analyzer, the critical component for secondary ion mass analysis, can be one of the following types: *magnetic sector analyzer*, *quadrupole analyzer* and *time-of-flight analyzer*.

Magnetic Sector Analyzer

This analyzer was the oldest type used for mass spectroscopy. Figure 8.10 illustrates the working principles of selecting ion masses using a magnetic field. Secondary ions are accelerated by an extraction potential of 4 kV before entering a magnetic field. The magnetic field in the magnetic sector will impose a field force in a direction orthogonal to the direction in which the ions travel. The ions with a given kinetic energy will change their travel path to a circular

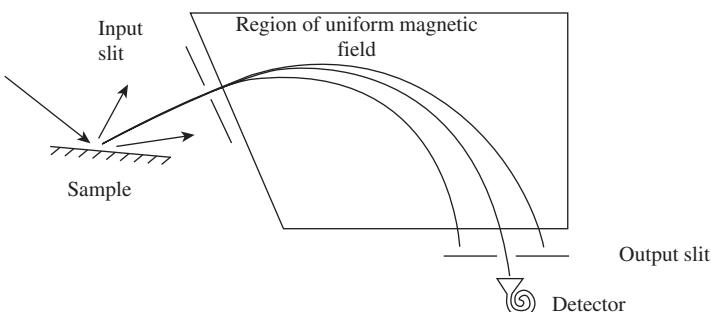


Figure 8.10 Magnetic sector analyzer. (Reproduced with permission from G.C. Smith, *Quantitative Surface Analysis for Materials Science*, Institute of Metals, London. © 1991 Institute of Metals.)

trajectory. The radius of path curvature R has the following relationship with the ion mass.

$$R = \frac{(2V)^{1/2}}{B} \left(\frac{m}{z} \right)^{1/2} \quad (8.2)$$

where the extraction potential (V) is constant. We can adjust the magnetic field strength B , to select ions with a certain *mass-to-charge ratio* (mz^{-1}) with a fixed radius of magnetic sector R . Thus, for given magnetic strength, only ions with selected mz^{-1} can travel through the exit slit of the magnetic sector. A shortcoming of the magnetic sector analyzer is that it can only analyze the secondary ions by sequential selection of mz^{-1} . The magnetic sector analyzer is suitable for dynamic SIMS experiments in which a few specific elemental ions are measured. The ability to resolve the ion mass is limited by the kinetic energy variations in the ions entering the sector. Secondary ions can obtain various amounts of kinetic energy when they are ejected from a surface. The initial kinetic energy of secondary ions may vary up to 100 eV for elemental ions. If taking this initial kinetic energy into account in Equation 8.2, we should replace V with ΔV . Consequently, a value of $\Delta(mz^{-1})$ will be selected with a given magnetic strength. The parameter, *mass resolution*, is used to evaluate the quality of mass analyzers in SIMS. The mass resolution is defined as the ratio of $m(\Delta m^{-1})$. An electrostatic field is used to reduce the kinetic energy bandwidth of ions before entering the magnetic sector analyzer. After kinetic energy correction, the mass resolution of a magnetic sector analyzer can reach a high 10^4 level.

Quadrupole Mass Analyzer

The quadrupole mass analyzer selects ions with a certain mz^{-1} by generating unstable oscillations for the non-selected ions. Oscillations of ion trajectories are created by an electric field combining a constant direct current (DC) and alternating current with a radio frequency (RF). Figure 8.11 illustrates the working principles of a quadrupole mass analyzer. The secondary ions are accelerated by an extraction field, and then the ions travel through the center of four circular rod electrodes. Combined DC and RF voltages are applied to one pair of rods and equal but opposite combined voltages are applied to another pair of rods. Such an arrangement of electric fields generates ion oscillation. The oscillation can be so severe that ion trajectories become unstable and ions strike the rods. Ions with unstable trajectories

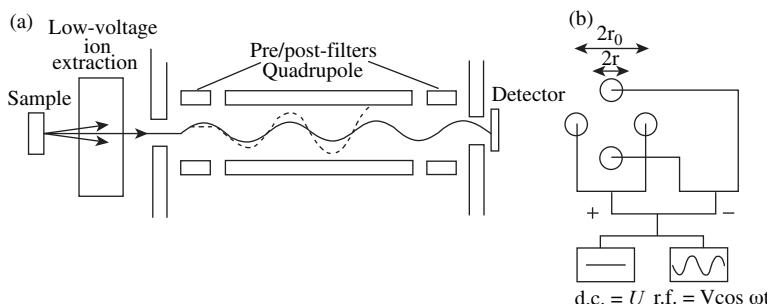


Figure 8.11 A quadrupole analyzer. The oscillations of ions are generated by combined DC and AC electric fields using four cylindrical rods: (a) travel path of secondary ions in the analyzer; and (b) electrode arrangement of the analyzer. (Reprinted by permission of Oxford University Press from J.C. Vickerman, A. Brown, and N.M. Reed, *Secondary Ion Mass Spectrometry: Principles and Applications*, Clarendon Press, Oxford. © 1989 Oxford University Press.)

cannot travel through the exit slit of the analyzer to reach the mass detector. Only the ions with a certain mz^{-1} have stable trajectories and can pass through the exit slit under a given ratio of DC : AC voltages. The analyzer is a sequential type that only allows the ions with single mz^{-1} values to reach the detector. Thus, the quadrupole analyzer is a device with low transmission, as less than 1% of ions can reach the detector at any time. To obtain a whole mz^{-1} spectrum, the analyzer increases the voltages but keeps the ratio of DC : AC voltages constant.

Time-of-Flight Analyzer

The time-of-flight (ToF) analyzer is the most widely used analyzer in static SIMS. As its name indicates, for this analyzer the flight time of an ion is the parameter for measurement. When ions are obtained with a constant kinetic energy from an acceleration potential (V) of 3–8 kV, the flight time of ions through a distance (L) of flight tube to reach a detector is calculated.

$$t = L(2V)^{-1/2} \left(\frac{m}{z} \right)^{1/2} \quad (8.3)$$

Thus, the mz^{-1} of ions is analyzed by measuring their flight time in the analyzer. Heavier ions will have longer flight times in the tube. To measure the time of flight, precisely pulsed primary ions should be used. The pulse is controlled by a highly accurate clock. The pulse periods are typically in the order of 10 ns. The flight time of ions to the detector is electronically measured and converted to mz^{-1} . In a pulse period, all the ions can be measured and a whole mass spectrum can be obtained almost simultaneously. Again, the initial kinetic energy of secondary ions will affect the resolution of mass analysis, because the velocity could be different when ions with the same mz^{-1} enter the flight tube. To overcome the problem of resolution of flight time, the ToF analyzer is constructed in such way that the ions will be reflected by a mirror as illustrated in Figure 8.12. The mirror is composed of a series of precisely spaced rings to which a gradually increasing electric field is applied. Ions with higher kinetic energy will penetrate further into the mirror before being reflected. Thus, ions with the same mz^{-1} will arrive at the detector at the same time if the mirror has been tuned correctly.

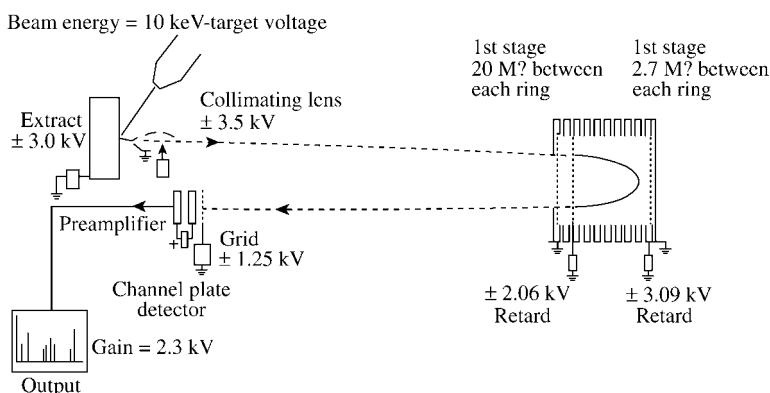


Figure 8.12 Time-of-flight mass analyzer. The secondary ion beam is reflected by a mirror to correct the flight time of ions with identical mz^{-1} . (Reproduced with permission from J.C. Vickerman, *Surface Analysis: The Principal Techniques*, John Wiley & Sons Ltd, Chichester. © 1997 J.C. Vickerman.)

Table 8.2 Comparison of SIMS Mass Analyzers

Mass Analyzer	Mass resolution	Range of Mass detection	Transmission	Manner of detection	Relative sensitivity
Quadrupole	$10^2\text{--}10^3$	$<10^3$	0.01–0.1	Sequential	1
Magnetic sector	10^4	$>10^4$	0.1–0.5	Sequential	10
Time-of-flight	$>10^3$	$10^3\text{--}10^4$	0.5–1.0	Parallel	10^4

(Data reproduced with permission from J.C. Vickerman, *Surface Analysis: The Principal Techniques*, John Wiley & Sons Ltd, Chichester. © 1997 John Wiley & Sons Ltd.)

Table 8.2 compares the main parameters of three mass analyzers. From Table 8.2 we can understand the reasons that the ToF analyzer has become so popular for static SIMS; it provides high resolution, high transmission and high sensitivity. The major shortcoming of the ToF analyzer is its use of pulse primary ions. The ratio of primary beam on- to off-time is only about 10^{-4} . Thus, it is not efficient for analysis such as depth profiling of chemical elements.

8.3 Surface Structure Analysis

Analysis of surface chemical structure requires use of the static SIMS technique to ensure that the major portion of the surface should not be affected by secondary ion emission. *Time-of-flight SIMS* (ToF SIMS) is the most widely used static SIMS technique. As its name indicates, ToF SIMS uses the ToF mass analyzer to measure mz^{-1} of secondary ions. ToF SIMS is a stand-alone instrument, not incorporated into or attached to other SIMS instruments as for dynamic SIMS. A typical structure is illustrated in Figure 8.13.

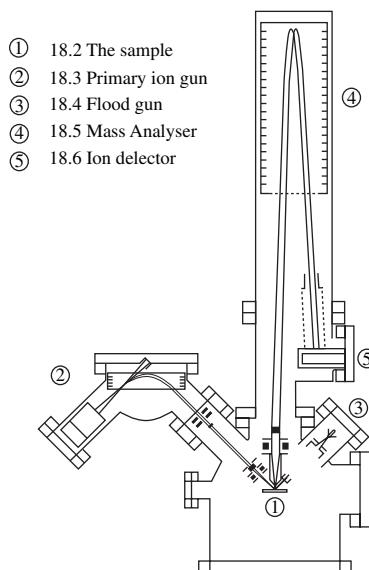


Figure 8.13 Time-of-flight SIMS instrumentation. (Reproduced with permission from J.C. Vickerman and D. Briggs, *ToF-SIMS Surface Analysis by Mass Spectrometry*, IM Publications and SurfaceSpectra, Chichester and Manchester. © 2001 IM Publications.)

8.3.1 Experimental Aspects

Primary Ions

The primary ions commonly used for ToF SIMS are heavy metals such as Ga^+ , Cs^+ and Bi^+ with atomic weights of 69.72, 132.91 and 208.98 amu (atomic mass units), respectively. The high-mass ions increase the yield of secondary ions with limited increase of the damaged cross-section (σ). The energy of the primary ion beam can be varied from 1 to 25 keV. Higher energy may not be favorable because it can cause excessive fragmentation and make SIMS spectra difficult to analyze. Thus, ToF analysis is better with heavy primary ions ‘gently’ striking a surface to be analyzed.

Flood Gun

ToF SIMS instruments are equipped with an *electron flood gun* as shown in Figure 8.13. The flood gun solves the problem of surface charging of insulating samples, similar to the surface charging problem in electron spectroscopy. A sample surface will be positively charged when bombarded by primary ions. The positive primary ions will be implanted into the sample and also secondary electrons will be ejected from the sample surface. The surface charge can rapidly build electric potential on a surface up to several hundred volts. Such potential will increase the kinetic energy of positive secondary ions emitted from surfaces; and on the other hand, it will retard negative ion emission. Between two pulses of primary ion beam bombardment, the flood gun provides electrons with relatively low energy to irradiate the sample surface and make it electrically neutral or even negative. Driving the surface potential down to negative is necessary to ensure release of negative secondary ions. Thus, the higher flux ($\sim 10 \times$ higher) of low energy electrons from the flood gun will be used for examining the negative secondary ions and fragments.

Sample Handling

The samples for ToF SIMS should be carefully handled to avoid chemical contamination and surface roughening. Because contamination can seriously distort the ToF SIMS spectrum of a sample surface, surface roughness and positioning in the instrument are critical. Regular cleaning of the sample surface ensures that the surface is free from contamination. ToF SIMS spectra are extremely sensitive to the atoms and molecules nearby, no matter whether they are in the air (and hence can be adsorbed) or whether they come from physically contacting other objects. A sample surface can even be contaminated by the plastic from a bag or box used to store the sample. Organic compounds from such containers may be released into the air and adsorbed on a cleaned sample surface. For inorganic samples, we may clean the surface by sputtering away the contamination using primary ions before acquiring secondary ions. Sputtering, however, is not suitable for organic samples because it may destroy the structure of an organic surface. Certain solvents to which the sample is inert, such as hexane, may be used to rinse organic surfaces.

Roughness or topographic features of a sample surface can cause a significant reduction in mass resolution. The main reason is that the electrostatic field on a rough surface is not even. The field strength is reduced in dips and troughs from the uppermost surface. The mass resolution will be degraded significantly when a surface has peak-to-trough roughness beyond 10 μm .

The positioning of the sample relative to the mass analyzer is also important. Inaccuracy in the distance to the secondary ion extractor associated with mass analyzer affects mass resolution, but does not strongly affect spectral intensity. Tilting the sample relative to the

extractor causes more problems because tilting will make the trajectories of the secondary ions curved when entering the mass analyzer and will, as a result, distort the spectra. Large molecular fragments are more affected by sample tilting. For example, a tilt angle of 4° will reduce the relative spectral intensity of C_7H_7 in polymer molecules by a factor of two, but will have little effect on the spectral intensity of CH_3 .

8.3.2 Spectrum Interpretation

The secondary ion mass spectra expresses the detected intensities of secondary ions versus their mass-to-charge ratio (mz^{-1}) where mass is in atomic mass units. Figures 8.14–8.16 show examples of mass spectra produced using ToF SIMS. We can identify individual ions, ion clusters and molecular fragments according to their mz^{-1} values. Two particular features of SIMS spectra are revealed in Figures 8.14–8.16. First, the spectra can be either that of positive ions or negative ions. For example, both Figure 8.14 and Figure 8.15

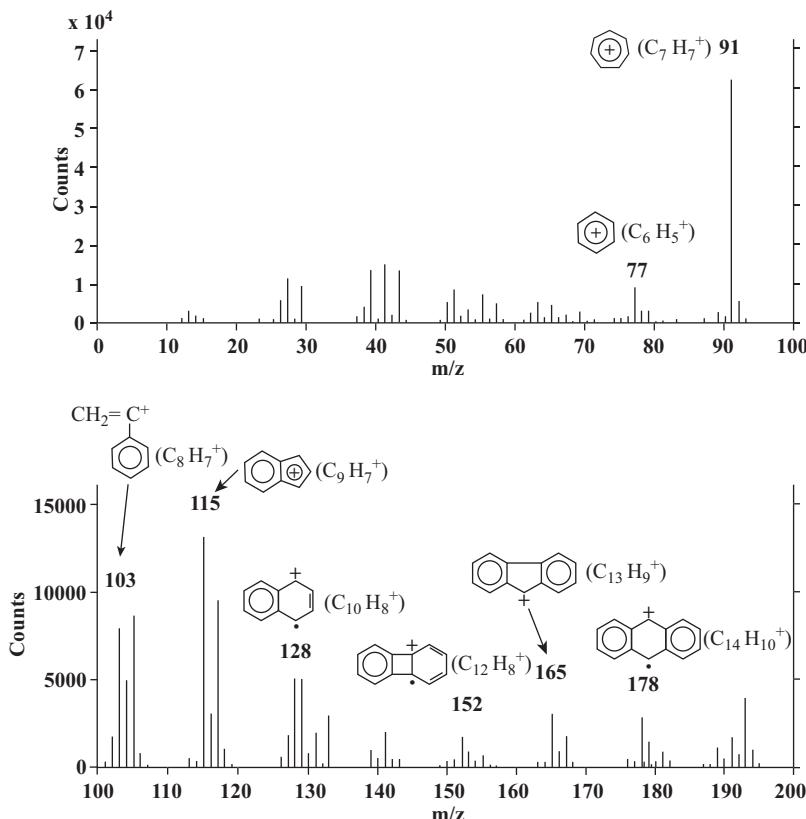


Figure 8.14 Positive ion ToF SIMS spectrum of polystyrene. (Reproduced with permission from D. Briggs, *Surface Analysis of Polymer by XPS and Static SIMS*, Cambridge University Press, Cambridge, UK. © 1998 Cambridge University Press.)

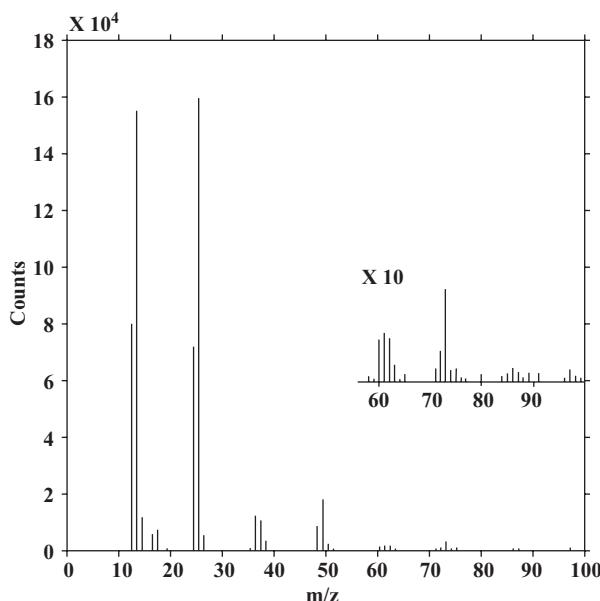


Figure 8.15 Negative ion ToF SIMS spectrum of polystyrene. (Reproduced with permission from D. Briggs, *Surface Analysis of Polymer by XPS and Static SIMS*, Cambridge University Press, Cambridge, UK. © 1998 Cambridge University Press.)

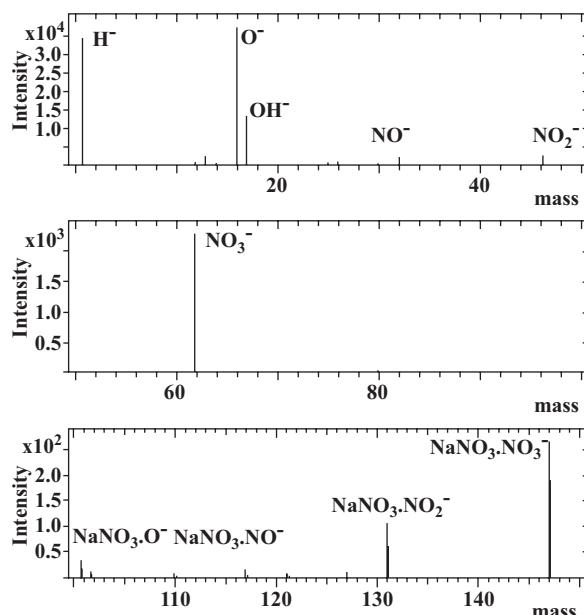


Figure 8.16 Negative ToF spectrum of NaNO_3 . (Reproduced with permission from J.C. Vickerman and D. Briggs, *ToF-SIMS Surface Analysis by Mass Spectrometry*, IM Publications and SurfaceSpectra, Chichester and Manchester. © 2001 IM Publications.)

are the mass spectra of polystyrene, but Figure 8.14 is of positive ions while Figure 8.15 is of negative ions. Second, SIMS spectra appear more complicated than those of AES and XPS because there are more peaks than the basic ion components. The SIMS spectra often exhibit secondary ions that are various combinations of atoms and ion fragments ejected from the sample. For example, the spectrum of a simple inorganic compound, NaNO_3 , includes not only peaks of the ions that comprise the molecule, but also peaks of adduct (unbonded associations of molecules) ions with high mz^{-1} values including $\text{NaNO}_3\cdot\text{O}^-$ and $\text{NaNO}_3\cdot\text{NO}^-$, as shown in Figure 8.16. The mass spectra of polymers are even more complicated, as shown in the example of polystyrene in Figure 8.14. Interpreting SIMS spectra, particularly those of polymers, requires experience and knowledge of materials.

Element Identification

Identifying elements from SIMS spectra is relatively simple. The elements can either be identified by their exact mz^{-1} or by detecting expected isotopes. To obtain accurate mass measurement, the mass correlation with flight time should be calibrated based on Equation 8.3. The calibration can be done using a known ion mass in the sample. There are general rules that are helpful for element identification, given as follows. The great majority of metals give exclusively positive ions in static SIMS conditions. The most electropositive elements, such as alkali and alkali earth metals, give intense positive ion peaks. The most electronegative elements, such as O, F, Cl, Br and I give intense negative ion peaks. Hydrogen gives a high yield of both H^+ and H^- ions. The CH^- peak is commonly more intense than the C^- peak. N is not usually detectable by an elemental peak. P is usually detected from peaks of PO_2^- and PO_3^- .

A polymer surface commonly has a complicated spectrum. The polymer spectrum can be divided into three regions: *sub-monomer*, *n-mer* and *oligomer* regions. The sub-monomer region is the region where mass values are less than the monomer mass. This is the main region of spectra to examine when identifying polymer molecules because it provides information on the basic chemical information of polymers. For example, the monomer mass of polystyrene is 104 amu. The spectral region at $mz^{-1} < 104$ in Figure 8.14 provides basic structural information on polystyrene. The negative ion spectrum of a polymer is simpler than that of its positive ion spectrum, as can be seen by comparing Figure 8.15 with Figure 8.14. The positive ion spectrum frequently includes hydrocarbon-originating ions, which do not provide much structure information for the polymer.

The *n-mer* region of a spectrum includes a small number of monomer units or monomer units plus fragments. This region may contain more information on the macromolecular structure. This region is more important for polymers containing more than one type of monomer (copolymers). Also, the region can reveal the state of cross-linking in polymers. The oligomer region is at high mz^{-1} , in which peaks are generated by a process called *cationization*. Cationization is the laying down of a thin layer of organic substance or polymer on a silver substrate. ToF SIMS is able to detect the peaks of organic molecules (M) attached to Ag as $[\text{M}_x^+\text{Ag}]^+$. The peaks of cationization can reach mz^{-1} levels at a scale of 10^3 . Figure 8.17 shows an example of cationization used to examine an organic residue on glass. The organic residue was deposited on a clean silver substrate by rubbing glass over the substrate. The monolayer of organic residue was sputtered with Ag. The oligomer region, in principle, can be used for estimating the average molecular weight distribution of polymeric materials.

The interpretation of polymer spectra has been the subject of extensive study; thus, published sources are available to help with interpretation. Databases of ToF SIMS spectra have been

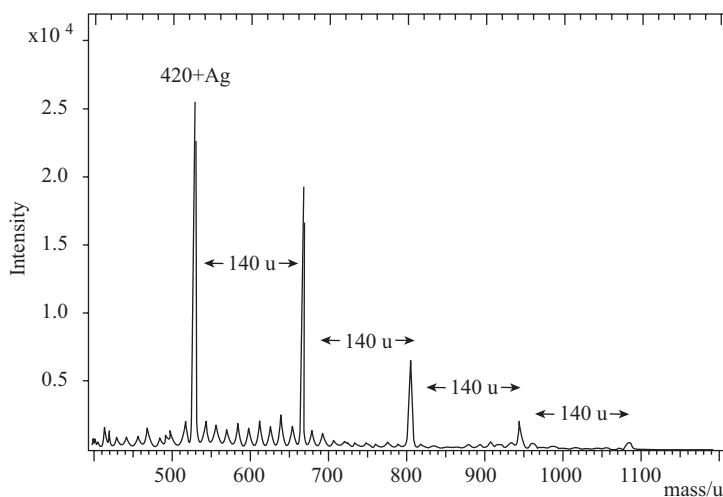


Figure 8.17 Positive ToF SIMS spectrum of a polymeric substance that has a repeat unit of $C_{10}H_{20}$. (Reproduced with permission from J.C. Vickerman and D. Briggs, *ToF-SIMS Surface Analysis by Mass Spectrometry*, IM Publications and SurfaceSpectra, Chichester and Manchester. © 2001 IM Publications.)

established in which more than a thousand spectra for surface analysis can be found. On the other hand, a regular cluster pattern of hydrocarbon polymers is possibly used to interpret some polymer spectra. The pattern is especially helpful when the standard spectra of samples to be analyzed are not available. For example, the high density polyethylene (HDPE) spectrum in Figure 8.18 shows a cluster pattern. There are several clusters of peaks and each of them represents a $C_nH_m^+$ ($m = n, \dots, 2n + 1$) cluster in the positive ion spectrum of HDPE. We can easily assign the clusters as C_1, C_2, C_3, \dots because the mass of each carbon atom is 12 amu. The peaks in each cluster represent the ions with the same number of carbon atoms but variation in hydrogen ions.

Generally, the intensity of ion peaks decreases with increasing ion mass. For hydrocarbon polymers containing cyclic ions (aromatic hydrocarbon polymers) such as polystyrene, which

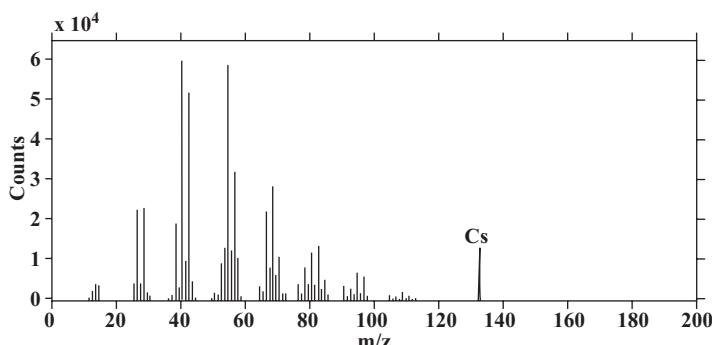


Figure 8.18 Positive ToF SIMS spectrum of high density polyethylene (HDPE). (Reproduced with permission from J.C. Vickerman and D. Briggs, *ToF-SIMS Surface Analysis by Mass Spectrometry*, IM Publications and SurfaceSpectra, Chichester and Manchester. © 2001 IM Publications.)

contains benzene rings, the $C_nH_m^+$ clusters in spectra extend to the high-mass region. The characteristic peaks of such polymers are for $C_7H_7^+$ and $C_6H_5^+$ clusters, which represent the cyclic ions with seven carbon atoms and six carbon atoms, respectively. Table 8.3 lists a few characteristic ion masses that are frequently encountered in surface analysis of polymeric materials.

Table 8.3 Characteristic Secondary Ions and Nominal Mass of Some Molecules Frequently Encountered in Polymer Surfaces

Molecule	Structure	Characteristic secondary ions (<i>m/z</i>)	
		Positive	Negative
Poly(dimethyl-siloxane) (silicone)	The structure shows a silicon atom (Si) bonded to two methyl groups (CH3) and one oxygen atom (O). This is part of a polymer chain where the silicon is also bonded to another oxygen atom in the next unit.	73,147,201, 221,281	59,60,75,119, 149,223
Poly(ethylene glycol)	The structure shows a carbon atom (CH2) bonded to another carbon atom (CH2) and an oxygen atom (O). This is part of a polymer chain where the carbons are also bonded to hydrogen atoms.	45,89,133, 177	61,105,149
Poly(propylene glycol)	The structure shows a carbon atom (CH) bonded to a methyl group (CH3) and another carbon atom (CH2), which is further bonded to an oxygen atom (O). This is part of a polymer chain where the carbons are also bonded to hydrogen atoms.	59,117,175, 233	57,75,113, 133
Stearic acid	$C_{17}H_{35}COOH$	267,285	283
Palmitic acid	$C_{15}H_{31}COOH$	239,257	255
Stearamide	$C_{17}H_{35}CONH_2$	284	42,282
Oleamide	$C_{17}H_{33}CONH_2$	282	42,280
Erucamide	$C_{21}H_{41}CONH_2$	338	42,336
Ethylene bis-stearamide	$(CH_2NHCOC_{17}H_{35})_2$	282,310	42
Di-octyl phthalate		149,167,261, 279	121
Nonyl phenol ethoxylate	$C_9H_{19}-\text{Phenyl group}-O(CH_2CH_2O)_nH$	247,419/435, 463/479 etc.	133,219
Glyceryl monostearate	$\begin{array}{c} CH_2OCOC_{17}H_{35} \\ \\ CHOH \\ \\ CH_2OH \end{array}$	267,341,359	71,283
Triphenyl phosphate	$(C_6H_5O)_3P=O$	77,251,327	63,79,93,249, 325

(Data reproduced with permission from D. Briggs, *Surface Analysis of Polymer by XPS and Static SIMS*, Cambridge University Press, Cambridge, UK. © 1998 Cambridge University Press.)

8.4 SIMS Imaging

SIMS instrumentation enables us to operate SIMS in a scanning mode because it is equipped with an X–Y deflector on the primary ion beam. In scanning mode, a focused primary ion beam scans over a surface area of the sample. Secondary ions emitted in each pixel generate a two-dimensional SIMS image that reveals the distributions of ions on a surface. Formation of SIMS images is quite similar to chemical element mapping by energy-dispersive spectroscopy (EDS) and AES.

8.4.1 Generation of SIMS Images

The liquid metal ion sources of ToF SIMS can provide a focused ion beam with a diameter in the order of 10^{-9} and 10^{-6} and even less, because the structure of such a primary ion source is similar to the electron field emission gun in electron microscopes. Gaseous ion sources used in dynamic SIMS can also provide a focused primary ion beam for scanning. The commonly used diameters of ion beams for scanning images are about 200 nm to 1 μm . The X–Y deflector attached to the primary ion source rasters the beam on the sample surface. Each pixel composing a SIMS image may comprise signals from a single selected mz^{-1} or a few mz^{-1} in dynamic SIMS, as well as from a whole mass spectrum in ToF SIMS.

Figure 8.19 shows an example of a SIMS image of a cross section of a coating on a titanium substrate. Images of positive ions in Figure 8.19 reveal the chemical gradient in the coating composed of calcium phosphate. The images of Ca and Ti ions clearly reveal the interface

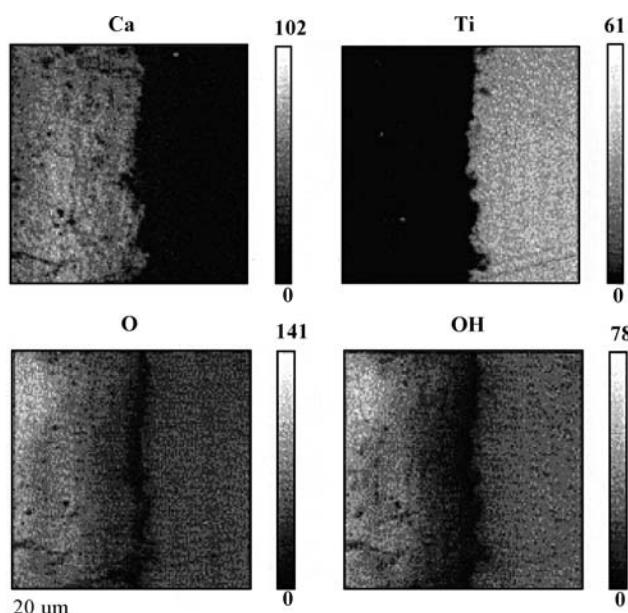


Figure 8.19 SIMS images of ceramic coating on titanium obtained with scanning ToF SIMS. Ga^+ primary ion beam generates the positive ion images of Ca, Ti, O and OH. Chemical gradients of O and OH in the coating layer (left-hand sides of images) are revealed.

between the calcium phosphate coating and the titanium substrate. The images of O and OH ions indicate that there is deficiency of O₂ and OH in the coating region near the interface because of lower ion intensities there.

8.4.2 Image Quality

Ideally, we wish to obtain SIMS images with high spatial resolution of secondary ion signals with high signal-to-noise ratios. However, reduction of pixel size causes a decreasing number of secondary ions in each pixel. Smaller pixel size means fewer atoms or molecules will be available to emit secondary ions in each pixel. This situation is especially true for ToF SIMS imaging. ToF SIMS limits the primary ion density to 10^{13} cm^{-2} . The number of primary ions in each pixel could be too small to generate sufficient secondary ions. The signal-to-noise ratio for SIMS imaging relies on counts of secondary ions received by a detector. To understand the problem, we can estimate the counts of secondary ions in each pixel by the ToF SIMS detector. Assuming a sputtering yield of 1, ionization probability of 10^{-3} and transmission of 0.5, we can only have about 50 secondary ions in each pixel of $1 \mu\text{m}^2$ based on Equation 8.1. For molecular fragment ions, the signals are even weaker because of their low sputtering yield.

One solution to the problem is to increase the ionization probability. This can be done by choosing primary ions with heavy mass, for example, Bi⁺ or even C₆₀⁺, which contains 60 carbon atoms. The noise level can also be reduced by techniques of digital image processing. For example, a fast Fourier transform technique has been used to remove noise from the image. This technique transforms an image from a space domain to a reciprocal domain by sine and cosine functions. Noise can be readily filtered out in such domain. After a reverse Fourier transform, filtered data produces an image with much less noise.

SIMS imaging can also be affected by surface topographic features. Particularly, for insulating samples, the electron field intensity varies with surface height. Correspondingly, the signals from secondary ions will vary even there is not chemical variation. We should be aware of this possible artifact when interpreting SIMS. Thus, to reduce this misleading variation, the sample for SIMS imaging should be as flat as possible. We may also check whether there is a topographic effect by comparing SIMS images for several types of secondary ions. For example, comparison of images of Ca and O ions in Figure 8.19 tells us that the oxygen gradient in the coating is real, not an artifact of surface topography. Note the contrast change near the interface region in the image of the O ions. This same contrast change would appear in the image of Ca ions if the contrast had resulted from the surface topography.

8.5 SIMS Depth Profiling

Early applications of SIMS were not for surface analysis, but for detecting trace elements in near-surface regions of samples. The high sensitivity of SIMS to all chemical elements enables us to trace any element with a concentration down to sub-parts per million (ppm) levels. This is the main advantage of SIMS depth profiling of trace elements with extremely low concentrations, compared with AES and XPS depth profiling. SIMS depth profiling relies on primary ions removing various elements from surface layers at a uniform sputtering rate. A steady state of sputtering can be achieved after a certain time, even though various elements show different sputtering rates initially. Depth profiling can be conducted in both dynamic and static SIMS. Depth profiling by ToF SIMS is less efficient than by dynamic SIMS because in

ToF SIMS separated ion beams are used for etching and analysis purposes. A depth profile is obtained by periodically switching between etching and analysis. In addition, the pulse mode operation in ToF SIMS further slows down the process. ToF SIMS, however, can be used to obtain more accurate depth profiles if time is not a concern. The next section introduces depth profiling based on the dynamic SIMS technique, which is very popular in analyzing electronic materials.

8.5.1 Generation of Depth Profiles

Depth profiling requires a focused ion beam raster over a rectangular area, similar to SIMS imaging. The area of depth profiles is usually in the range of $10\text{ }\mu\text{m} \times 10\text{ }\mu\text{m}$ to $500\text{ }\mu\text{m} \times 500\text{ }\mu\text{m}$. The beam energy is in the range 1–20 keV, and beam density is at the level of $1\text{ }\mu\text{A cm}^{-2}$. An O_2^+ ion beam is commonly used for positive ions profiles. The oxygen ions generate an oxide layer, and such a layer produces very high positive ion yields of electropositive elements. Cesium ions, Cs^+ , are widely used for generating high negative ion yields of the elements that exhibit low positive ion yields. The most widely used mass analyzer in dynamic SIMS is the magnetic sector analyzer. The high mass resolution and good transmission make the magnetic sector analyzer the top choice for depth profiling. However, the quadrupole analyzer is used when shallow profiles are needed, as these require only a low energy primary beam and a low incident angle.

Figure 8.20 shows an example of a depth profile obtained by dynamic SIMS; it reveals GaAs and Si layers coating an Al_2O_3 substrate. In a SIMS depth profile, the atomic density or secondary ion intensity of elements being analyzed is plotted against either depth (z) or sputtering time (t), and can be easily converted to depth at a constant sputtering rate. The depth profile shown in Figure 8.20 clearly indicates the location of GaAs–Si and Si– Al_2O_3 interfaces. In Figure 8.20, the Ga and As intensities do not drop to zero in the Si layer, even though such elements do not exist there. Instead, there is background with intensity fluctuation in the order of $10^{16}\text{--}10^{17}\text{ atoms cm}^{-3}$ in the profile. This background level represents the detection limit of depth profile, that is, the lowest atom concentration of Ga and As that can be detected.

8.5.2 Optimization of Depth Profiling

The best depth profiling detects extremely low concentrations of trace elements and their accurate depth positions. Thus, for depth profiling of elements, high *detection sensitivity*, low *detection limit* and high *depth resolution* are required. Detection sensitivity is the ability to detect a given element in a sample. The detection limit is referred to as the lowest atom concentration of a given element that can be detected in a sample. The detection limit is determined by the level of background signals. The ratio of peak intensity of a given element to the background intensity (called *dynamic range*) is often used to describe the quality of depth profiling. A good dynamic range means high detection sensitivity and low detection limit. The depth resolution is referred to as the accuracy of depth measurement for elemental concentrations. It is commonly defined as a depth range (Δz or Δt) in which the concentration (or secondary ion intensity) varies from 84.13–15.87% at a depth profile with a step-like chemical gradient as illustrated in Figure 8.21. Limitations in depth resolution will distort the profile at the interface from truly step-like to a gradual profile.

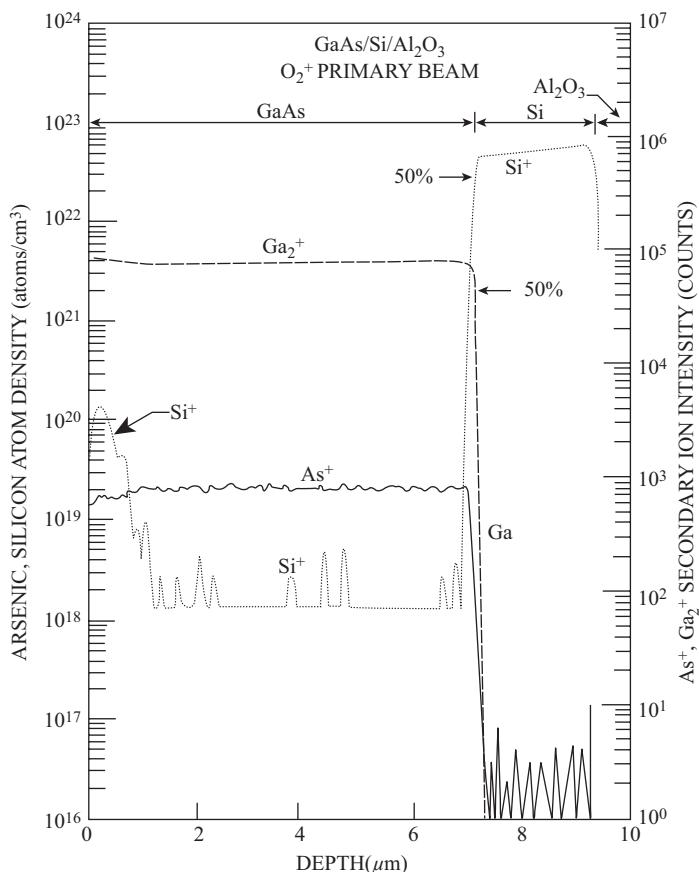


Figure 8.20 SIMS depth profile of the GaAs–Si– Al_2O_3 system. (Reproduced with permission from R.G. Wilson, F.A. Stevie and C.W. Magee, *Secondary Ion Mass Spectrometry: A Practical Handbook for Depth Profiling and Bulk Impurity Analysis*, John Wiley & Sons Inc., New York. © 1989 John Wiley & Sons.)

Optimization of depth profiling relies on selection of proper experimental conditions. Effects of processing parameters on the quality of profiles, including primary beam energy, incident angle of primary beam and the analysis area versus the sputtering area on a sample surface, are briefly introduced in the following section.

Primary Beam Energy

The beam energy affects the depth resolution and sputtering yield. Higher beam energy results in higher sputtering yield. However, a high energy ion beam can mobilize atoms in a sample, causing some atoms to move deeper into the sample while causing some atoms to move toward the surface. Consequently, a high-energy primary ion beam induces a mixing of atoms and distorts the depth profile. Figure 8.22 shows the effect of beam energy on depth resolution due to mixing, as As doping atoms in the SiO_2 are pushed into the Si by the high energy beam. Thus, lower beam energy is desirable to obtain better depth resolution.

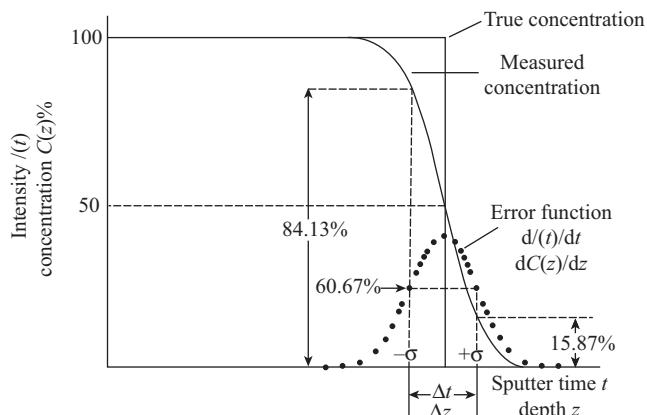


Figure 8.21 Definition of depth resolution and its effect on depth profiling. (Reproduced with permission from J.C. Vickerman, *Surface Analysis: The Principal Techniques*, John Wiley & Sons Ltd, Chichester. © 1997 John Wiley & Sons Ltd.)

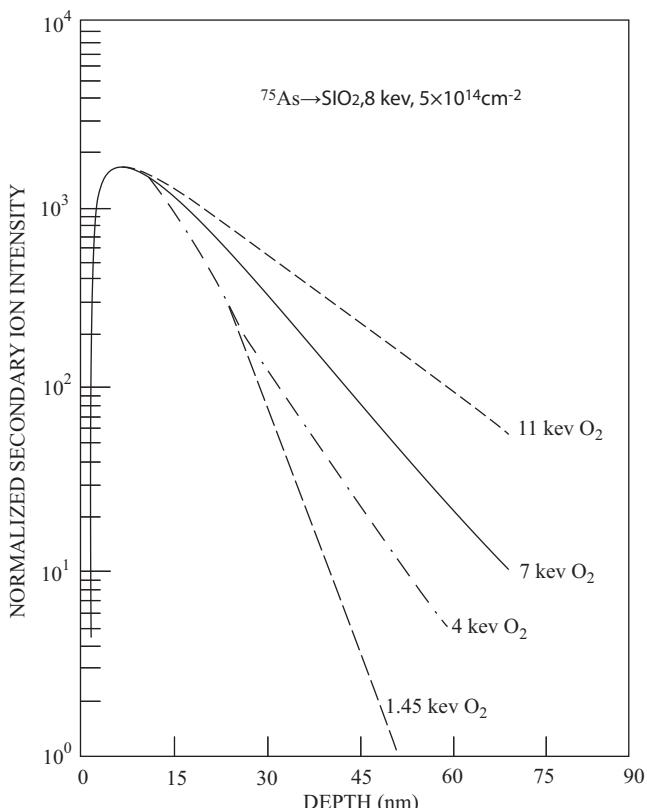


Figure 8.22 Primary O_2^+ beam energy effects on depth profile of As doped in SiO_2 substrate. (Reproduced with permission from R.G. Wilson, F.A. Stevie and C.W. Magee, *Secondary Ion Mass Spectrometry: A Practical Handbook for Depth Profiling and Bulk Impurity Analysis*, John Wiley & Sons Inc., New York. © 1989 John Wiley & Sons.)

Incident Angle of Primary Beam

The incident angle of the primary beam affects sputtering yield, ionization probability and depth resolution. Generally, the sputtering yield decreases with increasing incident angle and reaches a minimum at an angle normal to the surface. For example, a Si sputtering yield under O_2^+ beam with the incident angle of 30 degrees is 10 times higher than that of 90 degrees. However, the ionization probability increases with the incident angle and reaches a maximum at the angle normal to the surface. For example, the ionization probability of Si under O_2^+ beam with a incident angle of 60 degrees is two orders of magnitude higher than that of 30 degrees. The high incident angle may also affect the depth resolution because of surface roughening by the primary beam. The intensity inside a primary ion beam is not uniform, but a Gaussian distribution as the center of the beam has higher intensity than the edge. Such a Gaussian distribution of ion intensity inside the beam generates uneven sputtering when the beam rasters on the surface. The roughening by a beam normal to the surface is more severe than by a beam with small incident angle. Surface roughness will degrade the depth resolution. Thus, an optimized incident angle should be decided by the purpose of depth profiling.

Analysis Area

The primary ion beam etches a well defined crater in a sample by a beam deflector as illustrated in Figure 8.23. Ideally, only the secondary ions emitted from the crater bottom should be collected for depth profiling. However, secondary ions will be collected at the same time as when the primary ions etch the crater. Thus, the secondary ions from the crater walls may also contribute to the secondary ions in analysis. Such secondary ions from crater walls do not represent the true concentrations of elements at the crater bottom. To avoid this problem,

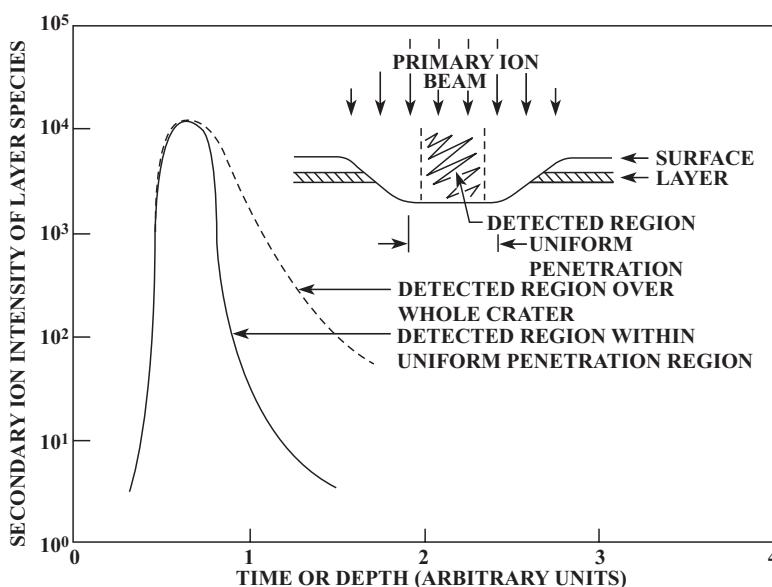


Figure 8.23 Schematic illustrating the importance of selecting an analysis area smaller than the crater bottom. (Reproduced with permission from R.G. Wilson, F.A. Stevie and C.W. Magee, *Secondary Ion Mass Spectrometry: A Practical Handbook for Depth Profiling and Bulk Impurity Analysis*, John Wiley & Sons Inc., New York. © 1989 John Wiley & Sons.)

the analysis area should be smaller than the size of the crater. Thus, the profile will not be affected by the crater walls, as illustrated in Figure 8.23. Commonly, an *electronic gate* is incorporated in the detection system. The electronic gate will switch off detection when the primary beam does not strike the analysis area in the middle of the crater bottom. The electronic gate and proper optical arrangement can significantly lower the detection limit to the order of 10^{16} atom cm $^{-3}$, which is equivalent to about 0.2 ppm.

References

- [1] Vickerman, J.C. (1997) *Surface Analysis: The Principal Techniques*, John Wiley & Sons, Chichester.
- [2] Vickerman, J.C., Brown, A., and Reed, N.M., (1989) *Secondary Ion Mass Spectrometry: Principles and Applications*, Clarendon Press, Oxford.
- [3] Briggs, D. (1998) *Surface Analysis of Polymer by XPS and Static SIMS*, Cambridge University Press, Cambridge, UK.
- [4] Briggs, D., Brown, A. and Vickerman, J.C. (1989) *Handbook of Static Secondary Ion Mass Spectrometry*, John Wiley & Sons, Chichester.
- [5] Smith, G.C. (1991) *Quantitative Surface Analysis for Materials Science*, The Institute of Metals, London.
- [6] Vickerman, J.C. and Briggs, D. (2001) *ToF-SIMS Surface Analysis by Mass Spectrometry*, IM Publications and SurfaceSpectra, Chichester and Manchester.
- [7] Wilson, R.G., Stevie, F.A. and Magee, C.W. (1989) *Secondary Ion Mass Spectrometry, a Practical Handbook for Depth Profiling and Bulk Impurity Analysis*, John Wiley & Sons, New York.

Questions

- 8.1 Figure 8.24 shows the sputtering yield of copper atoms as a function of bombarding energy from three types of primary ions: He, Ar and Xe. Determine which legend of the experimental data is for He, Ar and Xe, respectively; justify your decision.

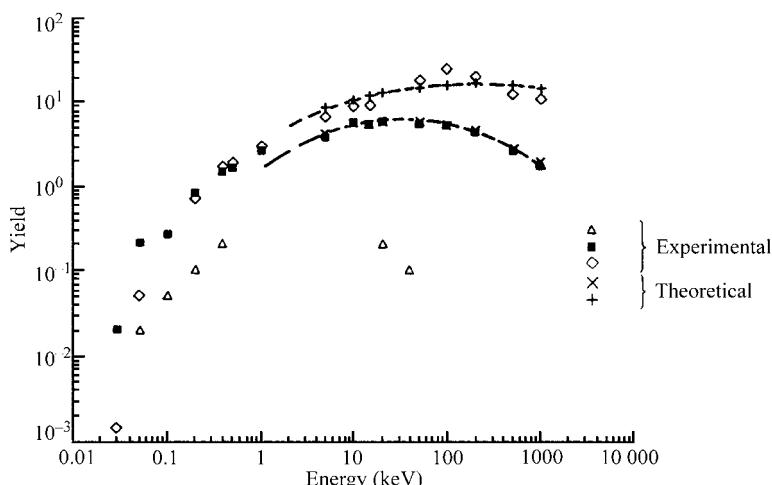


Figure 8.24 Sputtering yield data of copper atoms as a function of primary ion energy for different primary ions. (Reproduced with permission from J.C. Vickerman, *Surface Analysis: The Principal Techniques*, John Wiley & Sons Ltd, Chichester. © 1997 John Wiley & Sons Ltd.)

- 8.2 Why is O_2^+ commonly used as the source of primary ions for dynamic SIMS analysis?
- 8.3 What are the main differences between dynamic and static SIMS?
- 8.4 We may estimate the lifetime of a surface monolayer by assuming that each primary ion will knock out one atom in the monolayer. Estimate the lifetime of surface monolayer bombarded by primary ion beam densities of 10 nA cm^{-2} and 0.1 nA cm^{-2} , respectively. Why should the primary ion dose be even less than that based on the lifetime estimation?
- 8.5 The positive ion spectrum of TiO_2 exhibits peaks at 48, 49, 64 and $65 \text{ } m_z^{-1}$. Try to assign the ions or ion clusters to those peaks.
- 8.6 Try to determine the ions in the negative SIMS spectrum of polystyrene in Figure 8.15.
- 8.7 Surface charging affects the analysis in EDS, XPS and SIMS. What are the similarities and differences for problems of surface charge among those techniques?
- 8.8 Why does ToF SIMS use a pulsed, not continuous, primary ion beam?
- 8.9 Which type of primary ion source can provide better static SIMS analysis? Why do we not recommend using the smallest diameter of a primary beam for ToF SIMS imaging?
- 8.10 Can you determine the inter-diffusion distances of atom A and B at their interface region in a SIMS depth profile? Why?

9

Vibrational Spectroscopy for Molecular Analysis

Vibrational spectroscopy is a technique to analyze the structure of molecules by examining the interaction between electromagnetic radiation and nuclear vibrations in molecules. Vibrational spectroscopy is significantly different from the spectroscopic methods using interactions between materials and X-rays introduced in Chapters 6 and 7. Vibrational spectroscopy uses electromagnetic waves with much longer wavelengths, in the order of 10^{-7} m, than the X-rays which are also electromagnetic waves but with wavelengths in the order of 10^{-10} m. Typical electromagnetic waves in vibrational spectroscopy are infrared light. Energies of infrared light match with vibrational energies of molecules. Vibrational spectroscopy detects the molecular vibrations by the absorption of infrared light or by the inelastic scattering of light by a molecule. Vibrational spectroscopy can be used to examine gases, liquids and solids. It is widely used to examine both inorganic and organic materials. However, it cannot be used to examine metallic materials because they strongly reflect electromagnetic waves. This chapter introduces two spectroscopic methods: *Fourier transform infrared spectroscopy* (FTIR) and *Raman microscopy* (also called micro-Raman), which are the vibrational spectroscopy techniques most commonly used by scientists and engineers for materials characterization.

9.1 Theoretical Background

9.1.1 Electromagnetic Radiation

It is necessary to review electromagnetic radiation in general before discussing its interaction with molecular vibrations. Electromagnetic radiation, traveling at a constant speed of light, varies in wavelength over 10 orders of magnitude and includes radio waves ($\sim 10^2$ m) to γ rays ($\sim 10^{-12}$ m) as illustrated in Figure 9.1. Visible light occupies only a short range of wavelengths ($0.40\text{--}0.75 \times 10^{-6}$ m). The energies of molecular vibrations match with those of electromagnetic radiation in a wavelength range near visible light. The electromagnetic radiation in near visible light is able to change status of molecular vibrations and produce

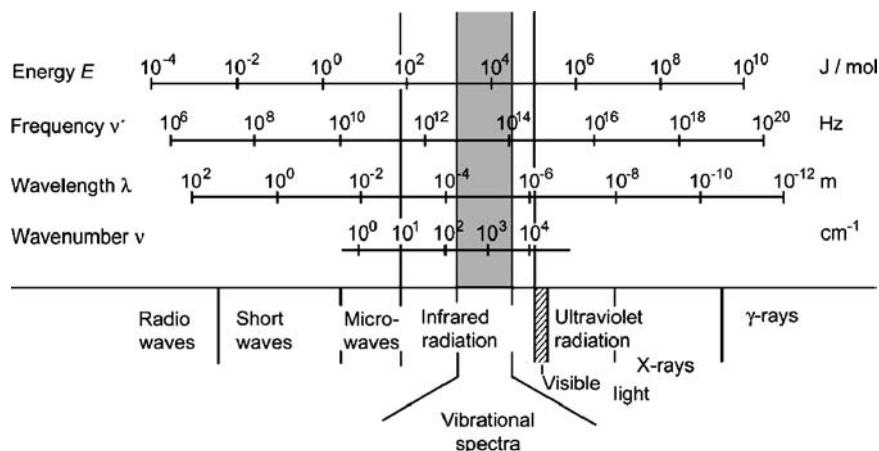


Figure 9.1 Energy, frequency, wavelength and wavenumber ranges of electromagnetic waves. Frequency range of molecular vibrations is in infrared region close to visible light. (Reproduced from A. Fadini and F-M. Schnepel, *Vibrational Spectroscopy: Methods and Applications*, Ellis Horwood, Chichester, 1989.)

vibrational spectra of molecules. Vibrational spectroscopy characterizes the electromagnetic waves in terms of *wavenumber* ($\bar{\nu}$), which is defined as the reciprocal of wavelength in the unit of cm^{-1} .

$$\bar{\nu} = \frac{1}{\lambda} \quad (9.1)$$

Thus, the wavenumber is number of waves in a 1 cm-long wavetrain. It is convenient for us to remember that wavenumber is proportional to the frequency of the electromagnetic wave (ν) with a constant factor that is the reciprocal of the speed of light (c).

$$\bar{\nu} = \frac{1}{c} \nu \quad (9.2)$$

The wavenumber represents radiation energy, as does wavelength. As we know, electromagnetic waves can be considered as photons. The photon energy is related to the photon frequency (ν_{ph}).

$$E = h\nu_{ph} \quad (9.3)$$

h is Planck's constant ($6.626 \times 10^{-34} \text{ J s}$). Thus, the photon energy can be represented as its wavenumber.

$$E = hc\bar{\nu} \quad (9.4)$$

The conversion constant (hc) is about $2.0 \times 10^{-23} \text{ J cm s}$. For a wavenumber of 1000 cm^{-1} , the corresponding energy should be only about $2.0 \times 10^{-20} \text{ J}$ or 0.12 eV . Note that this is much smaller than the photon energy of X-rays, which is in the order of $10,000 \text{ eV}$. As shown

in Figure 9.1, the vibrational spectra are in the wavenumber range from several hundreds to thousands. This indicates that the vibrational energy of molecules are only in the order of about 10^{-2} – 10^{-1} eV.

9.1.2 Origin of Molecular Vibrations

Molecules in solids are always in vibration at their equilibrium positions, except at the temperature of absolute zero (-273.15°C). Molecular vibrations can be simply modeled as massless springs connecting nuclei in a molecule. Figure 9.2 illustrates the simplest model of molecular vibrations: a diatomic molecule vibration by stretching or compressing the bond (a massless spring) between two nuclei. The vibrational energy can be calculated using the spring model. The force (F) due to linear elastic deformation of the bond is proportional to displacement of two nuclei from its equilibrium distance r_e .

$$F = -K\Delta r = -K(r - r_e) \quad (9.5)$$

The force constant K is a measure of bond strength, and r is the distance between two nuclei after displacement. Such vibration motion is considered as *harmonic*. The potential energy of a *harmonic vibration* is expressed as E_{vib} .

$$E_{\text{vib}} = \frac{1}{2}K(\Delta r)^2 = \frac{1}{2}K(r - r_e)^2 \quad (9.6)$$

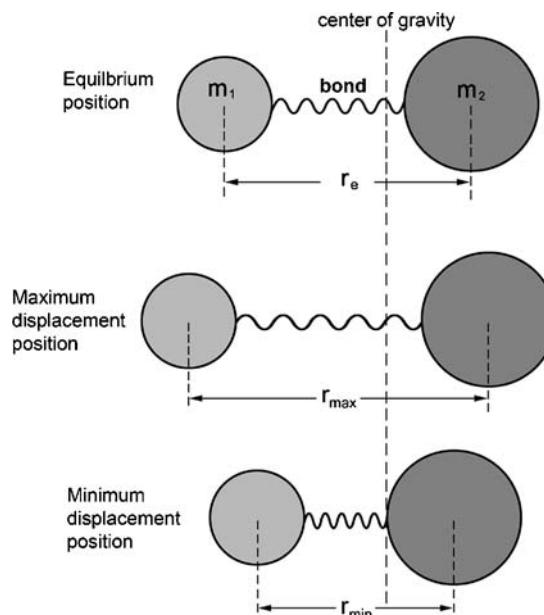


Figure 9.2 Diatomic model of molecular vibration. The centre of gravity does not change during the stretching vibration.

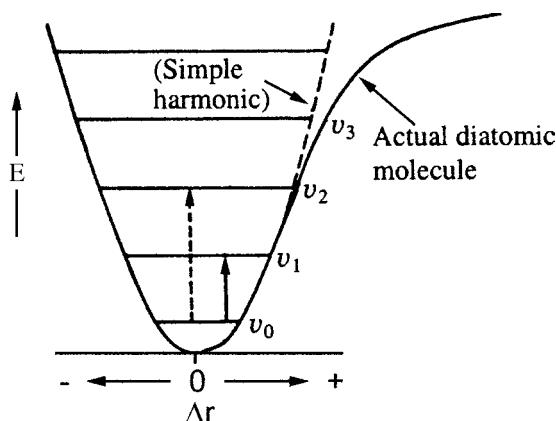


Figure 9.3 Potential energy (E) of diatomic molecule vibration. For a small displacement vibration (Δr), the vibration motion can be treated as harmonic. Quantum theory divides the potential energy of vibration into several levels as $v_0, v_1, v_2, v_3, \dots$

Figure 9.3 illustrates the potential energy of harmonic motion as a parabolic function of displacement from the equilibrium position of nuclei. The harmonic curve differs from a real relationship between energy and displacement in molecules. However, this harmonic assumption is a good approximation for molecular vibrations with small displacement. Although the potential energy is a continuous function of the nuclear displacement, quantum mechanics tell us that the real vibrational energy of a molecule should be quantified. For harmonic vibration, the vibrational energy can be described.

$$E_{vib} = h\nu_{vib} \left(v + \frac{1}{2} \right) \quad v = 0, 1, 2, \dots \quad (9.7)$$

v is the *vibrational quantum number*, which defines distinguishable vibrational levels as shown in Figure 9.3 and ν_{vib} is the vibrational frequency of a molecule. The vibrational frequency of molecules is in the range of middle infrared (IR) frequencies (6×10^{12} – 1.2×10^{14} Hz). The ground state corresponds to $v = 0$ with corresponding vibrational energy of $\frac{1}{2}h\nu_{vib}$; the first excited level ($v=1$) should have vibrational energy of $\frac{3}{2}h\nu_{vib}$; and so on.

Commonly, the vibrational spectroscopy covers a wavenumber range from 200 to 4000 cm^{-1} . We should know that crystalline solids also generate *lattice vibrations* in addition to molecular vibrations. The lattice vibrations refer to the vibrations of all the atoms in crystal lattice in a synchronized way. Such vibrations exhibit lower frequencies compared with those of common molecular vibrations and have a wavenumber range of about 20–300 cm^{-1} . Coupling between lattice and molecular vibrations can occur if the molecular vibrations lie in such a low wavenumber range. Molecular vibrations can be distinguished from the lattice vibrations because they are not as sensitive to temperature change as lattice vibrations.

9.1.3 Principles of Vibrational Spectroscopy

Infrared Absorption

Infrared spectroscopy is based on the phenomenon of infrared absorption by molecular vibrations. When a molecule is irradiated by electromagnetic waves within the infrared frequency range, one particular frequency may match the vibrational frequency of the molecule (ν_{vib}). Consequently, the molecular vibration will be excited by waves with the frequency $\nu_{ph} = \nu_{vib}$. The excitation means that the energy of molecular vibration will increase, normally by $\Delta\nu = +1$, as shown in Equation 9.7. In the meantime, the electromagnetic radiations with the specific frequency ν_{ph} will be absorbed by the molecule because the photon energy is transferred to excite molecular vibrations. The fundamental transition from $\nu = 0$ to $\nu = 1$ dominates the infrared absorption, although other transitions may be possible.

Figure 9.4 illustrates an example of the HCl diatomic molecule with $\nu_{vib} = 8.67 \times 10^{13}$ Hz. When it is excited by this frequency of electromagnetic radiation, the intensity of radiation at that frequency (8.67×10^{13} Hz) will be reduced (absorbed) in the infrared spectrum while the molecule itself will be moved to a higher vibrational energy level. The absorption intensity depends on how effectively the infrared photon energy can be transferred to the molecule. The effectiveness of infrared absorption is discussed in detail in Section 9.3 on Raman spectroscopy. Figure 9.5 shows an example of an infrared spectrum in which the intensity of transmitted infrared radiation is plotted over a range of radiation wavenumbers. In the figure, an individual deep valley represents a single vibration band that corresponds to a certain molecular vibration frequency.

Raman Scattering

Raman spectroscopy is based on the *Raman scattering* phenomenon of electromagnetic radiation by molecules. When irradiating materials with electromagnetic radiation of single frequency, the light will be scattered by molecules both elastically and inelastically. Elastic scattering means that the scattered light has the same frequency as that of the radiation. Inelastic scattering means that the scattered light has a different frequency from that of the radiation. Elastic scattering is called *Rayleigh scattering* while inelastic scattering is called Raman scattering.

Image Not Available In The Electronic Edition

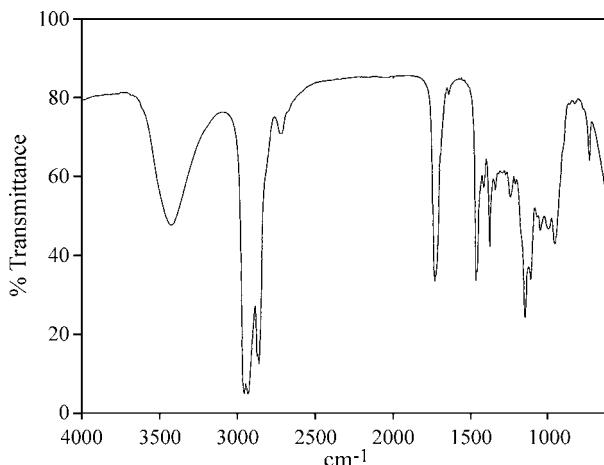


Figure 9.5 Typical IR absorption spectrum of hexanal. (Reproduced from P. Hendra, P.C. Jones, and G. Warnes, *Fourier Transform Raman Spectroscopy: Instrumentation and Chemical Applications*, Ellis Horwood, Chichester, 1991.)

Both elastic and inelastic scattering can be understood in terms of energy transfer between photons and molecules as illustrated in Figure 9.6. As mentioned before, photons of electromagnetic radiation can excite molecular vibration to a higher level when $\nu_{ph} = \nu_{vib}$. If the excited vibration returns to its initial level (Figure 9.6a), there is no net energy transfer from photons to molecular vibration. Thus, the scattered photons from molecules will have the same frequency as the incident radiation, similar to elastic collisions between photons and molecules. If the

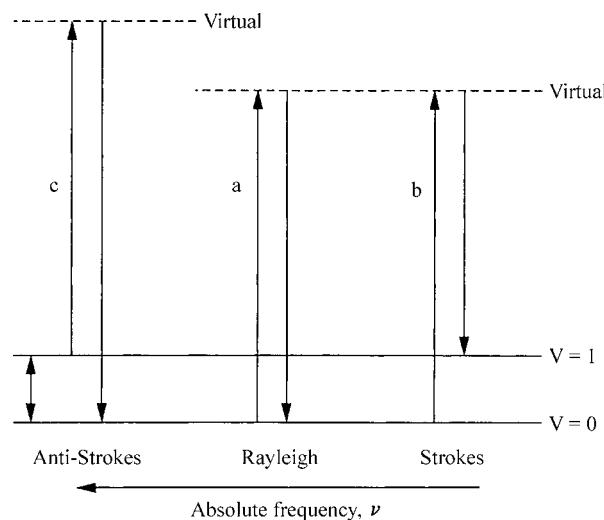


Figure 9.6 Elastic and inelastic scattering of incident light by molecules. Rayleigh, elastic scattering; Stokes and anti-Stokes, inelastic scattering. (Reproduced from P. Hendra, P.C. Jones, and G. Warnes, *Fourier Transform Raman Spectroscopy: Instrumentation and Chemical Applications*, Ellis Horwood, Chichester, 1991.)

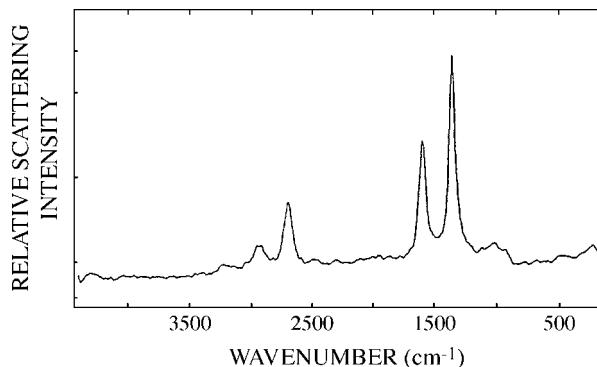


Figure 9.7 Typical Raman spectrum of polycrystalline graphite. (Reproduced with permission from G. Turrell and J. Corset, *Raman Microscopy, Developments and Applications*, Academic Press, Harcourt Brace & Company, London. © 1996 Elsevier B.V.)

excited molecular vibration does not return to its initial level, then scattered photons could have either lower energy or higher energy than the incident photons, similar to inelastic collision between photons and molecules. Understandably, the energy change corresponds to the energy difference between initial and final levels of molecular vibration energy. The energy changes in scattered photons are expressed as their frequency changes. If the photon frequency decreases from ν_{ph} to $(\nu_{ph} - \nu_{vib})$, the final energy level of molecular vibration is higher than the initial energy (Figure 9.6b). This is called *Stokes scattering*. If the photon frequency increases from ν_{ph} to $(\nu_{ph} + \nu_{vib})$, the final energy level is lower than the initial energy (Figure 9.6c). This is called *anti-Stokes scattering*. The intensity of anti-Stokes scattering is significantly lower than that of Stokes scattering. Thus, a Raman spectrum commonly records the frequency changes caused by the Stokes scattering by molecules. Such frequency change (difference between radiation and scattering frequency) is called the *Raman shift* in the spectrum, which should be in the same range as the infrared absorption spectrum. Figure 9.7 shows an example of the Raman spectrum in which the intensity of the Raman shift is plotted over a range of wavenumbers. An individual band of Raman shift corresponds to a molecular vibration frequency. Thus, both infrared and Raman spectra are plotted on the same scale of wavenumber.

9.1.4 Normal Mode of Molecular Vibrations

Understanding vibrational spectroscopy also requires basic knowledge of the vibrational behavior of molecules. The vibrations of nuclei in a molecule can be characterized with properties of the *normal mode* vibration. The normal mode vibration of molecules has the following characteristics:

- Nuclei in a molecule vibrate at the same frequency and in the same phase (passing their equilibrium positions at the same time); and
- Nuclear motion does not cause rigid body movement (translation) or rotation of molecules.

The normal mode of molecular vibrations ensures that a molecule vibrates at its equilibrium position without shifting its center of gravity. Each type of molecule has a defined number of vibration modes and each mode has its own frequency. We can examine the

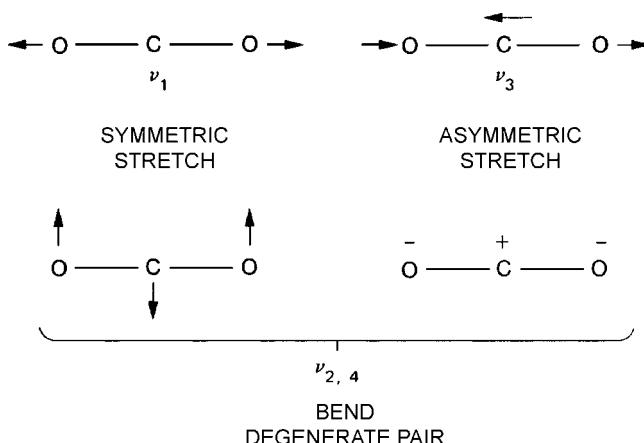


Figure 9.8 Normal modes of vibrations in a CO_2 molecule. (Reproduced from P. Hendra, P.C. Jones, and G. Warnes, *Fourier Transform Raman Spectroscopy: Instrumentation and Chemical Applications*, Ellis Horwood, Chichester, 1991.)

normal modes of molecular vibration by using simple molecule models. The simplest model of molecular vibrations is the diatomic model, which has only one stretching mode as shown in Figure 9.2. For polyatomic molecules, the vibrations are more complicated. For example, a linear molecule of three atoms connected by two bonds has a total of four vibration modes as illustrated in Figure 9.8. These vibration modes include two stretching modes and two bending modes. One stretching mode is *symmetric* (ν_1 in Figure 9.8) because the two bonds are stretched simultaneously. Another stretching mode is *asymmetric* (ν_3 in Figure 9.8) because one bond is stretched while the other is compressed. A bending mode changes the angle between two bonds, but not the bond lengths. The two bending modes are notated as ν_2 and ν_4 in Figure 9.8. One is bending in the plane of the figure and the other is in the horizontal plane.

Two features of bending modes can be revealed from the molecule model in Figure 9.8. First, the bending vibration can occur in an infinite number of planes that are in neither the figure plane nor the horizontal plane. However, we can ‘decompose’ any bending mode and express it in terms of ν_2 and ν_4 , as for decomposing a vector. Second, the two bending modes in Figure 9.8 are indistinguishable in terms of frequency, because the relative movements of nuclei in the two bending modes are physically identical. These two bending vibrations with the same physical identity are the *degenerate vibrations*.

Normal vibration modes of a triatomic molecule with a bending angle are illustrated in Figure 9.9. There are three normal modes in this molecule: symmetric stretching, asymmetric stretching, and bending in the plane of the figure. Bending in the horizontal plane, similar to Figure 9.8, will generate rotation of the molecule, which is not considered as a vibration.

Number of Normal Vibration Modes

It is necessary for us to know how to count the number of normal vibration modes in a molecule. The total number of vibration modes relates to the degrees of freedom in molecular motion. For N atomic nuclei in a molecule, there are $3N$ degrees of freedom because each nucleus can move in x , y or z directions in a three-dimensional space. Among the $3N$ degrees of freedom, however, there are three related to translation of a molecule along x , y or z direction

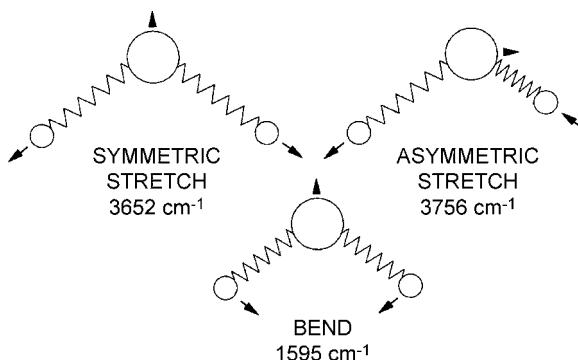


Figure 9.9 Normal modes of vibrations in a H_2O molecule. (Reproduced from P. Hendra, P.C. Jones, and G. Warnes, *Fourier Transform Raman Spectroscopy: Instrumentation and Chemical Applications*, Ellis Horwood, Chichester, 1991.)

as a rigid body without stretching any bond connecting nuclei or changing any bond angle. Also, there are three degrees of freedom related to rotate a molecule around the x , y or z axes as a rigid body. The molecular vibrations are related only to internal degrees of freedom in a molecule without translation and rotation of the rigid body. Thus, the total vibration modes of an N -atomic molecule should be $3N - 6$. For a linear molecule, the rotation around the bond axis is meaningless, considering the nuclei as points in space. Thus, there are only two rotational degrees of freedom for the molecule, and the total vibration modes should be $3N - 5$ for a linear molecule. For example, the linear molecule of CO_2 should have $9 - 5 = 4$ vibration modes as shown in Figure 9.8.

Classification of Normal Vibration Modes

It is possible to categorize the various modes of molecular vibration modes into four types:

1. Stretching vibration (ν);
2. In-planar bending vibration (δ);
3. Out-of-planar bending vibration (γ); and
4. Torsion vibration (τ).

Table 9.1 summarizes the features of the four vibration types. The last three types only change the bond angles, but do not stretch bonds. The torsion vibration can be considered as a special type of bending vibration as illustrated in Table 9.1. Thus, we can simply say that a vibration is either the stretching or the bending type. γ and τ only occur in molecules that have at least three bonds. Generally, the vibration frequency varies with vibration type. The frequency decreases with increasing number of bonds involved. Special types of vibrations may occur in a larger molecule like in a polymer; however, they all can be derived from these four basic types.

9.1.5 Infrared and Raman Activity

Among the total number of normal vibration modes in a molecule, only some can be detected by infrared spectroscopy. Such vibration modes are referred to as *infrared active*. Similarly, the

Table 9.1 Basic Types of Molecular Vibrations

Type	Stretching vibrations	Bending vibrations		Torsion vibrations
		In-plane	Out-of-plane	
Schematic				
Symbol	ν	δ	γ	τ
minimum number of bonds	1	2	3	3

(Data reproduced from A. Fadini and F-M. Schnepel, *Vibrational Spectroscopy: Methods and Applications*, Ellis Horwood, Chichester, 1989.)

vibration modes that can be detected by Raman spectroscopy are referred to as *Raman active*. Consequently, only active modes can be seen as vibration bands in their respective infrared or Raman spectra. For example, Figure 9.10 shows the infrared and Raman spectra of chloroform molecules. The infrared and Raman spectra complement each other as shown in Figure 9.10. The principles of infrared and Raman activity are complicated; thorough understanding requires *group theory* of mathematics. Nevertheless, it is possible to understand the basic ideas without group theory as presented in the following sections.

Infrared Activity

To be infrared active, a vibration mode must cause alternation of dipole moment in a molecule. A molecule has a center of positive charge and a center of negative charge. If these two centers

Image Not Available In The Electronic Edition

are separated by a distance (l), the dipole moment (μ) is defined.

$$\mu = el \quad (9.8)$$

e represents the amount of electrical charge at the charge center of the molecule. A molecule may exhibit a permanent dipole moment with separated positive and negative centers, or zero dipole moment with superimposed centers. The infrared activity requires changes in the magnitude of normal vibration during the vibration. The magnitude is commonly represented with a parameter q (which is equivalent to $|\Delta r|$ in Equation 9.5).

Mathematically, the requirement of infrared activity is expressed that the derivative of dipole moment with respective to the vibration at the equilibrium position is not zero.

$$\left(\frac{\partial \mu}{\partial q} \right)_{q=0} \neq 0 \quad (9.9)$$

It does not matter whether the molecule has permanent dipole moment, because the dipole moment can be induced by the electric field of an electromagnetic wave. For example, if plotting the changes in dipole moment versus vibration modes of CO₂ as in Figure 9.11a, we find that vibration modes v_2 , v_3 , and v_4 are infrared active, but not v_1 (see Figure 9.8 for the CO₂ vibration modes).

We can understand more about the infrared activity from Figure 9.12. The HCl molecule (Figure 9.12a) has a permanent dipole. Its stretching vibration is infrared active because the dipole distance changes during stretching vibration. The H₂ molecule (Figure 9.12b) has a center of symmetry and a zero dipole moment. It is infrared inactive, because the centers of positive and negative charge always remain as the center of symmetry and the dipole moment is always equal to zero during vibration. The CO₂ molecule also has a center of symmetry and zero dipole moment at equilibrium (Figure 9.12g). Its symmetric stretching is infrared inactive, similar to the case of H₂ (Figure 9.12c). However, its asymmetric stretching (Figure 9.12d) cannot retain the center of symmetry and change of dipole moment from 0 to

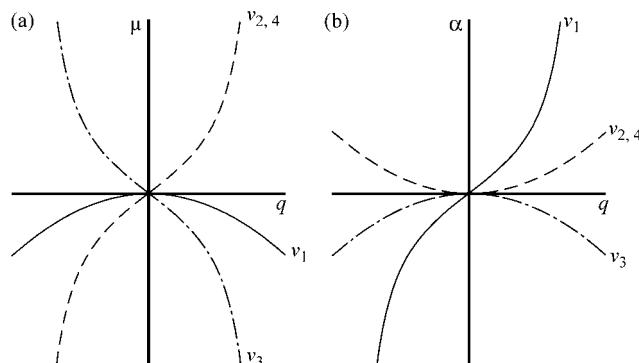


Figure 9.11 CO₂ molecule vibrations: (a) dipole moment (μ) as a function of vibration displacement (q); and (b) polarizability (α) as a function of vibration displacement (q). (Reproduced from P. Hendra, P.C. Jones, and G. Warnes, *Fourier Transform Raman Spectroscopy: Instrumentation and Chemical Applications*, Ellis Horwood, Chichester, 1991.)

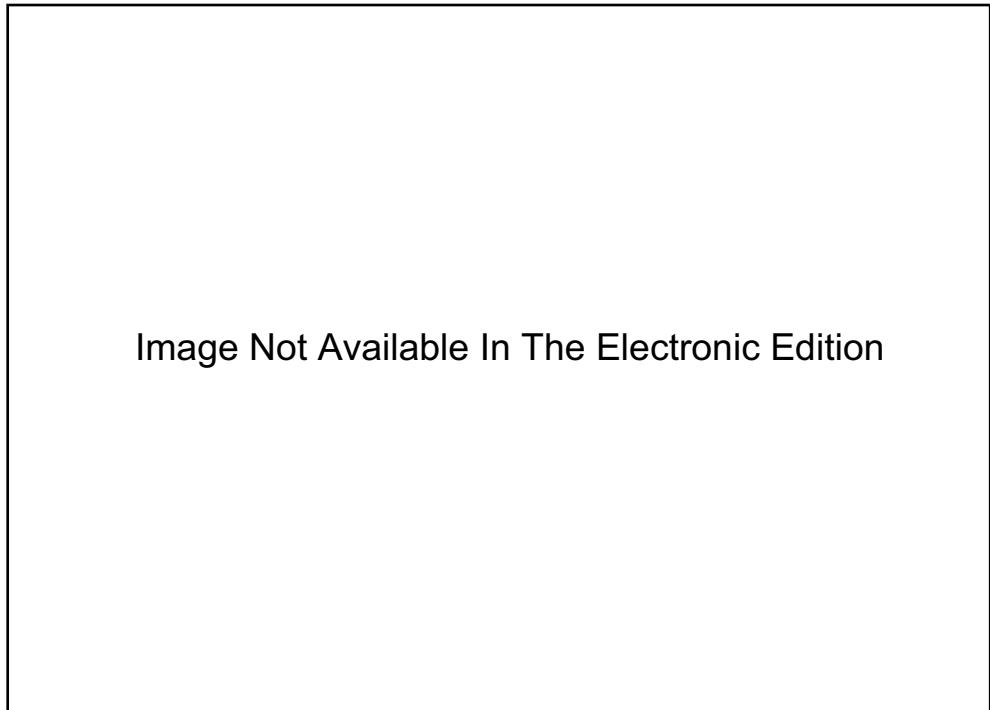


Image Not Available In The Electronic Edition

a certain value. Similarly, the bending vibration (Figure 9.12e) also induces a non-zero dipole moment during vibration. Thus, these two vibration modes are infrared active. The principle of vibration activity also applies to ion groups. For example, symmetric stretching of $(CO_3)^{2-}$ (Figure 9.12i) is infrared inactive because the centers of negative charge from three negatively charged oxygen ions always coincide with the center of positive charge from carbon. However, asymmetric stretching of $(CO_3)^{2-}$ (Figure 9.12h) is infrared active. The bending vibrations of $(CO_3)^{2-}$, which are not shown in Figure 9.12, should also be infrared active.

Raman Activity

To be Raman active, a vibration mode must cause *polarizability* changes in a molecule. When a molecule is placed in an electric field, it generates an induced dipole because its positively charged nuclei are attracted toward the negative pole of the field and its electrons are attracted toward the positive pole of the field. Polarizability (α) is a measure of the capability of inducing a dipole moment (μ) by an electric field. It is defined in the following equation.

$$\mu = \alpha E \quad (9.10)$$

E is the strength of electric field. Polarizability determines the deformability of the electron cloud of a molecule by an external electric field, similar to the compliance of elastic deformation. Mathematically, Raman activity requires that the first derivative of polarizability with

respect to vibration at the equilibrium position is not zero.

$$\left(\frac{\partial \alpha}{\partial q} \right)_{q=0} \neq 0 \quad (9.11)$$

Again, using CO₂ as the example, Figure 9.11b shows the change of α with respect to q and shows us that only ν_1 is Raman active because $\frac{\partial \alpha}{\partial q}$ at $q = 0$ is not zero. Polarization of molecules varies with directions in a three-dimensional space. For example, in a linear molecule such as CO₂, the electron cloud has the shape of an elongated watermelon. Such an electron cloud is more deformable along its long axis than the directions perpendicular to the long axis. The Raman activity can be illustrated using the changes of the *polarizability ellipsoid* that represents the polarizability variation in three-dimensional space. The ellipsoid has the dimension of $\alpha_i^{-1/2}$ (the distance between the ellipsoid origin to ellipsoid surface in direction i) as shown in Figure 9.13 and Figure 9.14. Generally, if the ellipsoid changes its size, shape and orientation with vibration, that vibration is Raman active.

Figure 9.13 illustrates the change of the polarizability ellipsoid in normal mode vibrations of the H₂O molecule. Its ν_1 , ν_2 and ν_3 modes are all Raman active because the ellipsoid changes size, shape or orientation as shown in Figure 9.13. We should interpret the ellipsoid change with caution. Some modes with apparent ellipsoid change with vibrations are not Raman active. For example, the change of the polarizability ellipsoid in CO₂ vibrations (Figure 9.14) does not mean Raman active in all modes. The symmetric stretching mode, ν_1 , of CO₂ is Raman active because the ellipsoid size changes during vibration. However, ν_2 and ν_3 of CO₂ are

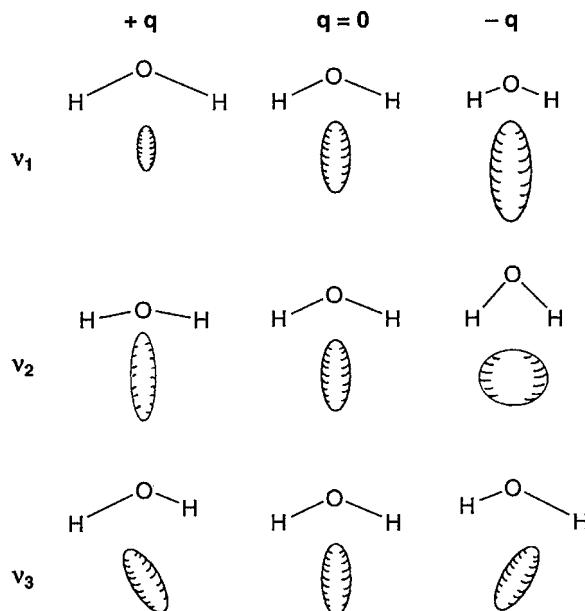


Figure 9.13 Normal modes of H₂O vibrations and changes in polarizability ellipsoids. ν_1 , ν_2 and ν_3 are all Raman active. (Reproduced with permission from J.R. Ferraro, K. Nakamoto, and C.W. Brown, *Introductory Raman Spectroscopy*, 2nd ed., Academic Press, San Diego. © 2003 Elsevier B.V.)

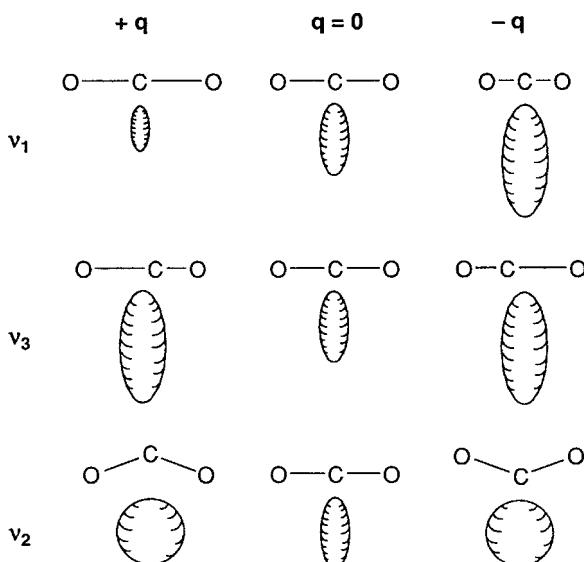


Figure 9.14 Normal modes of CO_2 vibrations and changes in polarizability ellipsoids. v_1 is Raman-active but v_3 , v_2 and v_4 are not. (Reproduced with permission from J.R. Ferraro, K. Nakamoto, and C.W. Brown, *Introductory Raman Spectroscopy*, 2nd ed., Academic Press, San Diego. © 2003 Elsevier B.V.)

not Raman active, even though the ellipsoid changes during vibrations. The problem is that ellipsoid change is symmetric in $+q$ and $-q$. Such displacement symmetry with respect to the equilibrium position makes the following expression zero.

$$\left(\frac{\partial \alpha}{\partial q} \right)_{q=0} = 0 \quad (9.12)$$

Figure 9.11b clearly illustrates such a situation of v_2 , v_3 and v_4 . The example of acetylene may further help our understanding. Figure 9.15 shows the two extreme positions of each vibration: 1, 2, 3, 4, and 5 in the linear acetylene molecule ($\text{H}-\text{C}\equiv\text{C}-\text{H}$). The vibrations 1, 2 and 3 are Raman active, but not 4 and 5. In vibrations 1 and 2, the molecular shape is different in the two extreme positions of vibrations. Thus, the polarizability ellipsoid will change its size or shape, similar to the v_1 mode of CO_2 . In vibration 3, the shape is identical at its two extreme positions, but there is an orientation change in its ellipsoid. The ellipsoid orientation change results from the clockwise and counterclockwise rotation of $\text{C}\equiv\text{C}$ and $\text{C}-\text{H}$ bonds. In vibrations 4 and 5, the polarizability ellipsoid does not change its size and shape because their two extreme positions are identical. Vibration 5 does not cause the ellipsoid orientation to change either, because the clockwise rotation of one $\text{C}-\text{H}$ bond is compensated for by the counterclockwise rotation of another $\text{C}-\text{H}$ bond.

In summary, we can highlight the features of infrared- and Raman-active vibrations:

- A vibration can be one of three cases: infrared-active, Raman-active, or both infrared- and Raman-active; and
- In molecules that have a center of symmetry, infrared- and Raman-active vibrations are mutually exclusive.

Image Not Available In The Electronic Edition

Further, it is helpful for us to know that vibrations of ionic bonds are strong in infrared spectroscopy (for example O–H), and vibrations of covalent bonds are strong in Raman spectroscopy (for example C=C).

9.2 Fourier Transform Infrared Spectroscopy

Fourier transform infrared spectroscopy (FTIR) is the most widely used vibrational spectroscopic technique. FTIR is an infrared spectroscopy in which the Fourier transform method is used to obtain an infrared spectrum in a whole range of wavenumbers simultaneously. It differs from the dispersive method, which entails creating a spectrum by collecting signals at each wavenumber separately. Currently, FTIR has almost totally replaced the dispersive method because FTIR has a much higher signal-to-noise ratio than that of dispersive method.

9.2.1 Working Principles

The key component in the FTIR system is the *Michelson interferometer*, as schematically illustrated in Figure 9.16. The infrared radiation from a source enters the Michelson interferometer. The interferometer is composed of one beam-splitter and two mirrors. The beam-splitter transmits half of the infrared (IR) beam from the source and reflects the other half. The two split beams strike a fixed mirror and a moving mirror, respectively. After reflecting from the mirrors, the two split beams combine at the beam-splitter again in order to irradiate the sample before the beams are received by a detector.

The function of the moving mirror is to change the optical path lengths in order to generate light interference between the two split beams. If the moving mirror is located at the same distance from the beam splitter as the fixed mirror, the optical paths of the two split beams are the same; thus, there is *zero path difference*. An optical path difference (δ) will be introduced by translating the moving mirror away from the beam-splitter. Changing the optical path difference is similar to what happens in diffraction of crystallographic planes (Figure 2.6). The two split beams will show constructive and destructive interference periodically, with continuous change of δ value. There will be completely constructive interference when $\delta = n\lambda$, but completely destructive interference when $\delta = (\frac{1}{2} + n)\lambda$. A change of δ value is realized by a change in the position of the moving mirror as illustrated in Figure 9.16.

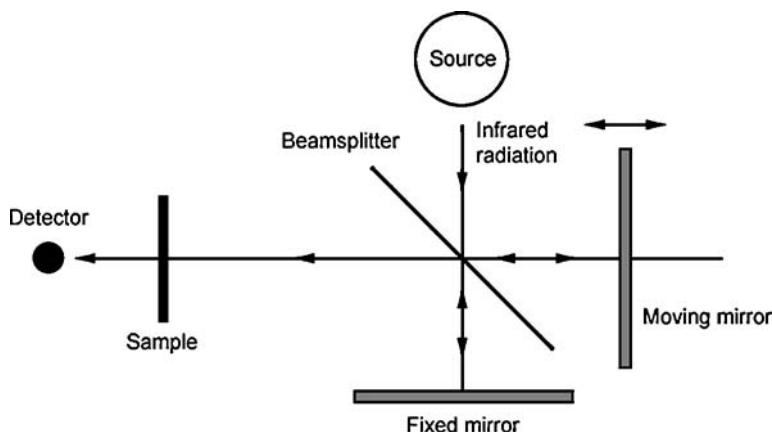


Figure 9.16 Optical diagram of a Michelson interferometer in FTIR.

A plot of light interference intensity as function of optical path difference is called an *interferogram*. Figure 9.17 illustrates interferograms of a radiation with two wavelengths λ and 3λ . The interferogram at the top is the sum of interferograms of two light waves with wavelength λ and 3λ . The IR radiation from the FTIR source is composed of numerous wavelengths. An interferogram of FTIR looks like that in Figure 9.18a. It has a sharp center burst that dies off quickly into two wings. The center burst corresponds to the position of the moving mirror generating zero path difference at which the interferogram has the maximum intensity because all the waves have completely constructive interference (in phase). An interferogram irradiating the sample is the sum of sinusoidal waves with a range of wavelengths. The FTIR detector receives interferogram signals which are transmitted to a sample (or reflected from a sample). The interferogram received by the detector is not an infrared spectrum.

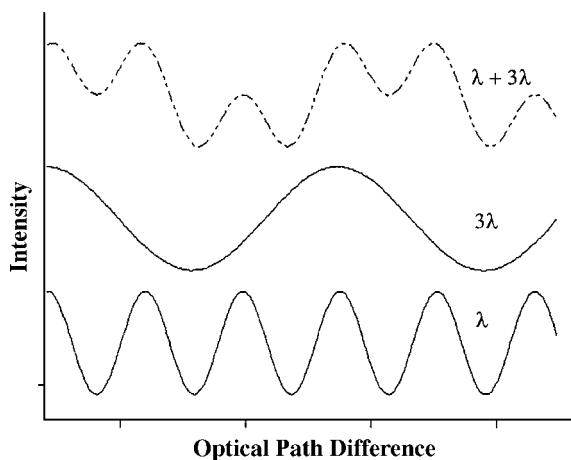


Figure 9.17 Interferograms: (a) sum of (b) and (c); (b) interferogram of light with wavelength 3λ ; and (c) interferogram of light with wavelength λ . (Reproduced with permission from B.C. Smith, *Fundamentals of Fourier Transform Infrared Spectroscopy*, CRC Press, Boca Raton. © 1996 CRC Press LLC.)

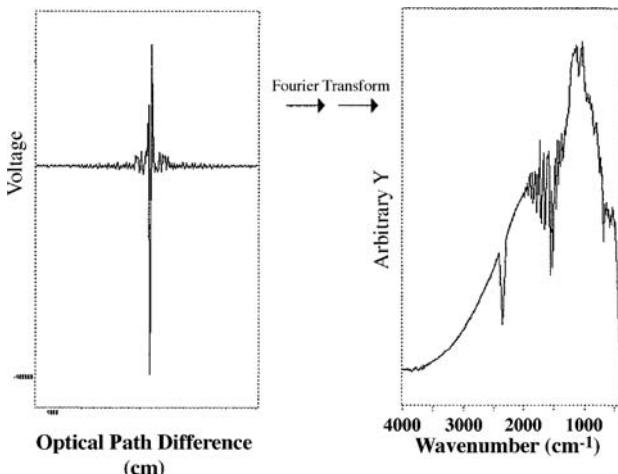


Figure 9.18 Plots of: (a) an interferogram; and (b) a Fourier transform from an interferogram to an IR spectrum. (Reproduced with permission from B.C. Smith, *Fundamentals of Fourier Transform Infrared Spectroscopy*, CRC Press, Boca Raton. © 1996 CRC Press LLC.)

Fourier transformation is necessary to convert an interferogram into an infrared spectrum, which is a plot of the light intensity versus wavenumber, as shown in Figure 9.18b. The Fourier transform is based on a fact that any mathematical function can be expressed as a sum of sinusoidal waves. All the information of wave intensity as a function of wavelength is included in the sum of sinusoidal waves. A computer equipped with FTIR constructs the infrared spectrum using a *fast Fourier transform* (FFT) algorithm which substantially reduces the computation time.

The Fourier transform as a mathematical tool has been mentioned in Chapter 3 when discussing the relationship between real space and reciprocal space. Here it is emphasized as a tool which transfers information between a function in the time (t) domain and its corresponding one in the frequency (ω) domain.

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \quad (9.13)$$

In an FTIR instrument, the Fourier transform converts the intensity versus optical path difference to the intensity versus wavenumber. The optical path difference can be considered to be in the time domain because it is obtained by multiplying time with the speed of a moving mirror. The wavenumber can be considered in the frequency domain because it is equal to frequency divided by the light speed.

9.2.2 Instrumentation

Infrared Light Source

The infrared light source generates wideband radiation by heating solid materials to incandescence using electric power. There are two commonly used IR sources: the *Nernst glower*, which is composed of mainly oxides of rare-earth elements; and the *Globar*, which is

A large rectangular box with a thin black border, centered on the page. Inside the box, the text "Image Not Available In The Electronic Edition" is printed in a black, sans-serif font.

composed of silicon carbide. The range of IR wavelength and energy distribution is highly temperature sensitive as shown in Figure 9.19. Most IR sources are operated at the temperature where the maximum energy of radiation is near the short wavelength limit of the IR spectrum.

Beam-Splitter

Beam-splitters should be made of material semi-transparent to infrared light. Beam-splitters should reflect one half portion of infrared light to the moving mirror while transmitting the rest infrared to a fixed mirror. The most common beam-splitter is a sandwich structure, with a thin layer of germanium (Ge) between two pieces of potassium bromide (KBr); it works well in the wavenumber range $4000\text{--}400\text{ cm}^{-1}$. Ge is able to split infrared light. KBr is a good substrate material because it is transparent to infrared light, yet has good mechanical strength. It functions as a protective coating for the Ge. The drawback of using KBr as a substrate is its hygroscopic nature. KBr tends to absorb water vapor from the atmosphere and fog readily forms on it. Thus, all FTIR instruments require low humidity or are sealed from atmosphere. The low humidity environment around the beam-splitter can be obtained by purging with dry air or nitrogen. More conveniently, the instrument is sealed such that the infrared beam passes through a window to reach the sample.

Infrared Detector

The infrared detector is a device to measure the energy of infrared light from the sample being examined. It functions as a transducer to convert infrared light signals to electric signals. There are two main types: the thermal detector and the semiconductor detector. The key component in a thermal detector is a pyroelectric crystal, of which the most commonly used type is *deuterated triglycine sulfate* (DTGS). Infrared radiation causes a temperature change in DTGS which, in turn, changes its dielectric constant. This dielectric constant affects its

capacitance and results in a voltage change across the DTGS. The DTGS detector operates in the wavenumber range $4000\text{--}400\text{ cm}^{-1}$. It is simple and inexpensive but less sensitive than the semiconductor detector. The most commonly used semiconductor detector is made of *mercury cadmium telluride* (MCT). An MCT detector absorbs infrared photons which in turn causes the electrons to migrate from the valence band to the conduction band of the semiconductor. The electrons in the conduction band generate electric current signals. The MCT detector is up to 10 times more sensitive than the DTGS type; however, it detects a narrower band of radiation ($4000\text{--}700\text{ cm}^{-1}$). The MCT detector needs to be cooled, commonly to liquid nitrogen temperature (-196°C), before operation. Another disadvantage of MCT is that it saturates easily with high intensity radiation. This can be a problem for quantitative analysis.

9.2.3 Fourier Transform Infrared Spectra

An infrared spectrum converted from an interferogram by Fourier transform is called a *single beam spectrum*. A single beam spectrum includes both spectra from the sample and background. The background spectrum contains only the information from the instrument and atmosphere, not from the sample being examined. The instrument contributions to background spectrum are from the detector, beam-splitter, mirror and the IR source. The atmospheric contributions are mainly from water vapor and carbon dioxide. Figure 9.20 demonstrates how

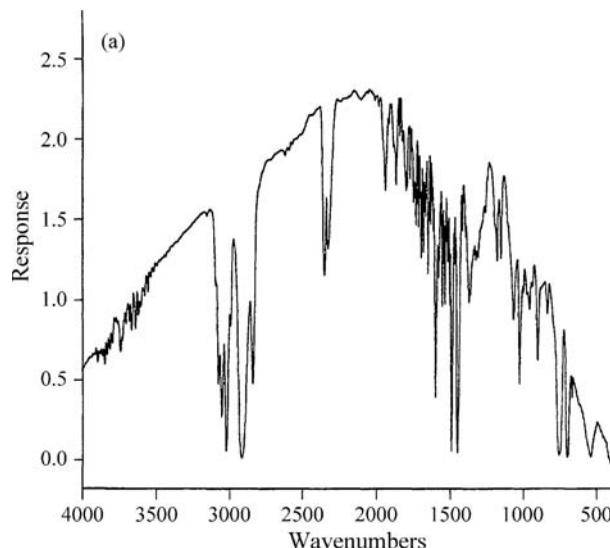


Figure 9.20 (a) Single beam FTIR spectrum of background: a plot of raw detector response versus wavenumber without sample; (b) a sample single beam FTIR spectrum of polystyrene (vibrational bands of polystyrene are superimposed on background spectrum); and (c) final FTIR spectrum of polystyrene that only contains the vibration bands from the polystyrene sample (absorption intensity is expressed as the transmittance). (Reproduced with permission from B.C. Smith, *Fundamentals of Fourier Transform Infrared Spectroscopy*, CRC Press, Boca Raton. © 1996 CRC Press LLC.)

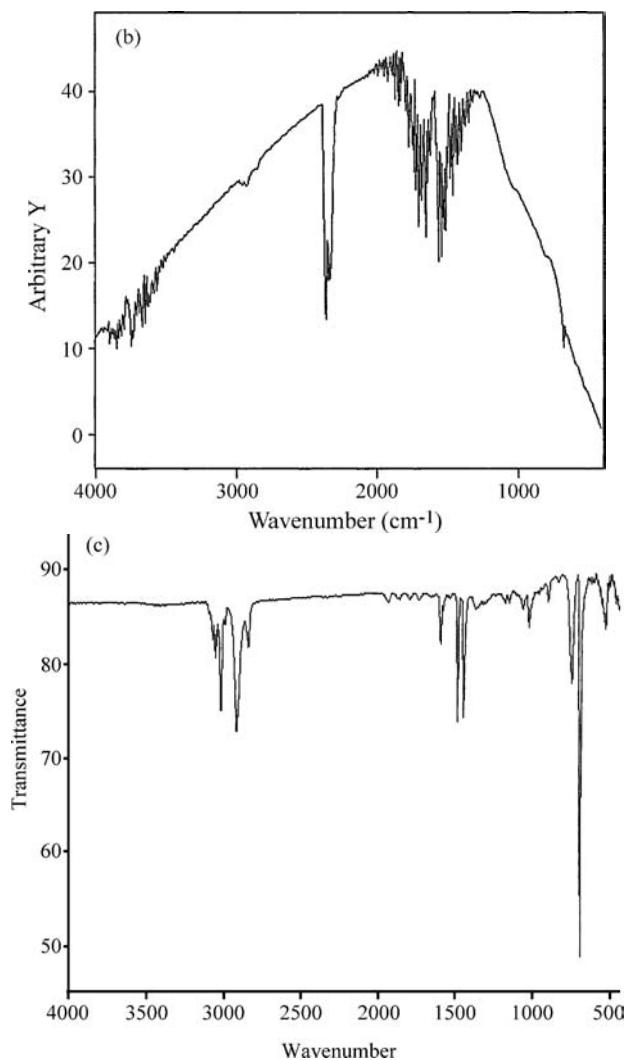


Figure 9.20 (Continued)

significant background can make to the IR spectrum. Figure 9.20a shows a simple beam spectrum of a polystyrene sample. Figure 9.20b shows a single beam spectrum of the chamber without the polystyrene sample, which is a background spectrum. In the background spectrum, typical water vapor bands at around 3500 and 1630 cm⁻¹ and CO₂ gas bands at 2350 and 667 cm⁻¹ are visible. To eliminate the background influence, the ratio of the single beam spectrum of a sample with the background spectrum should be made. This process results in a *transmittance spectrum* as shown in Figure 9.20c. *Transmittance* (*T*) is defined as the ratio of intensities.

$$T = I/I_o \quad (9.14)$$

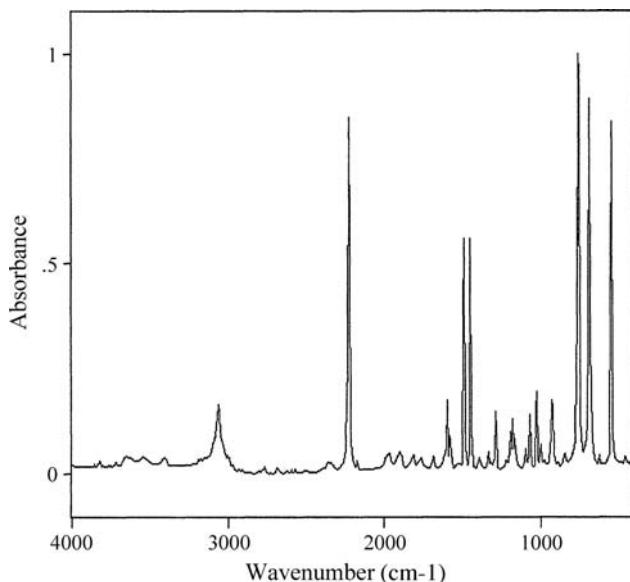


Figure 9.21 FTIR spectrum where the absorption intensity is expressed as absorbance. In this spectrum the absorbance scale is normalized to a range 0–1. (Reproduced with permission from B.C. Smith, *Fundamentals of Fourier Transform Infrared Spectroscopy*, CRC Press, Boca Raton. © 1996 CRC Press LLC.)

I is the intensity measured in a single beam spectrum of sample and I_o is the intensity measured in the background spectrum. Transmittance is often expressed as $\%T$, which forms the scale of the vertical axis in Figure 9.20c. The spectrum can also be presented as *absorbance* (A) versus wavenumber as shown in Figure 9.21. The absorbance is calculated from the transmittance.

$$A = -\log T \quad (9.15)$$

An FTIR spectrum is commonly expressed as a transmittance spectrum, in which vibration band peaks point downward. It can also be expressed as an absorbance spectrum, in which vibration band peaks point upward (Figure 9.21). For quantitative analysis, an absorbance spectrum should be used because the peaks of a transmittance spectrum are not linearly proportional to concentration.

9.2.4 Examination Techniques

Transmittance

The transmittance examination technique refers to the method of obtaining an infrared spectrum by passing an IR beam through a sample as illustrated in Figure 9.16. It is the most commonly used examination method in FTIR for two reasons. First, the transmittance technique generates high signal-to-noise ratios, and second, it is suitable for samples in any of solid, liquid or gas phases. Its main disadvantage is the thickness limitation of samples. A thick sample will absorb so much of the infrared radiation that detecting infrared transmission becomes impossible.

Generally, for transmission examination, the sample thickness should not be more than $20\text{ }\mu\text{m}$. On the other hand, a sample that is too thin ($<1\text{ }\mu\text{m}$) yields absorption bands too weak to be detected.

Solid Sample Preparation

Solid samples for transmittance examination can be in one of two forms: thin film or powder. Thin film samples are mainly polymeric materials. Casting films with polymer solutions is a commonly used method. Polymer films can also be made by mechanically pressing under elevated temperatures. Powder samples are made by grinding solid to powder and then diluting the powder with infrared-inert matrix materials. There are two typical methods for preparing powder samples: making KBr pellets and making *mulls*. As mentioned in the instrumentation section, KBr is transparent to infrared light. To make pellets, we grind a solid sample with KBr together to obtain powder particles with size less than $2\text{ }\mu\text{m}$ diameter. Then, the powder mixture is die-pressed to form a pellet. For the mull method, a solid sample is ground to a powder and then diluted with a mulling agent (usually mineral oil). The slurry of powder–oil is smeared on a KBr plate and sandwiched with another KBr plate. The mull method is simple and inexpensive; however, the mulling agent contains long and straight chains of hydrocarbons, which absorb strongly around 3000 and 1400 cm^{-1} , and hence it may complicate the sample spectrum.

Liquid and Gas Sample Preparation

Liquid and gas samples do not need much preparation, but special cells to contain the samples are often necessary. The simplest method to prepare a liquid sample is to make a capillary thin film of the liquid. The capillary thin film is made by placing a drop of liquid on a KBr plate and sandwiching it with another KBr plate. This method, however, is not suitable for volatile liquids. Liquid cells can be used for volatile liquid and toxic liquid samples, particularly for quantitative analysis. The spacing between the bottom and the top of liquid cell is typically from 1 to $100\text{ }\mu\text{m}$. The cell is made of an infrared-transparent material. Typically, KBr is used; however, KBr should not be selected as the material for holding samples containing water because water dissolves KBr. Instead, ZeSe or AgCl should be used because they are infrared-transparent but not water soluble. Cells for gas samples are structurally similar to cells for liquid but the dimension is much larger.

Reflectance

Reflectance examination techniques refer to methods for obtaining an infrared spectrum by reflecting IR radiation from a solid or liquid sample. The main advantage is that bulk and coating samples can be examined without destructive preparation. These techniques are especially attractive for solid samples that are difficult to grind into powder and for fast sample examination. However, reflectance techniques are less popular than transmission because of their following disadvantages. First, infrared radiation has limited penetration into the sample. A reflectance spectrum can only contain the infrared signals from the top 1 to $10\text{ }\mu\text{m}$ of a solid sample surface. This is a serious deficiency for a sample with possible chemical composition change near its surface, and for quantitative analysis for which the exact pass-length of infrared light in the sample must be known. Second, it is more difficult to capture reflected infrared light than transmitted light. A reflectance spectrum will show much more noise than the transmission spectrum for the same duration of examination. Third, reflectance

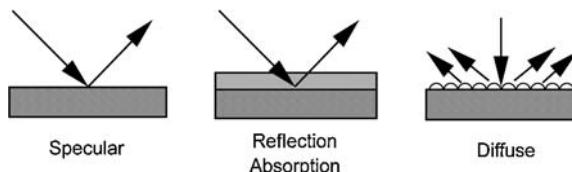


Figure 9.22 Reflectance types in FTIR spectroscopy.

techniques require special accessories, making the instrumentation more complicated and expensive.

There are three types of reflectance techniques: specular, diffuse and reflection-absorption as illustrated in Figure 9.22. Specular reflectance is applied to samples with smooth and polished surfaces, diffuse reflectance is applied to samples with rough surfaces, and reflection-absorption is applied to IR-transparent thin films on IR opaque substrates. The specular and diffuse techniques are more widely used and are introduced in more detail in the following text.

Specular reflectance occurs when an incident infrared beam strikes a smooth surface of a solid sample. The reflectance angle is the same as the incidence angle during specular reflectance. The optical arrangement of the specular reflectance accessory is schematically shown in Figure 9.23. It consists of two flat mirrors and a platform with a hole. The sample is placed over the hole in the platform. The background spectrum is obtained by placing an ‘ideal reflector’ of infrared light, such as a gold or aluminum mirror, over the hole. The optical arrangement is also applied to a reflection-absorption case where there is reflectance on a smooth substrate coated with a thin film. In such case, the infrared beam goes through a *double transmission* through the coating film. To obtain a background spectrum for examining the coating film, the smooth surface of substrate material is placed over the platform hole.

Diffuse reflectance occurs when an incident infrared beam strikes a rough surface; the surface could be solid, powder or even liquid. The reflectance angle spreads over a wide range during diffuse reflectance. The optical arrangement of accessories for measuring diffuse reflectance consists of a focusing mirror, two flat mirrors and a sample cup as shown in Figure 9.24. The focusing mirror collects infrared light reflected off the sample. The sample cup can hold either powder or liquid samples. Diffuse reflectance can be used to examine a sample of microgram mass. It can also be used to examine intractable objects such as a large piece of plastic. Another

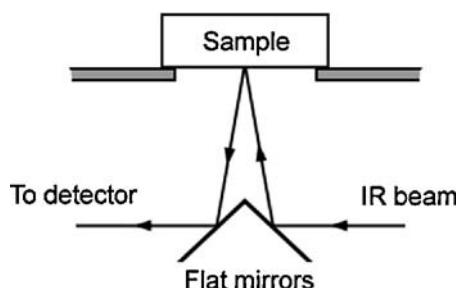


Figure 9.23 Optical diagram of a simple specular reflectance accessory for FTIR instrument.

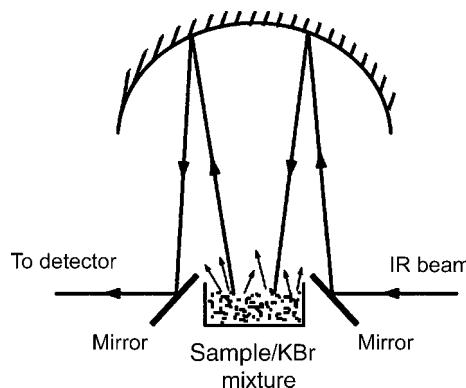


Figure 9.24 Optical diagram of a diffuse reflectance accessory for an FTIR instrument.

method for preparing samples for diffuse reflectance examination is to take a piece of abrasive paper with silicon carbide particles and rub the surface of the object to be examined. The abrasive paper will be covered with a certain amount of sample material. Then, we can place the abrasive paper in the sample cup for FTIR examination.

9.2.5 Fourier Transform Infrared Microspectroscopy

FTIR spectroscopy can combine with microscopy for generating FTIR spectra from microscopic volumes in materials. The instrument for FTIR microspectroscopy is simply called the *FTIR microscope*, which is often attached to the conventional FTIR instrument. The FTIR microscope is increasingly used for materials characterization because of its simple operation and FTIR spectra can be collected rapidly from microscopic volumes selected with the microscope.

Instrumentation

The FTIR microscope has an optical system that can easily switch between visible light observation and infrared light spectroscopy. Also, the microscope can be operated in either transmittance or reflectance modes in order to meet the sample conditions, either transparent or opaque to light. The microscope has two light sources: a visible light source and infrared light source. It also has an infrared detector and video camera or eyepieces, like a regular light microscope. However, there is only one system of condenser and objective lens.

Figure 9.25 illustrates the optical paths of the microscope in the modes for FTIR spectroscopy. Figure 9.25a shows the optical path of the infrared beam when the microscope operates in the transmittance mode, while Figure 9.25b shows the optical path in the reflectance mode. We should note that lenses used in FTIR differ from those used in conventional light microscope. The lenses used in FTIR microscopy are the *Cassegrain type*. The Cassegrain lens is composed of curved mirrors, not transparent glass pieces. Light passing the Cassegrain lens is reflected by a larger mirror with a parabolic shape and a smaller mirror with a hyperbolic shape for collection and focus of the beam, respectively. The Cassegrain lens is preferred for the reflecting lens in the FTIR microscope because it does not absorb the energy of infrared light as a conventional optical lens does. The lens does not possess

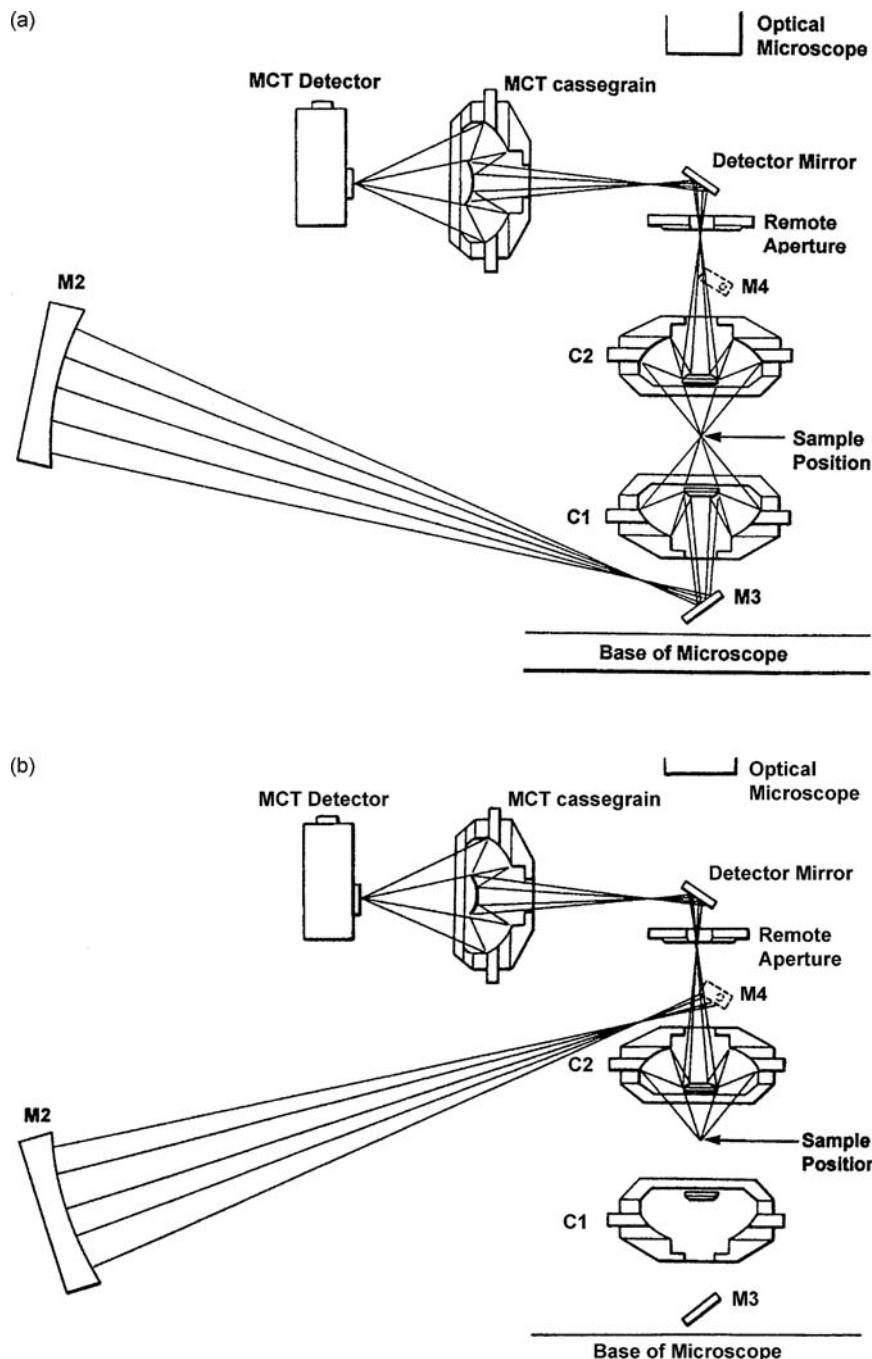


Figure 9.25 Optical paths of FTIR microscope with IR radiation: (a) transmittance; and (b) reflectance. M, mirror; C, Cassegrain lens. (Reproduced by permission of PerkinElmer Inc.)

chromatic aberration either. The power of the Cassegrain objective used in FTIR is relative low, usually less than $10\times$.

The selection of the microscopic area for FTIR microspectroscopy is achieved by a *remote aperture* located between the objective and detector. The remote aperture commonly has a rectangular opening with two pairs of knife-edged blades. The blades are often made from a material that is transparent to visible light but opaque to infrared light.

The infrared detector in the FTIR microscope must be highly efficient, such as the mercury cadmium telluride (MCT) detector. The IR radiation intensity from a microscopic area is significantly lower than from the sample in a conventional FTIR instrument. The MCT detector can ensure a sufficiently high signal-to-noise ratio for meaningful IR spectroscopy. The drawback of using the MCT detector is the requirement of low temperature operation at -196°C . It usually takes more than 8 hours to cool the MCT detector to the operating temperature using liquid nitrogen.

Applications

In principle, operation of the FTIR microspectroscopy is the same as for a conventional FTIR instrument except the spectrum is obtained from a microscopic area or intensity distribution is mapped in the sample plane. A spectrum from an area in the order of $10 \times 10 \mu\text{m}$ can be obtained. Mapping or FTIR imaging at micro-level resolution can be achieved by scanning a sample using a motorized sample stage. The resolution is primarily determined by the size of the focused IR beam and precision of motorized stage. Reflectance microspectroscopy is more widely used than the transmittance mode in FTIR microscopy because minimal sample preparation is required.

Figure 9.26 and Figure 9.27 illustrate an example of using the FTIR microscope to identify a micro-sized particle. A contaminant particle is isolated, as shown in the micrograph (Figure 9.26). The twisted fiber is cotton. The particle attached to the fiber was examined with reflectance FTIR microspectroscopy. Figure 9.27 shows the IR spectrum of the particle and

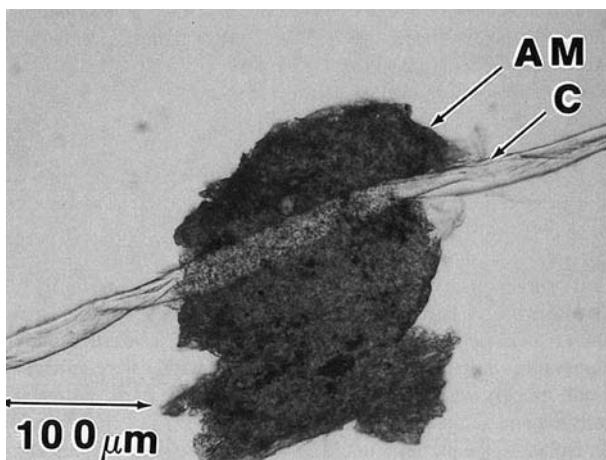


Figure 9.26 Micrograph of isolated particulate contamination (AM) and cotton fiber (C). (Reproduced with permission from H.J. Jumecki, *Practical Guide to Infrared Microspectroscopy*, Marcel Dekker, New York. © 1995 Taylor & Francis Group Ltd.)

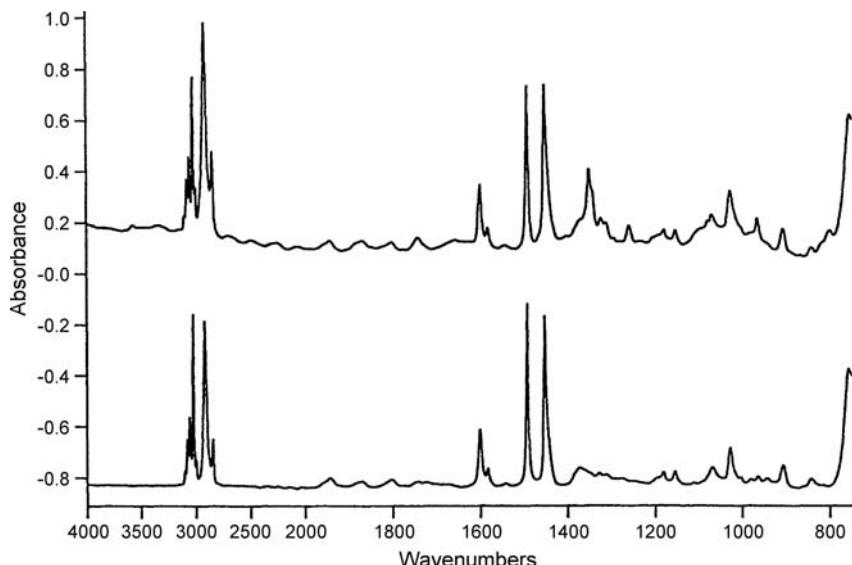


Figure 9.27 IR spectrum of the isolated particle (lower) and IR spectrum of polystyrene (upper). (Reproduced with permission from H.J. Jumecki, *Practical Guide to Infrared Microspectroscopy*, Marcel Dekker, New York. © 1995 Taylor & Francis Group Ltd.)

its identification as polystyrene by comparing the experimental spectrum with the standard polystyrene spectrum.

9.3 Raman Microscopy

Raman microscopes are more commonly used for materials characterization than other Raman instruments. Raman microscopes are able to examine microscopic areas of materials by focusing the laser beam down to the micrometer level without much sample preparation as long as a surface of the sample is free from contamination. This technique should be referred to as Raman microspectroscopy because Raman microscopy is not mainly used for imaging purposes, similar to FTIR microspectroscopy. An important difference between Raman micro- and FTIR microspectroscopies is their spatial resolution. The spatial resolution of the Raman microscope is at least one order of magnitude higher than the FTIR microscope.

Raman microspectroscopy (often called micro-Raman), like most Raman spectrometry, is of the dispersive type. It requires collecting a spectrum at each wavenumber separately, not like the FTIR type that collects a spectrum in a range of wavenumbers simultaneously. Although this chapter only describes the instrumentation for Raman microscopy, its working principles and spectra are basically the same as those of conventional dispersive Raman instruments, which consist of the following elements:

- Laser source;
- Sample illumination and collection system;
- Spectral analyzer; and
- Detection and computer control and processing system.

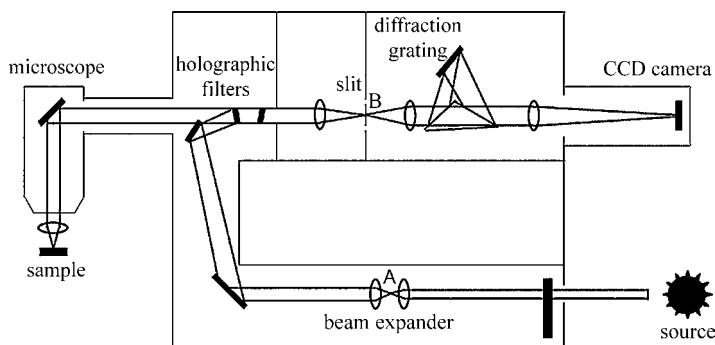


Figure 9.28 Optical diagram of a Raman microscope.

Raman spectroscopy requires highly monochromatic light, which can be provided only by a laser source. The laser source is commonly a continuous-wave laser, not a pulsed laser. The laser source generates laser beams with the wavelengths in the visible light range or close to the range. In a Raman microscope, sample illumination and collection are accomplished in the microscope. The microscope's optical system enables us to obtain a Raman spectrum from a microscopic area: this is the main difference between the micro-Raman and conventional Raman spectrometers.

9.3.1 Instrumentation

The optical arrangement of a Raman microscopic system is schematically illustrated in Figure 9.28. A laser beam passes through a filter to become a single wavelength beam, which is then focused on a sample surface by the microscope. The Raman scattered light reflected from a microscopic area of sample is collected by the microscope and sent to the detector. The Raman scattered light, which results from inelastic scattering, is weak compared with the incident laser light. Thus, a *holographic filter* has to be used in order to block the laser light entering the detector system. The wavelength of Raman scattering light is selected by a *diffraction grating* system before being recorded by a detector. More details of the optical components in a Raman microscopic system are presented below.

Laser Source

Commonly used laser sources are gas continuous-wave lasers such as Ar^+ , Kr^+ and He–Ne. Such laser sources are often capable of generating beams of multiple wavelengths. For example, Ar^+ generates a range of wavelengths with different intensities. The highest intensity wavelengths include 515.4, 488.0 and 350.7 nm. Thus, it is necessary to filter out wavelengths other than 515.4 nm when an Ar^+ laser is used. The He–Ne laser, however, generates beams of only one wavelength at 632.8 nm. Gas laser sources generate several tens of mW of laser power, but the laser power reaching the microscopic area of the sample is only about 5 mW.

Microscope System

The microscope in the Raman system is different from conventional microscopes used for observation of microstructure in two aspects: the microscope only needs to illuminate a

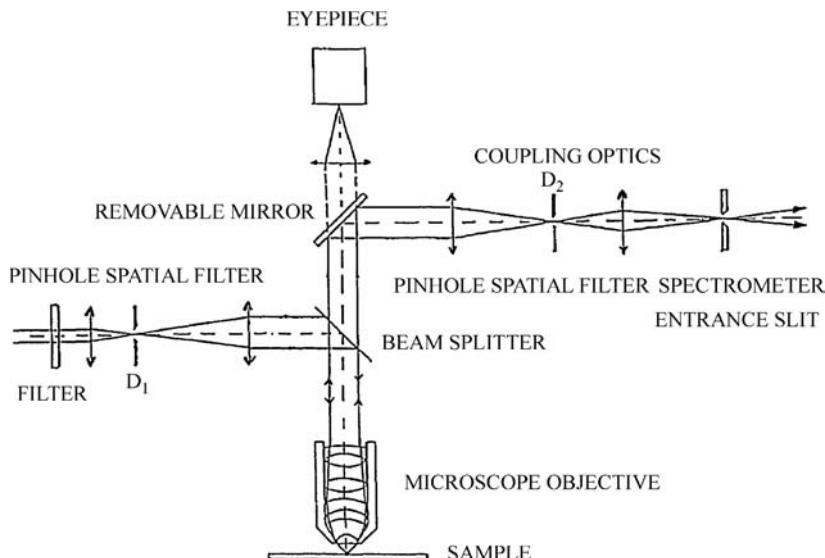


Figure 9.29 Optical diagram of a Raman microscope. A pinhole spatial filter consists of a pinhole confocal diaphragm (D₁ and D₂). (Reproduced with permission from G. Turrell and J. Corset, *Raman Microscopy, Developments and Applications*, Academic Press, Harcourt Brace & Company, London. © 1996 Elsevier B.V.)

microscopic area, not the entire field, and the microscope must have a high numerical aperture (*NA*) in order to collect the Raman-scattered light over a large solid angle effectively. The schematic layout in Figure 9.29 illustrates the optical features of the Raman microscope. The beam from the laser source is first filtered to obtain a monochromatic wavelength. Then, a pinhole spatial filter removes the appearance of the diffraction rings and speckle noise from around the focused spot in order to obtain a clean point laser beam to illuminate the sample. This pinhole spatial filter (illustrated around pinhole D₁ in Figure 9.29) is located at A in Figure 9.28. The clean laser beam is reflected by a beam-splitter and goes through an objective lens to illuminate the sample.

The Raman-scattered light is collected by a wide-aperture objective lens and focused in an adjustable pinhole D₂ placed in the image plane of the microscope. The second pinhole spatial filter around D₂ is located at B in Figure 9.28. The pinholes D₁ and D₂ are called confocal diaphragms because they are in exact optical conjugation positions with respect to the point source in the object plane, similar to the confocal aperture in confocal microscope introduced in Chapter 1. This confocal arrangement ensures that only light originating from the microscopic area of the sample is transmitted to the spectral analyzer and detector. These two pinhole spatial filters are important for illumination and Raman light collection. They multiply and increase spatial resolution by eliminating stray light coming from the out-of-focus region of the sample. By adjusting the pinhole diaphragm D₂, the spatial resolution of the sample can reach 1 μm when a 100 \times objective lens is used.

Prefilters

The scattered light from the microscope must be passed through special filters before reaching the spectral analyzer in order to remove elastically scattered light. The Raman light cannot

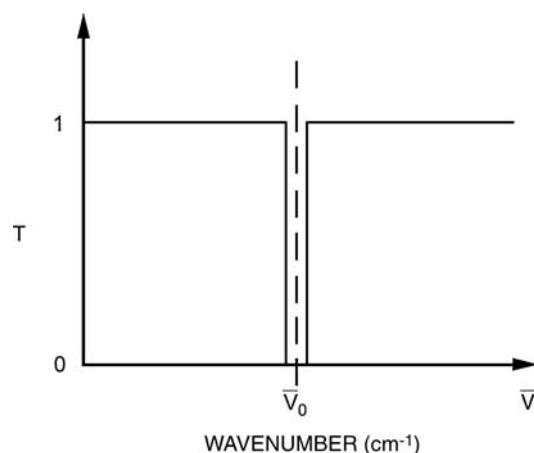


Figure 9.30 Spectral response of an ideal notch prefilter in micro-Raman spectroscopy to remove the elastically scattered light with wavenumber \bar{v}_o . (Reproduced with permission from G. Turrell and J. Corset, *Raman Microscopy, Developments and Applications*, Academic Press, Harcourt Brace & Company, London. © 1996 Elsevier B.V.)

be seen without the filters, because the elastically scattered light has much higher intensity than the inelastically scattered Raman light. The filter should ideally have a ‘notch’ feature, as shown in Figure 9.30. The notch is zero transmission in a narrow wavenumber range centered on a exciting laser wavenumber. As illustrated in Figure 9.28 two holographic filters used in the Raman microscope serve such a purpose.

Diffraction Grating

The key component of the spectral analyzer in the Raman microscope is the diffraction grating. It disperses Raman scattered light according to its wavenumbers. The surface of the diffraction grating contains fine parallel grooves that are equally spaced as illustrated in Figure 9.31. When Raman scattered light is incident on the diffraction grating, the grating disperses the light by diffracting it in a discrete direction. The light dispersion is based on Bragg’s Law

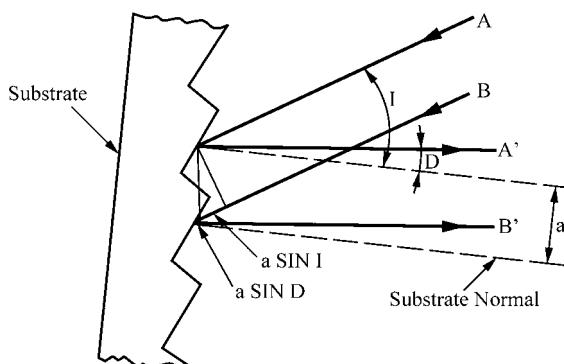


Figure 9.31 Principle of optical grating diffraction.

of diffraction angle and incident wavelength. The diffraction grating plays a role similar to that of equal-spaced atomic planes in crystals. It diffracts the different wavelengths in discrete angles (Figure 9.31). One diffraction grating can cover a range of 1000 wavenumbers. For a Raman spectrum with a range of 400 to 4000 cm^{-1} , in order to record the spectrum in the whole range, the grating should change its angle with incident Raman scattered light. For a modern Raman microscope, the diffraction grating can continuously rotate with respect to the incident Raman scattered light and separate the Raman scattering wavenumbers smoothly. The diffraction grating can provide spectral resolution of 1 cm^{-1} .

Figure 9.31 illustrates the working principles of a diffraction grating. When light rays A and B are incident on the grooves of grating at angle I with respect to the grating substrate normal, they will be reflected by the grooved surfaces as reflected rays A' and B' at angle D with respect to the grating substrate normal. The path difference between A' and B' is calculated.

$$a \sin I + a \sin D$$

a is the distance between grooves. Constructive interference occurs when the following condition is satisfied.

$$a (\sin I + \sin D) = m\lambda \quad (9.16)$$

m is an integer representing the order of diffraction. This basic grating equation is similar to Bragg's Law (Equation 2.3). Incident rays with different wavelengths are reflected at different angles (are dispersed) according to the grating equation.

Detector

The Raman scattered light separated according to wavelength is recorded by a detector which is made from photoelectric materials. The detector converts photon signals to electric signals. The detector for Raman scattered light is often multi-channel and solid-state. The *charged-coupled device* (CCD) is the most commonly used detector. A CCD is a silicon-based semiconductor arranged as an array of photosensitive elements. A one-dimensional array of a CCD can detect and record light intensity of discrete wavelengths separated by the diffraction grating. The Raman shift (mainly Stokes scattering) is calculated and plotted versus wavenumber in the Raman spectrum by computer processing.

9.3.2 Fluorescence Problem

Colored samples or impurities in polymer samples may absorb laser radiation and re-emit it as fluorescence. The intensity of fluorescence can be as much as 10^4 times higher than that of Raman scattered light. The fluorescence problem is the major drawback of using Raman spectroscopy. Thus, a Raman spectrum can be completely masked by fluorescence. Three main methods can be used to minimize fluorescence:

1. Irradiate the sample with high-power laser beams for a prolonged time to bleach out the impurity fluorescence;
2. Change the wavelength of laser excitation to a longer wavelength (near infrared). The chance of fluorescence is reduced because the excitation energy of the laser beam is lower; and

3. Use a pulsed laser source to discriminate against fluorescence because the lifetime of Raman scattering (10^{-12} – 10^{-10} s) is much shorter than that of fluorescence (10^{-7} – 10^{-9} s). Thus, an electron gate can be used to preferentially measure the Raman signals.

Using a long-wavelength laser, however, generates a problem of Raman intensity reduction, because the light intensity decreases with wavelength exponentially. A Fourier transform type of Raman spectroscopy (FT-Raman) has been developed that can solve fluorescence problems, particularly in conjunction with using a longer wavelength. FT-Raman uses a Nd-YAG laser source, which generates excitation of IR wavelength (1064 nm) that does not cause fluorescence in samples. Furthermore, the FT-Raman instrument collects the Raman signal using the Michelson interferometer, which can accommodate large beams. A larger beam can partially compensate for loss of using a longer wavelength. However, FT-Raman is not useful in Raman microscopy in which the pinhole diaphragm has to be used to ensure spatial resolution.

9.3.3 Raman Imaging

Raman imaging is a technique to obtain spatial distribution of specific molecules in a sample. It is similar to element mapping in X-ray, electron and secondary ion mass spectroscopy. The working principle is to record the specific Raman peak that represents the specific component in the sample, as schematically shown in Figure 9.32. Although it can be done with Raman microscopy, Raman imaging is more difficult than other mapping techniques because the Raman scattered light is inherently weak.

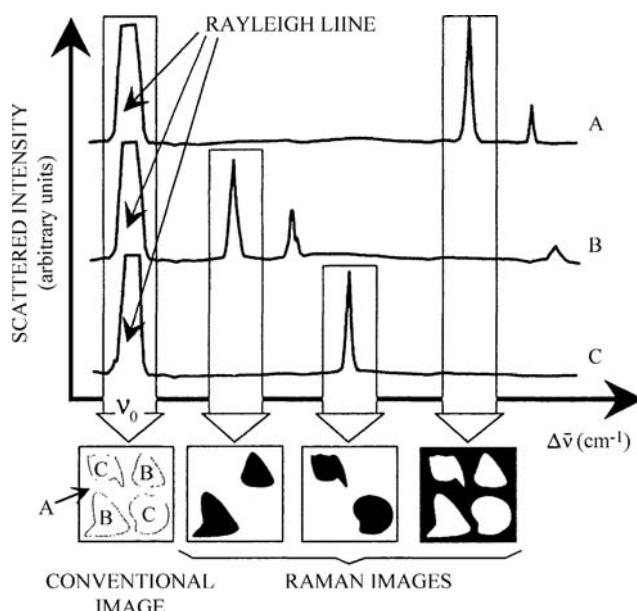


Figure 9.32 Principle of Raman imaging. The spectra of A, B, C represent those of substrate, molecule B and molecule C, respectively. (Reproduced with permission from G. Turrell and J. Corset, *Raman Microscopy, Developments and Applications*, Academic Press, Harcourt Brace & Company, London. © 1996 Elsevier B.V.)

Raman imaging can be either obtained by a scanning or a direct-imaging method in the Raman microscope. The scanning method requires a scanning mechanism of focused laser beam to generate a raster area in the sample, similar to scanning electron microscope (SEM) imaging. The scanning method is often not favorable because it is too time consuming. Unlike for SEM imaging, the intensity of Raman scattered light is so low that dwell time of the laser beam on each pixel should be long in order to obtain a satisfactory two-dimensional image.

The direct-imaging method is rather simple, obtaining a complete two dimensional image for a chosen wavenumber, which represents a molecule, by illuminating a whole area to be analyzed. The entire sample can be done by defocusing the laser beam. Simple defocusing of the laser beam, however, will generate uneven distribution of laser intensity as a Gaussian distribution with a maximum at the beam center and very low intensity near the beam edge. Special optical arrangement of the objective and condenser lenses in a Raman microscope can be used to eliminate this problem.

A Raman image of specific molecule is obtained by mapping its unique wavenumber of Raman scattering. Thus, only the scattered light with a wavenumber should be recorded by a two-dimensional detector. Raman wavenumber selection can be readily performed by the diffraction grating, which serves as a wavenumber filter. Good spatial resolution images of a whole illuminated area can be obtained by properly coupling the image plane of the objective to the surface grating. A high quality Raman image also requires good spectral resolution (narrow range of wavenumbers). To achieve both spatial and spectral resolution, more complicated filtering and optical systems are needed.

Figure 9.33 shows an example of Raman imaging with a direct-image method. The image of the epoxy film reveals the rubber particles in the epoxy matrix. The rubber exhibiting a Raman scattering at 1665 cm^{-1} generates the image contrast. The rubber particles can effectively increase the toughness of the epoxy.

9.3.4 Applications

Raman spectroscopy is attractive as an examination technique for ceramic and polymeric materials because it can simply examine them by illuminating their surfaces regardless of sample thickness and form. Raman microscopy is even more attractive because it can examine a microscopic area with diameters in the order of $1\text{ }\mu\text{m}$. Raman microscopy is increasingly used for materials characterization, including:

- Phase identification of polymorphic solids;
- Polymer identification;
- Composition determination;
- Determination of residual strain; and
- Determination of crystallographic orientation.

Phase Identification

Raman microscopy is able to determine phases for polymorphic solids at the microscopic level. This is its advantage over conventional X-ray diffraction spectrometry in which the sample volume cannot be too small. Phase identification with Raman spectroscopy uses

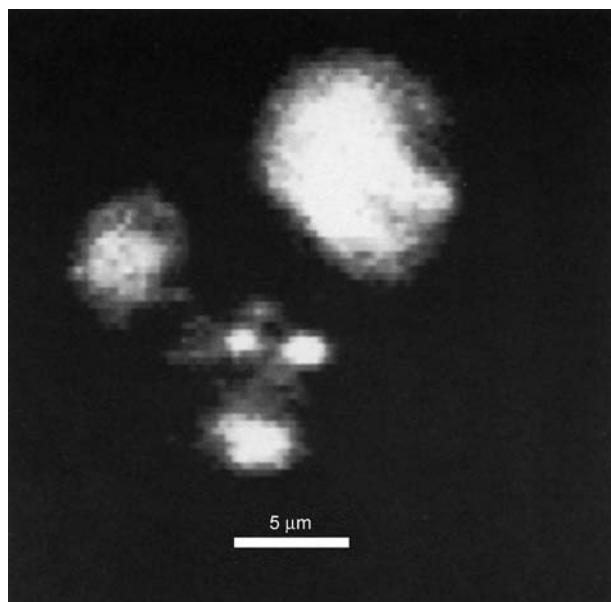


Figure 9.33 Raman image of rubber particles in epoxy film at 1665 cm^{-1} . (Reproduced with permission from G. Turrell and J. Corset, *Raman Microscopy, Developments and Applications*, Academic Press, Harcourt Brace & Company, London. © 1996 Elsevier B.V.)

the characteristic vibration band(s) associated with a certain phase in a solid. For example, Table 9.2 lists characteristic vibration bands in Raman spectroscopy in several polymorphic solids. This feature of Raman spectroscopy can be further demonstrated using carbon as an example. Figure 9.34 shows four types of carbon spectra which represent highly oriented graphite, polycrystalline graphite, amorphous carbon and diamond-like carbon. Significant differences appear in Raman spectra for graphite and diamond because carbon atoms are bonded with sp^2 -type orbitals in graphite while they are bonded in sp^3 -type orbitals in diamond.

Polymer Identification

Raman microscopy is able to identify the different type of polymers even though they all contain C, H and O. For example, Figure 9.35 shows the Raman spectra of a laminated polymer film. Individual layers can be identified with Raman microscopy by focusing the laser beam on

Table 9.2 Phase Identification with Raman Microscopy

Polymorphic Materials	Phase 1	Bands (cm^{-1})	Phase 2	Bands (cm^{-1})
Si_3N_4	α	262, 365, 514, 670, 850	β	185, 210, 230
BN	Cubic	1055	Hexagonal	1367
ZrO_2	Monoclinic	181, 192	Tetragonal	148, 264
Carbon	Diamond	1332	Graphite	1580

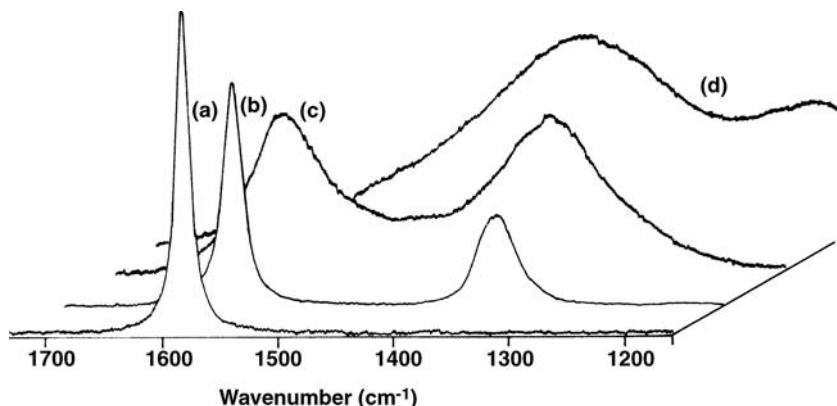


Figure 9.34 Characteristic spectra of carbon in different structures: (a) highly oriented graphite; (b) polycrystalline graphite; (c) amorphous carbon; and (d) diamond-like carbon. (Reproduced with permission from G. Turrell and J. Corset, *Raman Microscopy, Developments and Applications*, Academic Press, Harcourt Brace & Company, London. © 1996 Elsevier B.V.)

each of the layers in the cross-section of the laminated sample. The Raman spectra reveal the laminated sample is composed of polyester, polyethylene and a layer of paper.

Raman spectroscopy is sensitive to polymer conformation. For example, a polymer blend of polybutadiene–polystyrene in which polybutadiene is used to increase toughness of the polystyrene can be examined by Raman microscopy to identify its heterogeneity. Polybutadiene has three isomer conformations (*cis*-1,4, *trans*-1,4 and *syndiotactic*-1,2). These three types of isomers can be identified from C=C stretching modes as shown in Figure 9.36. The Raman spectra of the copolymer indicate the difference in amounts of isomer types at the edge and the center of the polybutadiene–polystyrene sample. Relative amounts of these isomer types affect the mechanical properties of the copolymer.

Composition Determination

Raman microscopy has been used to monitor the amount of chemical elements present in, or added to, a solid. For example, Figure 9.37 shows that the graphite spectrum changes according to the number of molecules intercalated into its layered crystal structure. Molecular intercalation into gaps between crystal layers changes the bonding strength between the layers, which in turn changes the vibrational frequency in the graphite. The degrees of intercalation are represented by the stage number, which indicates intercalation occurs in every stage number of layers; lower stage numbers mean higher degrees of intercalation. The graphite spectrum is highly sensitive to the stage number of intercalated graphite as shown in Figure 9.37.

Raman microscopy has also been used to identify the composition variation of $\text{Si}_{1-x}\text{Ge}_x$ semiconductor materials. Figure 9.38 shows the Raman spectrum of a $\text{Si}_{0.7}\text{Ge}_{0.3}$ film on Si substrate. The vibrations of Si–Si, Si–Ge and Ge–Ge bonding in the film generate three distinct Raman peaks. Note the Raman shift of Si–Si vibration in $\text{Si}_{1-x}\text{Ge}_x$ is different from that of Si–Si in pure Si. The composition variation of $\text{Si}_{1-x}\text{Ge}_x$ is related to Raman peak positions in the spectrum. Figure 9.39 shows that the Raman shift position is linearly related to the amount of Ge added in the Si.

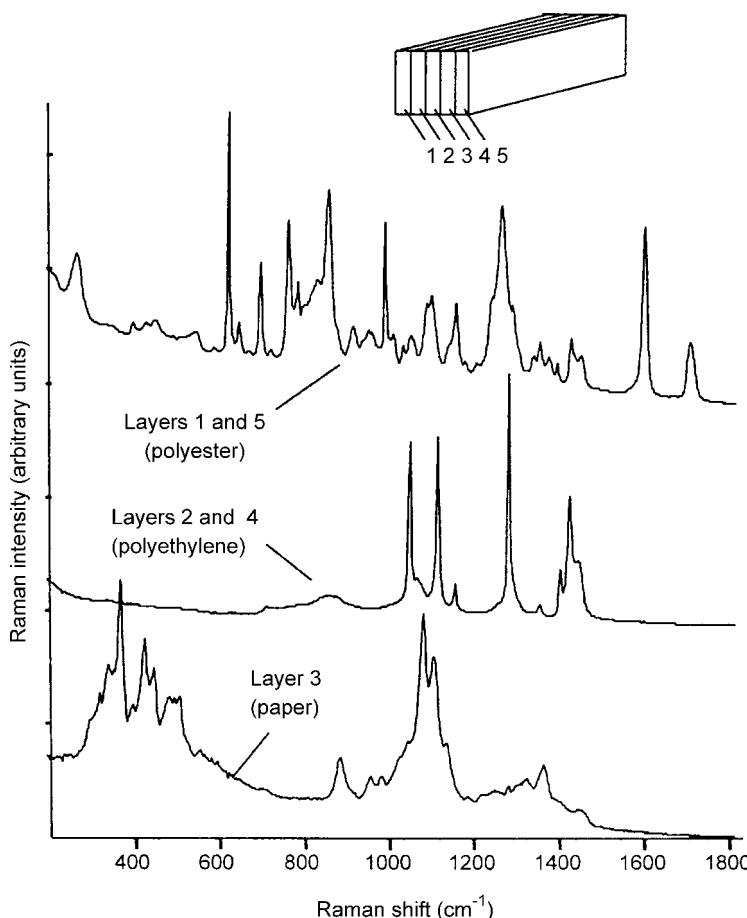


Figure 9.35 Raman spectra of a laminated polymer film obtained with excitation of a 785-nm laser. (Reproduced from M.J. Pelletier, *Analytical Applications of Raman Spectroscopy*, Blackwell Science, Oxford. © 1999 Blackwell Publishing.)

Determination of Residual Strain

Residual strain in a solid can also be determined by the shift of Raman bands because strain changes the length of a bond between atoms and vibrational frequency is exponentially affected by the bond length. For example, compressive strain in a microscopic area of a sample will reduce the bond length, and in turn, increase the corresponding vibrational frequency. Figure 9.39 clearly indicates the residual strain effect on Raman shift. A thin Si–Ge film on Si substrate contains residual stress because the lattice parameter of Ge is different from that of Si. The larger lattice parameter of Ge (0.566 nm) than that of Si (0.543 nm) generates biaxial compression in the film. The residual compression strain alters the relationship between the Raman shift and Ge content as indicated by a solid line, compared with the dotted line that represents the relationship when residual strain is not present in the Si–Ge film. In thicker films,

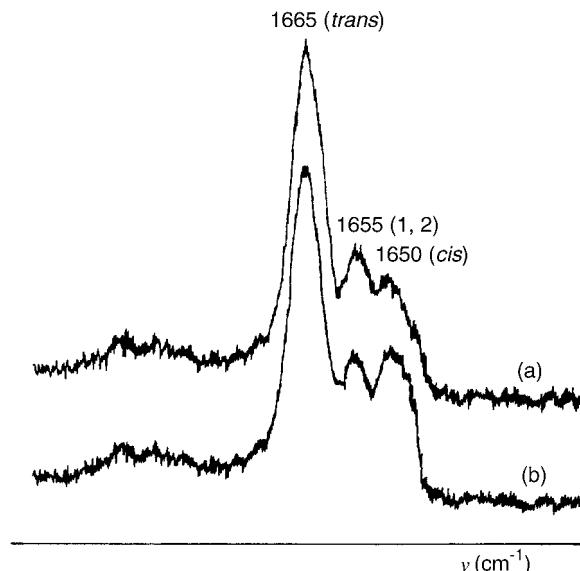


Figure 9.36 Raman spectra of polybutadiene–polystyrene copolymer in the C=C stretching region showing three distinct bands of polybutadiene isomers: *cis*-1,4 at 1650 cm^{-1} , *trans*-1,4 at 1665 cm^{-1} and *syndiotactic*-1,2 at 1655 cm^{-1} : (a) spectrum from the center of sample; and (b) spectrum from the edge of sample. (Reproduced with permission from G. Turrell and J. Corset, *Raman Microscopy, Developments and Applications*, Academic Press, Harcourt Brace & Company, London. © 1996 Elsevier B.V.)

the residual stress is relaxed by formation of cracks. Thus, the relation between Raman shift and Ge content is closer to the dotted line.

Figure 9.40 shows another example of residual strain effects on the Raman spectrum. The Si stretching vibration changes its frequency when Si is in the form of a thin film on sapphire. We can compare the Raman band positions between the strain-free sample and strained sample to evaluate the bond length change, which can be converted to residual strain. The residual strain is commonly elastic in nature. Thus, the residual stress in materials can be determined with the linear elastic relationship between strain and stress.

Determination of Crystallographic Orientation

Crystallographic orientation in a sample area can also be determined by measuring the intensity change of Raman scattered light in different polarization directions. The method of crystal orientation determination requires extensive theoretical treatment, which is beyond the scope of this book. However, its concept can be understood from the following description. When polarized light is scattered by a crystallographic plane of a sample, the polarization of light will be changed. The degree of change will depend on the angle between the polarization direction of incident light and direction of the crystallographic plane, and the angle between the polarization direction of scattered light and direction of the crystallographic plane. If we measure the scattered light intensity change with polarization direction using an analyzer, the direction of the crystallographic plane at the examined sample area can be determined.

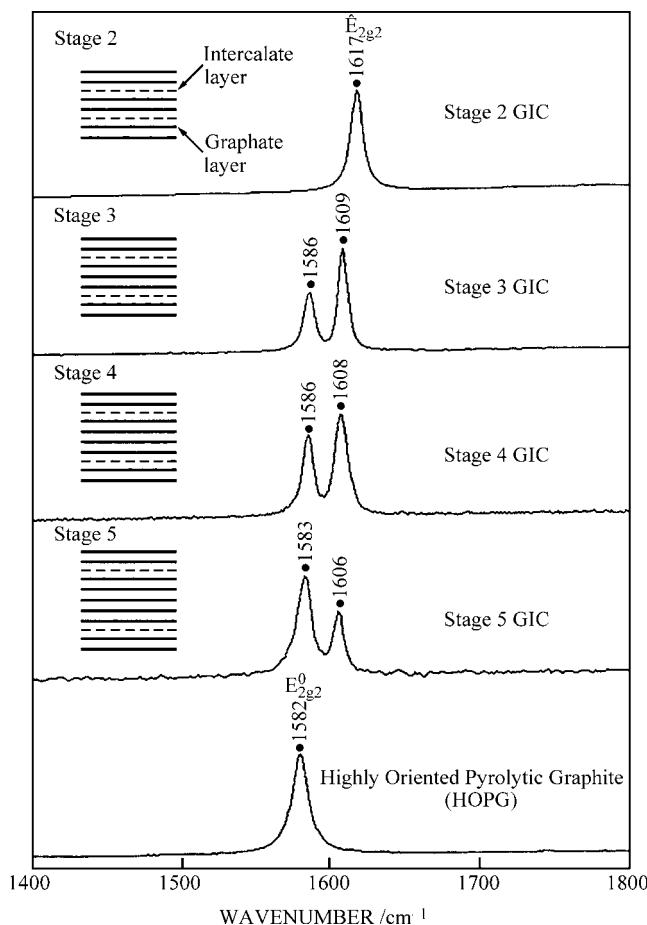


Figure 9.37 Raman spectra of graphite and graphite intercalation compounds (GIC) with FeCl_3 . A lower stage number indicates a higher degree of intercalation.

9.4 Interpretation of Vibrational Spectra

Theoretical interpretation of molecular vibration spectra is not a simple task. It requires knowledge of symmetry and mathematical group theory to assign all the vibration bands in a spectrum precisely. For applications of vibrational spectroscopy to materials characterization, we can still interpret the vibrational spectra with relatively simple methods without extensive theoretical background knowledge. Here, we introduce some simple methods of vibrational spectrum interpretations.

9.4.1 Qualitative Methods

A simple way to qualitatively interpret vibrational spectra is the ‘fingerprint’ method of identifying materials. The fingerprint method includes spectrum comparison with a reference, and identification of characteristic bands in spectra.

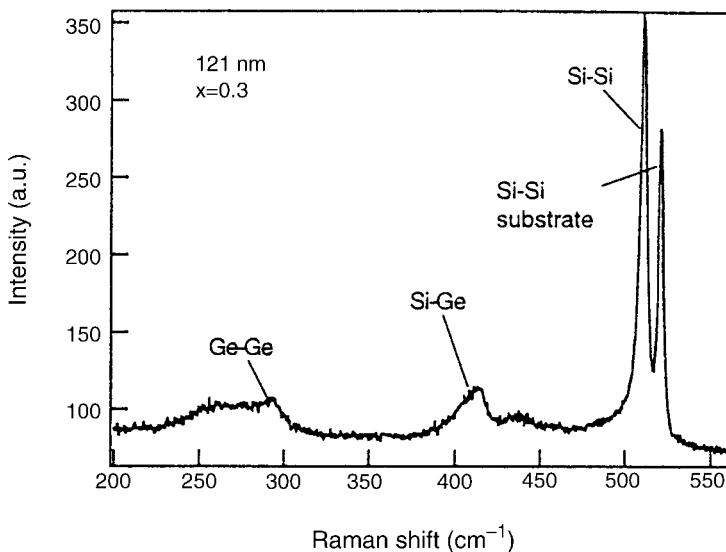


Figure 9.38 Raman spectrum of a $\text{Si}_{1-x}\text{Ge}_x$ thin film with thickness of 121 nm on Si substrate, with argon laser (457.9 nm) excitation. (Reproduced from M.J. Pelletier, *Analytical Applications of Raman Spectroscopy*, Blackwell Science, Oxford. © 1999 Blackwell Publishing.)

Spectrum Comparison

Spectrum comparison is the simplest method to interpret a vibration spectrum. We do not need to have much theoretical knowledge of spectrum interpretation for doing so. A sample can be identified if its spectrum matches a reference in both band positions and intensities.

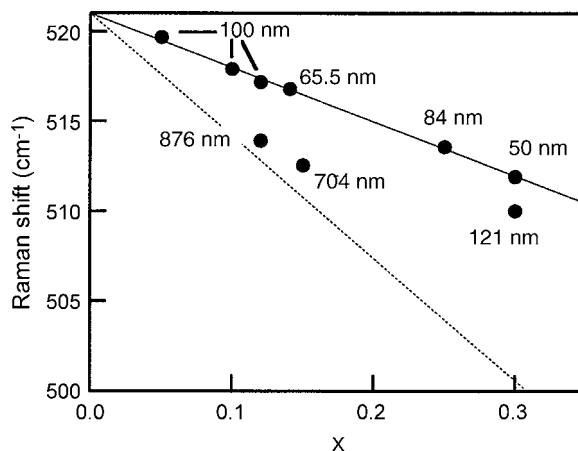


Figure 9.39 Raman peak shift relationship with the composition change in $\text{Si}_{1-x}\text{Ge}_x$ thin film. The plot also indicates the film thickness effect on the relationship. (Reproduced from M.J. Pelletier, *Analytical Applications of Raman Spectroscopy*, Blackwell Science, Oxford. © 1999 Blackwell Publishing.)

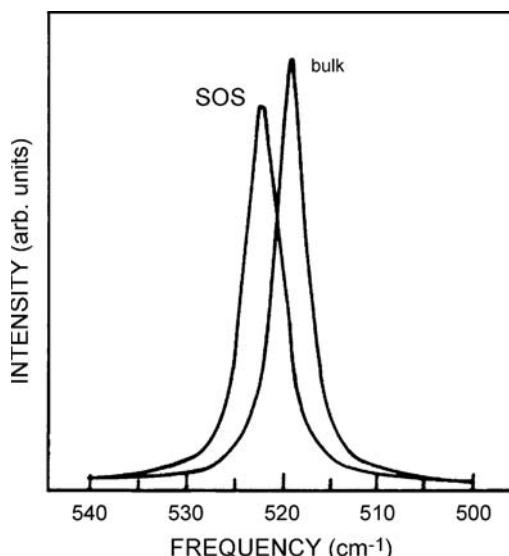


Figure 9.40 Comparison of Raman peak position in bulk Si and an Si film on sapphire (SOS). (Reproduced from M.J. Pelletier, *Analytical Applications of Raman Spectroscopy*, Blackwell Science, Oxford. © 1999 Blackwell Publishing.)

in the whole spectrum range. There is no need to assign vibration bands in a spectrum in spectrum comparison. Comparison of sample and reference spectra should satisfy the following conditions; otherwise the spectra are not comparable:

- Both sample and reference spectra must be in the sample physical state (for example, either both are liquid or solid); and
- Both sample and reference must be measured using the same techniques (for example, both are dispersed with KBr for FTIR measurement).

Vibration frequencies in the solid phase are generally lower than in the liquid phase. Certain bands of solid samples do not appear in liquid and gaseous samples. Physical and/or chemical interactions of the sample with a dispersing matrix can alter the spectrum and make comparison impossible.

Typical applications of spectrum comparison include verifying whether synthetic materials are identical to their natural counterparts, and determining the stability of materials over a period of time or in certain environments. For spectrum comparison, a region of spectrum below 1200 cm^{-1} is of importance, often called the ‘fingerprint region.’ The reason is that the region of relatively low wavenumber shows the vibrational bands involving large parts of molecules and crystals. The band positions and intensities in such a region are more sensitive to minor chemical variations. For example, exchange of the A^+ ion species in compounds of $\text{AHMo}_5\text{O}_{16}(\text{H}_2\text{O})\text{H}_2\text{O}_m$, can be detected by comparing fingerprint regions in the Raman spectra of the compounds as shown in Figure 9.41.

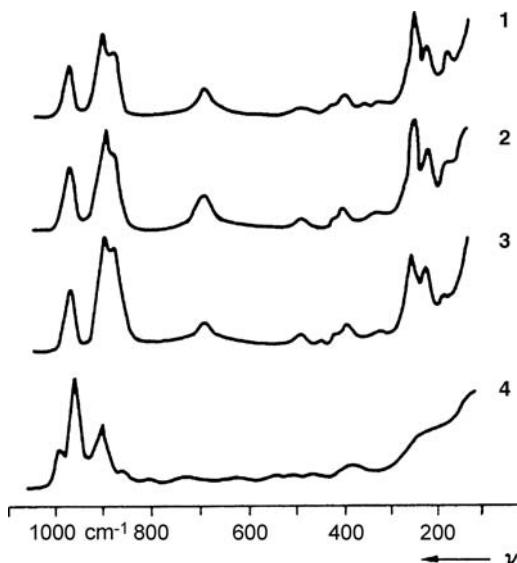


Figure 9.41 Comparison of Raman spectra of $\text{AHMo}_5\text{O}_{16}(\text{H}_2\text{O})\text{H}_2\text{O}_m$: A = (a) NH_4^+ ; (b) K^+ ; (c) Na^+ ; and (d) compound spectrum in aqueous solutions. (Reproduced from A. Fadini and F.M. Schnepel, *Vibrational Spectroscopy: Methods and Applications*, Ellis Horwood, Chichester, 1989.)

Identifying Characteristic Bands

Vibrational frequencies of some atom groups are quite independent even if they are parts of larger molecules and crystals. Such vibration bands serve as characteristic bands for atom groups. A characteristic band for an atom group should satisfy the following conditions:

- The band occurs in all spectra of materials containing the atom group; and
- The band is absent when materials do not contain the atom group.

For example, compounds containing the $-\text{CH}=\text{CH}_2$ group exhibit a band near 1650 cm^{-1} in their IR spectra; compounds containing O–H stretching vibration exhibit a band near 3500 cm^{-1} .

It is helpful to have basic knowledge of the relationship between vibrational frequencies and bond strength and atomic masses. We may use the relationship in understanding and assigning positions of characteristic bands in spectra. This relationship can be illustrated with the simplest vibration of a diatomic model. Classical mechanics shows the vibrational frequency relationship.

$$\nu_{vib} \propto \sqrt{f(m_1 + m_2)/(m_1 m_2)} \quad (9.17)$$

f is the bond strength of atom 1 with mass m_1 and atom 2 with mass m_2 . This equation indicates that vibrational frequency increases with the bond strength and decreases with atomic mass. Thus, it is understandable that the following relationships hold for vibration frequency

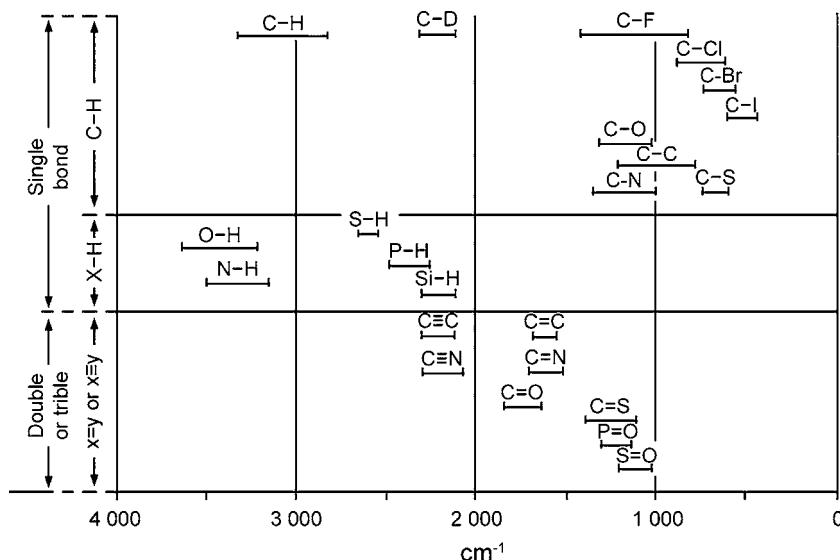


Figure 9.42 Positions of characteristic bands in vibrational spectra for some diatomic stretching vibrations. (Reproduced from A. Fadini and F-M. Schnepel, *Vibrational Spectroscopy: Methods and Applications*, Ellis Horwood, Chichester, 1989.)

increases.

$$\nu(C\equiv C) > \nu(C=C) > \nu(C-C) \quad (9.18)$$

Also

$$\nu(C-H) > \nu(C-O) > \nu(C-Cl) \quad (9.19)$$

Figure 9.42 illustrates approximate distributions of some diatomic stretching vibrational bands in spectra. It clearly demonstrates the relationship between wavenumber ($\propto \nu_{vib}$) and characteristics of bonds. Characteristic bands in Figure 9.42 show ranges instead of exact wavenumbers. This phenomenon reflects that the vibrational frequencies also depend on neighboring atoms in materials, even though the frequencies are mainly determined by bond strength and atomic masses.

In general, the analysis method based on identifying characteristic bands is applicable to polymers and ceramics. Theoretically, vibrational spectra of such materials are generally complicated in nature. Their complex nature arises from several aspects: the presence of new bonds coupling repeated units, the large number of vibration bands ($3N - 6$) due to a large number of atoms (N) in a polymer molecule, and vibration coupling of crystal lattices. Practically, vibration spectra of polymers and ceramics are often considered as a summation of the spectra of individual repeat units or functional groups, with slight modifications due to the presence of new bond coupling between the repeated units. The characteristic bands can still be identified; however, the bands may be broadened or slightly shifted. Relatively few extra vibration bands from molecular chain and crystal structures are observed in spectra of polymers and ceramics.

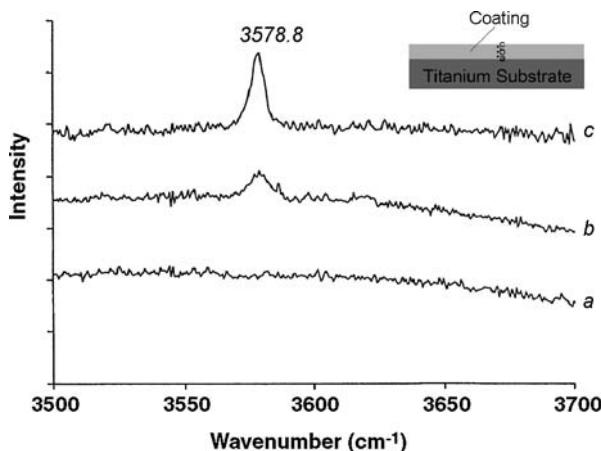


Figure 9.43 Raman spectra of plasma sprayed calcium phosphate coating shows the difference in hydroxyapatite content by comparing intensity of the OH stretching band at: (a) near substrate; (b) in the middle; and (c) near the surface of the coating layer.

We may use the characteristic bands to distinguish compounds that include or are composed of similar ion groups. For example, hydroxyapatite $[\text{Ca}_5(\text{PO}_4)_3\text{OH}]$ can be distinguished from other compounds of calcium phosphates, for example tricalcium phosphate $[\text{Ca}_3(\text{PO}_4)_2]$, in Raman spectroscopy by its hydroxyl group. Raman microscopy (Figure 9.43) demonstrates that a calcium phosphate coating layer produced by plasma sprayed on a titanium substrate exhibits compositional change in the coating layer. The relative amount of hydroxyapatite in different depths of the coating layer of calcium phosphate can be estimated by comparing the intensity of O–H stretching band at about 3600 cm^{-1} in different locations of a coating layer: near the surface, in the middle of layer and close to substrate interface.

On the other hand, limitations of using characteristic bands should not be ignored. For example, some characteristic bands may overlap each other. The vibrational bands for O–H and N–H stretching occur in the same region in the spectra, as shown in Figure 9.42. Also, certain factors can shift or eliminate characteristic bands. Commonly, characteristic bands are distinct for organic materials, but may not be for inorganic materials. One reason is strong vibrational coupling in inorganic materials. Such coupling occurs when bond strength and atomic masses vary slightly through inorganic molecules. For example, an isolated C=O stretching band is located at $1620\text{--}1830\text{ cm}^{-1}$. For CO₂ with two C=O bonds sharing one carbon atom, the stretching vibrations of C=O become one symmetric and one anti-symmetric vibration (Figure 9.8). The two distinctive vibrational bands are located at 1345 cm^{-1} and 2349 cm^{-1} , respectively.

Band Intensities

Interpreting the intensities of vibrational bands in a vibrational spectrum is not straightforward. The band intensity is not simply proportional to the number of atoms involved with such specific vibration. The intensity varies with types of bonds and types of spectroscopy, either IR or Raman. The intensity of an IR absorption band is proportional to the square of its change

in dipole moment during vibration.

$$I(IR) \propto \left(\frac{\partial \mu}{\partial q} \right)^2 \quad (9.20)$$

But the intensity of a Raman shift is proportional to the square of its change in polarizability during vibration.

$$I(Raman) \propto \left(\frac{\partial \alpha}{\partial q} \right)^2 \quad (9.21)$$

Table 9.3 Intensity Estimations of IR and Raman Bands for Certain Structure Elements

System [†]	Structure element	Vibration	Intensity [‡] IR	Intensity [‡] Raman
X-Y		v(C-C)	w	s
		v(C-X)	s	vs
	R-O-H	v(O-H)	s	w
X=Y		v(C=C)	w	s
		v(C=O)	s	s
X≡Y	-C≡C-	v(C≡C)	w	s
Y-X-Y	O=C=O	v _s (YXY)	w-m	s
		v _{as} (YXY)	s	w

[†] X = heavy atom.

[‡] w = weak, m = medium, s = strong, vs = very strong

(Data reproduced from A. Fadini and F.-M. Schnepel, *Vibrational Spectroscopy: Methods and Applications*, Ellis Horwood, Chichester, 1989.)

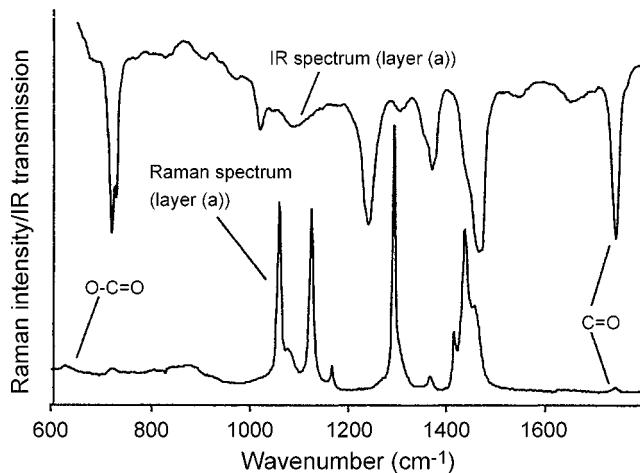


Figure 9.44 Comparison of IR and Raman spectra of copolymer ethylene–vinyl acetate (EVA). EVA is distinguished by the C=C bands. (Reproduced from M.J. Pelletier, *Analytical Applications of Raman Spectroscopy*, Blackwell Science, Oxford. © 1999 Blackwell Publishing.)

In practice, some simple rules are useful to estimate the band intensities:

- Polar bonds yield intense IR but weak Raman bands;
- Weak polar or non-polar bonds yield intense Raman but weak IR bands; and
- Intensity increases with number of identical structural elements in a molecule.

To illustrate, Table 9.3 lists several structural elements of molecules and their intensity in IR and Raman bands. The intensity of a vibration band is often more complicated than is indicated in this table. For example, Table 9.3 suggests that unsaturated bonds (= and ≡) exhibit higher band intensities in Raman than in IR spectra. However, this is not always true. For example, Figure 9.44 shows a comparison of IR and Raman spectra of a copolymer sample, EVA (ethylene–vinyl acetate). The C=O band shows high intensity in the IR spectrum, not the Raman spectrum. More detailed information about band intensities, including band contours and effects of vibration symmetry on intensity, can be found in the references listed at the end of this chapter.

9.4.2 Quantitative Methods

Quantitative Analysis of Infrared Spectra

The concentration of certain components in a multi-component sample can be determined by quantitative analysis of its vibrational spectrum. Similar to quantitative analysis of X-ray diffraction spectra, the concentration of a component can be evaluated from the intensities of its vibration bands. For an IR spectrum, concentration is proportional to absorbance (A) according to the absorption law, also called Beer's Law.

$$A = -\log T = \log \frac{I_o}{I} = alC \quad (9.22)$$

A large rectangular box with a thin black border, containing the text "Image Not Available In The Electronic Edition".

a is the absorptivity that is a constant at a particular wavenumber, and l is the sample thickness or the path length of radiation in the sample. I and I_o are defined in Equation 9.14. The concentration (C) can be determined by measuring the magnitude of absorbance from an IR spectrum, if a and l are known. These latter two parameters can be obtained by using a calibration graph of A versus C if we have several samples of known concentrations.

The absorbance A is commonly obtained by peak height measurement in quantitative IR analysis. Figure 9.45 illustrates how to measure I and I_o from a conventional IR spectrum in which $\%T$ is plotted against the wavenumber. We should note that the transmittance (T), cannot be directly used for quantitative analysis. The figure shows a popular method for measuring I_o using a baseline method. The baseline is drawn tangentially to the spectrum at the wings of the analytical band. The baseline intersection with a vertical line at the wavenumber of band peak is used as I_o .

We should know that the peak height is sensitive to instrument resolution. The peak area measurement under a vibration band shows much less instrumentation dependence. The peak area represents the integrated intensity of the vibration band and is proportional to the square of the change in dipole moment with respect to the normal coordinate as Equation 9.20.

For analysis of two-component materials, a ratio method can be used for quantitative analysis based on Beer's Law. For example, there is a mixture of two components and two components have their own characteristic bands, which do not interfere with each other. Then, the ratio of their absorbance will be calculated.

$$\frac{A_1}{A_2} = \frac{a_1 l_1 C_1}{a_2 l_2 C_2} \quad (9.23)$$

The path lengths should be the same in the same sample. We can determine the ratio of $a_1:a_2$ using a standard sample in which C_1 is known.

$$\frac{a_1}{a_2} = \left(\frac{1}{C_1} - 1 \right) \frac{A_1}{A_2} \quad (9.24)$$

For analysis of multi-component materials with overlapping vibration bands, Beer's Law can still be used because absorbances of components are additive. Combining a and l as a single parameter k , the absorbance of components is expressed.

$$A = k_1C_1 + k_2C_2 + k_3C_3 + \dots k_nC_n \quad (9.25)$$

There will be n number of equations and $n \times n$ number of k parameters for a sample with n number of components. For example, a three-component system should have the following expressions.

$$\begin{aligned} A_1 &= k_{11}C_1 + k_{12}C_2 + k_{13}C_3 \\ A_2 &= k_{21}C_1 + k_{22}C_2 + k_{23}C_3 \\ A_3 &= k_{31}C_1 + k_{32}C_2 + k_{33}C_3 \end{aligned} \quad (9.26)$$

The best way to write a group of linear equations like Equation 9.26 is to use the matrix form.

$$\mathbf{A} = \mathbf{KC} \quad (9.27)$$

The concentrations of components should be resolved as \mathbf{C} .

$$\mathbf{C} = \mathbf{PA} \quad (9.28)$$

where $\mathbf{P} = \mathbf{K}^{-1}$, the inverse of \mathbf{K} matrix. This method is required to determine all the k values in the \mathbf{K} matrix for standard samples.

Quantitative Analysis of Raman Spectra

Raman spectroscopy is a scattering, not an absorption technique as FTIR. Thus, the ratio method cannot be used to determine the amount of light scattered unless an internal standard method is adopted. The internal standard method requires adding a known amount of a known component to each unknown sample. This known component should be chemically stable, not interact with other components in the sample and also have a unique peak. Plotting the Raman intensity of known component peaks versus known concentration in the sample, the proportional factor of Raman intensity to concentration can be identified as the slope of the plot. For the same experimental conditions, this proportional factor is used to determine the concentration of an unknown component from its unique peak. Determining relative contents of Si and Ge in Si–Ge thin films (Figure 9.38 and Figure 9.39) is an example of quantitative analysis of a Raman spectrum.

References

- [1] Fadini, A. and Schnepel, F-M. (1989) *Vibrational Spectroscopy. Methods and Applications*, Ellis Horwood, Chichester.
- [2] Smith, B.C. (1996) *Fundamentals of Fourier Transform Infrared Spectroscopy*, CRC Press, Boca Raton.
- [3] Ferraro, J.R., Nakamoto, K. and Brown, C.W. (2003) *Introductory Raman Spectroscopy*, 2nd edition, Academic Press, San Diego.

- [4] Turrell, G. and Corset, J. (1996) *Raman Microscopy, Developments and Applications*, Academic Press, Harcourt Brace & Company, London.
- [5] Hendra, P., Jones, P.C. and Warnes, G. (1991) *Fourier Transform Raman Spectroscopy, Instrumentation and Chemical Applications*, Ellis Horwood, Chichester.
- [6] Colthup, N.B., Daly, L.H. and Wiberley, S.E. (1990) *Introduction to Infrared and Raman Spectroscopy*, 3rd edition, Academic Press, San Diego.
- [7] Nakamoto, K. (1986) *Infrared and Raman Spectra of Inorganic and Coordination Compounds*, John Wiley & Sons, New York.
- [8] Mohan, J. (2000) *Organic Spectroscopy, Principles and Applications*, CRC Press, Boca Raton.
- [9] Chase, D.B. and Rabolt, J.F. (1994) *Fourier Transform Raman Spectroscopy, From Concept to Experiment*, Academic Press, San Diego.
- [10] Pelletier, M.J. (1999) *Analytical Applications of Raman Spectroscopy*, Blackwell Science, Oxford.
- [11] Jumecki, H.J. (1995) *Practical Guide to Infrared Microspectroscopy*, Marcel Dekker, New York.

Questions

- 9.1 Can we identify atoms or monatomic ions of materials using vibrational spectroscopy? Why?
- 9.2 Determine the normal modes of planar molecule XY_3 with a central symmetric configuration, where X is located at the center of the molecule.
- 9.3 Find out the normal modes of tetragonal molecule XY_4 .
- 9.4 List the similarities of, and differences between, IR and Raman spectroscopy in terms of working principles, spectra and detectable vibration bands.
- 9.5 What is/are the advantage(s) of using an FTIR instead of dispersive IR spectrometer?
- 9.6 Why does transmission IR require samples to have thickness in the range 1–20 μm ? How can we examine materials in bulk form?
- 9.7 Why does IR spectroscopy require radiation containing continuous wavelengths while Raman spectroscopy requires a monochromatic source?
- 9.8 What are the special requirements for a microscope in a Raman microscope compared with a conventional light microscope?
- 9.9 Where is the ‘fingerprint region’ in a vibration spectrum?
- 9.10 What kinds of vibrational bands are considered as characteristic bands?

10

Thermal Analysis

Thermal analysis (TA) is a group of analytical techniques that measure properties or property changes of materials as a function of temperature. Changes in materials properties with temperature are considered as thermal events. The properties that can change include dimension, mass, phase and mechanical behavior. Thermal analysis methods are relatively simple because changing the temperature of a sample is less complicated than probing a sample using high-energy X-ray, electron or ion beams in techniques of spectroscopy. Many TA methods have been developed for a variety of examination purposes. Table 10.1 lists commonly used TA techniques. This chapter introduces the most commonly used thermal analysis techniques for materials characterization: thermogravimetry, differential thermal analysis and differential scanning calorimetry. Thermogravimetry (TG) is mainly used to examine the decomposition of materials by monitoring mass change with temperature. Differential thermal analysis (DTA) and differential scanning calorimetry (DSC) are widely used for examining the phase changes of materials.

10.1 Common Characteristics

10.1.1 Thermal Events

Thermal analysis relies on material reactions to thermal energy flow in or out of solids. Such reactions are referred to as *thermal events*. Possible thermal events of a solid can be reviewed by increasing its temperature from absolute zero (zero Kelvin). At absolute zero, a solid does not have any thermal energy, and it means that its atoms are static: there is no vibration or rotation of atomic bonds. With increasing temperature, thermal energy gains in the solid will increase vibrations and rotations of atomic bonds. When the amplitudes of vibrations reach a certain level, the following changes in the solid may occur if the solid is in an inert atmosphere: solid phase transformation, glass transition, melting, sublimation and thermal decomposition.

Solid phase transformation may occur for solids that have different equilibrium crystalline phases in different temperature ranges. For example, pure iron changes its crystal structure from body-centered cubic (BCC) to face-centered cubic (FCC) when the temperature increases to 912 °C. Glass transition occurs in non-crystalline solids. With increasing temperature, a rigid non-crystalline solid in a glass phase may change to a rubber-like solid. The glass transition

Table 10.1 Commonly Used Thermal Analysis Techniques

Technique	Abbreviation	Measurement
Thermogravimetry	TG	Mass change
Differential thermal analysis	DTA	Temperature difference
Differential scanning calorimetry	DSC	Heat flow
Dilatometry	—	Length or volume change
Thermomechanical analysis	TMA	Deformation
Dynamic mechanical analysis	DMA	Deformation

is referred to as a second-order phase transition. It is distinguishable from first-order phase transitions in the solid state, which are phase transformations in crystalline solids. The glass transition is an important property of polymeric materials. Melting is simply the phase transformation from solid to liquid, and sublimation is the phase transformation from solid to gas. Thermal decomposition is referred to as a change in a solid containing more than one type of atom or molecule. With increasing temperature, the bonding between atoms becomes so weak that certain constituents will be released from the solid into the atmosphere and the remaining constituents will form a new type of solid. For example, solid CaCO_3 will decompose to form solid CaO and gaseous CO_2 over a certain temperature range.

Table 10.2 lists the thermal events that may occur when a solid (A) is heated in an inert atmosphere. The heat flow between a solid and its surrounding atmosphere at a certain temperature is indicated as *enthalpy change* (ΔH) in Table 10.2. ΔH quantifies the heat flowing into or out of a solid under constant pressure. Commonly, thermal events occurring under a constant-pressure atmosphere are analyzed in thermal analysis instruments.

Enthalpy Change

Enthalpy is the thermodynamic parameter that is useful for describing the thermal events of a material under constant pressure, for example, one atmosphere. We can understand the meaning of enthalpy and enthalpy change by reviewing the first law of thermodynamics.

$$\Delta U = Q - W \quad (10.1)$$

The internal energy change ΔU of a system is a sum of the heat flowing into it (Q) and the work done by the system ($-W$). Heat is defined as the energy transfer between the system

Table 10.2 Thermal Events During Heating of a Solid in an Inert Atmosphere

Thermal Event	Reaction	Heat flow ΔH	Mass change in solid
Solid phase transformation	$A(\text{solid } \alpha) \rightarrow A(\text{solid } \beta)$	+ or -	no
Glass transition	$A(\text{glass}) \rightarrow A(\text{rubber})$	no	no
Melting	$A(\text{solid}) \rightarrow A(\text{liquid})$	+	no
Sublimation	$A(\text{solid}) \rightarrow A(\text{gas})$	+	yes
Thermal decomposition	$A(\text{solid}) \rightarrow B(\text{solid}) + C(\text{gas})$	+ or -	yes

and its surroundings due to their temperature difference. Work includes all types of energy transfer other than heat. If we consider that the work is mechanical work of volume change in the system under constant pressure, then the first law should be written as follows.

$$\Delta U = Q_P - P\Delta V \quad (10.2)$$

Q_P represents the heat required for internal energy change under constant pressure. Enthalpy is defined as H .

$$H \equiv U + PV \quad (10.3)$$

Thus, under constant pressure, the enthalpy change will be written.

$$\Delta H = \Delta U + P\Delta V$$

or

$$\Delta U = \Delta H - P\Delta V \quad (10.4)$$

Comparing Equation 10.2 and Equation 10.4, we can conclude the following.

$$\Delta H \equiv Q_P \quad (10.5)$$

Thus, the heat flowing in or out of a system under constant pressure is equal to its enthalpy change. In the rest of Chapter 10, the heat flow under constant pressure will be measured as the enthalpy change, ΔH . Quite often we simply refer to ΔH as ‘enthalpy’, not ‘enthalpy change’, although the latter is technically correct.

10.1.2 Instrumentation

All TA techniques have certain features in common in their instrumentation: a furnace in which sample is heated (or cooled) with a controllable environment, and a transducer by which the property changes in materials are monitored. Figure 10.1 illustrates general TA instrumentation. The sample is analyzed under a designed temperature profile. The temperature profile may take several forms such as a constant heating rate, a modulated heating rate (for example, a sinusoidal heating curve), or even an isothermal profile. The measurements are displayed as thermal analysis curves. The features of the curves (peaks, discontinuities and slope changes) are related to thermal events.

10.1.3 Experimental Parameters

TA results can be significantly affected by experimental parameters such as sample dimension and mass, heating (or cooling) rates, atmosphere surrounding the sample and even thermal and mechanical history of the sample. Reproducible results of a chemical species

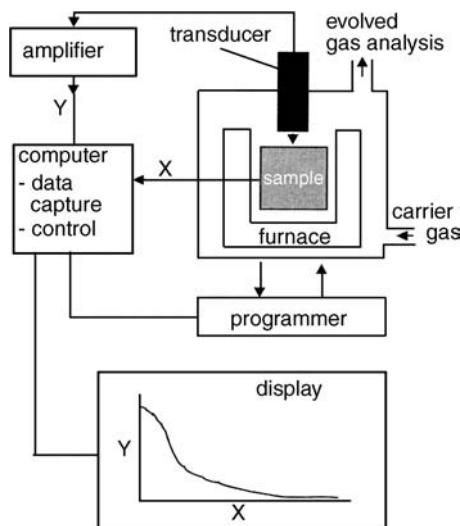


Figure 10.1 General instrumentation for thermal analysis. (Reproduced with kind permission of Springer Science and Business Media from M. Brown, *Introduction to Thermal Analysis*, Kluwer Academic Publishers, Dordrecht. © 2001 Springer Science.)

may not be obtained when the instrumental and experimental parameters are not identical. The main reason is that TA is highly sensitive to heat transfer conditions and to accuracy of temperature measurements. The following considerations can help ensure reliable TA data.

Sample dimension and mass should be small. For most TA techniques, samples in the form of a powder with sample mass less than 10 mg are preferred. Heat transfer between the sample and atmosphere will be faster in such a sample than in a lump; thus, thermal equilibrium is more likely to be achieved between sample and atmosphere during analysis. Samples to be analyzed should have the same thermal and mechanical history. Thermal events are affected by such history and different results for the same chemical species are likely to be generated if the samples have different histories. The main reason for this is that thermal measurement is affected by the internal energy of samples, and the internal energy can be changed by thermal and mechanical processes.

Heating rate probably is the most important parameter to affect TA results. A slow heating rate is often favorable in order to approach thermal equilibrium during TA analysis. A fast heating rate will generate a thermal lag between the heating source and sample. The thermocouple for temperature measurement in a TA instrument is rarely in direct contact with sample. A fast heating rate can generate a temperature gradient surrounding the sample and errors in measurement of sample temperature result.

The atmosphere surrounding the sample serves to transfer heat and supply or remove gaseous reactants or products. Thus, its chemical nature and flow affect the TA data. In most TA methods, we need an inert atmosphere to prevent the sample from oxidation reactions. Maintaining a certain flow rate is important for TA involving a gaseous product; for example, examination of thermal decomposition. The flow will provide a stable partial pressure of products.

10.2 Differential Thermal Analysis and Differential Scanning Calorimetry

Differential thermal analysis (DTA) and differential scanning calorimetry (DSC) are the most widely used thermal analysis techniques. Both techniques have the same objective: to examine thermal events in a sample by heating or cooling without mass exchange with its surroundings. The thermal events examined by DTA and DSC include solid phase transformation, glass transition, crystallization and melting. ‘Differential’ emphasizes that analysis is based on differences between sample material and a reference material in which the examined thermal events do not occur.

10.2.1 Working Principles

Differential Thermal Analysis

A DTA instrument is designed to measure temperature differences between sample and reference as illustrated in Figure 10.2. A sample and reference are symmetrically placed in a furnace. The temperature difference between the sample and reference are measured by two thermocouples: one is in contact with the underside of the sample holder (also called the *cruible*), the other is in contact with the underside of the reference holder. The reference should be made from a material that satisfies the following conditions: it does not undergo thermal events over the operation temperature range, does not react with any component in the instrument, and has similar thermal conductivity and heat capacity to the sample being examined.

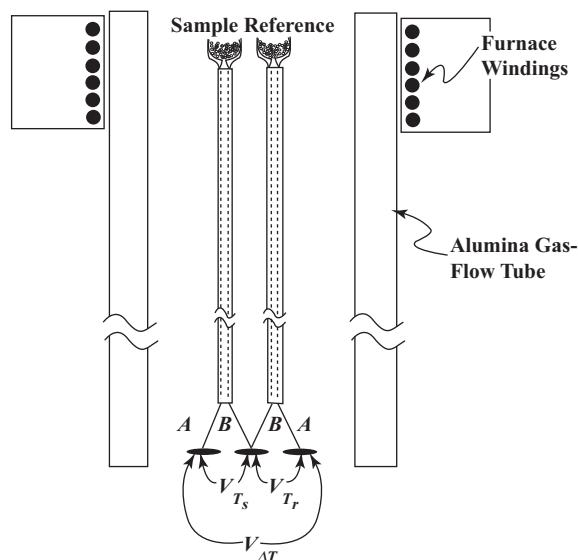


Figure 10.2 Differential thermal analysis (DTA) instrumentation. V_{T_s} and V_{T_r} are the thermocouple voltages for measuring sample and reference temperatures, respectively. (Reproduced with permission from R.F. Speyer, *Thermal Analysis of Materials*, Marcel Dekker, New York. © 1993 Taylor & Francis Group Ltd.)

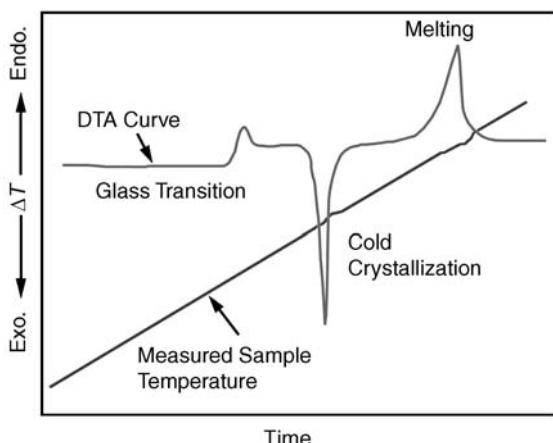


Figure 10.3 DTA curve for a polymer sample under a constant heating rate. (Reproduced with permission from T. Hatakeyama and F.X. Quinn, *Thermal Analysis: Fundamentals and Applications to Polymer Science*, 2nd ed., John Wiley & Sons Ltd, Chichester. © 1999 John Wiley & Sons Ltd.)

The temperature difference between sample and reference (ΔT) should be zero when no thermal event occurs in the sample, because of similar thermal conductivity and heat capacity between sample and reference. When a thermal event occurs in the sample, nonzero ΔT will be generated. ΔT will be negative if the thermal event is *endothermic* (absorbing heat) or positive if the thermal even is an *exothermic* (releasing heat). An endothermic event makes the sample temperature lower than the reference temperature that follows the heating program closely. On the other hand, an exothermic event makes the sample temperature higher than the reference temperature. A DTA curve will reveal thermal events over a certain temperature range as illustrated in Figure 10.3. The DTA curve is a plot of ΔT versus reference temperature or heating time when heating rate is constant.

Differential Scanning Calorimetry

A DSC instrument is designed to measure the heat flow difference between sample and reference. There are two widely used DSC systems: the *heat flux DSC* and the *power-compensated DSC*. The heat flux DSC is also called ‘quantitative DTA’ because it measures the temperature difference directly and then converts it to heat flow difference as illustrated in Figure 10.4a. This conversion is accomplished by an algorithm in computer software installed in the system. The power-compensated DSC directly measures the enthalpy change of a sample during a thermal event. Its design is very different from DTA. In power-compensated DSC, there are two separate chambers to contain the sample and reference. Each chamber has its own individual heat element to control temperature (Figure 10.4b). The instrument always maintains a thermal null state ($\Delta T=0$). When a thermal event occurs in the sample, the power to the heating element has to change in order to maintain the sample temperature the same as that of the reference. An endothermic event causes an increase in power to heat the sample; an exothermic event causes a reduction in power to cool the sample. The amount of power change should equal the energy of heat flow to compensate for the heat release or gain of the sample.

A DSC curve is illustrated in Figure 10.5 in which heat flow is plotted versus temperature. The heat flow has a unit of energy per unit time per unit mass, usually in units of W g^{-1} .

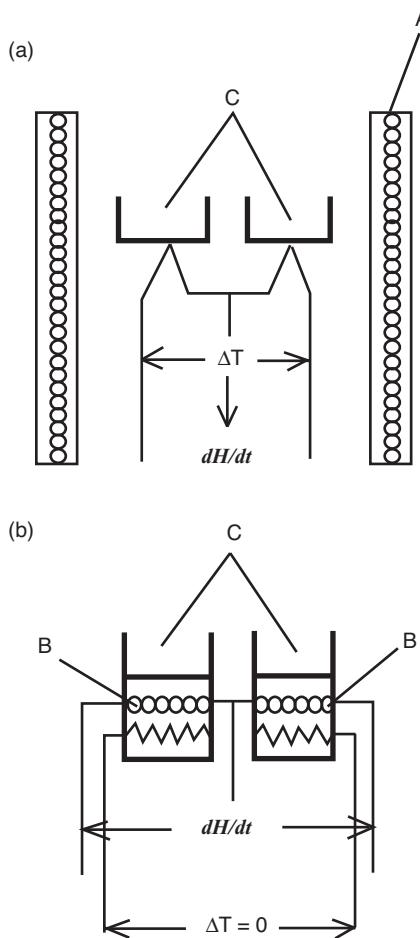


Figure 10.4 Differential scanning calorimetry (DSC) instrumentation design: (a) heat flux DSC; and (b) power compensation DSC. A, furnace; B, separate heaters; and C, sample and reference holders. (Reproduced with permission from E.L. Charsley and S.B. Warrington, *Thermal Analysis: Techniques and Applications*, Royal Society of Chemistry, Cambridge, UK. © 1992 Royal Society of Chemistry.)

Commonly, heat flow into a sample is indicated as an upward feature of the DSC curve. This differs from a DTA curve in which an endothermic event is indicated as a downward feature. The DSC curves are commonly recorded over a temperature range by heating or cooling a sample with a constant rate, similar to DTA curves.

Although DTA and DSC are similar in thermal event measurements and instrumentation, they are different in the following aspects. DTA is a qualitative technique because measured temperature differences do not provide any quantitative data for energy. In contrast, DSC is a quantitative technique because measured heat flow provides enthalpy changes in the sample during a thermal event. DTA can operate over a wider temperature range than DSC. The DTA can reach a temperature of greater than 1500 °C, while the power-compensated DSC is restricted to a maximum temperature of about 750 °C. This DTA feature is important for examining materials with high melting temperature such as ceramics and some metals.

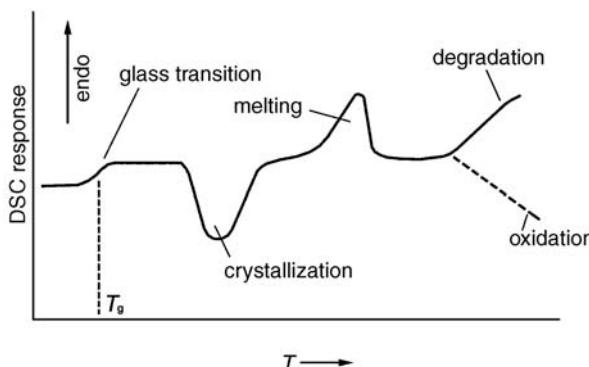


Figure 10.5 Schematic DSC curves for a polymeric sample. T_g , glass transition temperature. (Reproduced with kind permission of Springer Science and Business Media from M. Brown, *Introduction to Thermal Analysis*, Kluwer Academic Publishers, Dordrecht. © 2001 Springer Science.)

Temperature-Modulated Differential Scanning Calorimetry

In conventional DSC measurements, the enthalpy change of thermal events in the sample is measured at a constant heating rate. It means the sample temperature change with time as written as follows.

$$T = T_o + \beta t \quad (10.6)$$

T_o is the starting temperature, t is time and β is heating rate.

Temperature modulated DSC (TMDSC) is an advanced DSC technique in which the heating rate is modulated by superimposing a cyclic heating rate on the constant rate; for example, a sinusoidal temperature modulation is superimposed on the temperature profile.

$$T = T_o + \beta t + B \sin(\omega t) \quad (10.7)$$

B is amplitude of temperature modulation that is commonly in the range of ± 1 to ± 10 K and $\omega = 2\pi p^{-1}$, in which p is the modulation period that is commonly in the range 10–100 s. Figure 10.6 illustrates a TMDSC heating profile. The constant heating profile to the sample can provide total heat flow information while the sinusoidal heating profile gives information corresponding to the rate of temperature change. Raw TMDSC data are rather complex compared with conventional DSC data. TMDSC data require deconvolution to generate standard DSC curves such as the one shown in Figure 10.5.

To understanding the applications of TMDSC, we need to know the heat flow features of DSC. At any moment of a DSC experiment, the heat flow can be expressed as a differential equation.

$$\frac{dH}{dt} = C_P \beta + f(T, t) \quad (10.8)$$

C_P is the sample heat capacity. The first term on the right hand side of Equation 10.8 represents the heat flow of reversing thermal events and the second term represents the heat flow of non-reversing thermal events. $f(T, t)$ is also referred as to the kinetic component of heat flow.

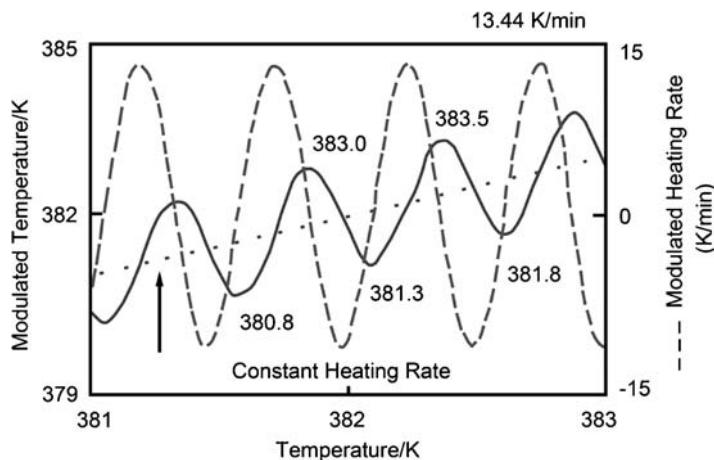


Figure 10.6 Schematic temperature modulated DSC (TMDSC) curves for a polymeric sample. (Reproduced with permission from T. Hatakeyama and F.X. Quinn, *Thermal Analysis: Fundamentals and Applications to Polymer Science*, 2nd ed., John Wiley & Sons Ltd, Chichester. © 1999 John Wiley & Sons Ltd.)

The main advantage of TMDSC is its ability to separate the reversing and non-reversing thermal events. Reversing thermal events include glass transition and fusion (melting). Non-reversing thermal events include oxidation, curing, relaxation and cold crystallization (glass–crystal transition below the melting temperature).

Separating the reversing and non-reversing heat flow components, TMDSC can distinguish overlapping thermal events. Figure 10.7 illustrates the example of using TMDSC to examine a poly(ethylene terephthalate)–poly(acrylonitrile–butadiene–styrene) (PET–ABS) polymer blend. Figure 10.7a shows conventional DSC curves of the polymer blend. The first heating curve reveals the PET glass transition at 340 K, the PET cold crystallization at 394 K and PET fusion at 508 K (not shown). The ABS glass transition at 379 K cannot be observed in the first heating curve because it is overlapped with the PET cold crystallization peak. The ABS glass transition can only be revealed by a second heating after the first heating and cooling. However, TMDSC clearly shows both PET and ABS glass transitions in its reversing curve (Figure 10.7b). The TMDSC non-reversing curve not only reveals the PET cold crystallization, but also a relaxation in the polymer blend at 343 K, which cannot be seen in Figure 10.7a.

10.2.2 Experimental Aspects

Sample Requirements

Samples for DTA or DSC should be in the form of dense powder or small discs. Film, sheets and membranes are often cut into discs fitting into the sample pans. Large shear forces during cutting samples should be avoided because shear-force may induce plastic deformation in a sample and that can affect DTA and DSC curves. Low-mass samples are also preferred because a low-mass sample can quickly reach temperature equilibrium throughout its volume. A large-mass sample will have undesirable internal temperature gradient and this will affect the accuracy of the curve. Commercial DSC instruments are able to measure phase change

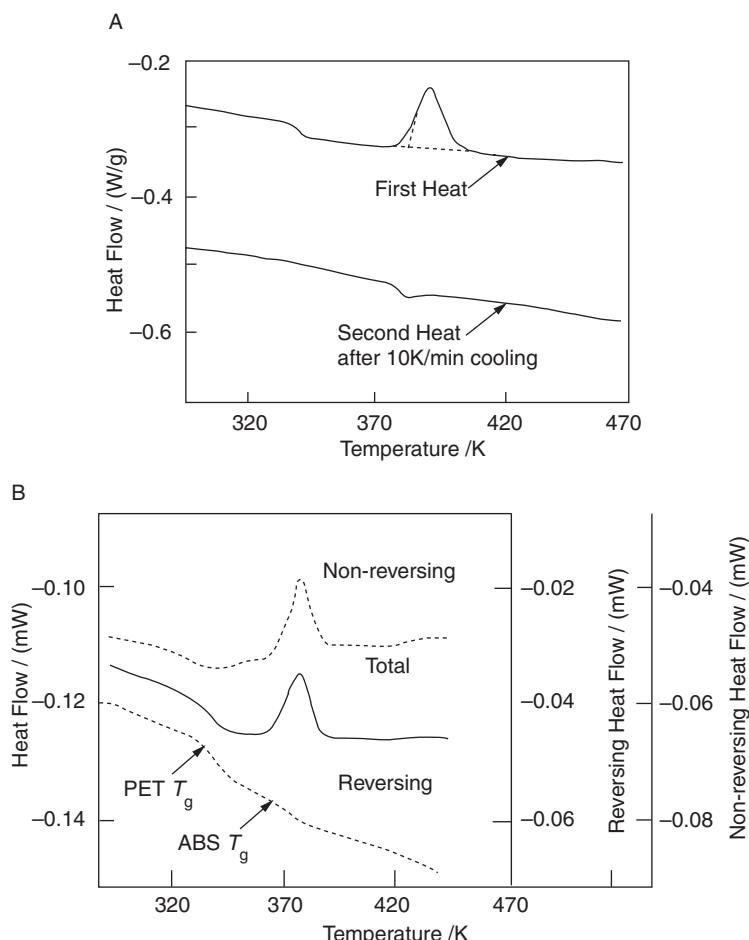


Figure 10.7 DSC curves of poly(ethylene terephthalate)-poly(acrylonitrile-butadiene-styrene) (PET-ABS) blends: (a) conventional DSC first and second heating curves with heating and cooling rate of 10 K min^{-1} ; and (b) temperature modulated DSC (TMDSC) first heating curves with $\beta = 2\text{ K min}^{-1}$, $p = 60\text{ s}$ and $B = \pm 1\text{ K}$. T_g , glass transition temperature. (Reproduced with permission from T. Hatakeyama and F.X. Quinn, *Thermal Analysis: Fundamentals and Applications to Polymer Science*, 2nd ed., John Wiley & Sons Ltd, Chichester. © 1999 John Wiley & Sons Ltd.)

in microgram samples. In practice, the lower limit of sample size relates to the nature of sample materials. For composites and polymer blends, large sample size of about 10 mg may be required.

DTA is usually operated in a high temperature range. Platinum and gold crucibles are commonly used as sample and reference holders. DSC is usually operated in a low temperature range ($<500^\circ\text{C}$), and thus aluminum pans are commonly used as sample and reference holders. The pans often need to be sealed to avoid sample mass change due to evaporation. A special press can be used to mechanically weld a lid and a pan together.

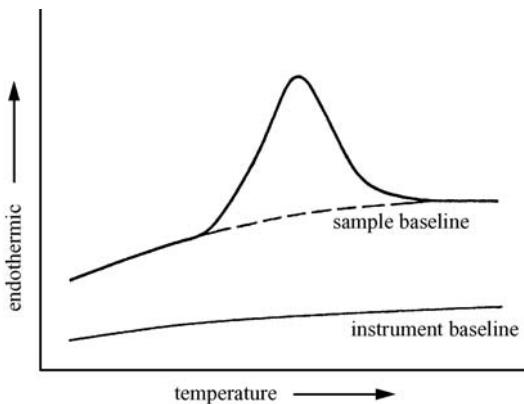


Figure 10.8 Illustration of sample baseline and instrument baseline.

Baseline Determination

Thermal event measurement from a DTA or DSC curve relies on determining the deviation of the experimental curve from the curve baseline as shown by a DSC curve (Figure 10.5). There are two baselines: the *instrument baseline* (also called the *zero line*) and the *sample baseline*, as illustrated in Figure 10.8. The instrument baseline is the curve recorded without the sample over the instrument's operational temperature range. Ideally, the instrument baseline is a straight line over a temperature range. In practice, however, the baseline may be curved and/or display a high level of noise. For example, the instrument may be contaminated with trace amounts of residue from a previous experiment. Decomposition or sublimation of such a residue will distort the instrument baseline. The instrument baseline will also be affected if the flowrate of purge gas in the instrument is not constant or if the purge gas contains a large amount of water vapor. Thus, we should always check the instrument baseline by first running the instrument without a sample.

The sample baseline of an experimental curve, as illustrated in Figure 10.8, is the line that connects the curve before and after a peak, as if no peak had developed. Accuracy of enthalpy change measurement relies on correctly constructing the sample baseline from the curve. Construction of the sample baseline from the measured curve may not be straightforward. Figure 10.9 shows examples of how to construct a baseline using interpolation methods for various curve shapes. The procedure of baseline subtraction is often done automatically as it can be incorporated in the software provided with the instrument.

Effects of Scanning Rate

DTA and DSC experiments are commonly run through a temperature range with a constant heating or cooling rate. This rate should be reported with the DTA and DSC results for reasons that are described here. The rate of scanning through a temperature range affects thermal event measurement. More specifically, the scanning rate can change the curve shape and characteristic temperature indicated in a curve. Figure 10.10 shows scanning rate effects on the DSC curve of PET. Figure 10.10a reveals that the peak area of PET melting increases with increasing heating rate. Figure 10.10b reveals that the crystallization temperature of PET decreases and peak width increases with increasing cooling rate. From Figure 10.10, we note that the scanning rate effects are even more obvious in cooling than in heating. The rate effects on the curves

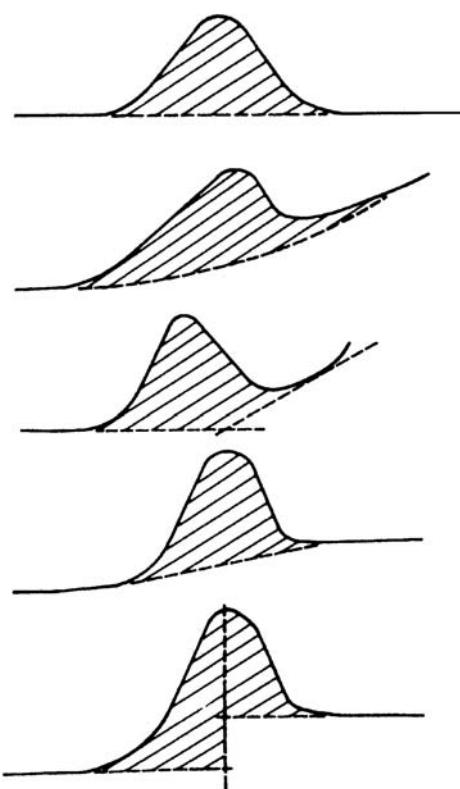


Figure 10.9 Examples of sample baseline interpolation for different types of differential thermal analysis (DTA) and DSC curves. (Reproduced with kind permission of Springer Science and Business Media from M. Brown, *Introduction to Thermal Analysis*, Kluwer Academic Publishers, Dordrecht. © 2001 Springer Science.)

simply reflect that thermal events are kinetic processes. The processes of material change, such as melting, require time for heat transfer and atom or molecule diffusion.

Increase of scanning rate will result in increase of temperature gradient in the sample and the instrument, thus making it difficult for the sample to reach equilibrium. Although a slower rate can make a measurement more accurate, a fast heating rate often makes thermal event peaks more obvious in the curve. The scanning rate of DTA and DSC instruments commonly varies from 0.1 to $40\text{ }^{\circ}\text{C min}^{-1}$. We also should note that the programmed scanning rate may not represent the true rate of sample heating or cooling, because of the possible temperature gradient between sample and thermocouple. Scanning rate effects are difficult to eliminate, particularly for polymeric samples. Thus, the scanning rate should be reported with the DTA and DSC results.

10.2.3 Measurement of Temperature and Enthalpy Change

Transition Temperatures

Measuring transition temperatures is the primary task of DTA and DSC. The common transitions are melting, crystallization and glass transition. Figure 10.11 demonstrates how to

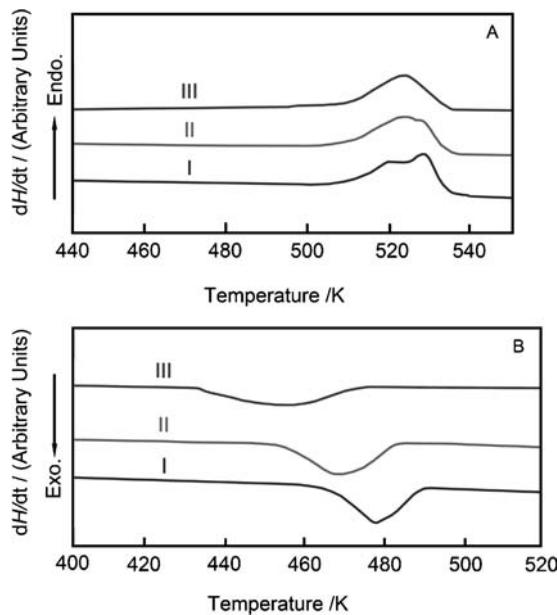


Figure 10.10 Scanning rate effects on DSC curves of a poly(ethylene terephthalate) (PET) sample: (a) heating; and (b) cooling. I, 10 K min^{-1} ; II, 20 K min^{-1} ; III, 40 K min^{-1} . (Reproduced with permission from T. Hatakeyama and F.X. Quinn, *Thermal Analysis: Fundamentals and Applications to Polymer Science*, 2nd ed., John Wiley & Sons Ltd, Chichester. © 1999 John Wiley & Sons Ltd.)

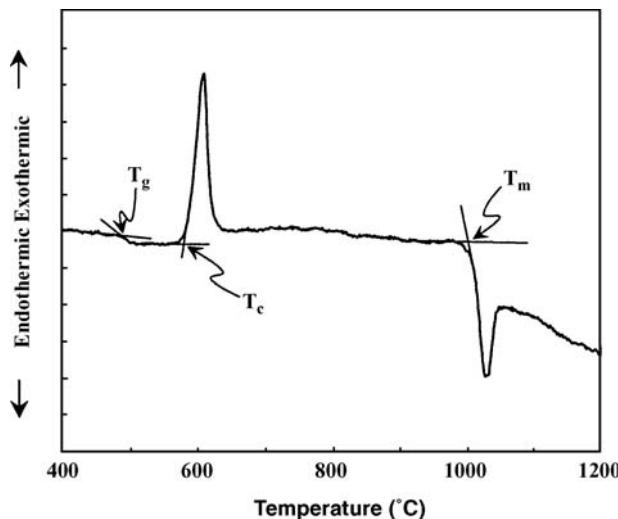


Figure 10.11 DTA curve of glass-ceramic sample ($\text{Li}_2\text{O}-\text{Al}_2\text{O}_3-6\text{SiO}_2$). Temperatures of glass transition (T_g), crystallization (T_c) and melting (T_m) are determined from the curve. (Reproduced with permission from R.F. Speyer, *Thermal Analysis of Materials*, Marcel Dekker, New York. © 1993 Taylor & Francis Group Ltd.)

determine the temperatures of thermal events from a DTA or a DSC curve. The DTA curve of a glass–ceramic sample shown in Figure 10.11 exhibits three thermal events during heating: glass transition, crystallization (or devitrification) and melting. Melting and crystallization exhibit endothermic and exothermic peaks, respectively. The glass transition exhibits only a change in the slope of the curve. The glass transition temperature (T_g) can be determined at the point where the curve deviates from the linearity of the sample baseline. For DSC curves which are not linear, T_g should be determined at the temperature where the curve deviates from the interpolated sample baseline, known as the onset temperature. T_g can more reliably be defined as the intersection between two tangents to the sample baselines before and after slope change when the point of the curve deviating from the sample baseline is not certain, as shown in Figure 10.11. The temperature of the first-order phase transition such as crystallization temperature (T_c) and melting temperature (T_m) should be defined at the intersection of the tangent to the maximum rising slope of the peak and the extrapolated sample baseline. The method to define T_m or T_c can be understood by analyzing the change of sample temperature during DTA scanning at a constant heating rate. When a solid melts, the sample temperature (T_s) remains at the melting temperature while the reference temperature (T_r) continuously increases with heating. Thus, $\Delta T (=T_s - T_r)$ starts to decrease, deviating from the baseline, until melting is completed; thus, ΔT reaches a minimum. The temperature of the melting peak represents the temperature of melting completion. After that, the sample temperature starts to increase to catch up with that of the reference during heating, and the curve increases exponentially after melting is completed. The reason for the exponential curve is that the sample temperature increases more rapidly at the moment right after melting completion but the temperature increase will gradually slow when the temperature gradient between sample and surroundings decreases with increasing sample temperature. Figure 10.12 illustrates an ideal DTA melting peak when the x -axis is the true sample temperature. In a modern DTA instrument, the sample temperature is not directly measured by thermocouple junction in contact with the sample, but in contact with the sample holder. Thus, the DTA peak of melting has a shape as shown in Figure 10.12. In any case, the sharp turning point from the baseline at the DTA curve, shown in Figure 10.12, should represent the melting temperature. Comparing peaks in Figure 10.11 and Figure 10.12, we can conclude that the intersection point best represents the melting temperature. A similar argument should apply to the T_c determination, as well as to the T_m and T_c determinations in

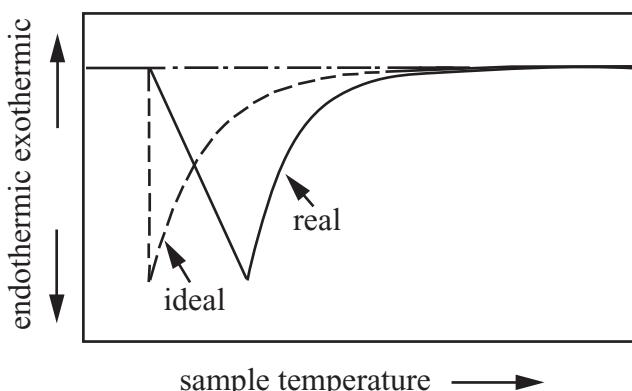


Figure 10.12 Comparison of real and ideal melting peaks in DTA curve at a constant heating rate.

Table 10.3 Standard Materials for Calibration

Standard	Melting Temperature (°C)	Enthalpy change of melting (J g ⁻¹)
Biphenyl	68.93	120.6
Benzil	94.85	110.6
Indium	156.61	28.71
Tin	231.92	60.55
Lead	327.47	23.00
Zinc	419.53	108.6
Aluminum	660.33	401.3

DSC curves. The only difference is that the heat flow in DSC replaces temperature change in DTA.

Measurement of Enthalpy Change

The enthalpy change of phase transformation can be directly measured from the corresponding peak area of a DSC curve. The DSC is commonly plotted as $\frac{dH}{dt}$ per mass unit versus temperature. The total enthalpy change ΔH should be proportional to the peak area (A_p).

$$\Delta H/\text{mass} = K_c A_p \quad (10.9)$$

where K_c is a calibration factor that includes contributions from the experimental conditions and the thermal conductivity of the system. We can obtain K_c by measuring the peak area of the standard sample, for which we know enthalpy change, such as those listed in Table 10.3.

For the power-compensated DSC, K_c is virtually independent of temperature. For the heat flux DSC, K_c becomes temperature dependent and has to be determined at a temperature close to that of the peak to be measured. The A_p measurement relies on interpolation of the sample baseline as illustrated in Figure 10.9.

Calibration of Temperature and Enthalpy Change

Besides the correction procedure in determining temperature and enthalpy change, measurement accuracy relies on whether the instrument is well calibrated for the given experimental conditions. Calibration should be done frequently with standard materials. Standard materials are selected based on their well-characterized melting temperatures and enthalpy changes of melting (also called *latent heat*). Table 10.3 lists commonly used standard materials for temperature and enthalpy calibration. The listed materials cover the temperature range of about 69–660 °C. It is important to calibrate measurements of temperature and enthalpy change by selecting a proper standard material for which the melting temperature falls within the melting temperature range of the sample to be examined.

10.2.4 Applications

DTA and DSC are so similar in their working principles and in characteristics of their curves that we often do not distinguish between these two techniques in their applications for materials characterization. These two techniques are especially useful in characterization of polymeric materials, as well as in characterization of inorganic materials. Some typical applications of DTA and DSC are introduced in this section.

Determination of Heat Capacity

The principle of heat capacity measurement is based on proportionality of the DSC response to the heat capacity. By definition, the heat capacity of materials under constant pressure is the following.

$$C_P \equiv \left(\frac{\partial H}{\partial T} \right)_P \quad (10.10)$$

Figure 10.13 illustrates the DSC curves for heat capacity measurement. Figure 10.13a shows an ideal instrument baseline and a DSC line with sample. The displacement between the baseline

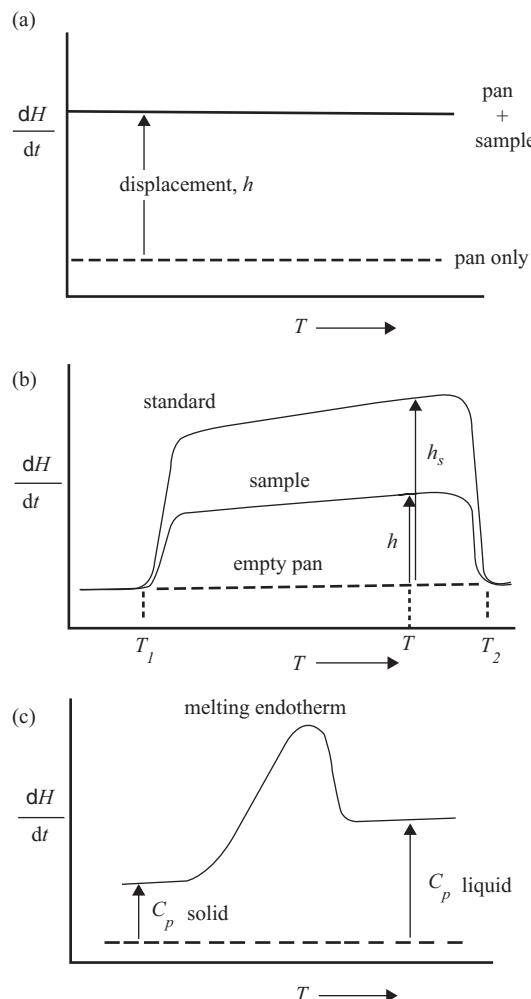


Figure 10.13 Heat capacity measurement: (a) proportionality of heat capacity and DSC response; (b) two displacements required for heat capacity calculation; and (c) heat capacity as a function of temperature. (Reproduced with kind permission of Springer Science and Business Media from M. Brown, *Introduction to Thermal Analysis*, Kluwer Academic Publishers, Dordrecht. © 2001 Springer Science.)

and sample line (h) should be proportional to sample heat capacity.

$$h = B\beta C_p \quad (10.11)$$

The proportional factors are determined by the heating rate (β) and by the calibration factor (B). We can eliminate these factors if a DSC curve of a standard sample is recorded under exactly same experimental conditions. Sapphire is commonly used as the standard with known heat capacity. Figure 10.13b illustrates DSC curves with an empty sample holder (pan), with a sample, and with a standard sample recorded on the same chart. The heat capacity of the sample can be calculated with the heat capacity of the standard (C_{Ps}) and displacements measured from Figure 10.13b.

$$C_P = C_{Ps} \left(\frac{hM_s}{h_s M} \right) \quad (10.12)$$

M_s and M are the mass of standard and sample, respectively, h_s is the displacement between the baseline and the standard line as marked in Figure 10.13b. The displacements h and h_s can be determined graphically.

Correct determination of the instrument baseline is important for accurate heat capacity measurement. The instrument baseline is not likely to be recorded as a horizontal line in experiments. The displacements closed to T_1 and T_2 in Figure 10.13b should not be used for measurement because the heating conditions are not stable there. We should note that the measured heat capacity is more likely to be a function of temperature, particularly when a thermal event occurs in the measured temperature range (Figure 10.13c).

Determination of Phase Transformation and Phase Diagrams

DTA and DSC curves are sensitive to the first-order transition, which exhibits an enthalpy change. First-order transitions include melting and solid-state phase transformations. The curve peaks of DTA and DSC readily reveal solid-state phase transformations over a temperature range. Figure 10.14 shows an example of phase transformation detection. The DSC curve of carbon tetrachloride reveals the solid-state transformation from a monoclinic to rhombohedral then cubic crystal structure during temperature change. Figure 10.14 also tells us that the operating temperature of a DSC can be extended to the cryogenic range. A cooling medium such as liquid nitrogen has to be used for measurement at cryogenic temperatures. We can determine a phase diagram of a materials system by determining transformation temperatures as a function material composition. Figure 10.15 shows an example of a binary phase diagram determination. A series of cooling curves starting from the liquid state of materials are recorded. The whole composition range of a binary system, from 100% component A to 100% component B, is measured. In Figure 10.15, the liquidus lines (the curved lines in Figure 10.15), and the eutectic temperature and composition are determined from the transformation temperatures as revealed by DSC or DTA curves.

Applications to Polymers

Polymeric materials are more sensitive to temperature than metals and ceramics. A large number of polymers exhibit physical and chemical property changes with changes in temperature as

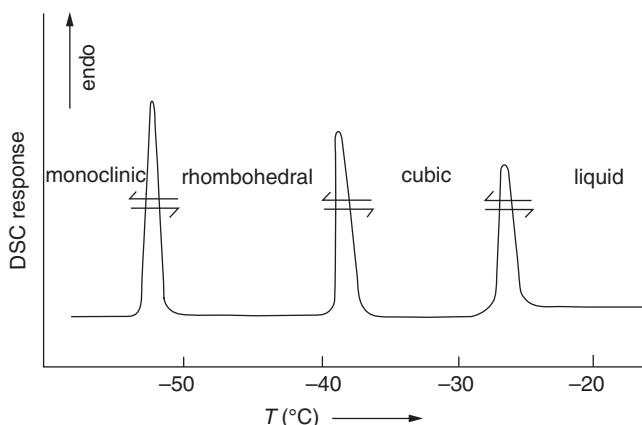


Figure 10.14 The DSC curve of carbon tetrachloride exhibits three solid-state phase transformations before melting. (Reproduced with kind permission of Springer Science and Business Media from M. Brown, *Introduction to Thermal Analysis*, Kluwer Academic Publishers, Dordrecht. © 2001 Springer Science.)

illustrated in Figure 10.5, a DSC curve of a quenched crystalline polymer. Besides determining the temperatures of glass transition, crystallization and melting, we can also determine special properties of polymers such as crystallinity, curing status, polymer content and stability.

A large number of polymeric materials have both amorphous and crystalline structures. The melting temperature is that of a crystalline solid to liquid transformation. We can estimate the crystallinity of polymer (X_c) by comparing its melting ΔH with that of its fully crystalline counterpart (ΔH_{100}).

$$X_c = \frac{\Delta H}{\Delta H_{100}} \quad (10.13)$$

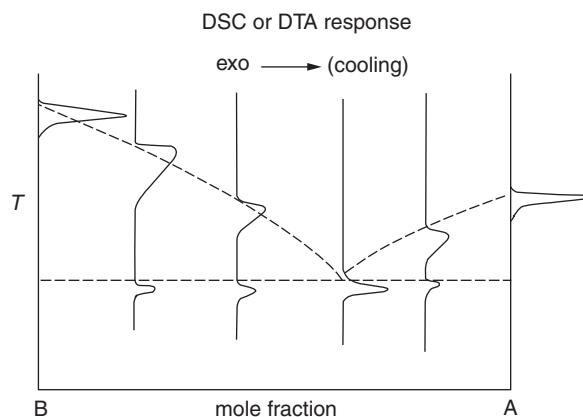


Figure 10.15 Determining a binary phase diagram containing a eutectic reaction by DSC or DTA. (Reproduced with kind permission of Springer Science and Business Media from M. Brown, *Introduction to Thermal Analysis*, Kluwer Academic Publishers, Dordrecht. © 2001 Springer Science.)

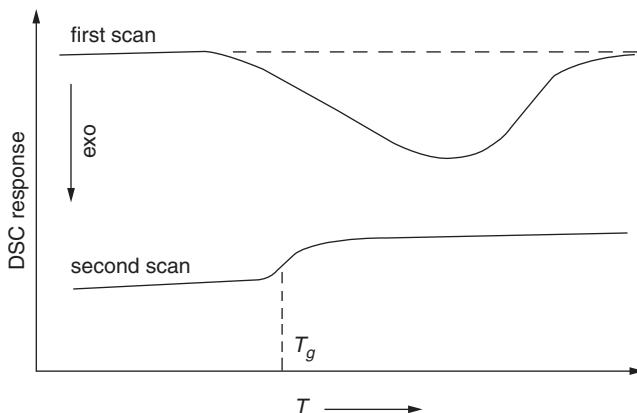


Figure 10.16 Examining the curing status of epoxy resins. The first scan reveals the curing process and the second scan shows a completion of curing. (Reproduced with kind permission of Springer Science and Business Media from M. Brown, *Introduction to Thermal Analysis*, Kluwer Academic Publishers, Dordrecht. © 2001 Springer Science.)

The simple method based on Equation 10.13, however, is not commonly used because 100% pure crystalline samples for most polymers are not available. One alternative approach is to use the *fusion enthalpy*, the latent heat of melting, of chemical repeated units (ΔH_u) to replace ΔH_{100} in the calculations. ΔH_u can be calculated using the Flory relationship for the depression of the equilibrium melting temperature of a homopolymer due to the presence of low molecular mass diluents. The ΔH_u values of some polymers are available in literature.

A DSC curve can reveal the curing status of thermosetting polymers as illustrated in Figure 10.16. During curing, polymer chains cross-link; this process releases heat, which is reflected in the DSC curve. When a polymer is going through a curing process, its DSC curve will show a broad peak at the curing temperature, as shown in the curve of the first scan in Figure 10.16. When cross-linking is complete, the curing peak disappears and is replaced by a feature of glass transition as shown in the curve of the second scan in Figure 10.16.

DSC curves can also be used to identify individual polymers in a polymer mixture. This is a rather unique capability of DSC or DTA. Figure 10.17 shows DSC curves of plastic waste. The individual melting peaks reveal the waste's polymer content. By comparing these curves with DSC curves of pure polymers, we can assign the melting peaks as those of low density polyethylene (LDPE), high density polyethylene (HDPE), polypropylene (PP), Nylon-6, Nylon-66 and polytetrafluoroethylene (PTFE).

A DSC curve can also reveal the thermal stability of polymers. For example, a polyethylene sample is examined in isothermal conditions at 200 °C with an isothermal DSC curve versus time (Figure 10.18). The time when polyethylene oxidation and degradation occurs is revealed by deviation of the DSC curve in Figure 10.18. Comparing the times of oxidation for polyethylene with and without using a stabilizer during processing, we know that the stabilizer has significantly increased polyethylene's resistance to oxidation at elevated temperatures.

10.3 Thermogravimetry

Thermogravimetry (TG) is a technique for measuring mass change of a sample with temperature. A sample to be measured is placed in a furnace and its mass change is monitored

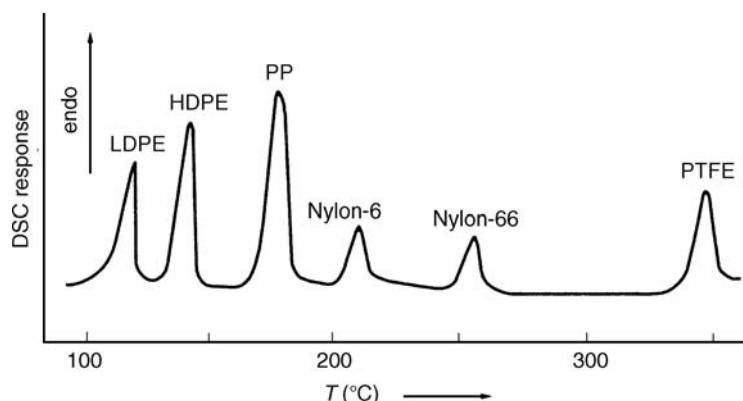


Figure 10.17 DSC curves of plastic waste containing several polymers: low density polyethylene (LDPE), high density polyethylene (HDPE), polypropylene (PP), Nylon-6TM, Nylon-66TM and polytetrafluoroethylene (PTFE). (Reproduced with kind permission of Springer Science and Business Media from M. Brown, *Introduction to Thermal Analysis*, Kluwer Academic Publishers, Dordrecht. © 2001 Springer Science.)

by a *thermobalance*. The main application of TG is to analyze material decomposition and thermal stability through mass change as a function of temperature in scanning mode or as a function of time in the isothermal mode. TG curves are plotted as mass change expressed in percent versus temperature or time. Figure 10.19 illustrates a typical TG curve in which the mass change is plotted against increasing temperature. Decomposition of a sample is represented by two characteristic temperatures: T_i and T_f . T_i is the lowest temperature when the onset of mass change is detected and T_f is the lowest temperature when the mass change is completed.

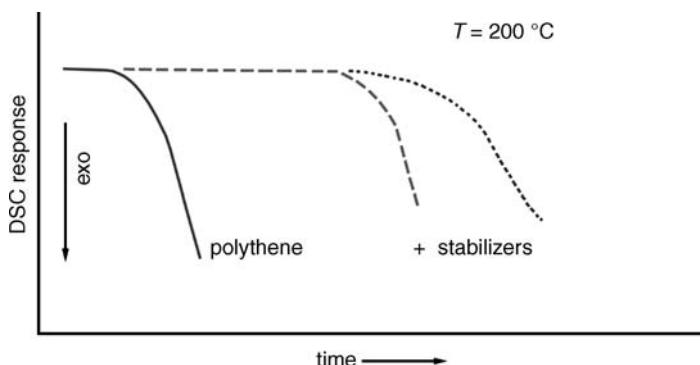


Figure 10.18 Isothermal DSC curve of polyethylene demonstrating the effectiveness of stabilizer to improve resistance to oxidation. (Reproduced with kind permission of Springer Science and Business Media from M. Brown, *Introduction to Thermal Analysis*, Kluwer Academic Publishers, Dordrecht. © 2001 Springer Science.)

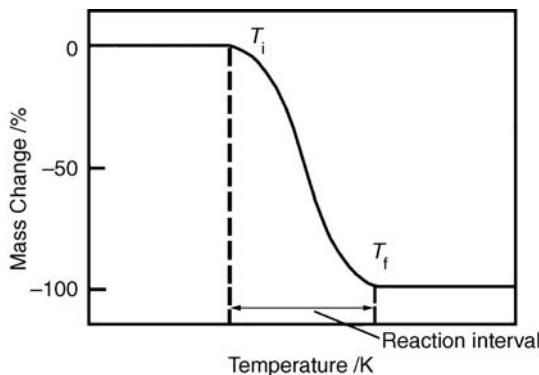


Figure 10.19 Thermogravimetric curves that exhibit decomposition starting temperature T_i and finish temperature T_f . (Reproduced with permission from T. Hatakeyama and F.X. Quinn, *Thermal Analysis: Fundamentals and Applications to Polymer Science*, 2nd ed., John Wiley & Sons Ltd, Chichester. © 1999 John Wiley & Sons Ltd.)

10.3.1 Instrumentation

The TG thermobalance structure is illustrated in Figure 10.20: it includes a microbalance, furnace, temperature programmer and computer. The key component is the microbalance, which measures the mass change. A typical microbalance is able to measure mass change of $\pm 1 \mu\text{g}$ with maximum mass of 100 mg. A commonly used microbalance is the *null-point* type. The null-point microbalance can maintain the sample in a vertical position when its mass changes. Figure 10.21 shows the structure of the commonly used null-point type microbalance, the *Cahn microbalance*.

The Cahn microbalance senses the vertical displacement of a sample due to mass change using an optical system. The optical system includes a light source, a flag, a light tube and a photodiode. The flag beneath the balance arm interferes with light propagating from the source to the light detector (photodiode) when the mass change is sensed by the balance beam.

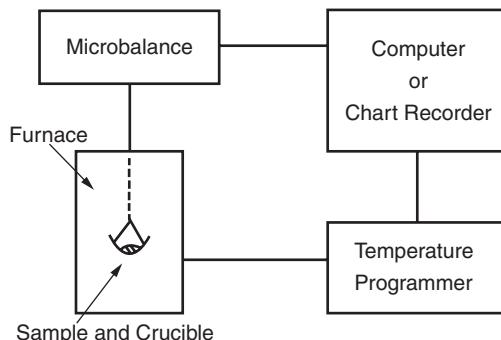


Figure 10.20 Structure of instrumentation. (Reproduced with permission from T. Hatakeyama and F.X. Quinn, *Thermal Analysis: Fundamentals and Applications to Polymer Science*, 2nd ed., John Wiley & Sons Ltd, Chichester. © 1999 John Wiley & Sons Ltd.)

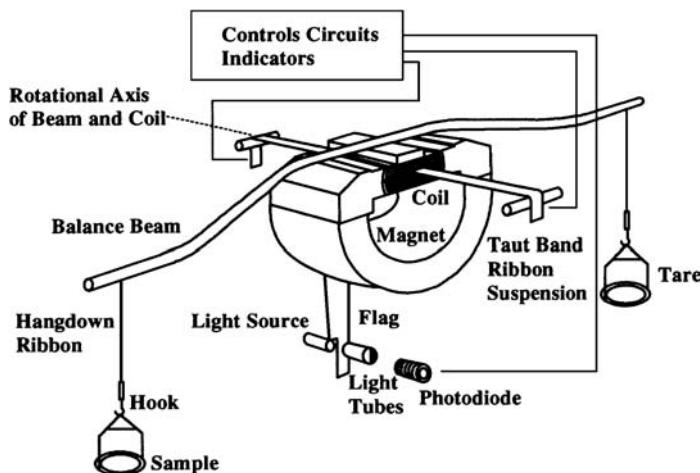


Figure 10.21 Structure of a Cahn microbalance with capability to measure the mass loss of a sample without the sample moving downward. (Reproduced with permission from R.F. Speyer, *Thermal Analysis of Materials*, Marcel Dekker, New York. © 1993 Taylor & Francis Group Ltd.)

A feedback control system adjusts the current in a magnetic-coil system and maintains the balance beam in its original horizontal position even if the sample mass keeps changing.

The most common arrangement in TG instruments is when the sample is connected to the microbalance by a suspension wire. The sample is located in the centre of a furnace as illustrated in Figure 10.22. The furnace is designed as a cylindrical tube with heating elements on its wall. The diameter of the tube is such that it can accommodate a sample with relatively little open space. A thermocouple is located in the tube axis and under the sample holder with a small gap in between. Such a design can largely eliminate any temperature gradient among the sample, thermocouple and heating elements; it enables them all to reach thermal equilibrium quickly. During operation, protective gas will flow into and out of the furnace tube to maintain an inert atmosphere during heating.

10.3.2 Experimental Aspects

Samples

Sample mass, volume and form are important for recording accurate and reproducible TG curves. Reliable TG curves rely on minimization of deviation between sample temperature and programmed temperature. Deviation typically results from either endothermic or exothermic reactions in the sample and subsequent heat transfer between heat source and sample. Sample mass is the most important parameter affecting TG curves. In general, small sample mass is better than large mass for minimizing temperature deviation. TG sample mass is usually about several milligrams. The low limit of sample mass depends on the resolution limit of the microbalance.

Samples can be in block, flake, fiber or powder form. Sample form is the second most important parameter affecting TG curves. The sample form effect on a TG curve is demonstrated in Figure 10.23, which shows the TG curves of different forms of polymethyl methacrylate

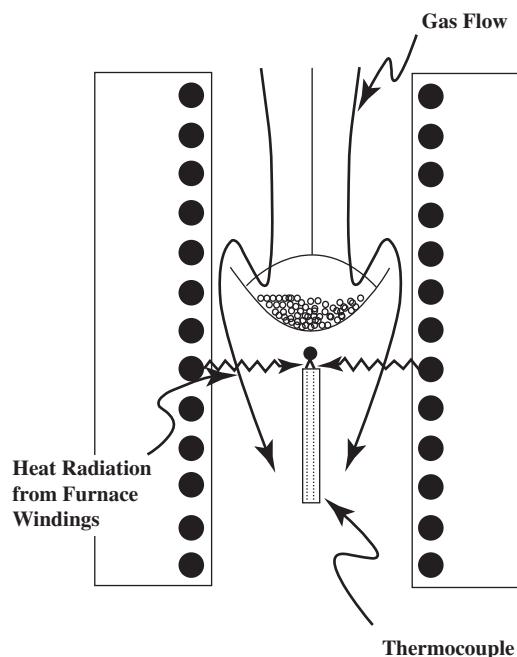


Figure 10.22 Sample position in furnace in a TG system with protective gas flow downward. (Reproduced with permission from R.F. Speyer, *Thermal Analysis of Materials*, Marcel Dekker, New York. © 1993 Taylor & Francis Group Ltd.)

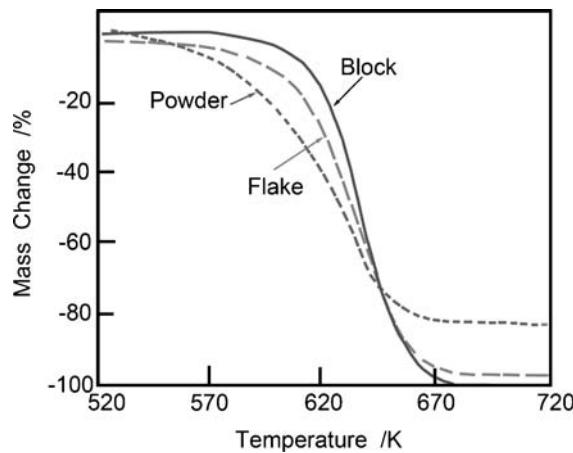


Figure 10.23 Effects of sample form on TG curve of a polymethyl methacrylate (PMMA) sample. Experimental conditions: 5 mg PMMA and a heating rate of 5 K min^{-1} . (Reproduced with permission from T. Hatakeyama and F.X. Quinn, *Thermal Analysis: Fundamentals and Applications to Polymer Science*, 2nd ed., John Wiley & Sons Ltd, Chichester. © 1999 John Wiley & Sons Ltd.)

(PMMA). Decomposition of the powdered PMMA sample occurs at lower temperature than it does with other forms, as shown in the TG curves. A block sample should be ground or sliced to obtain a suitable form for TG examination. We should avoid excessive force during such mechanical processing, because mechanical deformation of the sample may also affect the TG curves.

Atmosphere

TA can be run in either a reactive or non-reactive atmosphere. Reactive atmospheres include corrosive, oxidizing and reducing gases. Non-reactive atmospheres should be an inert gas with little water vapor. Dry Ar and N₂ are commonly used for non-reactive atmospheres. Gas flows through the furnace tube around the sample and carries volatile products out. A flow rate of 15–25 ml min⁻¹ is recommended for a sample mass of about 2–10 mg.

Gas flow, however, generates concerns about possible thermal shielding and/or buoyancy effects. Thermal shielding can be understood from Figure 10.22. Gas flow downward to the sample holder will make the holder shield the thermocouple junction from convective heat transfer through the flowing gas. Consequently, the thermocouple junction temperature will not be the same as for a sample with flowing gas. The buoyancy effect is illustrated in Figure 10.24. Gas flow upward will generate a lifting force on the sample holder. This effect is similar to weight loss of a sample body in liquid. Such gas flow effects are difficult to eliminate. Careful calibration and good instrument design reduce such problems to a minimum.

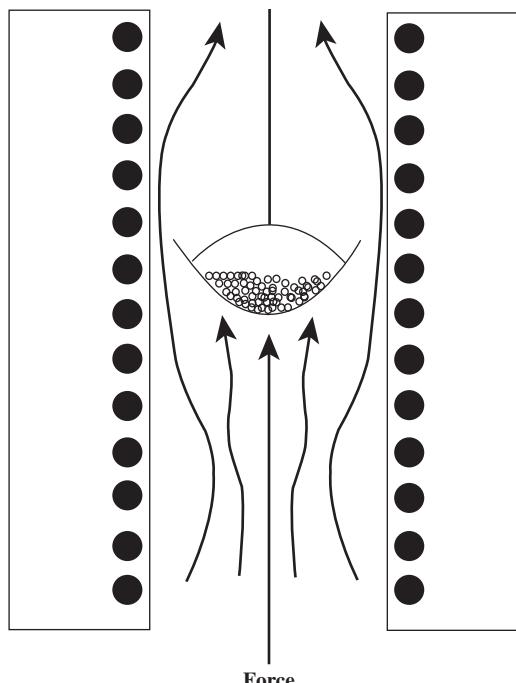


Figure 10.24 Buoyancy effects of upward protective gas flow. (Reproduced with permission from R.F. Speyer, *Thermal Analysis of Materials*, Marcel Dekker, New York. © 1993 Taylor & Francis Group Ltd.)

Temperature Calibration

Temperature calibration is more complicated in TG instruments than in other type of TA instruments because the thermocouple junction cannot touch either the sample or sample holder. Good TA results can only be ensured with careful temperature calibration. A *Curie point* method has been used to calibrate the temperature of TG instruments. The Curie point is the transition temperature of ferromagnetic materials where they lose ferromagnetism. Figure 10.25 illustrates the Curie point method. A magnet is placed below the furnace. The total weight, as measured by the microbalance, is the sum of sample weight plus the magnetic downward force below the Curie point. When the Curie point of a ferromagnetic sample is reached, the magnetic downward force disappears and weight loss will be recorded in the TA curve as illustrated in Figure 10.25b. The TG temperature can be calibrated with several ferromagnetic materials with the Curie point ranging 163–1000 °C as shown in Figure 10.25c.

The *fusible line method* can also be used as a calibration temperature. A fusible wire with known melting temperature connects the microbalance suspension wire and inert mass pieces. The fusible wire should locate the position of sample in the furnace. Melting of the wire at its melting temperature causes mass loss, which is recorded in the TG curve.

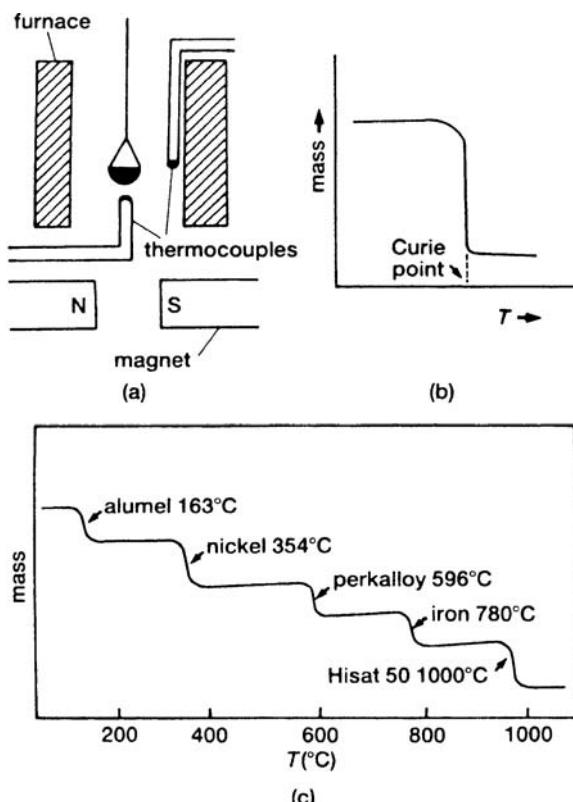


Figure 10.25 The Curie point method for temperature calibration. (Reproduced with kind permission of Springer Science and Business Media from M. Brown, *Introduction to Thermal Analysis*, Kluwer Academic Publishers, Dordrecht. © 2001 Springer Science.)

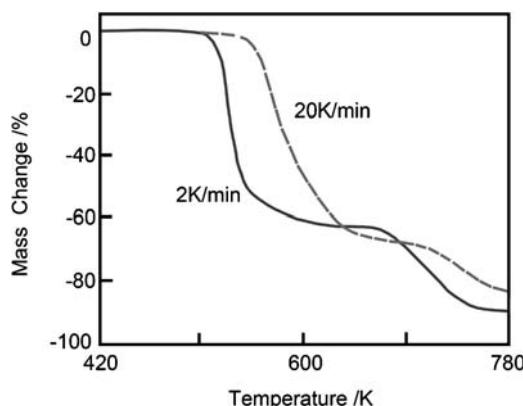


Figure 10.26 Effects of heating rates on TG curves of powdered polyvinyl chloride (PVC). Experimental conditions: 5 mg in dry N_2 gas with a flow rate of 20 ml min^{-1} . (Reproduced with permission from T. Hatakeyama and F.X. Quinn, *Thermal Analysis: Fundamentals and Applications to Polymer Science*, 2nd ed., John Wiley & Sons Ltd, Chichester. © 1999 John Wiley & Sons Ltd.)

Heating Rate

The heating rate affects TG curves considerably, similar to DTA and DSC. For example, with endothermic decomposition, a high heating rate will increase the starting and finishing temperatures of decomposition. Also, the temperature range from start to finish will be wider at a higher heating rate than a lower heating rate. Figure 10.26 shows an example of the effect of heating rate on the TG curve of polyvinyl chloride (PVC). There are two reasons for heating rate effects. First, high heating rate is more likely to generate a temperature difference between the sample and thermocouple junction. The real sample temperature may lag behind that of the thermocouple. Second, in decomposition with volatile products, it takes time for those products to diffuse out of the sample and to be carried away by flowing gas. A low heating rate is more likely to generate thermal equilibrium and give a reproducible result for the analysis. A heating rate of about $5\text{--}10 \text{ }^{\circ}\text{C min}^{-1}$ is recommended for TG examination. For high resolution TG examination, the rate can be reduced to $1 \text{ }^{\circ}\text{C min}^{-1}$.

We can also use stepwise heating and dynamic heating instead of linear heating in TG. Stepwise heating increases the temperature in small increments and then maintains the temperature for a certain period. This heating manner may generate a quasi-isothermal condition that provides sufficient time for thermal equilibrium and diffusion. Dynamic heating is flexible heating. The heating rate changes smoothly and continuously in response to the rates of sample decompositions. For example, the heating rate is fast in the temperature regions without thermal events, but the rate is slow in the temperature regions where sample decompositions occur.

10.3.3 Interpretation of Thermogravimetric Curves

Types of Curves

TG curves can be classified into seven types as illustrated in Figure 10.27. Type (i) is a nearly horizontal line that indicates there is no decomposition with mass loss of volatile products

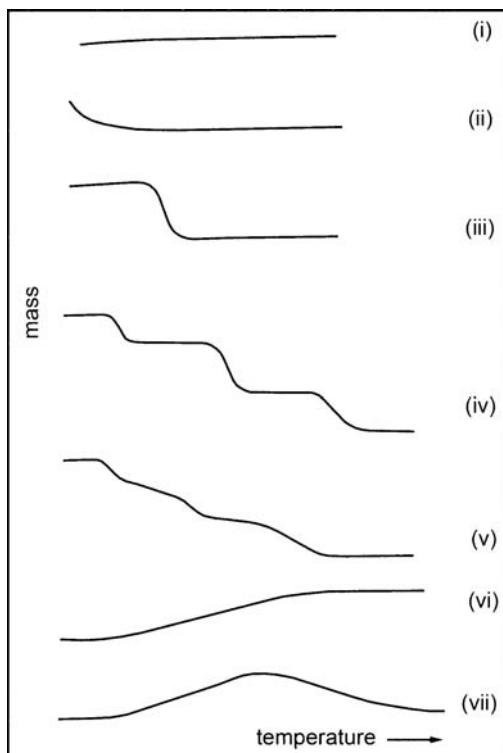


Figure 10.27 Classification of thermogravimetric curves into seven types, each described in the text. (Reproduced with kind permission of Springer Science and Business Media from M. Brown, *Introduction to Thermal Analysis*, Kluwer Academic Publishers, Dordrecht. © 2001 Springer Science.)

over the temperature range. Use of other types of TA techniques may be necessary to find out whether a thermal event has occurred in the temperature range. Type (ii) indicates a rapid mass loss at the initial stage of a TA curve. It is likely that the sample has gone through drying or desorption. Type (iii) is a one-stage decomposition curve which is typical in TG curves. It can define the stability limit of a sample. Type (iv) is a curve of multi-stage decomposition with stable intermediates. Type (v) is also a curve of multi-stage decomposition, but there is no stable intermediate. Type (v) may be a high heating-rate version of type (iv). Rerunning a TG analysis of a sample showing a type (v) curve with a low heating rate is necessary. Type (vi) indicates that a chemical reaction with mass gain has occurred in the sample. A typical example is oxidation of metal samples. Type (vii) indicates a mass-gain reaction occurs and then a mass-loss reaction occurs at a higher temperature in the sample, which is rarely seen.

A slope change of a TG curve is the main feature used to analyze a sample. Sometimes, the slope change is uncertain; in this case, a derivative TG curve can be used. The derivative TG (DTG) curve is a plot of $\frac{dm}{dT}$ versus temperature. Figure 10.28 shows comparison of TG and corresponding DTG curves. A peak in a DTG curve represents a maximum of mass change rate. DTG does not contain any new information other than the original TG curve; however, it clearly identifies the temperature at which mass loss is at a maximum.

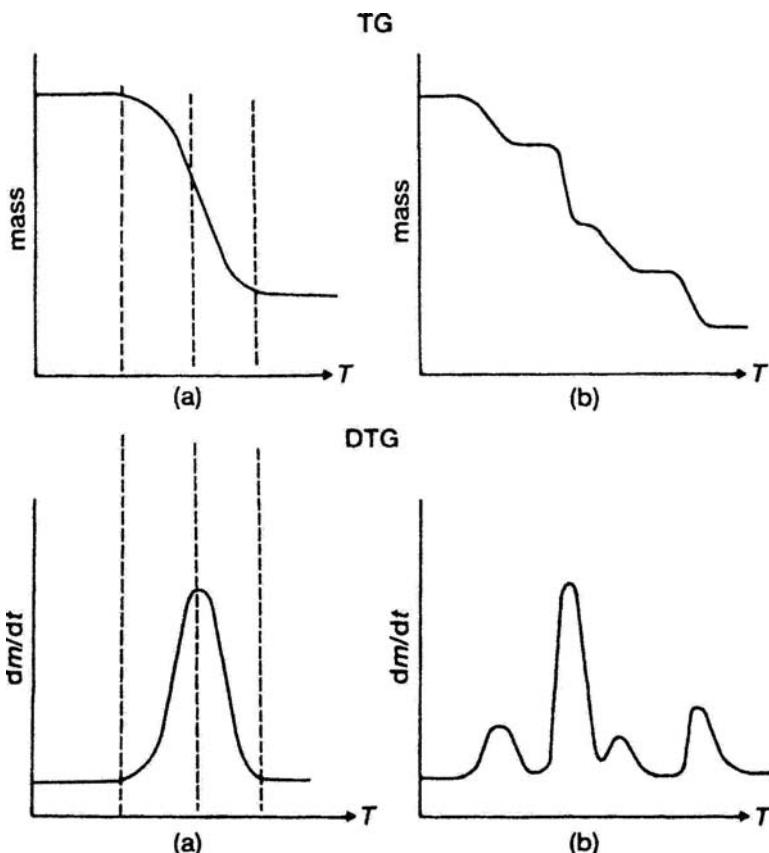


Figure 10.28 Schematic comparison of (a) TG curves; and (b) corresponding DTG curves. (Reproduced with kind permission of Springer Science and Business Media from M. Brown, *Introduction to Thermal Analysis*, Kluwer Academic Publishers, Dordrecht. © 2001 Springer Science.)

Temperature Determination

The determination of characteristic temperatures in TG curves is similar to that in DTA and DSC curves. As illustrated in Figure 10.29, the temperature at which decomposition starts is defined as the intersection of initial line tangent and tangent of line portion when the slope changed. The finishing temperature of decomposition is defined in a similar manner, as shown in Figure 10.29. A mid-point temperature between the starting and finishing temperature can be defined as T_B , as shown in Figure 10.29.

10.3.4 Applications

The TG technique is simple but effective for assessing thermal stability and chemical reactions by monitoring mass change in materials. Several reactions involve mass change, including dehydration, desorption, decomposition and oxidation. Figure 10.30 shows an example of TG analysis of the thermal stability of $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$. Crystal water loss occurs at separated

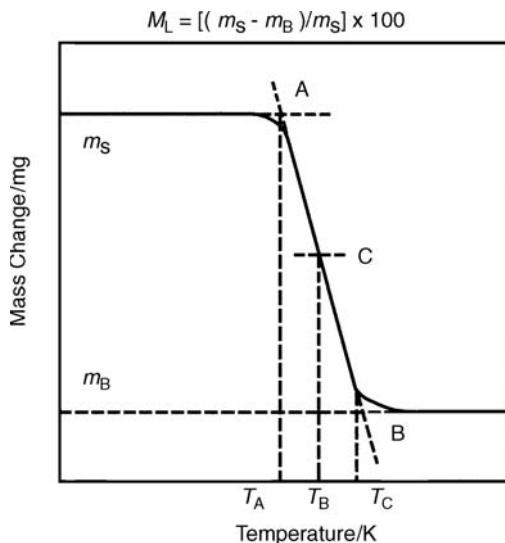


Figure 10.29 Temperature determination from a single-stage TG curve. (Reproduced with permission from T. Hatakeyama and F.X. Quinn, *Thermal Analysis: Fundamentals and Applications to Polymer Science*, 2nd ed., John Wiley & Sons Ltd, Chichester. © 1999 John Wiley & Sons Ltd.)

temperatures. Correspondingly, the structure of a sample goes through several stages of change during crystal water loss. The thermal stability of polymeric materials is often a concern. The way in which TG characterizes thermal stability of polymers is like a ‘fingerprint’ method. Other characterization methods are often required to help us determine the exact nature of reactions revealed by a TG curve.

TG curves not always show obvious decomposition temperatures like Figure 10.30. It is often desirable to plot DTG curves with TG curves to reveal the decomposition

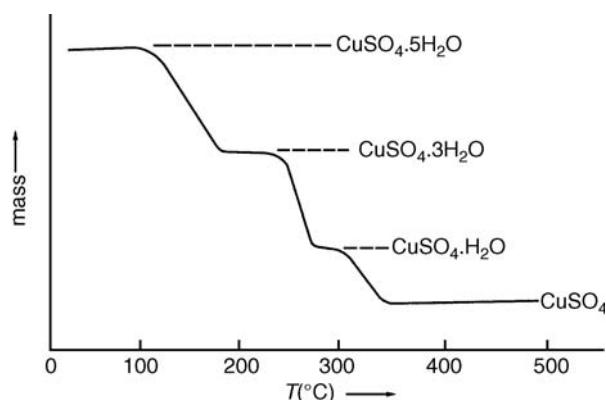


Figure 10.30 TG curves of $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ in a range of ambient temperature to 500°C . (Reproduced with kind permission of Springer Science and Business Media from M. Brown, *Introduction to Thermal Analysis*, Kluwer Academic Publishers, Dordrecht. © 2001 Springer Science.)

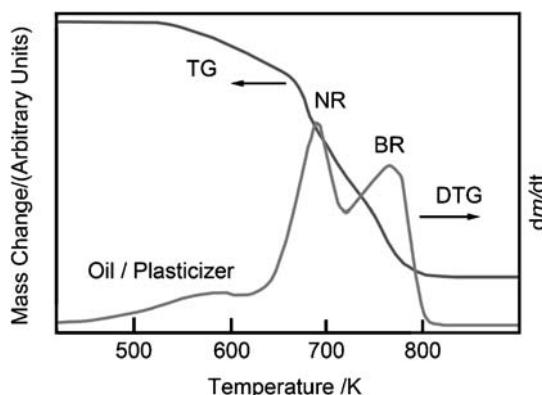


Figure 10.31 TG and DTG curves of a natural rubber–butadiene rubber blend. (Reproduced with permission from T. Hatakeyama and F.X. Quinn, *Thermal Analysis: Fundamentals and Applications to Polymer Science*, 2nd ed., John Wiley & Sons Ltd, Chichester. © 1999 John Wiley & Sons Ltd.)

temperatures of polymers. Figure 10.31 shows an example of polymer blend decomposition. Although TG curves revealed the decomposition of the polymer blend of a natural rubber and a butadiene rubber, only the DTG curves could clearly indicate the decomposition temperature of natural rubber, as well as the higher decomposition temperature of butadiene rubber.

TG curves can also be used for quantifying compositions of composites containing thermally decomposable components. Figure 10.32 shows an example of using the TG curves to determine the weight fraction of ceramic content in a polymer matrix. It is quite often that the real content of particulates in a composite is not the same as the nominal content. Figure 10.32 demonstrates that the amount of ceramics (hydroxyapatite) in composites can be accurately

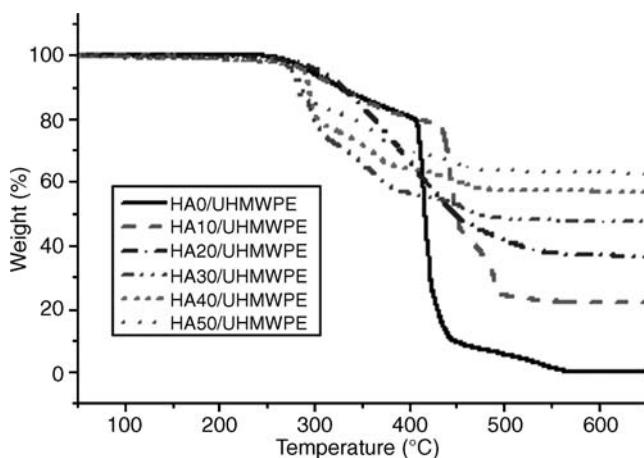


Figure 10.32 TG curves of hydroxyapatite (HA)–ultra-high molecular weight polyethylene (UHMWPE) composites with different HA content. The legends indicate the nominal volume fraction of HA in the UHMWPE matrix.

determined by measuring the weight differences in composite samples before and after the polymer (ultra-high molecular weight polyethylene) matrix is completely decomposed during heating in TG analysis.

References

- [1] Brown, M. (2001) *Introduction to Thermal Analysis*, Kluwer Academic Publishers, Dordrecht.
- [2] Hatakeyama, T. and Quinn, F.X. (1999) *Thermal Analysis: Fundamentals and Applications to Polymer Science*, 2nd edition, John Wiley & Sons, Chichester.
- [3] Charlsley, E.L. and Warrington, S.B. (1992) *Thermal Analysis: Techniques and Applications*, Royal Society of Chemistry, Cambridge, UK.
- [4] Speyer, R.F. (1993) *Thermal Analysis of Materials*, Marcel Dekker, New York.
- [5] Hemminger, W. and Höhne, G. (1984) *Calorimetry*, Verlag Chemie, Weinheim.
- [6] Haines, P.J. (2002) *Principles and Thermal Analysis and Calorimetry*, Royal Society of Chemistry.
- [7] Wunderlich, B. (2005) *Thermal Analysis of Polymeric Materials*, Springer, Berlin.
- [8] Groenewoud, W.M. (2001) *Characterization of Polymers by Thermal Analysis*, Elsevier, Amsterdam.
- [9] Scheirs, J. (2000) *Compositional and Failure Analysis of Polymers*, John Wiley & Sons, New York.

Questions

- 10.1 What are thermal events of materials?
- 10.2 List similarities and differences of DTA and DSC.
- 10.3 What thermal events of a semi-crystalline polymer sample will be detected by DSC? List them from low to high temperature.
- 10.4 Why should sample mass be at the microgram level?
- 10.5 Why is a powder sample favorable in TA?
- 10.6 Sketch a DSC curve that shows a melting peak with a low and a high heating rate.
- 10.7 What material property does the slope change of a DSC curve indicate?
- 10.8 Why does a melting temperature not correspond to a peak temperature in a DTA curve?
- 10.9 People often run a sample twice in DSC or DTA. First, the sample is heated and cooled at a constant rate, then the sample is heated again to obtain the DSC curve for analysis. What are the reasons for this?
- 10.10 Does TG always measure mass loss of sample? Why?
- 10.11 Does the Cahn microbalance balance mass change of a sample using a counter weight? Why?
- 10.12 How can TA be used to detect inorganic impurities in a polymer sample? How can it be used to detect organic impurities in a polymer sample?

Index

- aberration, 6
 - astigmatism, 6, 138
 - chromatic, 6
 - curvature of field, 6
 - spherical, 6
- absorbance, 273
- absorption coefficient, 48
- absorption contrast, 91
- absorption edge, 49
- absorptivity, 297
- acceleration voltage, 81, 121
- achromat, 13
- adatom, 150
- adhesive mounting, 18
- Airy disk, 4
- analyzing crystal, 180
- anti-Stokes scattering, 259
- aperture diaphragm, 10
- apoachromat, 13
- atomic number effect, 193
- atomic scattering factor, 57
- Auger electron, 172, 198
- axial-divergence error, 60
- backscattered electrons, 123
- backscattering coefficient, 133
- barrier filter, 37
- baseline, 310
- beam brightness, 125
- beam splitter, 270
- Beer's law, 297
- bending contours, 113
- bending mode, 260
- beryllium window, 184
- binding energy, 197
- birefringence, 30
- Bragg angle, 64, 72
- Bragg conditions, 51
- Bragg's law, 49, 72
- Bragg-Brentano arrangement, 60
- bright-field, 25, 93
- brightness, 5
- buoyancy effect, 324
- Cahn microbalance, 321
- camera constant, 102
- camera length, 102
- capillary force, 153
- Cassegrain lens, 276
- cathode ray tube (CRT), 122
- cationisation, 241
- characteristic X-rays, 46
- charge neutralization, 215
- charged-coupled device (CCD), 283
- chemical shift, 210, 212
- cold crystallization, 309
- collision cascade, 226
- collision sputtering, 226
- color temperature, 9
- compositional contrast, 133
- concentric hemispherical analyzer, 204
- condenser annulus, 29
- confocal diaphragm, 281
- confocal plane, 39
- conjugate aperture planes, 10
- conjugate field planes, 10
- conjugate focal planes, 10
- constant analyzer energy, 205
- constant current mode, 149
- constant height mode, 150
- constant retarding ratio, 205

- constructive interference, 28, 49
contact mode, 156
continuous X-rays, 46
contrast, 5, 90
 threshold, 5
convergence angle of probe, 125
critical angle of total reflection, 185
critical-point drying, 142
crossed position, 32
crucible, 305
crystallinity of polymer, 318
crystallization temperature, 313
Curie point, 324
curing temperature, 319
- dark-field, 25, 94
dead time, 185
Debye rings, 55
deconvolution, 308
defocus distance, 99
degenerate vibration, 260
dehydration, 142
depth of field, 5, 134
depth of focus, 6
depth resolution, 246
destructive interference, 28, 49
detection limit, 246
detection sensitivity, 246
deuterated triglycine sulfate, 270
deviation vector, 112
devitrification, 313
diamond knife, 89
diatomic model, 255
dichroic mirror, 37
differential interference contrast,
 13, 34
diffraction contrast, 92
diffraction grating, 280, 282
diffuse reflectance, 275
diffusion pump, 139, 201
dimple grinder, 88
dipole moment, 262
dislocations, 115
double-tilt, 85
DSC
 heat flux, 306
 power-compensated, 306
 temperature modulated, 308
d-spacing, 67, 72
duoplasmatron source, 232
- dwell time of probe, 127
dynamic mode, 155
dynamic range, 246
- effective magnification, 5
elastic scattering, 123
electric discharge machine, 15
electrolytic polishing, 21
electrolytic thinning, 87
electromagnetic lens, 79, 83
electron sourcebombardment source, 231
electron channeling, 134
electron energy analyzer, 204
electron flood gun, 215, 238
electron gun
 field emission, 82
 thermionic emission, 82
electron number effect, 130
electron retardation, 205
electronic gate, 250
electrostatic force, 152
elliptically polarized light, 31
emission volume, 187
energy dispersive spectroscopy, 171, 176, 183
enthalpy change, 302
epi-illumination, 11
equatorial plane, 72
etchants, 22
etching, 22
eutectic temperature, 317
Everhart-Thornley detector, 124
Ewald sphere, 53
excited filter, 37
external standard method, 70
- Fano factor, 184
Faraday cage, 124
far-field interaction, 145
fast Fourier transform, 269
field diaphragm, 10
field emission, 82
filament current, 82
flat-specimen error, 60
fluorchromes, 37
fluorescent labeling, 37
fluorescent yield, 175
focal length, 2
force modulation, 159, 161
force sensor, 154
Fourier transform, 98

- frame time, 128
freeze-drying, 142
friction force microscopy, 158
FTIR microscope, 276
FT-Raman, 284
full width at half-maximum, 63
fundamental parameter method, 193
fusion enthalpy, 319

glass knife, 89
glass transition, 301
goniometer circle, 61
grinding, 19
grinding jig, 86
group theory, 262

Hamaker constant, 152
harmonic vibration, 255
heat capacity, 308, 316
heat event, non-reversing, 309
heat event, reversing, 309
heat tinting, 25
hemispherical sector analyzer, 204
high resolution transmission electron microscopy, 95
high resolution X-ray diffractometry, 61
holographic filter, 280
hysteretic behavior, 167

inelastic scattering, 123
infrared absorption, 257
infrared active, 261
infrared light source, 269
 Globar, 269
 Nernst Glower, 269
instrument baseline, 310
instrument factor, 191
interferogram, 268
intermittent contact mode, 155
internal standard method, 70, 192,
 299
interpretable structure image, 99
infrared detector, 270
ion gun, 204
ion milling, 87
Ionization probability, 227

jet-polishing, 87
jump-off-contact, 157
jump-to-contact, 156

K α doublet, 47
Kikuchi line, 108
Köhler system, 10

lanthanum hexaboride, 82
latent heat, 315
lateral force microscopy, 158
lattice vibration, 256
layered synthetic microstructure,
 180
lens
 collector, 10
 condenser, 10
 eyepiece, 1, 13
 intermediate, 83
 objective, 1, 13, 83
 projector, 1, 83
 relay, 11
lifetime of surface, 230
light dispensor, 6
light filter, 11
 colored, 11
 heat, 11
 interference, 11
 neutral, 11
light guide, 124
linear absorption coefficient,
 48, 69
liquid metal ion source, 232
local density of states, 149

magnetic sector analyzer, 234
manipulation mode, 150
mass absorption coefficient, 48,
 194
mass resolution, 235
mass-density contrast, 90
mass-to-charge ratio, 235
matrix effect, 193
matrix factor, 191
mechanical clamping, 18
mercury cadmium telluride (MCT),
 271
meridional plane, 72
Michelson interferometer, 267
microbalance, 320
micro-Raman, 279
microtomy, 16
Miller index, 50
monochromatic filter, 59

- Moseley's law, 172
mounting, 16
mull, 274
multiplet splitting, 208

near-field force, 152
near-field interaction, 145
n-mer, 241
Nomarski contrast, 25, 34
non-contact mode, 156
normal mode vibration, 259
numerical aperture, 4

object function, 96
oligomer, 241
optical path difference, 267

parafocusing, 60
pass energy, 204
phase contrast, 25, 28
phase imaging, 160
phase plate, 29
phase shift, 99
phase transition, 302
photodiode, 154
photoelectron, 197
photomultiplier tube, 124
piezoelectric materials, 147
pinhole spatial filter, 281
Planck's constant, 254
plane polarized light, 30
plasma ion source, 231
plasmon loss, 208
pleochroism, 32
polarizability, 264
polarizability ellipsoid, 265
polarization factor, 56
polarized light, 25, 30
polarizer, 30
polishing, 20
 cloth, 21
position vector, 96
potassium bromide (KBr), 270
powder diffraction file, 66
prefilter, 281
primary absorption, 192
primary ion, 225
probe current, 125
probe size, 126
probe tip, 149

quadrupole analyzer, 235
quantum number, 173

Raman active, 262
Raman scattering, 257
Raman shift, 259
raster, 40
Rayleigh scattering, 257
real image, 1
receiving slit, 59
reciprocal lattice, 51
reciprocal unit vector, 51
reflectance, 274
remote aperture, 278
residual stress, 64
Rose visibility criterion, 127

sample baseline, 310
scanner, 147
scintillator, 124
secondary absorption, 192
secondary electrons, 123
secondary fluorescence, 192
secondary ion, 225
secondary mode, 185
sectioning, 15
selected-area diffraction, 93, 101
selection rules, 173
sensitivity factor, 219
shake-up satellites, 206
short-range force, 152
Si(Li) detector, 183
Siegbahn notation, 174
single beam spectrum, 271
Soller slits, 59
specimen-transparency error, 60
spectroscopic mode, 150
specular reflectance, 275
spherical aberration coefficient, 99
sputter depth profiling, 220
sputter ion pump, 201
sputter yield, 221, 228
sputtering, 141
 plasma-magnetron, 141
stacking fault, 113
staining, 142
standardless quantitative analysis, 195
static mode, 155
step width for scanning, 63
stigmator, 138

- Stokes scattering, 259
stopping factor, 194
stretching mode, 260
structure extinction, 57
structure factor, 58
sub-monomer, 241
surface charging, 139
surface ionization source, 232
- take-off angle, 186
tapping mode, 155, 159
Taylor cone, 232
temperature calibration, 324
temperature factor, 70
thallium acid phthalate, 180
thermal decomposition, 302
thermal drift, 167
thermal event, 301
thermal event, endothermic, 306
thermal event, exothermic, 306
thermal lag, 304
thermal null state, 307
thermal shielding, 324
thermal sputtering, 226
thermionic emission, 82
thermobalance, 320
thermocouple, 304
thermogravimetry, 301
thin film X-ray diffractometry, 60
time-of-flight, 233, 236
 analyzer, 236
tint etching, 25
topographic contrast, 130
total reflection, 185
trajectory effect, 130
transmission, 228
transfer function, 98
- transmittance, 272
tungsten filament gun, 82
tunneling current, 147, 148
turbomolecular pump, 201
two-beam condition, 118
- ultra-microtomy, 88
unit cell volume, 70
- vacuum impregnation, 18
van der Waals force, 152
vibrational frequency, 256
vibrational quantum number, 256
virtual image, 1
- wavelength dispersive spectroscopy, 171, 176–180
wavenumber, 254
wavetrain, 254
weak-beam dark-field, 94
wedge fringe, 111
Wehnelt electrode, 81
white X-rays, 46
wide angle X-ray diffraction, 71
wide angle X-ray scattering, 74
Wien filter, 233
Wollaston prisms, 34
working distance, 125
- X-ray absorption, 48
 factor, 194
X-ray fluorescence, 172
X-ray photoelectron, 197
X-ray powder diffractometry, 58
X-ray tube, 46
- ZAF method, 194
zero baseline, 310