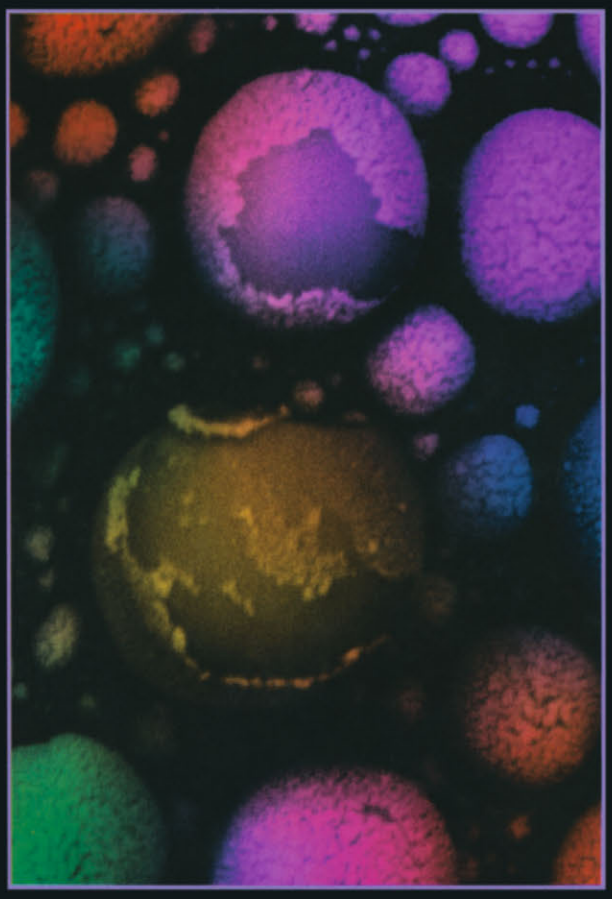
Scanning Electron Microscopy and X-Ray Microanalysis

THIRD EDITION



Joseph Goldstein, Dale Newbury,
David Joy, Charles Lyman,
Patrick Echlin, Eric Lifshin,
Linda Sawyer, and Joseph Michael

Scanning Electron Microscopy and X-Ray Microanalysis

THIRD EDITION

Scanning Electron Microscopy and X-Ray Microanalysis

THIRD EDITION

Joseph I. Goldstein
University of Massachusetts
Amherst, Massachusetts

Dale E. Newbury
National Institute of Standards and Technology
Gaithersburg, Maryland

Patrick Echlin
Cambridge Analytical Microscopy Ltd.
Cambridge, England

David C. Joy
University of Tennessee
Knoxville, Tennessee

Charles E. Lyman
Lehigh University
Bethlehem, Pennsylvania

Eric Lifshin
State University at Albany
Albany, New York

Linda Sawyer
Ticona LLC
Summit, New Jersey

Joseph R. Michael
Sandia National Laboratories
Albuquerque, New Mexico

Springer Science+Business Media, LLC

Additional material to this book can be downloaded from http://extras.springer.com.

ISBN 978-1-4613-4969-3 ISBN 978-1-4615-0215-9 (eBook)
DOI 10.1007/978-1-4615-0215-9

Originally published by Kluwer Academic / Plenum Publishers, New York in 2003
Softcover reprint of the hardcover 3rd edition 2003

9 8 7 (Corrected at 6th printing 2007)

springer.com

Preface to the Third Edition

In the decade since the publication of the second edition of *Scanning Electron Microscopy and X-Ray Microanalysis,* there has been a great expansion in the capabilities of the basic scanning electron microscope (SEM) and the x-ray spectrometers. The emergence of the variable-pressure/environmental SEM has enabled the observation of samples containing water or other liquids or vapor and has allowed for an entirely new class of dynamic experiments, that of direct observation of chemical reactions *in situ*. Critical advances in electron detector technology and computer-aided analysis have enabled structural (crystallographic) analysis of specimens at the micrometer scale through electron backscatter diffraction (EBSD). Low-voltage operation below 5 kV has improved x-ray spatial resolution by more than an order of magnitude and provided an effective route to minimizing sample charging. High-resolution imaging has continued to develop with a more thorough understanding of how secondary electrons are generated. The field emission gun SEM, with its high brightness, advanced electron optics, which minimizes lens aberrations to yield an effective nanometer-scale beam, and "through-the-lens" detector to enhance the measurement of primary-beam-excited secondary electrons, has made high-resolution imaging the rule rather than the exception. Methods of x-ray analysis have evolved allowing for better measurement of specimens with complex morphology: multiple thin layers of different compositions, and rough specimens and particles. Digital mapping has transformed classic x-ray area scanning, a purely qualitative technique, into fully quantitative compositional mapping. New x-ray detectors using significant departures in detector concepts and design (x-ray microcalorimetry, silicon drift detectors, and flat crystal spectrometers in conjunction with x-ray capillary optics) have begun to emerge. These detectors will allow new approaches to solving problems because of enhanced resolution and x-ray throughput.

During these 10 years we have taught over 1500 students at the Lehigh Microscopy Summer School. The short course in basic SEM and x-ray microanalysis forms the basis for this book. The course has continuously evolved over this past decade to meet changes in the field as well as in the

needs of our students, whose backgrounds span a huge range of the physical and biological sciences and diverse technologies. The course evolution reached the point that the instructors felt that a completely rewritten book was necessary. In this book, we have incorporated information about the new capabilities listed above and we have added much new material, including specimen preparation for polymers, the use of the focused ion beam instrument (FIB) for preparing specific areas of specimens for imaging and analysis, and eliminating charging on nonconducting specimens. We have retained the key features of the first and second editions, with separate discussions of the principles of SEM imaging and x-ray spectrometry and analysis. New chapters (Chapters 5 and 10) contain specialized material on SEM imaging and x-ray microanalysis for practical specimens. Moreover, we have decided to move some of the detailed and equation-dense discussions to an accompanying CD, a feature that was not even technically possible in 1992! In this way, we believe that the true introductory student will be able to grasp the basic principles of SEM and x-ray microanalysis in a more readable form. For the advanced student, detailed x-ray microanalysis formulations can be found on the CD which accompanies the book. The capacity of the CD has allowed us to gather all of the data that formed the last chapter of the 1992 edition as well as new databases to create a readily available computer-compatible resource for the reader. The CD also contains numerous color image examples (e.g., stereomicroscopy, compositional mapping) that could not be reproduced in the printed volume without greatly increasing the cost.

The authors wish to thank their many colleagues who contributed to this volume by allowing us to use material from their research, by their criticism of drafts of the chapters, and by their general support. Although we are grateful for the vital contributions of the new members of the teaching/author team, 10 years has not softened the loss of Charles Fiori, one of the most creative individuals in the history of our field, who died a few months after the publication of the 1992 edition. One of the authors (JIG) wishes to acknowledge to continuing support of the NASA Cosmochemistry Program. Another author (JRM) wishes to acknowledge the support of the United States Department of Energy through the Sandia Corporation, a Lockheed Martin Company operated for the United States Department of Energy. Special thanks go to Jane Martel for her help with the preparation of the manuscript and to Sharon Coe and Dr. David Williams of Lehigh University for continuing support as the book was developed.

Finally, the authors hope that the book not only will communicate the critical information necessary to successfully apply scanning electron microscopy and x-ray microanalysis to the reader's particular challenges, but will also convey the excitement that this field offers to those who are privileged to work in it.

J. I. Goldstein
D. E. Newbury

Contents

1

Introduction

The scanning electron microscope (SEM) permits the observation and characterization of heterogeneous organic and inorganic materials on a nanometer (nm) to micrometer (μm) scale. The popularity of the SEM stems from its capability of obtaining three-dimensional-like images of the surfaces of a very wide range of materials. SEM images are used in a wide variety of media from scientific journals to popular magazines to the movies. Although the major use of the SEM is to obtain topographic images in the magnification range 10–10,000×, the SEM is much more versatile, as we shall now see.

In the SEM, the area to be examined or the microvolume to be analyzed is irradiated with a finely focused electron beam, which may be swept in a raster across the surface of the specimen to form images or may be static to obtain an analysis at one position. The types of signals produced from the interaction of the electron beam with the sample include secondary electrons, backscattered electrons, characteristic x-rays, and other photons of various energies. These signals are obtained from specific emission volumes within the sample and can be used to examine many characteristics of the sample (surface topography, crystallography, composition, etc.).

The imaging signals of greatest interest are the secondary and backscattered electrons because these vary primarily as a result of differences in surface topography. The secondary electron emission, confined to a very small volume near the beam impact area for certain choices of the beam energy, permits images to be obtained at a resolution approximating the size of the focused electron beam. The three-dimensional appearance of the images is due to the large depth of field of the scanning electron microscope as well as to the shadow relief effect of the secondary and backscattered electron contrast.

In the SEM, characteristic x-rays are also emitted as a result of electron bombardment. The analysis of the characteristic x-radiation emitted from samples can yield both qualitative identification and quantitative elemental information from regions of a specimen nominally 1 μm in diameter and

1 μm in depth under normal operating conditions. The evolution of the SEM and the specific capabilities of modern commercial instruments are discussed below.

1.1. Imaging Capabilities

The scanning electron microscope is one of the most versatile instruments available for the examination and analysis of the microstructural characteristics of solid objects. A major reason for the SEM's usefulness is the high resolution which can be obtained when bulk objects are examined; instrumental resolution on the order of 1–5 nm (10–50Å) is now routinely quoted for commercial instruments. The high-resolution micrographs shown in Figs 1.1 and 1.2 were taken with a field emission gun SEM under routine operating conditions. The image in Fig. 1.1 is of a cross-sectioned semiconductor device coated with a very thin layer of Pt. The image has a measured spatial resolution of 0.9 nm and represents the current state of the art in high-resolution imaging. The high-resolution image in Fig. 1.2 is of Celgard, a stretched and annealed polypropylene microporous membrane. The details of the porous structure and the size and shape of the individual pores are shown.

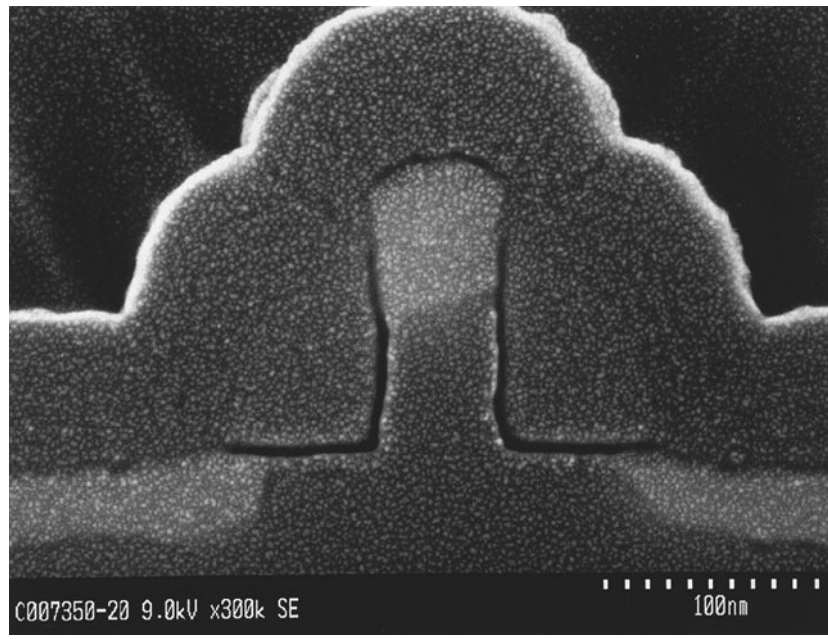


Figure 1.1. Ultrahigh-resolution SEM image taken at 9 keV of a cross-sectioned semiconductor device coated with Pt. The measured spatial resolution is 0.9 nm and represents the current state of the art in high-resolution imaging. The ultra-fine-grain Pt particles used for coating make it possible to focus and stigmate this image very accurately. The image was taken on a Hitachi S5200 at an original magnification of 300,000×. (Courtesy of Bryan Tracy of AMD.)

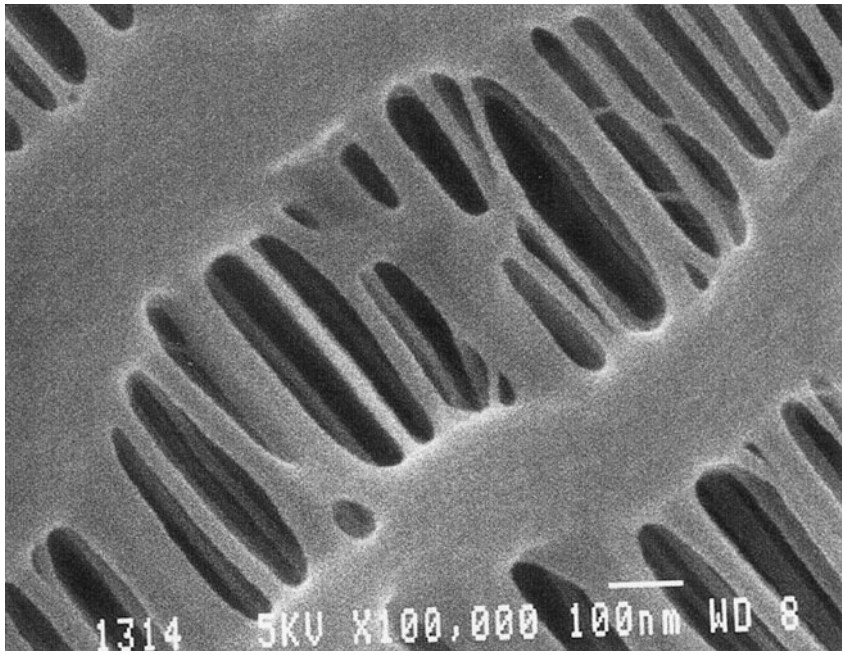
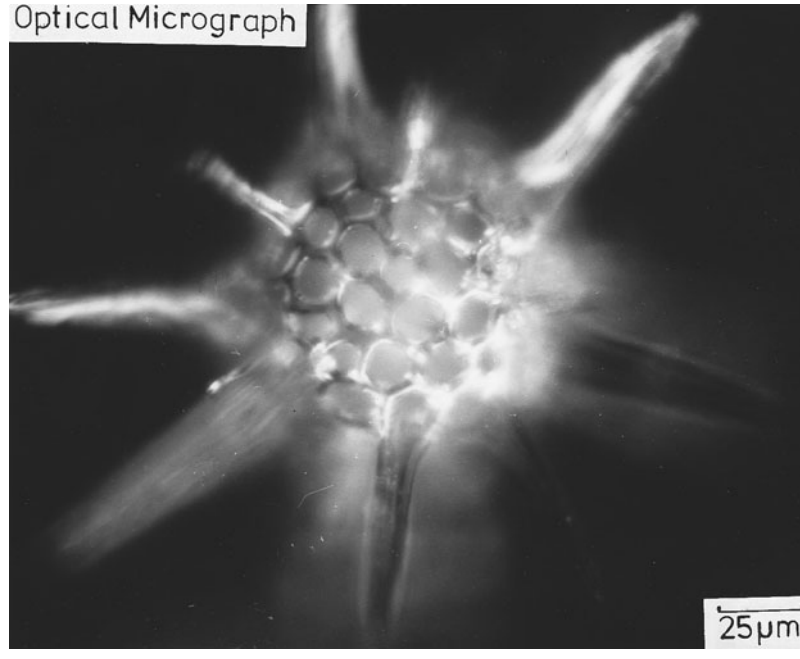


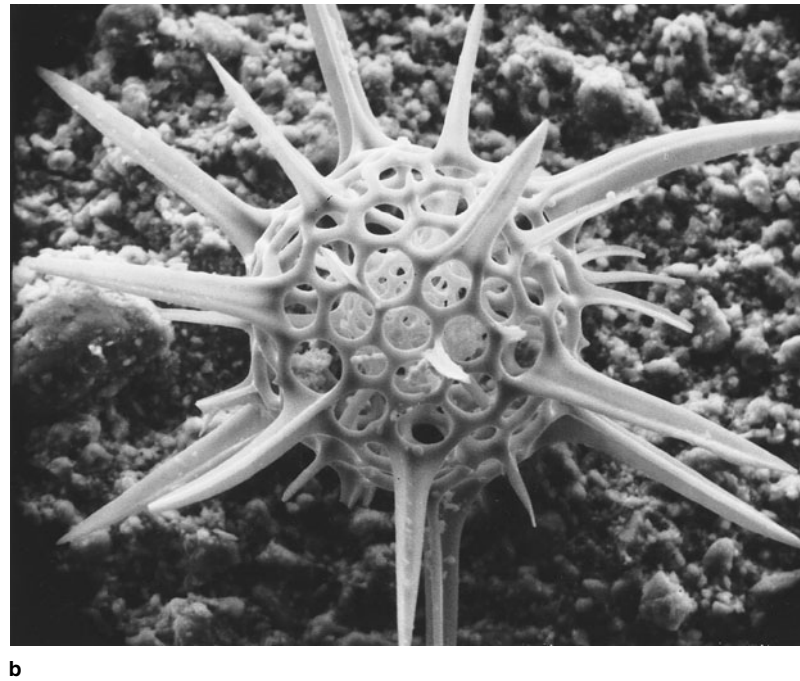
Figure 1.2. High-magnification SEM image of Celgard, a stretched and annealed polypropylene microporous membrane. The sample was ion-beam sputter-coated with about 5 nm of Pt and viewed in a field emission scanning electron microscope at an original magnification of 100,000×. The image shows details of the porous structure and the size and shape of the individual pores.

Another important feature of the SEM is the large depth of field, which is responsible, in part, for the three-dimensional appearance of the specimen image. Figure 1.3a shows the skeleton of a small marine organism (the radiolarian *Trochodiscus longispinus*) viewed with a light microscope and Fig. 1.3b shows the same object viewed with the SEM. The greater depth of field of the SEM provides much more information about the specimen. Examination of the literature indicates that it is this feature of a high depth of field that is of the most value to the SEM user. Most SEM micrographs, in fact, have been produced with magnifications below 8000 diameters (8000×). At these magnifications the SEM is operating well within its resolution capabilities. In addition, the SEM is also capable of examining objects at very low magnification. This feature is useful in forensic studies as well as other fields because the SEM image complements the information available from the light microscope. An example of a low-magnification micrograph of an archaeological subject is shown in Fig. 1.4.

The basic components of the SEM are the lens system, the electron gun, the electron collector, the visual and photorecording cathode ray tubes (CRTs), and the associated electronics. The earliest recognized work describing the concept of a scanning electron microscope is that of Knoll (1935). Subsequently von Ardenne (1938) constructed a scanning transmission electron microscope (STEM) by adding scan coils to a transmission



a



b

Figure 1.3. (a) Optical micrograph of the radiolarian *Trochodiscus longispinus*. (b) SEM micrograph of same radiolarian. The greater depth of focus and superior resolving capability of the SEM are apparent.

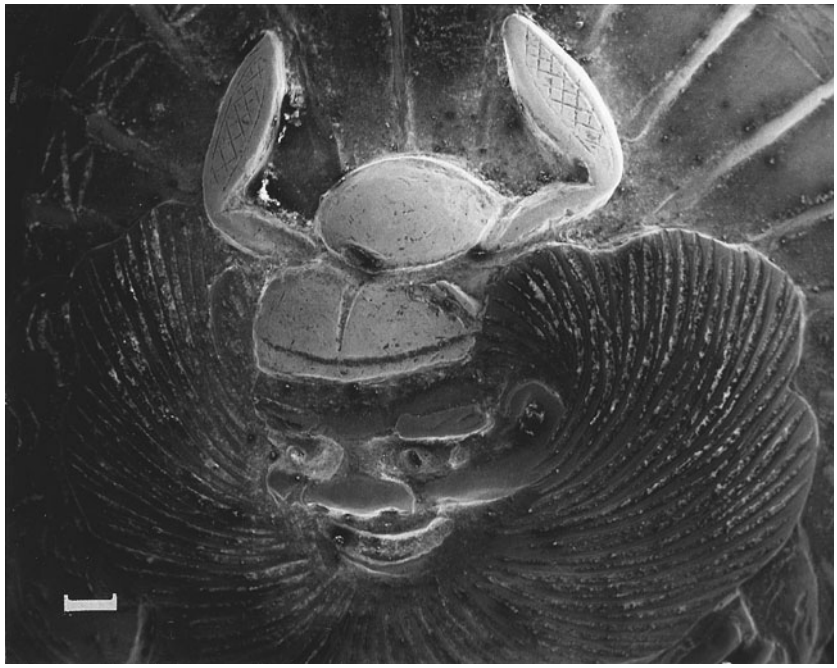


Figure 1.4. SEM image of the face of a helmeted, demon-like warrier from the rear of the handle of an 18th-century Japanese sword. The region of the helmet is gilt. Marker = 1 mm. (Courtesy of M. Notis, Lehigh University.)

electron microscope (TEM). Both the theoretical basis and practical aspects of STEM were discussed in fairly complete detail by von Ardenne (1938) and his instrument incorporated many features which have since become standard.

The first SEM used to examine thick specimens was described by Zworykin *et al*. (1942). The authors recognized that secondary electron emission would be responsible for topographic contrast. The collector was biased positive relative to the specimen by 50 V and the secondary electron current collected on it produced a voltage drop across a resistor. A detailed analysis of the interrelationship of lens aberrations, gun brightness, and spot size resulted in a method for determining the minimum spot size as a function of beam current (Zworykin *et al*., 1942). Their next contribution was the use of an electron multiplier tube as a preamplifier for the secondary emission current from the specimen. Resolution of about 50 nm (500 Å) was achieved with this, the first modern SEM, but, by comparison with the performance then obtainable from the rapidly developing TEM, this figure was considered unexciting and further development languished.

During the next few years C. W. Oatley and his student D. McMullan built their first SEM, and by 1952 this unit had achieved a resolution of 50 nm (500Å) (McMullan, 1952). McMullan was followed by Smith (1956), who, recognizing that signal processing could improve micrographs,

introduced nonlinear signal amplification. He also improved the scanning system by introducing double deflection scanning, and was also the first to insert a stigmator into the SEM (Smith, 1956).

The next step forward was the improvement of the secondary electron detector by Everhart and Thornley (1960), who employed a scintillator to convert the electrons to light, which was then transmitted by a light pipe directly to the face of a photomultiplier. Replacement of the electron multiplier by the more efficient photomultiplier increased the amount of signal collected and resulted in an improvement in signal-to-noise ratio. Hence, weak contrast mechanisms could be better investigated. Other uses of the SEM were developed in this time period. For example, Oatley and Everhart (1957) were able to observe, for the first time, the phenomenon known as voltage contrast. Later, Wells (1959) made the first studies of the effects of beam penetration on image formation in the SEM, and was the first to use stereographic pairs to produce three-dimensional SEM images (Wells 1960).

Pease (1963) built a system, known as SEM V, with three magnetic lenses, which also employed the Everhart–Thornley detector system. This instrument became the prototype of the first commercial instrument, the Cambridge Scientific Instruments Mark I "Stereoscan" (Pease, 1963; Pease and Nixon, 1965). A. D. G. Stewart and co-workers at the Cambridge Scientific Instrument Co. carried out the commercial design and packaging of the instrument. [For a detailed history of the development of the SEM, see Oatley (1972).] In the ensuing years, more than 50,000 SEM units have been sold by a dozen or more manufacturers in the U.S., U.K., Holland, Japan, Germany, and France.

Since the first commercial instrument of 1965, many advances have been made. One of these was the development of high-brightness electron sources such as the lanthanum hexaboride (LaB_6) electron cathode. With this source more electron current can be concentrated into a smaller beam spot and an effective improvement in resolution can be obtained. The field emission electron source, first used in the SEM in 1942, has now been developed to the point where it can be routinely used for the highest resolution imaging. The advantage of the field emission gun is that the source is very small, so that a probe of nanometer size with a very high brightness can be obtained. This brightness allows about 1000 times more current in the smallest electron probes than the conventional tungsten filament gun. Field emission sources, however, require clean, high vacuums of the order of 10^{-8} Pa (10^{-10} torr) or better to operate reliably, and such pressures require special attention to vacuum technology. Despite the increased cost and complexity of field emission SEMs, these microscopes have found widespread application in many research and technology fields where higher resolution and low-beam-energy capabilities are important.

Most modern SEMs are equipped to store images digitally, although in older instruments images are still made by photographing cathode ray tube images directly. Images can be observed on a computer screen, printed

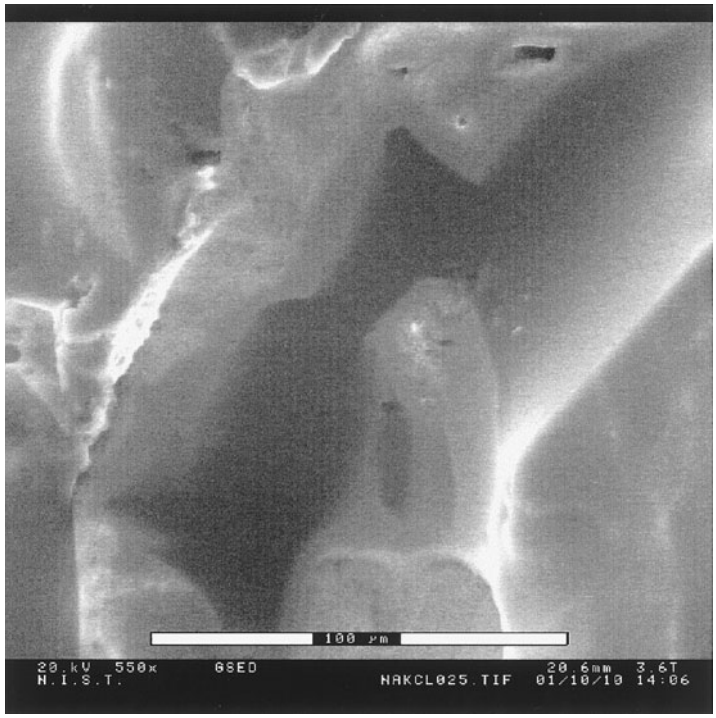
directly, and/or retained on a CD or other storage device for later observation or as a permanent record. Once in digital form the image can be processed in a variety of ways, such as nonlinear amplification, differentiation, and many other new and productive ways. The availability of powerful and inexpensive computers equipped with large storage capacity, high-resolution displays, and software packages capable of a full range of processing and quantitative functions on digital images gives the microscopist an unprecedented degree of flexibility and convenience in using the output of the SEM.

Other advances in the use of the SEM involve contrast mechanisms not readily available in other types of instrumentation, such as electron channeling contrast, produced by variations in crystal orientation, and magnetic contrast from magnetic domains in uniaxial and cubic materials.

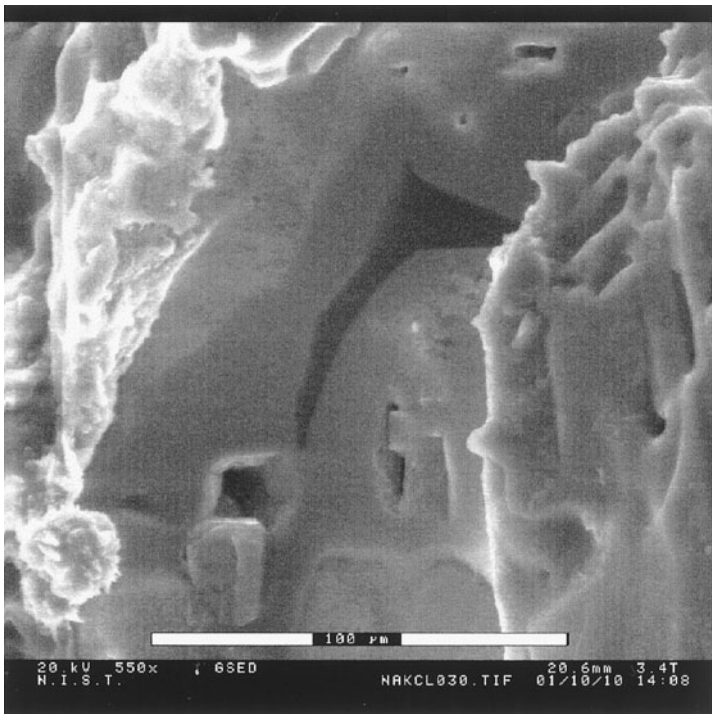
Some of the first scanning micrographs were of biological materials, such as hard-shell insects and wood fibers, that were strong enough to withstand the process of being dried in air without distortion (Smith and Oatley, 1955). Later Thornley (1960) showed SEM images of freeze-dried biological material examined at 1 keV to avoid charging. However, advances in the biological field have depended to some extent on needed advances in specimen preparation. Most biological specimens are wet, radiation-sensitive, thermolabile samples of low contrast and weak electron emissivity and are invariably poor conductors. Much attention has been paid to stabilizing delicate organic material, removing or immobilizing the cellular fluids, and coating the samples with a thin layer of a conducting metal. The development of low-temperature stages has helped to reduce the mass loss and thermal damage in sensitive specimens (Echlin, 1992).

One of the most important recent developments is the variable-pressure scanning electron microscope (VPSEM). This type of instrument allows the examination of surfaces of almost any specimen, wet or dry, because the environment around the specimen no longer has to be at high vacuum (see reviews by Danilatos, 1991, 1993). The environment can be water vapor or other gases in the pressure range from 25 to 2,500 Pa(0.2 to 20 torr). A differential pumping system is used to maintain the electron gun at high vacuum while the specimen is under a much higher pressure. In its modern form, the VPSEM can detect secondary and backscattered electrons as well as x-rays. Figure 1.5 shows the crystal growth of mixed NaCl and KCl from solution. The crystal growth is shown in a series of four images. This instrument can be used to study insulating, uncoated, wet, and generally untreated specimens. Biological samples are ideal for study in the VPSEM. In addition, experiments can be performed in the instrument in which the specimen surface may interact with the gaseous atmosphere.

The large depth of field available in the SEM makes it possible to observe three-dimensional objects using stereoscopy. Three-dimensional images allow different morphological features to be correctly interpreted and definitively measured. Equipment has been developed which allows quantitative evaluation of surface topography and for direct, real-time, high-resolution, stereo viewing in the SEM.

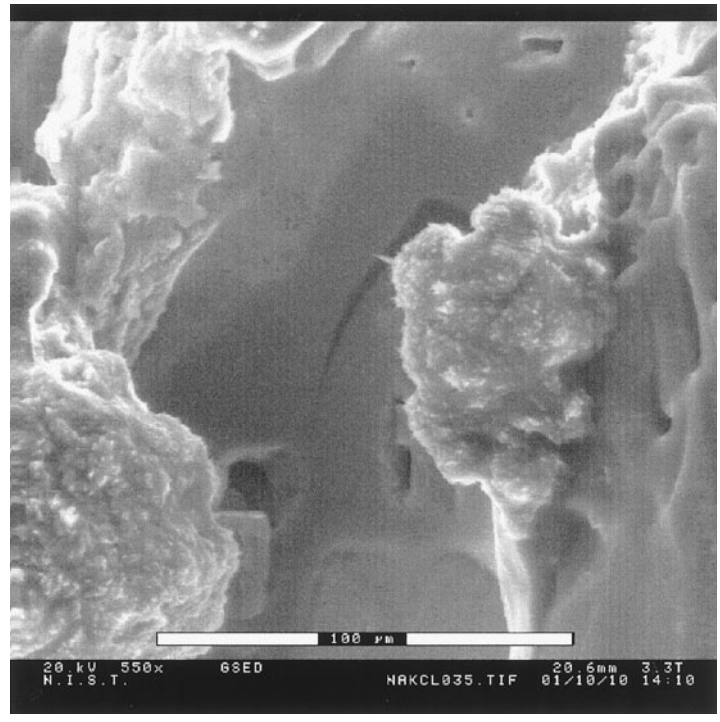


a

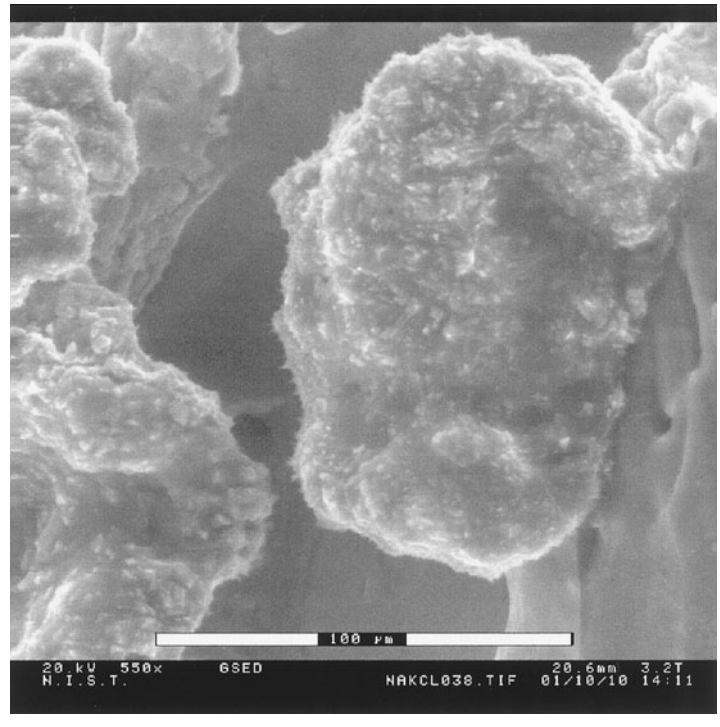


b

Figure 1.5. Crystal growth of mixed NaCl and KCl growing out of solution. The water was condensed from the water vapor after lowering the specimen temperature to 6°C with a Peltier stage in the instrument. After condensing the water and dissolving the salts, the process was reversed by lowering the pressure of the water vapor below saturation. The crystal growth is shown in a series of four images.



c



d

Figure 1.5. (*Continued*)

1.2. Structure Analysis

One of the most promising advances in the development of the SEM is the capability to determine the crystal structure and grain orientation of crystals on the surface of prepared specimens. This capability makes use of diffraction of the backscattered electrons emerging from the specimen surface (see review by Schwarz et al. 2000) and is known as electron backscattering diffraction (EBSD). To collect maximum intensity in the diffraction pattern, the specimen surface is steeply tilted at an angle of typically 70° from the horizontal. The intensity of backscatter Kikuchi patterns is rather low, as is the contrast of the signal, so extremely sensitive cameras and contrast enhancement facilities are required. The recent breakthrough in the application of Kikuchi patterns to bulk samples was marked by the development of a system for recording backscatter Kikuchi patterns using a highly sensitive video camera or more recently a very sensitive charge-coupled-device (CCD) camera. These patterns are then analyzed with a computer-assisted indexing method. Automated indexing of patterns and computer-automated crystal lattice orientation mapping allow this technique to identify phases and show misorientation across grain boundaries, respectively. Figure 1.6 shows the EBSD pattern of hematite, Fe_2O_3.

1.3. Elemental Analysis

The scanning electron microscope can also be used to obtain compositional information using characteristic x-rays. The development of an instrument for obtaining localized chemical analysis of solid samples (called an electron probe microanalyzer, EPMA) occurred at the same time as the development of the SEM.

The concept of the electron probe microanalyzer was patented in the 1940s (Hillier, 1947). It was not until 1949 that R. Castaing, under the direction of A. Guinier, described and built an instrument called the "microsonde éléctronique" (Castaing, 1951). In his doctoral thesis Castaing not only demonstrated that a localized chemical analysis could be made on the surface of a specimen, but also outlined the approach by which this information could be quantified. Recognition of the complexity of converting x-ray intensities to chemical composition has led numerous investigators over the past 50 years to refine the theoretical treatment of quantitative analysis first proposed by Castaing.

During the early 1950s several EPMA instruments were developed in laboratories in Europe and the United States. The first commercial EPMA instrument was introduced by CAMECA in France in 1956. The electron optics consisted of an electron gun followed by reducing lenses that formed an electron probe with a diameter of approximately 0.1–1 μm on the specimen. A light microscope for accurately choosing the point to be analyzed and one or more wavelength-dispersive spectrometers (WDS) for analyzing the intensity of x-ray radiation emitted as a function of energy are also part of the instrument.

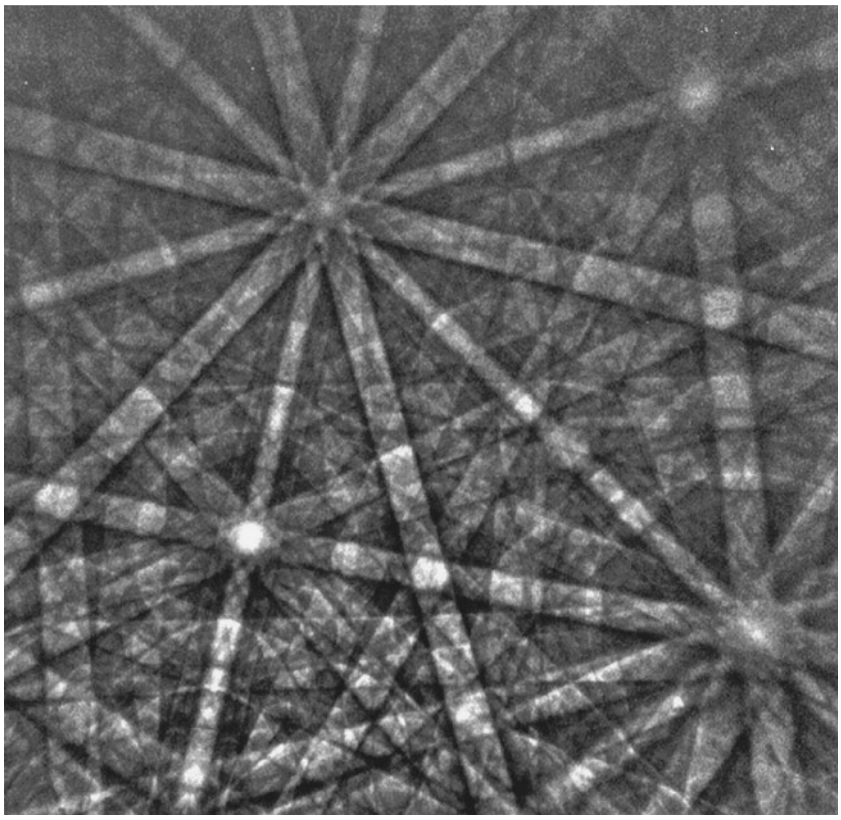


Figure 1.6. BSED pattern of hematite, Fe_2O_3. The image was taken at 20 keV.

Cosslett and Duncumb (1956) designed and built the first scanning electron probe microanalyzer at the Cavendish Laboratories in Cambridge, England. Whereas all previous electron microprobes had operated with a static electron probe, Cosslett and Duncumb swept the beam across the surface of a specimen in a raster, as is done in current SEMs. Although the concept of a local x-ray analysis is in itself a strong incentive for the utilization of a microprobe, the addition of the scanning concept was an extremely significant contribution. The SEM and EPMA were considered separate instruments for another decade. Today we consider the EPMA as a specially outfitted SEM with light optics and one or more WDS units.

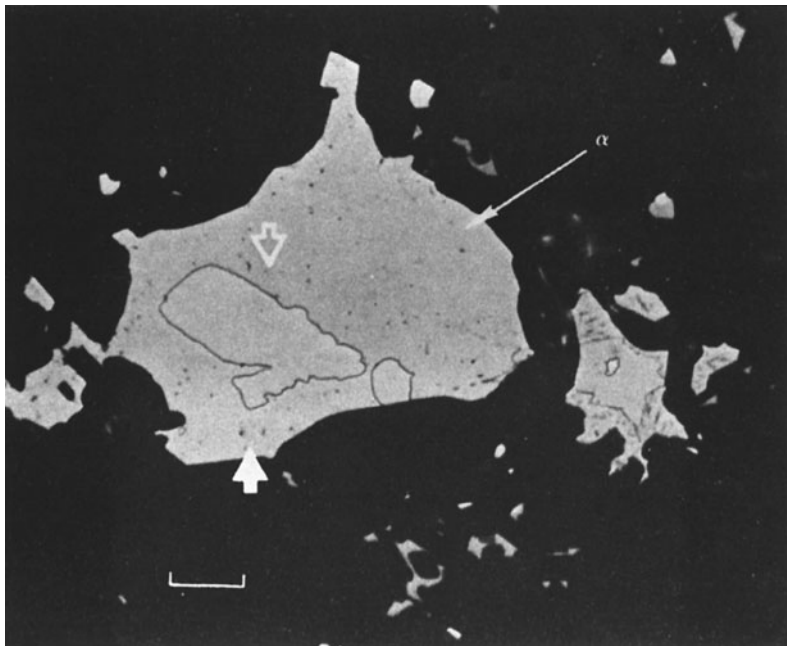
The addition of an energy-dispersive spectrometer (EDS) to an electron probe microanalyzer (Fitzgerald *et al*., 1968) to measure x-rays signaled the eventual coupling of such instrumentation to the SEM. These x-ray detectors were based upon the lithium drifted Silicon [Si(Li)] solid-state detector. Modern energy-dispersive spectrometers are capable of detecting characteristic x-rays of all elements above atomic number 4 at typical beam currents used for secondary electron imaging in the SEM. The vast majority of SEMs sold today are equipped with EDS capabilities. The EDS system offers a means of rapidly evaluating the elemental constituents of

a sample. In addition to rapid qualitative analysis, accurate quantitative analysis can also be achieved with EDS x-ray spectrometry. In a typical SEM equipped with EDS and WDS x-ray detectors, the EDS is used to measure characteristic x-rays from major elements (>10 wt%) in a sample, whereas the WDS is used to measure characteristic x-rays from minor or even trace elements (<0.1 wt%) in a sample. After many years where the principal focus in spectrometer development concentrated on incremental improvements in existing technologies, a number of recent developments have resulted in more significant departures in detector concepts and design. These recent developments include x-ray microcalorimetry, silicon drift detectors, and the use of flat crystal spectrometers in conjunction with capillary x-ray optics. These new x-ray spectrometers will find increasing use on SEM instruments.

A modern EPMA typically has the normal capabilities of the SEM, two or more wavelength spectrometers (WDS), a light-optical microscope to observe the specimen surface, and an electron beam current which is stabilized for x-ray analysis. The sample is analyzed nondestructively, and quantitative analysis can be obtained with a spatial resolution of the order of 1 μm on the sample. For flat-polished samples analyzed normal to the electron beam, an accuracy of the order of 1–2% of the amount present for a given element (5–10% in biological materials) can be obtained.

Figure 1.7 shows an example of a quantitative microprobe analysis of a flat-polished sample. In this example, the concentrations of C, Fe, Ni, and Co are measured across a 20-μm-wide carbide phase (cohenite). This phase nucleated and grew by a diffusion-controlled process in a ferrite α-bcc-phase matrix during cooling of a lunar metal particle on the moon's surface (Goldstein *et al*., 1976). In this example, the ability of the EPMA to measure the concentration of low-atomic-number elements such as carbon is also demonstrated. Figure 1.8 shows an example of an analysis of Mg, Al, and Si in the leaf of the tea plant *Camellia sinensis* (L). The analysis was carried out at various places in the upper epidermis cells of the tea leaf. Figure 1.8a shows the SEM image of a frozen-hydrated fracture face of the leaf. The upper epidermis cells are in a single layer at the bottom of the image. Figures 1.8b and 1.8c show EDS spectra from an analysis point within the walls of the upper epidermal cells. The Mg and Al $K\alpha$ peaks are visible in Fig. 1.8c. The Mg, Al, and Si concentrations are low, 725, 250, and 300 ppm, respectively.

Another important feature of the SEM with EDS and/or WDS is its capability of compositional mapping using characteristic x-rays. Figure 1.9 shows the distribution of Fe, Ni, and P among the major phases of the Goose Lake iron meteorite. The Widmanstatten structure is shown in the optical micrograph (Fig. 1.9a). The major phase of the structure, kamacite, is Ni-poor and Fe-rich. Magnifications up to 2500× are possible without exceeding the x-ray spatial resolution of the instrument. The attractiveness of this form of data gathering is that detailed microcompositional information can be directly correlated with light-optical and electron metallography. In this example, all metal regions with Ni contents less than



a

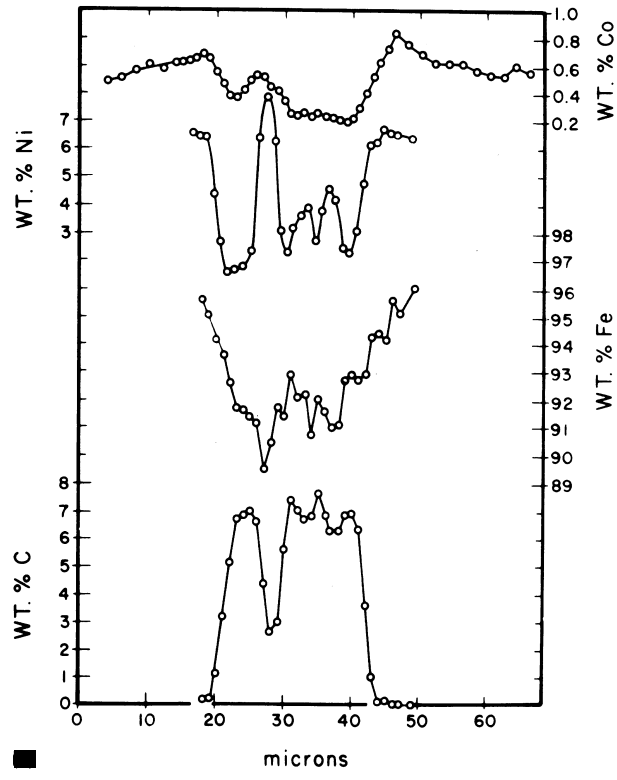
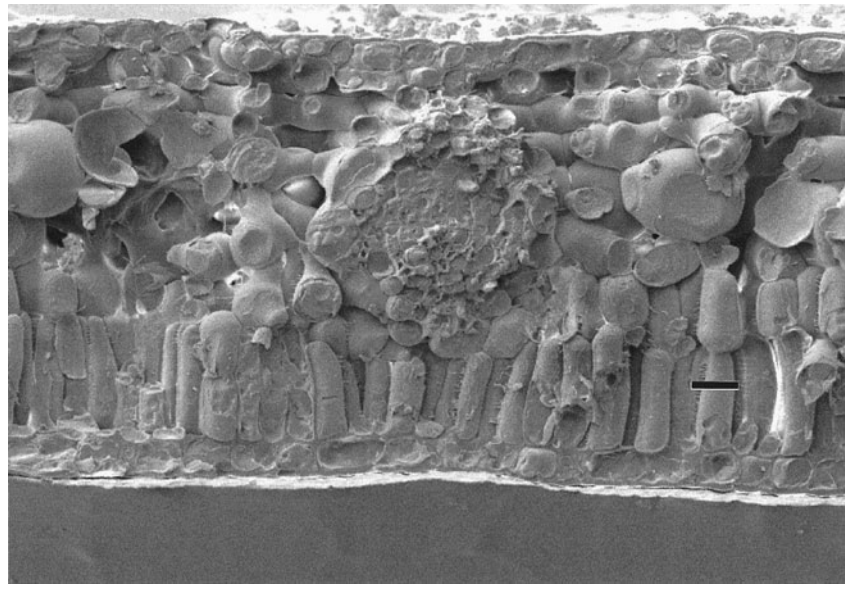
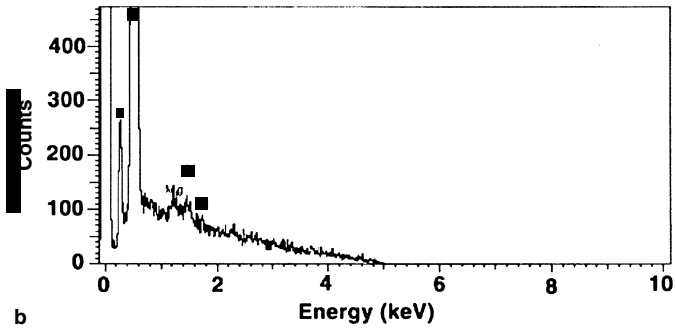


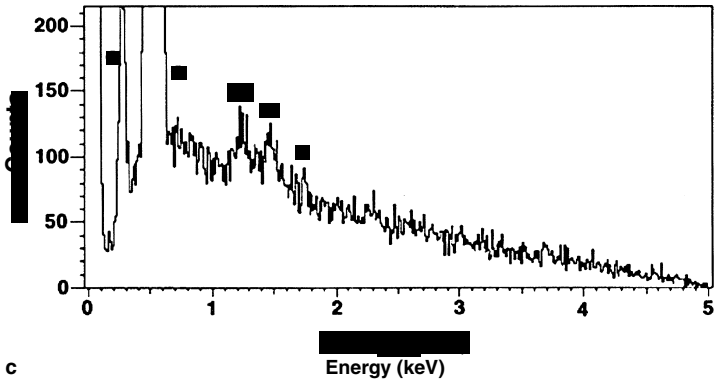
Figure 1.7. Analysis of metal particles in lunar rock 73275. (a) A reflected light micrograph of metal particles in lunar rock 73275.68. The α phase surrounds the cohenite and the arrows indicate the extent of the microprobe scan shown in (b) (Goldstein *et al*., 1976). Marker = 16 μm. (b) Variation of C, Fe, Ni, and Co concentrations across a cohenite (Fe_3C) particle in lunar rock 73275 as shown in (a).



a



b



c

Figure 1.8. Analysis of Mg, Al, and Si in the leaf of the tea plant *Camellia sinensis* (L). (a) A low-temperature (120 K) secondary SEM micrograph of a frozen-hydrated fracture face of a young leaf of the tea plant *Camellia sinensis* (L). The sample was plasma-magnetron sputter-coated with 2 nm of Cr. The leaf is made up of several distinct cell types. Analyses were carried out in the upper epidermal cells. A single layer of upper epidermal cells is shown at the bottom of the image. The image was taken at 5 keV and 35 pA beam current. The magnification marker is 5 μm. (b) EDS spectrum of a typical analysis area. (c) Detail of the EDS spectrum showing the Mg, Al and Si x-ray peaks. The spectra were taken with an ultrathin window detector. Because the sample is fully hydrated, the predominant elemental constituents are oxygen and carbon. The Mg, Al, and Si contents are of relatively low concentrations, 727, 250, and 300 ppm, respectively.

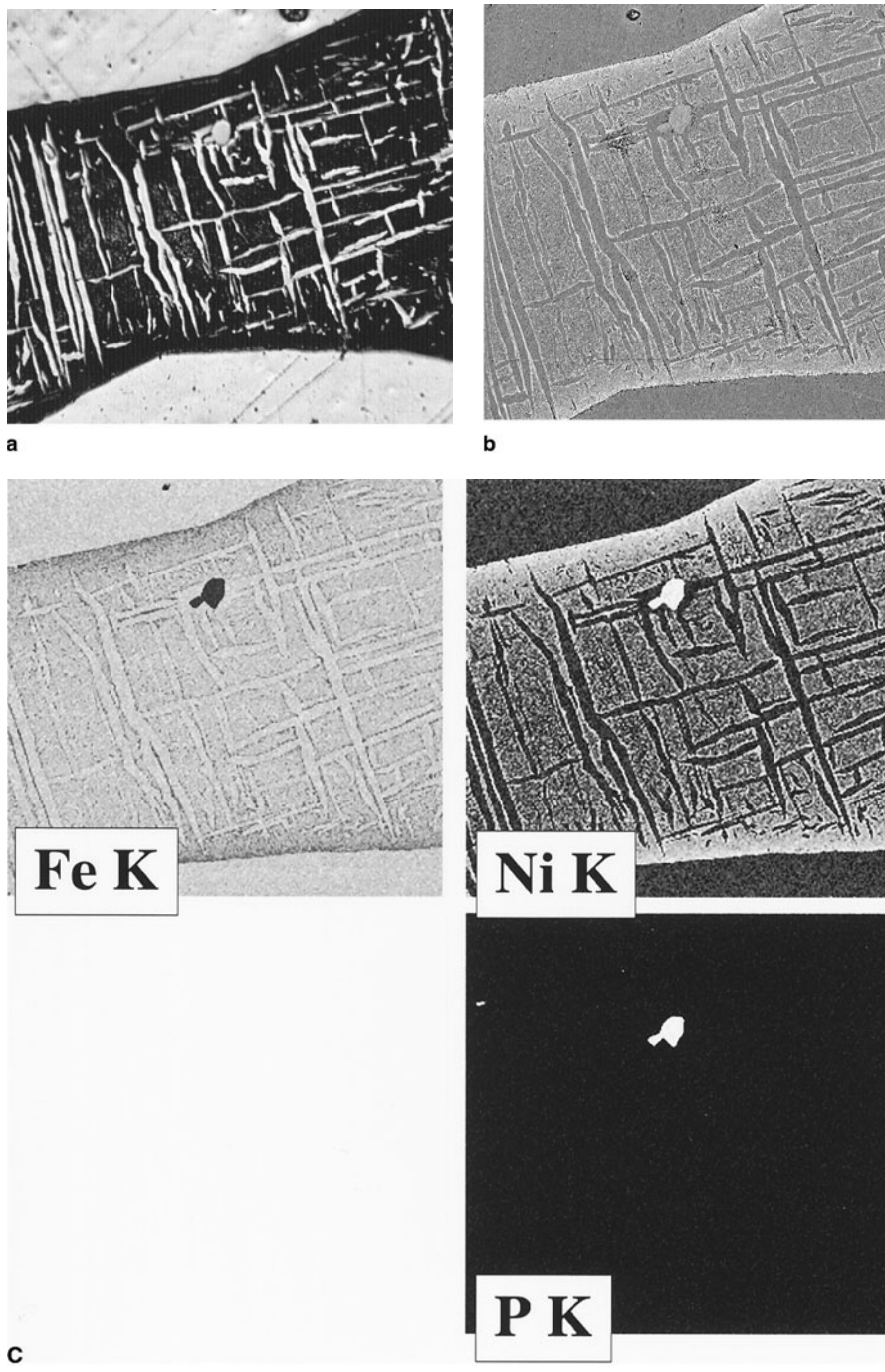


Figure 1.9. Distribution of Fe, Ni, and P between the major phases of the Goose Lake iron meteorite. (a) The Widmanstatten structure in the etched optical micrograph, (b) the BSE signal from a nonetched flat-polished region, and (c) the Fe, Ni, and P distributions. The field of view is 0.5 × 0.5 mm. The major phase of the structure, kamacite, is Ni-poor and Fe-rich. All metal regions with Ni contents less than 10 wt% are black in the Ni $K\alpha$ compositional map. The phosphide inclusion can be observed in the optical and BSE micrographs. The inclusion is relatively Fe-poor and Ni-rich. (Courtesy of D. Newbury, NIST.)

10 wt% are black in the Ni $K\alpha$ compositional map. Phase identification is made much easier in this application. The phosphide phase, schreibersite [$(Fe, Ni)_3P$], can be observed in the optical and backscattered electron images and clearly identified because of its relatively low Fe and high Ni content. In addition, the variety of signals available from the standard SEM (backscattered electrons, secondary electrons, etc.) can provide useful information about surface topography and composition in small regions of the specimen.

In the years since the development of the first EPMA instrument many advances have been made. Of particular importance was the development of diffracting crystals of organic molecules having large interplanar spacings (Henke, 1965). These crystals enable long-wavelength x-rays from low-atomic-number elements (B, C, N, O) to be measured with wavelength-dispersive spectrometers. More recently large-interplanar-spacing diffractors have been developed using physical vapor deposition of alternating layers of heavy and light elements. The ability to detect the low-atomic-number elements enables users of the EPMA to investigate many new types of problems with the instrument (see Fig. 1.7).

Recently, instruments have been developed allowing reliable analysis at low-electron-beam and low x-ray energies. The advantage of low electron beam energy is that one can minimize the x-ray source size and can minimize the effect of absorption in the quantitation procedure. Surface effects can be more easily measured, although carbon contamination of the sample surface due to the modest vacuum of the SEM is sometimes of concern.

The x-ray microanalysis of biological and polymeric materials is beset by the same problems associated with the examination of these samples in the SEM. In addition, the experimenter has to be very careful that the elements being measured remain at the specimen analysis point during measurement and are not removed or relocated by the preparative procedure. Although specimen preparation for biological material is more exacting than the procedures used in the material sciences, the quantitation methods are, in principle, somewhat simpler. These methods are based on the analysis of thin sections, where absorption and fluorescence effects may be neglected. However, biological and organic material are all too easily damaged by the electron beam and much effort in the past decade has gone into devising instrumentation as well as preparative and analytical procedures, such as low-temperature analysis, to limit specimen damage in such samples.

After application of the SEM with x-ray measuring capability was extended to nonmetallic specimens, it became apparent that other types of excitation phenomena might also be useful. For example, the color of visible light (cathodoluminescence) produced by the interaction of the electron probe with the specimen has been associated with the presence of certain impurities in minerals (Long and Agrell, 1965). In addition, visible photon emission produced from the recombination of excess electron–hole pairs in a semiconductor can be studied (Kyser and Wittry, 1966). Measurement of cathodoluminescence has now been developed as another important use of the SEM, particularly in the microelectronics industry.

Increased use of computer automation in conjunction with the SEM has greatly improved the quality and quantity of the data obtained. Computer programs have been developed to convert x-ray intensity ratios into chemical compositions, primarily because some of the correction parameters are functions of concentration and hence make successive approximations necessary. These programs enable the calculation of compositions within a few seconds so that the operator has greater flexibility in carrying out analyses. In addition, computer automation can be used to control the electron beam, specimen stage, and spectrometers. Automation greatly facilitates repetitive-type analysis, increases the amount of quantitative analysis performed, and leaves the operator free to concentrate on evaluating the analysis.

The development of quantitative compositional mapping provides another strong link between the imaging, the quantitative elemental analysis capability, and the use of computer automation in the SEM. In quantitative compositional mapping, a complete quantitative analysis is carried out under computer control at every discrete beam location in a scanned field. Arrays of numerical concentration values corresponding to the beam positions on the specimen are assembled into images with a digital image processor by encoding the concentration axes with an appropriate gray or color scale. The resulting images, or compositional maps, are supported at every picture element (pixel) by the complete numerical concentration values. In this way, the analyst can readily recover the analysis corresponding to any single pixel or array of pixels, and the compositional maps can be correlated with SEM images prepared from any of the available signals.

1.4. Summary and Outline of this Book

In its current form, the SEM is both competitive with and complementary to the capabilities offered by other microscopes. It offers much of the same ease of use and image interpretation found in the conventional light microscope while providing an improved depth of field, higher magnification, analytical capabilities, and the benefits of image processing. Although the SEM lacks the three-dimensional optical sectioning abilities of confocal microscopes, the analytical capabilities of the SEM provide elemental and crystallographic information that cannot be obtained on light-optical instruments. In fact, SEMs with field emission guns have lateral image resolutions in the nanometer range, which in many cases are directly comparable with the transmission electron microscope. The SEM has the added advantage that the specimen need not be made thin enough to be transparent to electrons. In some cases, the lateral resolution of a high-performance SEM on solid specimens is even comparable with that obtained by the scanning tunneling microscope or the atomic force microscope.

The SEM is also capable of being used for quantitative chemical measurements in solid samples at the micrometer level. EDS and WDS detectors supply the necessary x-ray data. Better spatial resolution approaching the nanometer level can be obtained with the analytical electron microscope,

but this capability requires preparing electron-transparent samples. The SEM is complementary to the light microscope, enabling the analyst to obtain detailed chemical information from phases and phase boundaries observed with the optical microscope. An overview of the SEM technique is given in Appendix A and an overview of x-ray microanalysis in Appendix B to this chapter. It is clear, therefore, that the SEM is a versatile and powerful instrument and, with topographic, crystallographic, and chemical information from the same instrument, is a major tool in research and technology.

The SEM community consists of users with an enormous diversity in technical backgrounds. The material of this book is primarily introductory. The authors have attempted to lay out how the instrument works and how to make optimum use of the instrument. The number of equations is kept to a minimum and the important concepts are also explained in a qualitative manner. It is also a comprehensive text. In some cases, where more complex concepts are of importance, these concepts are discussed in Enhancement sections, which are placed in the CD that accompanies the book. Advanced texts on SEM and x-ray microanalysis are planned to replace the 1986 Publication, Advanced SEM and x-ray Microanalysis.

The text is developed as follows: The electron-optical system and the electron signals produced during electron bombardment in the SEM are discussed in Chapters 2 and 3. Following these chapters is a discussion of image formation in the SEM, Chapter 4, and a discussion of special SEM topics including the VPSEM and electron backscattering diffraction (EBSD), Chapter 5. The next five chapters discuss x-ray microanalysis: x-ray production, Chapter 6, x-ray spectral measurement (EDS and WDS), Chapter 7; qualitative x-ray analysis, Chapter 8; quantitative x-ray analysis, Chapter 9; and special x-ray techniques, Chapter 10. The last four chapters deal with specific types of samples and their preparation for SEM and x-ray analysis: specimen preparation for solid materials, Chapter 11; polymer materials, Chapter 12; biological and hydrated materials, Chapters 13 and 14; and coating and conductivity techniques, Chapter 15.

Appendix A. Overview of Scanning Electron Microscopy

Basic Information. Size, shape, fine structure at the micrometer to nanometer scale.

Specimen Types:

- General: Thick, bulk (millimeter to centimeter dimensions), solid, low vapor pressure (no water); conductors (at least semiconducting)
- Special: Microscopic particles; film(s)-on-substrate; fixed biological specimens; wet specimens (variable-pressure or environmental SEM); nonconductors with conductive coating (conventional high beam energy and high vacuum) or uncoated (low beam energy, high vacuum; or high beam energy, variable-pressure or environmental SEM)

Signals Detected:
- Backscattered electrons (BSE)
- Secondary electrons (SE)
- Specimen, absorbed, induced currents (SC)

Resolution (Lateral):
- Conventional electron source: 10–50 nm
- Field emission electron source: 1–5 nm

Resolution (Depth):
- 10–1000 nm (BSE)
- 1–10 nm (SE)

Depth of Field. Selectable with final aperture: 0.1–1 unit of the image field width; high value enables stereomicroscopy for three-dimensional recognition.

Types of Image Information:
- Topography (SE, BSE)
- Composition (BSE)
- Crystal orientation (BSE)
- Electrical field (SE)
- Magnetic field (SE, BSE)
- Beam-induced current (SC)

Appendix B. Overview of Electron Probe X-Ray Microanalysis

Basic Information. Elemental identification and quantification.

Specimen Types:
- Ideal: Bulk specimen (millimeter to centimeter dimensions) polished flat to mirror finish, conductive
- Special cases: Particles, foils, film-on-substrate, rough surfaces, beam-sensitive specimens (especially biological), nonconductors

Spectrometer Types:
- Energy-dispersive x-ray spectrometer (130-eV resolution at Mn $K\alpha$)
- Wavelength-dispersive spectrometer (8-eV resolution at Mn $K\alpha$)

Signals Detected:
- Characteristic x-rays (identify elements)
- Continuum x-rays (background)

Speed:
- Qualitative analysis: 10–100 s (EDS)
- Quantitative analysis: 100–500 s
- Mapping: 1000–10,000 s

Elements Detected. $Z \geq 4$ (Be).

Accuracy (95% of Analyses):

- Flat, bulk target: ±5% relative for pure-element standards and matrix correction calculations, and ±25% relative for "standardless" analysis
- Particles, rough surfaces: ±50% relative

Limits of Detection:

- WDS: 10–100 parts per million (ppm)
- EDS: 1000–3000 ppm

Analytical Resolution (Lateral):

- Low Z: 1–5 μm
- High Z: 0.2–1 μm

References

Castaing, R. (1951). Thesis, University of Paris, ONERA Publication No. 55.
Cosslett, V. E., and P. Duncumb (1956). *Nature* **177,** 1172.
Danilatos, G. D. (1991). J. Microsc. **162,** 391.
Danilatos, G. D. (1993). *Microsc. Rese. Technique* **25,** 354.
Echlin, P. (1992). *Low Temperature Microscopy and Analysis,* Plenum Press, New York.
Everhart, T. E., and R. F. M. Thornley (1960). *J. Sci. Instr*. **37,** 246.
Fitzgerald, R., K. Keil, and K. F. J. Heinrich (1968). *Science* **159,** 528.
Goldstein, J. I., R. H. Hewins, and A. D. Romig, Jr. (1976). In *Proceedings of the 7th Lunar Science Conference* Pergamon Press, New York, p. 807.
Henke, B. L. (1965). In *Advances in X-Ray Analysis,* Vol. 9, Plenum Press, New York, p. 269.
Hiller, J. (1947). U.S. Patent 2,418,029.
Knoll, M. (1935). *Z. Tech. Phys*. **11,** 467.
Kyser, D. F., and D. B. Wittry (1966). In *The Electron Microprobe* (T. D. McKinley, K. F. J. Heinrich, and D. B. Wittry, eds.), Wiley, New York, p. 691.
Long, J. V. P., and S. O. Agrell (1965). *Min. Mag*. **34,** 318.
McMullan, D. (1952). Ph.D. Dissertation, Cambridge University, Cambridge, England.
Oatley, C. W. (1972). *The Scanning Electron Microscope,* Cambridge University Press, Cambridge.
Oatley, C. W., and T. E. Everhart (1957). *J. Electron*. **2,** 568.
Pease, R. F. W. (1963). Ph.D. Dissertation, Cambridge University, Cambridge, England.
Pease, R. F. W., and W. C. Nixon (1965). *J. Sci. Instr*. **42,** 81.
Schwarz, A. J., M. Kumar, and B. L. Adams, eds. (2000). *Electron Backscatter Diffraction in Materials Science,* Kluwer Academic/Plenum Publishers, New York.
Smith, K. C. A. (1956). Ph.D. Dissertation, Cambridge University, Cambridge, England.
Smith, K. C. A., and C. W. Oatley (1955). *Br. J. Appl. Phys.* **6,** 391.
Thornley, R. F. M. (1960). Ph.D. Dissertation, Cambridge University, Cambridge, England.
von Ardenne, M. (1938). *Z. Tech. Phys*. **109,** 553.
Wells, O. C. (1959). *J. Electron. Control*. **7,** 373.
Wells, O. C. (1960). *Br. J. Appl. Phys*. **11,** 199.
Zworykin, V. K., J. Hiller, and R. L. Snyder (1942). *ASTM Bull.* **117,** 15.

2

The SEM and Its Modes of Operation

Obtaining a low-magnification ($<1000\times$) image of a rough three-dimensional object is remarkably easy with an SEM. To obtain all the information the SEM can provide, however, requires an understanding of the major modes of microscopy and the electron beam parameters that affect them. We will discuss the following microscopy modes: resolution mode, high-current mode, depth-of-focus mode, and low-voltage mode. The electron beam diameter at the specimen limits the image resolution, and the amount of electron current in the final probe determines the intensity of the secondary and backscattered electron and x-ray signals. Unfortunately, the smaller the electron probe, the lower is the probe current available and the poorer is the visibility of image features. The angle of the conical beam impinging on the specimen governs the range of heights on the specimen that will simultaneously be in focus. The accelerating voltage (kilovolts) of the beam determines how faithful the image will be in representing the actual surface of the specimen. The operator must control these beam parameters to achieve optimum results in each microscopy mode. In this chapter we will describe the electron beam optical column, the modes of microscopy, and the important relationship between electron probe current and electron probe diameter (spot size).

2.1. How the SEM Works

The basic concepts given throughout this chapter apply to all SEMs even though the details of electron-optical design vary from manufacturer to manufacturer. This section provides a brief overview of the operation of the SEM by describing the functions of the various subsystems.

2.1.1. Functions of the SEM Subsystems

The two major components of an SEM are the electron column and the control console (Fig. 2.1). The electron column consists of an electron gun and two or more electron lenses, which influence the paths of electrons

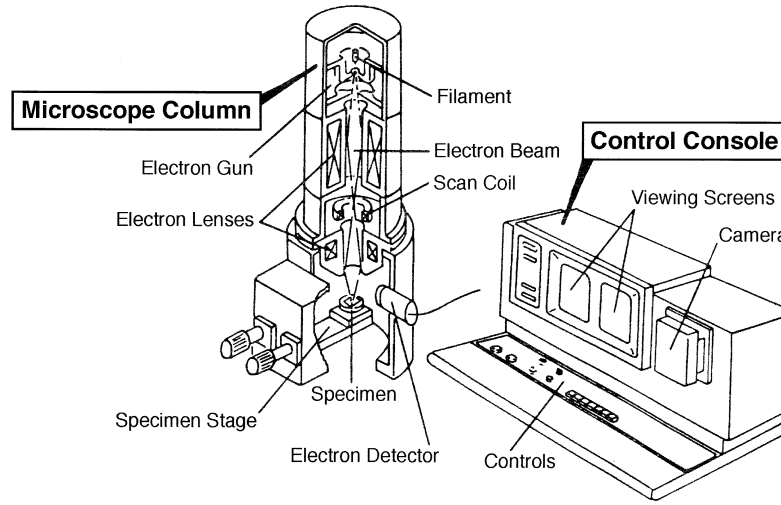


Figure 2.1. The two major parts of the SEM, the electron column and the electronics console. (Reprinted with permission of *The Morning Call*, Allentown, PA.)

traveling down an evacuated tube. The base of the column is usually taken up with vacuum pumps that produce a vacuum of about 10^{-4} Pa (about 10^{-6} torr, or roughly one billionth of atmospheric pressure). The control console consists of a cathode ray tube (CRT) viewing screen and the knobs and computer keyboard that control the electron beam.

2.1.1.1. Electron Gun and Lenses Produce a Small Electron Beam

The electron gun generates electrons and accelerates them to an energy in the range 0.1–30 keV (100–30,000 electron volts). The spot size from a tungsten hairpin gun is too large to produce a sharp image unless electron lenses are used to demagnify it and place a much smaller focused electron spot on the specimen, as shown schematically in Fig. 2.2. Most SEMs can produce an electron beam at the specimen with a spot size less than 10 nm (100Å) that contains sufficient probe current to form an acceptable image. The beam emerges from the final lens into the specimen chamber, where it interacts with the specimen to a depth of approximately 1 μm and generates the signals used to form an image.

2.1.1.2. Deflection System Controls Magnification

The scanned image is formed point by point. The deflection system causes the beam to move to a series of discrete locations along a line and then along another line below the first, and so on, until a rectangular 'raster' is generated on the specimen. Simultaneously, the same scan generator creates a similar raster on the viewing screen. Two pairs of electromagnetic deflection coils (scan coils) are used to sweep the beam across the specimen. The first pair of coils deflects the beam off the optical axis of the microscope and the second pair bends the beam back onto the axis at the pivot point of the scan (Fig. 2.3). The magnification M of the image is the ratio of the

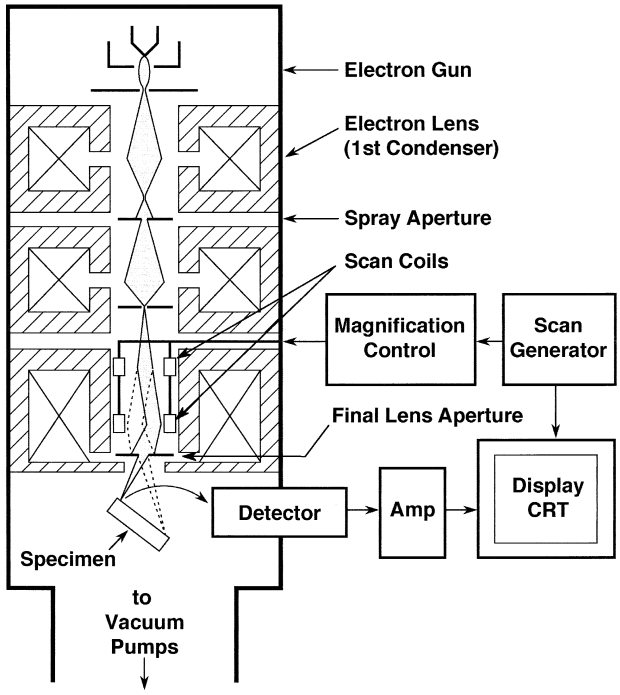


Figure 2.2. Schematic drawing of the electron column showing the electron gun, lenses, the deflection system, and the electron detector.

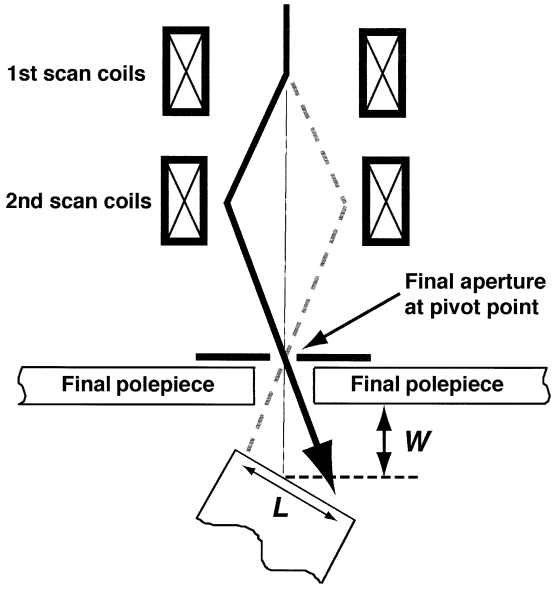


Figure 2.3. Deflection system inside the final lens. Working distance *W* is the distance between the specimen and the bottom of the final lens polepiece.

length of the raster on the viewing screen to the corresponding length of the raster on the specimen (L). For example, a 100-μm-wide raster on the specimen displayed on a 10-cm-wide viewing screen generates an image of 1000× magnification. When the operator requests an increase in image magnification, the scan coils are excited less strongly, so that the beam deflects across a smaller distance on the specimen. Note that the raster size on the specimen also depends on the working distance, the distance from the specimen to the bottom of the final lens. In a modern SEM the magnification is automatically compensated for each working distance to assure that the indicated magnification is correct.

2.1.1.3. Electron Detector Collects the Signal

Contrast in an image arises when the signal collected from the beam–specimen interaction varies from one location to another. When the electron beam impinges on the specimen, many types of signal are generated (see Chapter 3 for a detailed discussion of beam–specimen interactions), and any of these can be displayed as an image. The electronics of the detector system converts the signals to point-by-point intensity changes on the viewing screen and produces an image (see Chapter 4 for a detailed discussion of image formation). The two signals most often used to produce SEM images are secondary electrons (SE) and backscattered electrons (BSE). The standard Everhart–Thornley (E–T) detector collects both secondary and backscattered electrons as shown in Fig. 2.4. Both SE and BSE signals are collected when a positive voltage is applied to the collector screen in front of the detector. With a negative voltage on the collector screen, a pure BSE signal is captured because the low-energy SEs are repelled. Electrons captured by the scintillator/photomultiplier are then amplified for display on the viewing CRT. Although SEs and BSEs have been discussed here, any signal that can be collected and amplified may be used to form an image in the SEM.

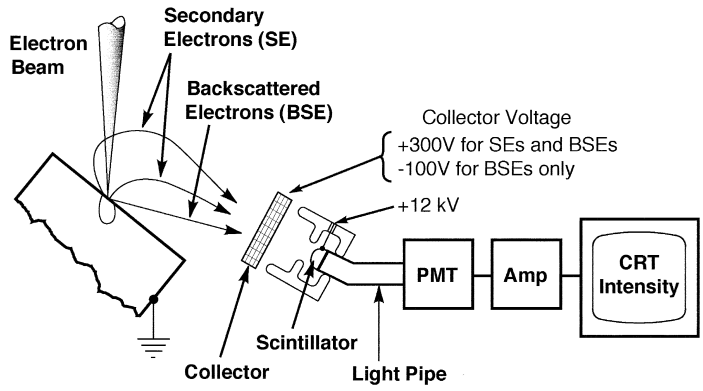


Figure 2.4. Diagram showing backscattered and secondary electron collection of these signals by the Everhart–Thornley (E–T) detector.

2.1.1.4. Camera or Computer Records the Image

Older SEMs have a separate CRT for slow-scan recording of images onto photographic film or, alternatively, a video printer to produce a hard copy of the CRT image. Modern SEMs store their images in digital form in a computer for processing and printing at a later time.

2.1.1.5. Operator Controls

The first controls that an operator must master are those that control the creation of the electron beam: the accelerating voltage and the emission current. Next come the lens controls: the condenser lens control determines both the amount of beam current available and the minimum beam size, and the objective lens control allows the beam to be focused so that the smallest diameter is located exactly at the specimen surface. Finally, because some locations on the specimen surface may appear too light or too dark on the viewing screen, signal controls, labeled contrast (or "dark level") and brightness, are provided to trim or expand the displayed intensity until a reasonable image is shown on the viewing screen. Observing the signal trace (line scan waveform) on an oscilloscope is a valuable means of understanding the action of these controls. Note that whereas focusing is an activity similar to that on the light microscope, frequent adjustments of the contrast and brightness between focusing operations are also required on the SEM to produce an image on the screen.

2.1.2. SEM Imaging Modes

Images can provide much more information about a specimen if we understand the conditions under which they are taken. The limiting sharpness and feature visibility of SEM micrographs are dependent upon four parameters: the electron probe size d_p, the electron probe current i_p, the electron probe convergence angle α_p, and the electron beam accelerating voltage V_0 (kV) as shown in Fig. 2.5. The probe or spot size d_p is defined as

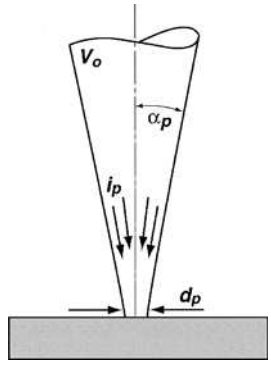
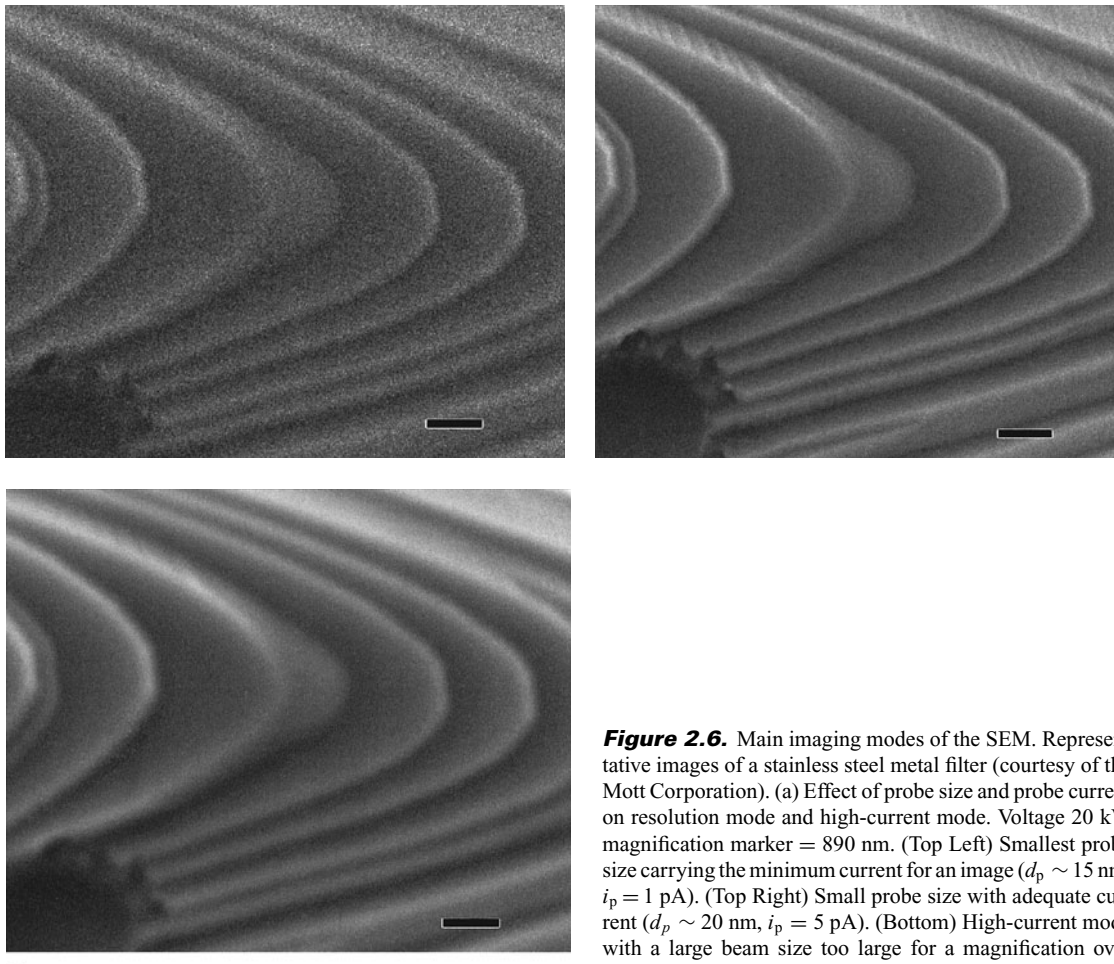


Figure 2.5. Four major electron beam parameters are defined where the electron probe impinges on the specimen: electron probe diameter d_p, electron probe current i_p, electron probe convergence α_p, and electron beam accelerating voltage V_0.

the diameter of the final beam at the surface of the specimen. In this book the term electron beam refers to electrons at any point in the column beyond the electron gun, whereas the term electron probe refers to the focused electron beam at the specimen. The probe current i_p is the current that impinges upon the specimen and generates the various imaging signals. The electron probe convergence angle α_p is the half-angle of the cone of electrons converging onto the specimen. We use the term convergence angle when we are specifically dealing with the beam at the specimen where it converges; at other locations in the column the beam angle may be diverging. The final beam parameter we will discuss here is the accelerating voltage V_0 of the electron gun. Each of these four beam parameters dominates one of the four major SEM imaging modes: resolution mode, high-current mode, depth-of-focus mode, and low-voltage mode. The effects on each image made when key beam parameters are varied are shown in Fig. 2.6.



a

Figure 2.6. Main imaging modes of the SEM. Representative images of a stainless steel metal filter (courtesy of the Mott Corporation). (a) Effect of probe size and probe current on resolution mode and high-current mode. Voltage 20 kV; magnification marker = 890 nm. (Top Left) Smallest probe size carrying the minimum current for an image ($d_p \sim 15$ nm, $i_p = 1$ pA). (Top Right) Small probe size with adequate current ($d_p \sim 20$ nm, $i_p = 5$ pA). (Bottom) High-current mode with a large beam size too large for a magnification over 10,000× ($d_p \sim 130$ nm, $i_p = 320$ pA).

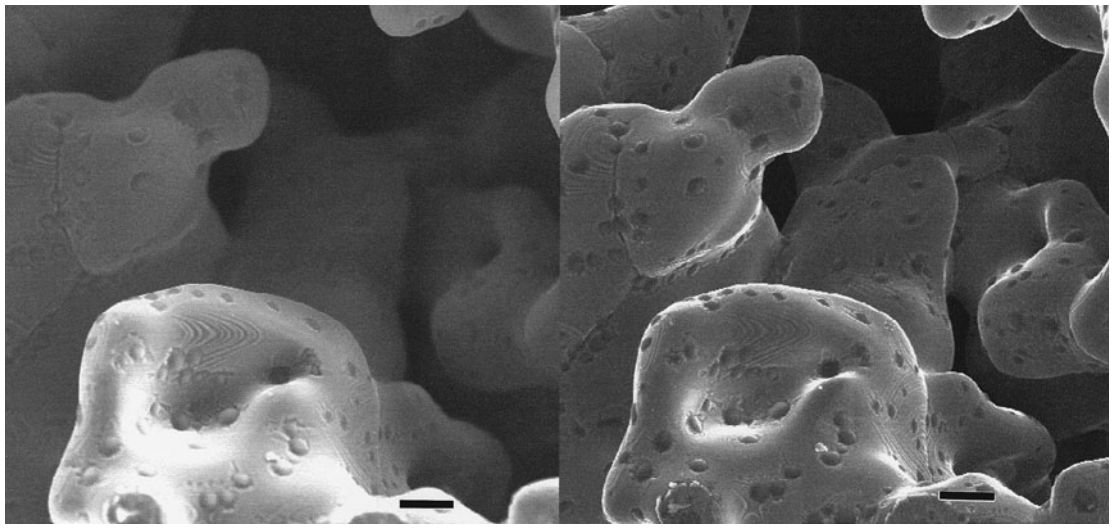
2.1.2.1. Resolution Mode

For the highest resolution image, d_p must be as small as possible while at the same time containing sufficient beam current to exceed the visibility threshold (discussed in Chapter 4) for the contrast produced by the features of interest. Resolution refers to the size of the finest details that can be observed. To image the finest details of the specimen surface, the probe diameter must be comparable with or smaller than the feature itself. Resolution mode is only meaningful at high image magnifications ($>10{,}000\times$) where these small details can be observed because at low magnifications image sharpness is controlled by the display parameters (see Chapter 4).

2.1.2.2. High-Current Mode

For the best image visibility and quality, large beam currents i_p are required. Unless the contrast between an image feature and background is distinguishable above random signal fluctuations (noise), detail cannot be observed even if the spot size is small enough for it to be easily resolved. Large beam currents are also necessary for successful x-ray microanalysis because only a small fraction of beam interactions result in x-ray emission.

Figure 2.6a illustrates the interplay of the high-resolution mode and high-current mode. The top left photo is taken with a small probe size ($d_p \sim 15$ nm, $i_p = 1$ pA) and the original magnification of the image was $11{,}200\times$. Note that, although the beam size is small and the resolution of the image is adequate, there is considerable noise in the image and the feature contrast is low. Some features are resolved, but they cannot be



b

Figure 2.6. (*Continued*) (b) Effect of convergence angle on depth-of-focus mode. Voltage, 20 kV; magnification marker = 11.6 μm. Varying aperture size. (Left) Large beam convergence angle (α_p = 15 mrad). (Right) Small beam convergence angle ($\alpha_p = 1$ mrad). A loss of backround features occurs as the convergence angle (and aperture size) increases.

observed easily because of the poor visibility. As the probe size and current are increased ($d_p \sim 20$ nm, $i_p = 5$ pA), as in the top right photo, the resolution is about the same as the top left image, but the image is no longer noisy. In this photo the feature visibility and resolution are excellent. Increasing the probe size and current further ($d_p \sim 130$ nm, $i_p = 320$ pA), as in the lower photo, blurs the image. The resolution is inadequate, that is, the probe size d_p is too large at the magnification of the image. Only at lower magnification would the image appear sharp.

2.1.2.3. Depth-of-Focus Mode

For the best depth of focus α_p must be as small as possible. By making the beam convergence angle low, the beam diameter changes only a little

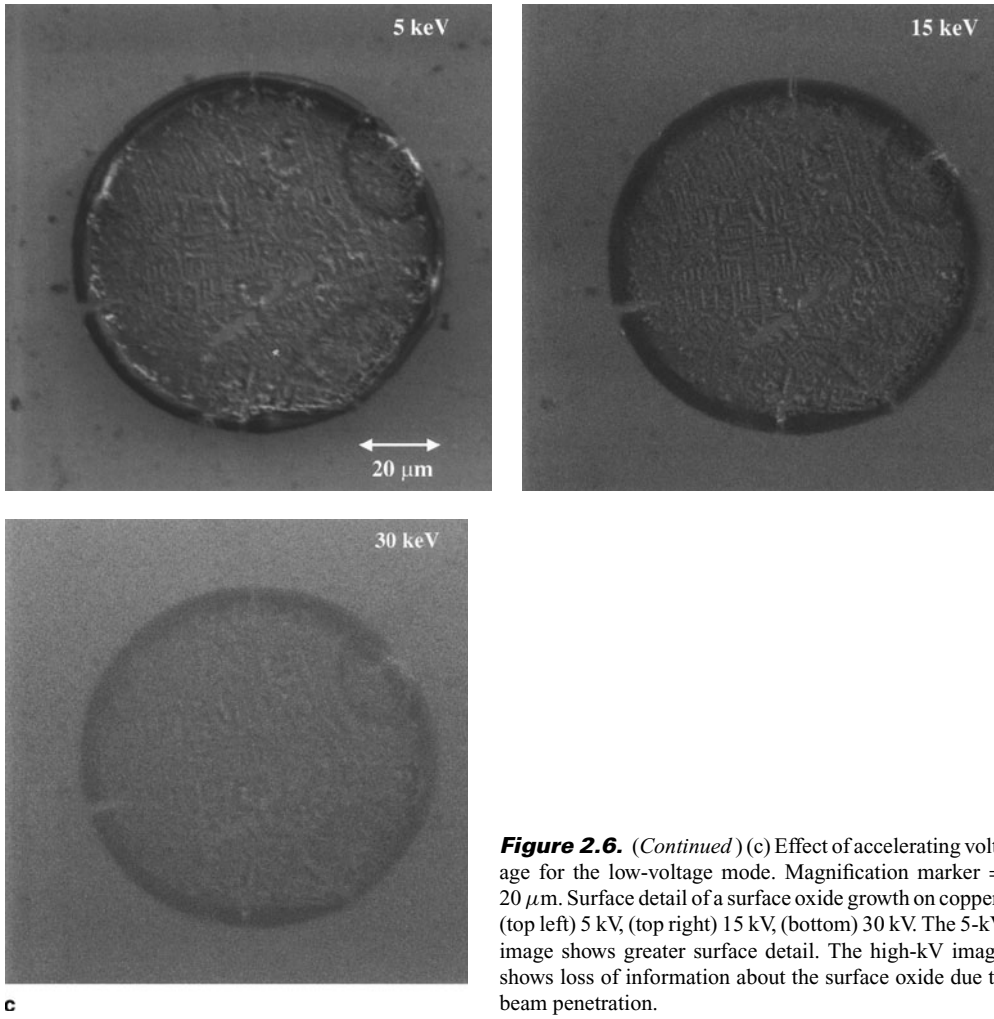


Figure 2.6. (*Continued*) (c) Effect of accelerating voltage for the low-voltage mode. Magnification marker = 20 μm. Surface detail of a surface oxide growth on copper: (top left) 5 kV, (top right) 15 kV, (bottom) 30 kV. The 5-kV image shows greater surface detail. The high-kV image shows loss of information about the surface oxide due to beam penetration.

over a long vertical distance and so features on the surface at different heights will all appear to be in focus at the same time. Figure 2.6b shows a stainless steel metal filter viewed with a large beam convergence angle ($\alpha_p = 15$ mrad, left) and a small beam convergence angle ($\alpha_p = 1$ mrad, right). For the large beam convergence angle, the central feature below the top surface is out of focus and no surface detail can be seen. For the small beam convergence angle, both the top surface and the lower surface are in focus. The large depth of focus of the SEM is one of its greatest strengths.

2.1.2.4. Low-Voltage Mode

At low accelerating voltages ($\leq$5 kV), the beam interaction with the specimen is confined to regions very close to the surface. This provides an image which is rich in surface detail compared to those obtained at higher accelerating voltages (15–30 kV), where the beam penetrates beneath the surface and the emerging signal electrons mostly carry information about the interior of the sample. Figure 2.6c shows the surface detail of a surface oxide growth on copper: top, 5 kV; middle, 15 kV; bottom, 30 kV. The 5-kV image shows greater surface detail. The high-kilovolt images show decreasing amounts of information about the surface of the sample. Although the low-voltage mode provides enhanced surface detail, image resolution may be poorer.

2.1.3. Why Learn about Electron Optics?

In most cases, we would like to have the benefits of all these imaging modes simultaneously; however, each mode requires particular settings of the electron gun, lens strength, working distance, and aperture size. Some modes tend to be mutually exclusive: i_p is inescapably reduced when d_p and α_p are made small. When the voltage is lowered, d_p increases and i_p decreases. The four beam parameters must be chosen to obtain appropriate imaging conditions in each situation. Therefore, an elementary understanding of electron optics inside the electron column is necessary for intelligent SEM operation.

2.2. Electron Guns

The purpose of the electron gun is to provide a stable beam of electrons of adjustable energy. Several types of electrons guns are used on commercial SEMs. These vary in the amount of current they can produce into a small spot, the stability of the emitted current, and the lifetime of the source. Most older SEMs use tungsten or LaB_6 thermionic emitters, but, increasingly, new microscopes are equipped with cold, thermal, or Schottky field emission sources because these provide enhanced performance, reliability, and lifetime.

2.2.1. Tungsten Hairpin Electron Guns

The most common electron gun consists of three components: a tungsten wire filament ("hairpin") serving as the cathode (negative electrode), the grid cap or Wehnelt (control electrode), and the anode (positive electrode) as schematically shown in Fig. 2.7. These components are maintained at different electrical voltages by appropriate connections to the high-voltage supply, which is variable in the range 0.1–30 kV. For example, if the accelerating voltage is set to 20 kV, the filament will be placed at –20,000 V with respect to the anode, which is at ground potential. The grid cap is set to a voltage slightly more negative than the filament to provide a focusing effect on the beam.

2.2.1.1. Filament

The tungsten hairpin filament gun is now over 70 years old, but it is reliable, well-understood, and inexpensive (Haine and Einstein, 1952; Haine and Cosslett, 1961), and for many SEM applications, such as low-magnification imaging or x-ray microanalysis, still remains the best choice. The cathode is a tungsten wire, about 100 μm in diameter, bent into a V-shaped hairpin with a tip radius about 100 μm (Fig. 2.8). Thermionic emission produces significant beam current only when the filament is at white heat. Thus the filament is heated resistively by a current i_f to a temperature of 2000–2700 K. Thermionic electrons are then emitted from the tip

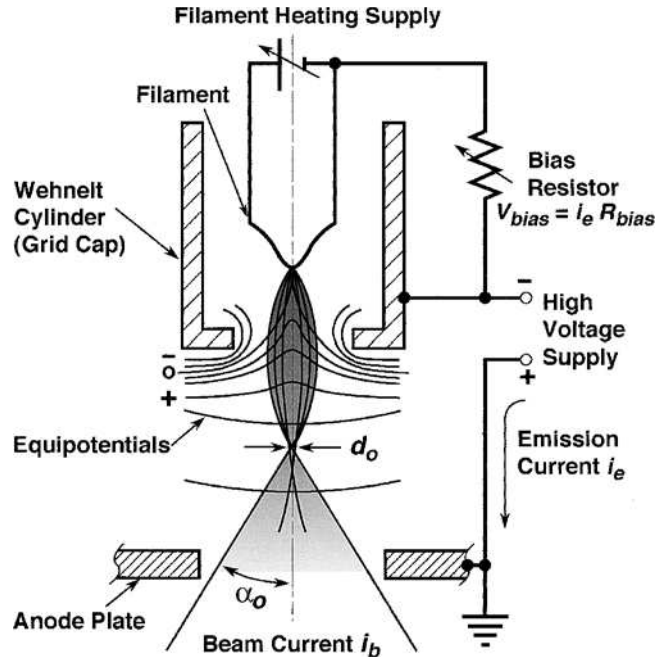


Figure 2.7. Schematic diagram of the conventional self-biased thermionic tungsten hairpin electron gun. (Adapted from Hall, 1966.)

over an area about 100 μm $\times$ 150 μm. A further discussion of thermionic electron emission is given in the Enhancements CD, Section E 2.1.

2.2.1.2. Grid Cap

Emitted electrons spread out into a broad cone from the filament emission area (Fig. 2.7). The grid cap (Wehnelt cylinder) acts to focus electrons inside the gun and to control the amount of electron emission. The grid cap is connected to the filament by a variable resistor. The emission current leaving the filament is replaced by an equal current to the filament through this resistor. This generates a negative bias between the grid cap and the filament. In Fig. 2.7 lines of constant electrostatic field potential (equipotentials) varying from negative through zero to positive are plotted. Electrons move toward positive potentials, so they leave the filament only where positive electrostatic field lines meet the surface of the filament. The bias voltage causes a high curvature in the equipotentials near the hole in the grid cap, producing a focusing action on the emitted electrons. This forces the electrons to a crossover of diameter d_0 and divergence angle α_0 between the grid cap and the anode as shown in Fig. 2.7. As we shall see, the condenser and objective lenses then produce a demagnified image of this crossover on the specimen, the final electron probe d_p.

2.2.1.3. Anode

Emitted electrons in the gun are accelerated from the high negative potential of the filament (e.g., $-20{,}000$ V) to ground potential (0 V) at the anode. A hole in the anode allows a fraction of these electrons to proceed down the column toward the lenses. Electrons collected on the anode and elsewhere down the column return via the electrical ground to the high-voltage power supply.

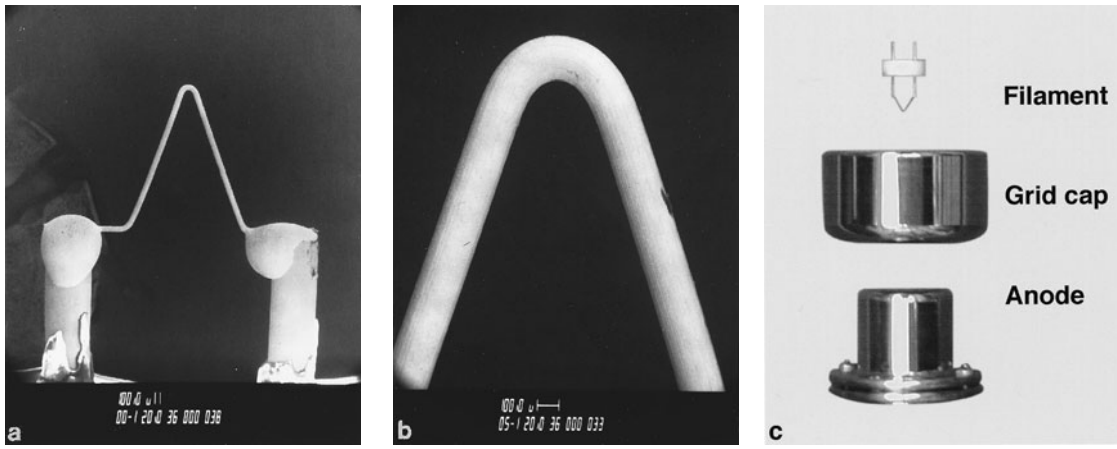


Figure 2.8. Conventional tungsten hairpin filament electron gun. (a) Filament wire spot-welded to support posts, (b) tip of wire showing wire-drawing marks, and (c) exploded view of electron gun showing filament assembly, grid cap, and anode displaced from one another.

2.2.1.4. Emission Current and Beam Current

The important parameters for any electron gun are the amount of current it produces and the stability of that current. The total current emitted from the filament is called the "emission current" i_e and the portion of electron current that leaves the gun through the hole in the anode is called the "beam current" i_b. At each lens and aperture along the column the beam current becomes smaller and is several orders of magnitude smaller when it is measured at the specimen as the "probe current" i_p.

2.2.1.5. Operator Control of the Electron Gun

Because information from the SEM is recorded serially as a function of time, the probe current must be the same for each point in the image. To ensure a stable beam current, the filament heating current i_f is adjusted to achieve the condition known as "saturation" where small increases or decreases in the filament current i_f will not change the electron beam current i_b because a plateau has been reached (Fig. 2.9). At saturation electrons are only emitted from the tip of the filament and are focused into a tight bundle under the influence of the negative bias voltage.

Saturation is a self-regulating negative feedback process that ensures a stable beam current. If the filament heating current i_f increases, then an increase in electron emission from the filament results. The increased emission current flowing through the bias resistor increases the negative bias voltage on the grid cap, which opposes the increase in emission. This

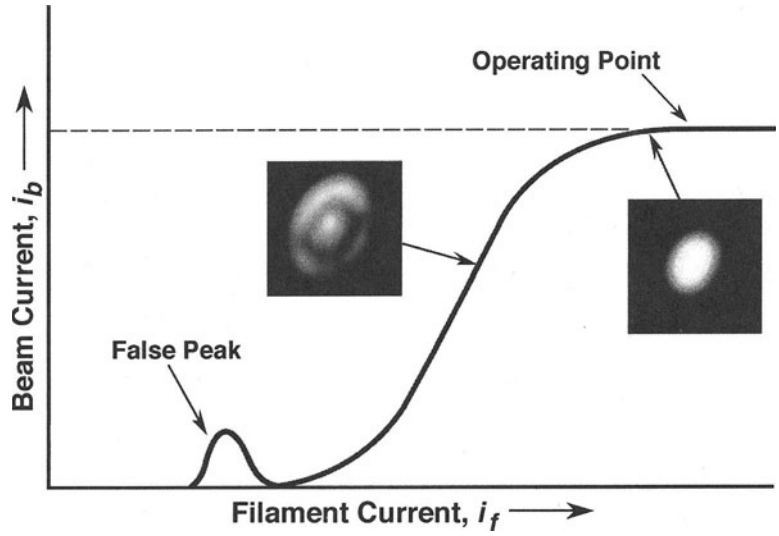


Figure 2.9. Saturation of a tungsten hairpin electron gun. At low filament heating some electrons are emitted into various directions from the undersaturated filament (left filament image). At higher filament heating, near saturation, electrons are concentrated into a tight bundle (right filament image). Further filament heating provides little increase in beam current. The gun should be operated just at the saturation point.

self-regulating feature of the autobiased electron gun (Fig. 2.7) guarantees the desired stable beam current. Each accelerating voltage requires a different filament setting to achieve saturation, and many modern SEMs compensate for this automatically.

2.2.2. Electron Gun Characteristics

There are several measures of electron gun performance: electron emission current, brightness, lifetime, source size, energy spread, and stability. Brightness is the most important of these because image quality at high magnification is almost entirely dependent on this parameter.

2.2.2.1. Electron Emission Current

The emission current i_e of a tungsten hairpin gun is typically about 100 μA, whereas field emitter guns generate emission currents of 10–30 μA. However, as mentioned above, only a small portion of the gun emission current passes through the anode aperture and the beam current i_b further decreases down the electron column as electrons are intercepted by various apertures. At each aperture the cross-sectional area and the convergence of the beam change, and this renders the emission current or beam current by itself an insufficient parameter by which to evaluate electron gun performance.

2.2.2.2. Brightness

Electron-optic brightness β incorporates not only the beam current i_b, but also the cross-sectional area of the beam d and the angular spread α of the electrons at various points in the column. Brightness is defined as the beam current per unit area per solid angle. It is important because it is "conserved," that is, the brightness at any point down the column is the same as the brightness measured at the electron source itself even as the individual values of i_b, d, and α change. For example, if α changes because a different-size aperture is inserted, the other two variables change so that β remains constant. Thus measurements of brightness made at the specimen level should be an estimate of the electron gun brightness. We can determine the brightness from the following equation:

$$\beta = \frac{\text{current}}{\text{area}\cdot\text{solid angle}} = \frac{i_p}{\left(\frac{\pi d_p^2}{4}\right)\cdot \pi\alpha_p^2} = \frac{4i_p}{\pi^2 d_p^2 \alpha_p^2}. \qquad (2.1)$$

[Note that if α is the beam convergence (or divergence) angle in radian measure, then, for small values of α, the solid angle is $\pi\alpha^2$ steradians (sr).] The "brightness equation" (2.1) is the first of two important equations that help the microscopist develop a practical understanding of electron-optical limitations on SEM images (the second is the "threshold equation," which is developed in Chapter 4).

By substituting experimental values of beam parameters into Eq. (2.1), one can make an estimate of gun brightness. Lens defects, called aberrations, cause this estimate to be lower than the actual gun brightness. Because brightness is the most important performance indicator for an electron gun, even such relative estimates are valuable. Theoretical estimates of brightness for a tungsten hairpin gun yield a value about 10^5 A/cm^2 sr at 20 kV.

For all electron guns, brightness increases linearly with accelerating voltage, so every electron source is 10 times as bright at 10 keV as it is at 1 keV. For thermionic guns, the brightness is optimized at the saturation point and varies with both the bias voltage and the filament-to-grid cap spacing, which is why SEM manufacturers provide detailed instructions for optimizing these adjustments. More information on brightness is given in the Enhancements CD, Section E 2.2.

2.2.2.3. Lifetime

Because the tungsten filament operates at white-hot temperatures, it gradually evaporates with time. Eventually the tungsten wire becomes thin and fails. Increasing the filament temperature (oversaturating) increases the evaporation rate, causing premature failure of the filament.

2.2.2.4. Source Size, Energy Spread, Beam Stability

The source size at the crossover of a tungsten hairpin gun is typically about $d_0 \sim 50$ μm, depending on the gun configuration and operating conditions. This relatively large size means that considerable electron-optic demagnification is required to achieve the small electron probe needed for good resolution in the SEM. More advanced electrons guns achieve higher brightness by providing a smaller effective source size, about 5 μm for LaB_6 and between 5 and 25 nm for field emitters.

The electron beam energy spread ΔE is the spread in electron energies leaving the filament. For the tungsten hairpin gun ΔE is 3.0 eV, compared to 1.5 eV for LaB_6 emitters and 0.3–0.7 eV for field emitters. This parameter is the major limiting factor in low-voltage operation, to be discussed later.

The stability of an electron gun is a measure of how constant the electron emission is over periods of minutes or hours. The most stable sources (about 1%/h) are Schottky field emitters, but all thermionic emitters display good performance. Cold field emitters have poor stability.

2.2.2.5. Improved Electron Gun Characteristics

To improve imaging performance in the SEM, the electron probe size must be reduced without causing a loss of probe current. The only way to accomplish this is to increase the electron gun brightness. Improvements in source brightness can be obtained by changing the filament material (to LaB_6) or the mechanism of emission (to field emission). These more advanced electron guns are rapidly displacing the conventional tungsten filament for applications where the highest performance is required. Table 2.1 compares the major electron gun characteristics for the various electron sources.

Table 2.1. Comparison of Electron Sources at 20 kV

Source	Brightness (A/cm^2 sr)	Lifetime (h)	Source size	Energy spread ΔE (eV)	Beam current stability (%/h)	Ref.
Tungsten hairpin	10^5	40–100	30–100 μm	1–3	1	[a,b]
LaB_6	10^6	200–1000	5–50 μm	1–2	1	[b,c]
Field emission						
Cold	10^8	>1000	<5 nm	0.3	5	[d,e]
Thermal	10^8	>1000	<5 nm	1	5	[e]
Schottky	10^8	>1000	15–30 nm	0.3–1.0	~1	[e]

[a] Haine and Cosslett (1961).
[b] Troyon (1987).
[c] Broers (1974).
[d] Crewe et al. (1971).
[e] Tuggle et al. (1985).

2.2.3. Lanthanum Hexaboride (LaB_6) Electron Guns

2.2.3.1. Introduction

Lanthanum hexaboride (LaB_6) electron guns provide 5–10 times greater brightness and a longer lifetime compared to tungsten (Lafferty, 1951; Broers, 1975). This material has a lower work function than tungsten, meaning that more electrons are emitted for the same heating temperature. The work function of a material is the energy barrier that must be exceeded for electrons to escape into the vacuum. The emitter is a tiny block of single-crystal LaB_6 about 100 μm in diameter and about 0.5 mm long (see Fig. 2.10). The crystal is both supported and resistively heated by either graphitic carbon or rhenium, two materials that do not chemically combine with LaB_6. The resistance of the supports is chosen so that the filament

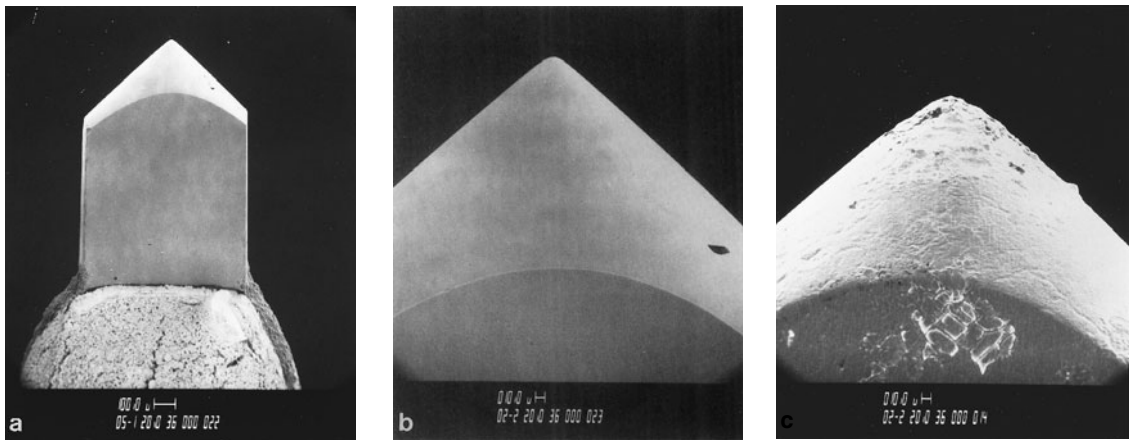


Figure 2.10. Lanthanum hexaboride source. (a) LaB_6 block mounted on support. (b) Higher magnification showing the finely ground tip with 10-μm radius. (c) Failed LaB_6 tip with evidence of evaporation and oxide formation.

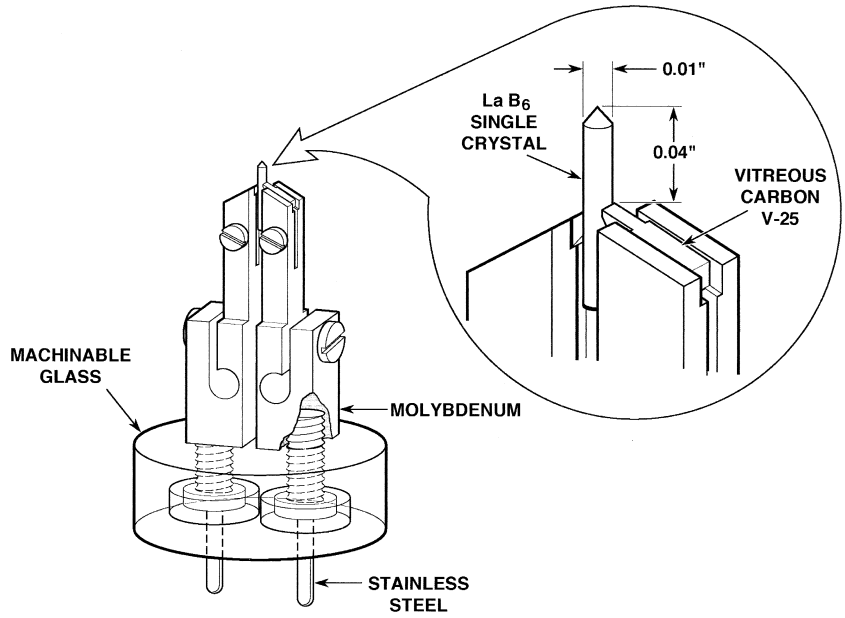


Figure 2.11. Schematic diagram of a directly heated LaB_6 emitter that directly replaces a tungsten filament assembly.

heating supply can provide enough current to raise the temperature of the LaB_6 to emission (Fig. 2.11). The sharpness of the tip also determines the emission, so the crystal is polished to about 1-μm tip radius to accentuate the electric field in front of the tip, making it easier for electrons to escape. Commercial LaB_6 guns are made to be direct plug-in replacements for the tungsten hairpin filament; only the bias voltage and the grid cap shape need to be modified. However, the vacuum in the gun chamber must be better than 10^{-4} Pa (about 10^{-6} torr), so differential pumping of the gun region with an additional ion pump is usually required.

Typical brightness values for LaB_6 emitters range from 5×10^5 to 5×10^6 A/cm^2 sr at 20 keV, depending upon the sharpness of the tip. Sharp emitters have the highest brightness, but the shortest lifetime; blunt or truncated tips exhibit slightly lower brightness, but longer lifetimes. LaB_6 emitters are generally more expensive to operate than the conventional tungsten hairpin gun because replacement LaB_6 emitters cost 10 times more than tungsten hairpins. Nevertheless the increased current into a given probe size and the significantly longer emitter lifetime ($>$1000 h) may justify the increased cost for many applications.

2.2.3.2. Operation of the LaB_6 Source

Because of the high electric field at the emitter tip, there is no saturation condition (Fig. 2.9). To avoid running the cathode at too high a temperature (i.e., above 1800 K) the safest approach is to use the "filament

imaging" mode provided on many SEMs and to raise the temperature until this image coalesces into a well-defined disk. Because LaB_6 cathodes are easily contaminated, the temperature must be raised slowly after any exposure to air to allow the emitter to clean itself. To avoid such problems some manufacturers recommend leaving the emitter operating 24 h a day at a low temperature, even when not in use. This avoids thermal shock from heating/cooling cycles, reduces thermal drift, improves the vacuum, and extends emitter lifetime.

2.2.4. Field Emission Electron Guns

The electron sources described so far rely on the use of high temperatures to enable electrons in the cathode to overcome the work function energy barrier and escape into the vacuum. These thermionic sources are relatively inexpensive and need no special vacuum, but have the disadvantages of low brightness, limited lifetime, and large energy spread. Field emission is an alternative way of generating electrons which is free of these disadvantages.

The field emission cathode is usually a wire fashioned into a sharp point (100 nm or less in radius) and supported by a tungsten hairpin as shown in Fig. 2.12. When a negative potential is applied to the cathode, the electric field is concentrated by the tip. When the field at the tip reaches a magnitude of about 10 V/nm, the potential barrier is lowered in height and also becomes so narrow that electrons can "tunnel" directly through it and leave the cathode (Gomer, 1961). Tungsten is usually used as the cathode material because it is strong enough to withstand the high mechanical stress produced on the tip by the electric field, but carbon and silicon nanotubes have also been employed successfully. A cathode current density of as high

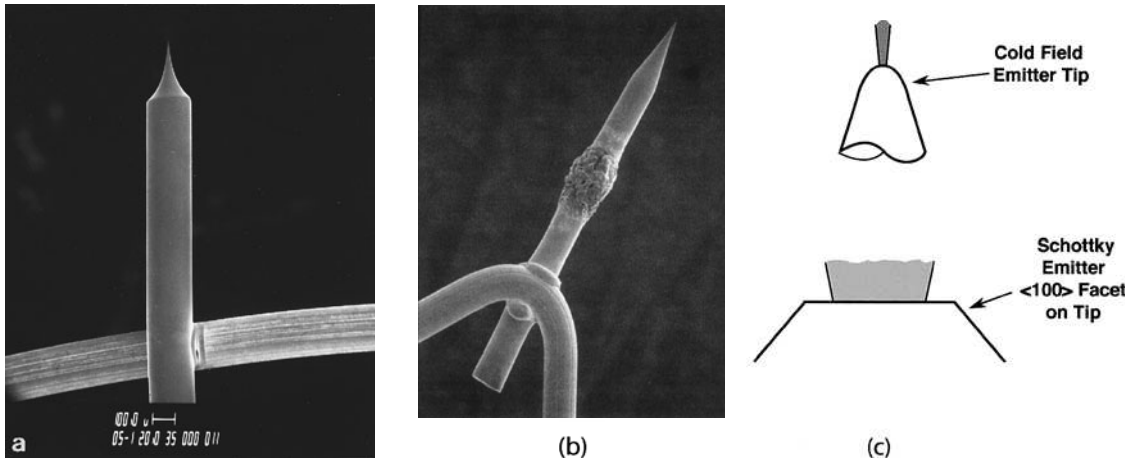


Figure 2.12. Examples of field emission sources. (a) Cold field emission tip of ⟨310⟩ single-crystal wire spot-welded to a tungsten wire. (b) Schottky field emission tip (courtesy of FEI Beaverton, OR). (c) Schematic diagram of emission patterns from a cold field emitter and a Schottky emitter (adapted from Tuggle *et al*., 1985).

as 10^5 A/cm^2 may be obtained from a field emitter as compared with about 3 A/cm^2 from a thermionic source.

Two forms of field emitter are now in common use in the SEM. The first is known as a cold field emitter (CFE). The CFE relies purely on the high applied field to cause electrons to tunnel out of the cathode wire and has its name because the magnitude of emission is independent of the temperature of the tip. The emitted electrons appear to come from an area which is only a few nanometers in diameter, and leave as a cone of electrons several degrees in width. Although the total emitted current is quite small, typically 1–10 μA, the source brightness is of the order of 10^8 A/cm^2 sr at 20 keV because of the confinement of the beam and the small source size. The CFE is invariably used in a triode field emission gun (Butler, 1966), which can act as a high-performance electrostatic lens to form a real crossover from the tip (Fig. 2.13). The voltage V_1 between the first anode and the tip determines the field strength available to extract electrons. Typically this voltage is in the range 3–5 kV to produce 10 μA of emission. The voltage V_0 between the tip and the grounded second anode determines the accelerating voltage of the gun, in the range between a few hundred volts and about 30 kV. The ratio of V_1/V_0 changes the focal properties of the gun lens, but the computer control typically provided on field emission-gun (FEG)-SEMs excuses the operator from the need to compensate for this behavior.

Cold field emission requires that the cathode surface must be atomically clean. Since even at a vacuum of 10^{-5} Pa a monolayer of gas is deposited every second, operation for sustained periods requires a vacuum level of 10^{-8}–10^{-9} Pa. Before use the tip is "flashed" clean by heating it for a few seconds to a temperature of about 2500 K. Following a flash (Fig. 2.14) the tip initially emits strongly, but the emission decays away over a period of 10–15 min as gas again builds up on the tip. Once a monolayer

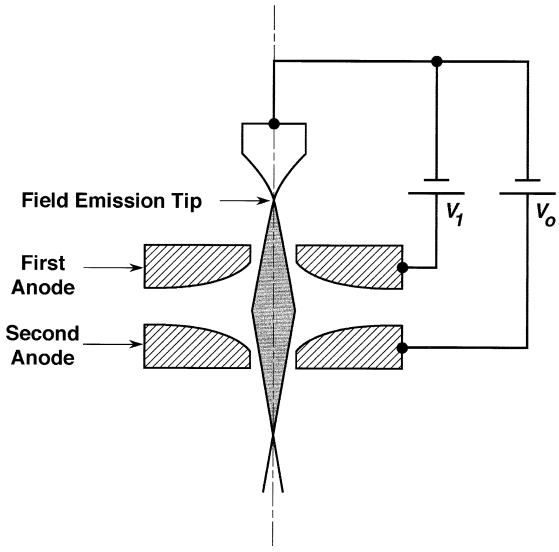


Figure 2.13. Schematic diagram of the Butler triode field emission source. V_1 is the extraction voltage of a few kilovolts and V_0 is the accelerating voltage. (Adapted from Crewe and Wall, 1970.)

of gas molecules has formed on the tip, the emission stabilizes and remains essentially constant for several hours until eventually instabilities again occur and the tip must be reflashed and the process repeats. Each time the tip is flashed it becomes slightly more blunt, and after many thousands of flashes the system may have insufficient V_1 voltage to achieve the desired emission current. The tip must then be replaced with a new, sharper emitter. Because a tip is usually only flashed once a day, the blunting process may take many years and so the useful life of a CFE is very long. The advantages of CFE are (1) the virtual source size of 3–5 nm, which means that little demagnification is required to get a 1-nm-diameter probe, (2) the very small energy spread, which enhances operation at low accelerating voltages, and (3) the fact that the tip lasts years, provided that the electron-optical column is kept intact, clean, and aligned, which aids reliability and reproducibility.

The second class of sources includes Schottky (SFE) and thermal (TFE) field emitters. A TFE has the same properties as a CFE, but is operated at an elevated temperature. This helps to keep the tip clean, reducing noise and instability even in degraded vacuum conditions. The emitter tip also sharpens itself in the extraction field, enhancing the performance. Thermal field emitters can even be operated for brief periods as cold field emitters by lowering the temperature.

In the Schottky emitter the field at the tip is mostly used to reduce the effective work function barrier. To still further lower the work function, ZrO_2 is deposited on the flattened tip from a small dispenser (Swanson et al., 1981; Tuggle *et al*., 1985). As a result, although the SFE is a thermionic source, its brightness and emission density are comparable with those of a CFE. SFE guns are generally similar to those for other field emitters, but include a suppressor grid to eliminate unwanted thermionic emission from regions outside of the tip. Emission currents in the range 30–70 μA are available. A Schottky gun is run continuously even when no current is being drawn from it, ensuring that it stays clean and stable and that no flashing is required. Because the ZrO_2 reservoir is finite in size, the useful lifetime of

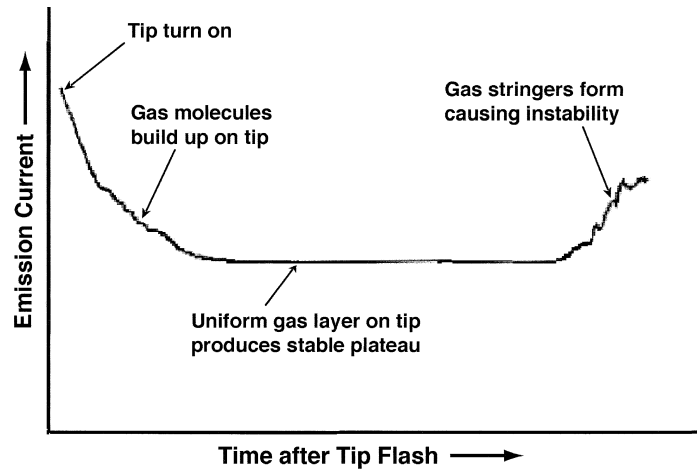


Figure 2.14. Emission current intensity versus time after a cold field emission tip has been flashed (rapidly heated) to remove adsorbed gas atoms.

a SFE is about 12–15 months, so it must be replaced on a regular basis. In principle the vacuum level required for successful Schottky operation is not as demanding as that for a CFE, but in practice an ultrahigh vacuum aids long-term stability, prevents poisoning of the ZrO_2 cathode, and maximizes brightness.

The virtual source size of a TFE is the same as for a CFE. The virtual source size of a SFE is larger, 20–30 nm, because the tip end radius is quite large. Thus, more demagnification is required for the SFE. This larger virtual source size may be advantageous, however, in microscopes where a wide range of spot sizes is required. The energy spread of the SFE is also higher than that of the CFE because the tip is warm, and this may be significant at low beam energies. However, the high current output, the ability to operate continuously for very long periods without adjustment, and the excellent stability of the SFE, make it the emitter of choice for many demanding applications.

Without exception all field emission systems are fully computerized, requiring only that the operator set an accelerating voltage and emission current. This convenience, the high performance, and the long-term reliability of such sources have made field emitters the preferred choice for many users. More information on field emission is given in the Enhancements CD, Section 2.3.

2.3. Electron Lenses

Electron lenses are used to demagnify the image of the beam crossover in the electron gun (e.g., $d_0 \sim 50\ \mu$m for a heated tungsten gun) to the final spot size on the specimen ($\sim$10 nm). This represents a demagnification of 5000$\times$. In a field emission system the source size is small already, so the FEG requires only a demagnification of 10–100 times to produce a 1- to 2-nm probe size. Electrons can be focused by electrostatic or electromagnetic fields, but SEMs employ only electromagnetic lenses because these lenses have smaller aberrations. Even with magnetic focusing, electron lenses perform poorly compared to typical glass lenses for focusing light. This is because some aberrations that are easily corrected in glass lenses cannot be corrected in electron lenses (except with additional complicated aberration correctors).

2.3.1. Making the Beam Smaller

2.3.1.1. Electron Focusing

An electromagnetic lens for focusing electrons is shown in Fig. 2.15. It consists of an iron case enclosing a coil of wire which generates a magnetic field across the lens gap between the polepieces. When energized by a lens current I through N turns of wire in the coil, fringing magnetic field lines extend across the gap. Off-axis electrons interact with these fringing fields

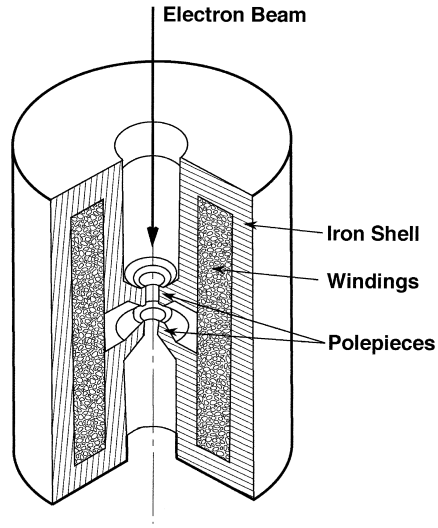


Figure 2.15. A rotationally symmetric electron lens where the coil windings are inside the iron shroud and the field is produced across the lens gap between polepieces. (Adapted from Hall, 1966.)

and begin to spiral through the lens and move toward the optic axis. The distance from the point where an electron starts to bend toward the axis to the point where it crosses the axis is called the lens focal length f. The focal length is a measure of lens strength and determines how much the beam crossover can be demagnified by passing through the lens. Unlike glass lenses, we can change f, and thus the demagnification, by changing the current I in the lens coil. For all but the strongest lenses the focal length f is nearly proportional to $V_0/(NI)^2$, where V_0 is the beam accelerating voltage. Because beams at higher accelerating voltages require more lens current to keep the same focal length, modern SEMs automatically compensate each lens current as the accelerating voltage is changed. More information on magnetic lenses is given in the Enhancements CD, Section E 2.4.

2.3.1.2. Demagnification of the Beam

The beam crossover in the electron gun d_0 must be demagnified to produce the desired spot d_p on the specimen. Neglecting lens aberrations, the demagnification of the electron beam can be described by the same simple geometrical equations used for light lenses. For weak lenses, the focal length and demagnification may be calculated from

$$\frac{1}{f} = \frac{1}{p} + \frac{1}{q}$$

$$\text{Magnification} = M = \frac{q}{p} \tag{2.2}$$

$$\text{Demagnification} = m = \frac{p}{q},$$

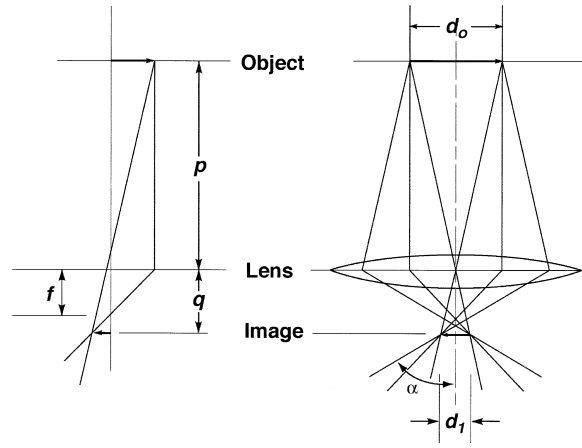


Figure 2.16. Geometric optics for a thin lens operated in demagnifying mode. (Left) Geometric ray diagram showing the graphical construction for determining the focal length ***f***. (Right) The geometric demagnification of the gun crossover d_0 for forming the intermediate crossover d_1 after the first condenser lens. Note that the image is inverted compared to the object.

where p is the distance from the object (beam crossover) to the center of the lens and q is the distance from the center of the lens to the image as shown in Fig. 2.16. By scaling these distances from Fig. 2.16 we find the demagnification through the lens is 3.4. Thus, the image of a tungsten gun crossover of $d_0 = 50\ \mu$m, after passing through this lens, would be $d_1 = d_0/m = d_0/3.4 = 14.7\ \mu$m. Using two or three lenses to demagnify, one can reduce the probe size for a typical tungsten hairpin SEM to about $0.01\ \mu m = 10$ nm, excluding the effects of aberrations. Note that the image is inverted and rotated compared to the object.

2.3.2. Lenses in SEMs

2.3.2.1. Condenser Lenses

One to three magnetic lenses are usually required to demagnify the electron beam in the SEM. The first condenser lens is variable and controls the amount of demagnification for a particular imaging mode. In microscopes with a second condenser lens, the two are often "ganged" so that a single condenser control, usually labeled "spot size," simultaneously adjusts both lenses.

2.3.2.2. Objective Lenses

The final lens in the column focuses the electron probe onto the sample surface and contributes additional demagnification. The knob controlling the current flowing through the winding of this "probe-forming" or "objective" lens is usually labeled "focus." Because the objective lens is much stronger than the condensers, with more current flowing through its windings, it usually must be water-cooled. The design of this lens must also

provide space for the scanning coils, the stigmator, and the beam-limiting aperture. As shown in Fig. 2.17, there are three basic designs of objective lens: (1) the pinhole lens, where the specimen is placed outside the lens and its magnetic field; (2) the immersion lens, where the specimen is placed inside the lens and its magnetic field; and (3) the snorkel lens, where the magnetic field reaches out to the specimen below the physical lens.

Pinhole Lens. The most common objective lens is called the asymmetric pinhole lens or conical lens, as shown in Fig. 2.17a. The final polepiece has a very small bore that keeps the magnetic field within the lens; this provides a field-free region above the specimen for the collection of secondary electrons (Oatley, 1972). Because both the objective lens focal length and its aberrations increase with the working distance, the specimen must be kept as close as possible to the polepiece to minimize lens aberrations that will enlarge the final probe size. There are two main advantages of the

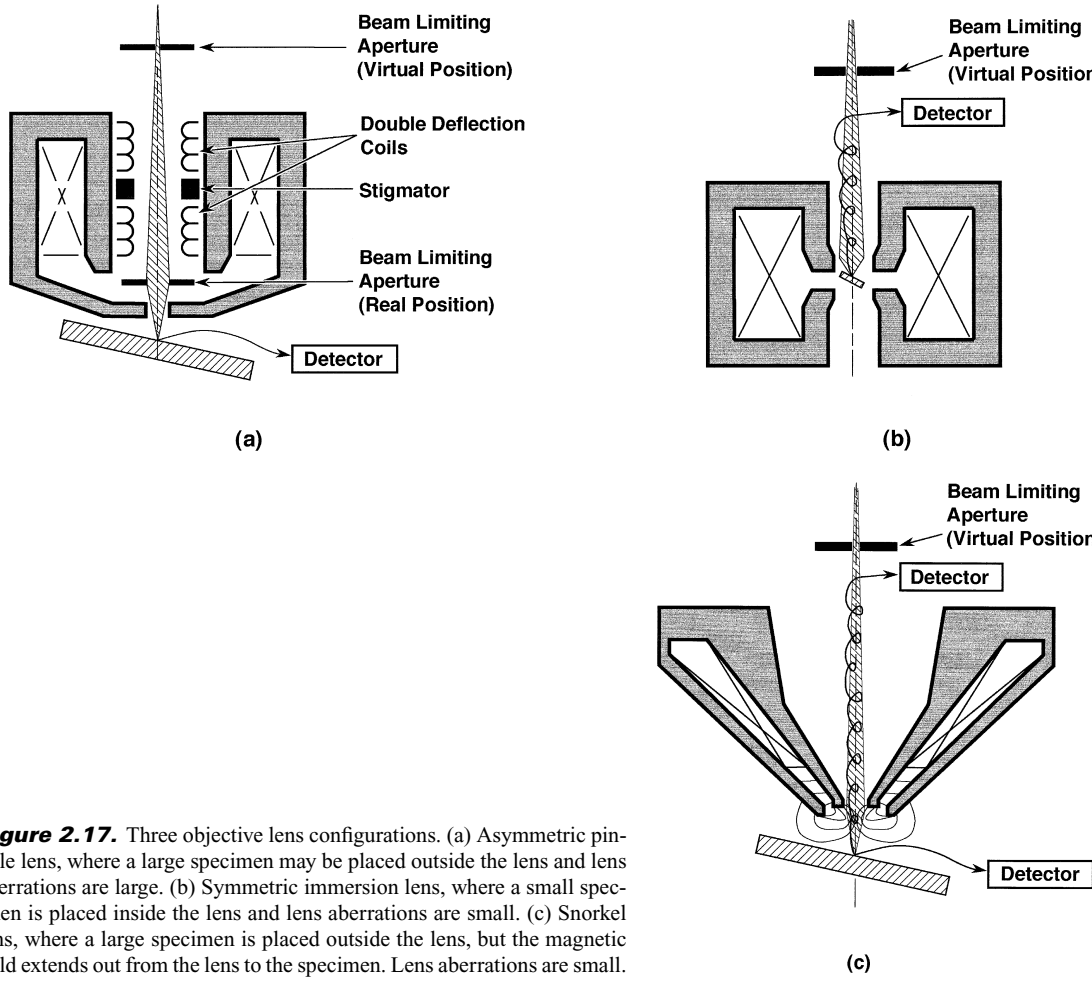


Figure 2.17. Three objective lens configurations. (a) Asymmetric pinhole lens, where a large specimen may be placed outside the lens and lens aberrations are large. (b) Symmetric immersion lens, where a small specimen is placed inside the lens and lens aberrations are small. (c) Snorkel lens, where a large specimen is placed outside the lens, but the magnetic field extends out from the lens to the specimen. Lens aberrations are small.

pinhole lens. First, specimen size is limited only by the size of the specimen chamber. Second, working distances up to 40 mm can be employed to produce images with high depth of field.

Immersion Lens. In the immersion lens, a small specimen (<5 mm) is placed directly inside the lens gap (Fig. 2.17b), giving a focal length in the range 2–5 mm. Because lens aberrations scale with focal length, this design yields the lowest aberrations, the smallest probe size, and the highest image resolution. Collection of secondary electrons in this lens takes advantage of the fact that the SEs can spiral upward in the strong magnetic field to a detector above the lens (Koike *et al*., 1971). With this "through-the-lens" detector, secondaries can be separated from backscattered electrons; the highly energetic backscattered electrons are less strongly confined by the magnetic field and do not reach the detector.

Snorkel Lens. The snorkel, or single-pole, lens (Mulvey, 1974) is designed such that a strong magnetic field extends from the polepiece directly to the specimen (Fig. 2.17c). This lens combines the best features of pinhole and immersion lenses. It has low aberrations like the immersion lens, but can accommodate large specimens like the pinhole lens. In addition, it can simultaneously accommodate both a lower (i.e., an E–T) and a through-the-lens (or upper) secondary electron detector, providing valuable flexibility for imaging.

2.3.2.3. Real and Virtual Objective Apertures

The beam-limiting aperture limits the convergence angle α of the beam to control the aberrations of the final lens. An objective aperture is called "real" if it is located inside the gap of the probe-forming lens (Fig. 2.17a). This location is near the poorer vacuum of the specimen chamber, so the aperture risks becoming contaminated. However, in many newer SEMs the beam-limiting aperture is located between the last condenser and the objective lens, and is referred to as a "virtual objective aperture." Virtual apertures are required for the immersion and the snorkel lenses, and tend to stay clean because they are further from the specimen chamber (Fig. 2.17b,c). For the same beam convergence angle the opening required in the virtual aperture is equal to the size of the equivalent real aperture divided by the demagnification of the objective.

2.3.3. Operator Control of SEM Lenses

The operator has three options for controlling the characteristics of the electron probe: changing the size of the objective aperture, setting the working distance using the specimen stage *z*-control, and adjusting the first condenser lens (C1). To see the effects of these controls on the electron beam parameters i_p, d_p, and α_p, we will use a schematic SEM column with a single condenser and a probe-forming objective lens (Fig. 2.18).

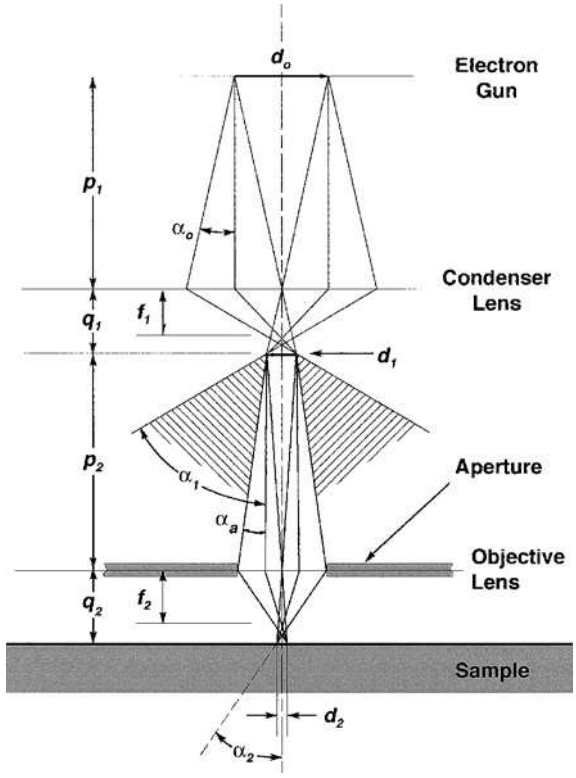


Figure 2.18. Electron ray traces through a schematic SEM column with a condenser lens and a probe-forming or objective lens. Lens distances p and q are marked for each lens.

2.3.3.1. Effect of Aperture Size

A real objective aperture placed in the gap of the probe-forming lens (usually 50–500 μm in diameter) is shown in Fig. 2.18. This aperture decreases the beam angle α_1 diverging from the condenser lens to a smaller angle α_a for the electrons entering the objective lens. The final aperture has three important effects on the final electron probe. First, as we will see below, there is an optimum aperture angle that minimizes the detrimental effects of aberrations on final probe size. Second, the final convergence angle controls the image depth of focus (see Chapter 4). Third, the aperture determines the current in the final probe because only a fraction of the current sprayed out to angles α_1 passes within the aperture angle α_a.

2.3.3.2. Effect of Working Distance

The effect of the working distance on beam parameters may be understood by changing this parameter alone, leaving the aperture size and condenser lens strength the same. The working distance W (the distance from the center of the lens to the specimen plane) may be increased by lowering the specimen stage with the z-control and refocusing the beam on the

specimen. From Fig. 2.19a to Fig. 2.19b the working distance W has been increased, making q_2 larger and the demagnification $m_2 = p_2/q_2$ smaller [see Eq. (2.2), and Fig. 2.18]. Increasing the working distance produces a larger spot size d_2 at the specimen and a consequent degradation of the image resolution, although the beam current remains about the same. The convergence angle α_2 decreases, giving an improved depth of focus. Weakening the objective to focus at a long W increases both the focal length f_2 and the aberrations of the lens. The longer working distance also increases the scan length that the beam traverses on the specimen (see Fig. 2.3) and in this way reduces the magnification. If one wishes to operate at a particular working distance W, one sets the current in the objective lens to a specific value. The specimen is then moved vertically with the z-control until it comes into focus.

2.3.3.3. Effect of Condenser Lens Strength

The strength of the first condenser lens controls both the final probe size and the final probe current. In Fig. 2.20a the condenser lens is weak

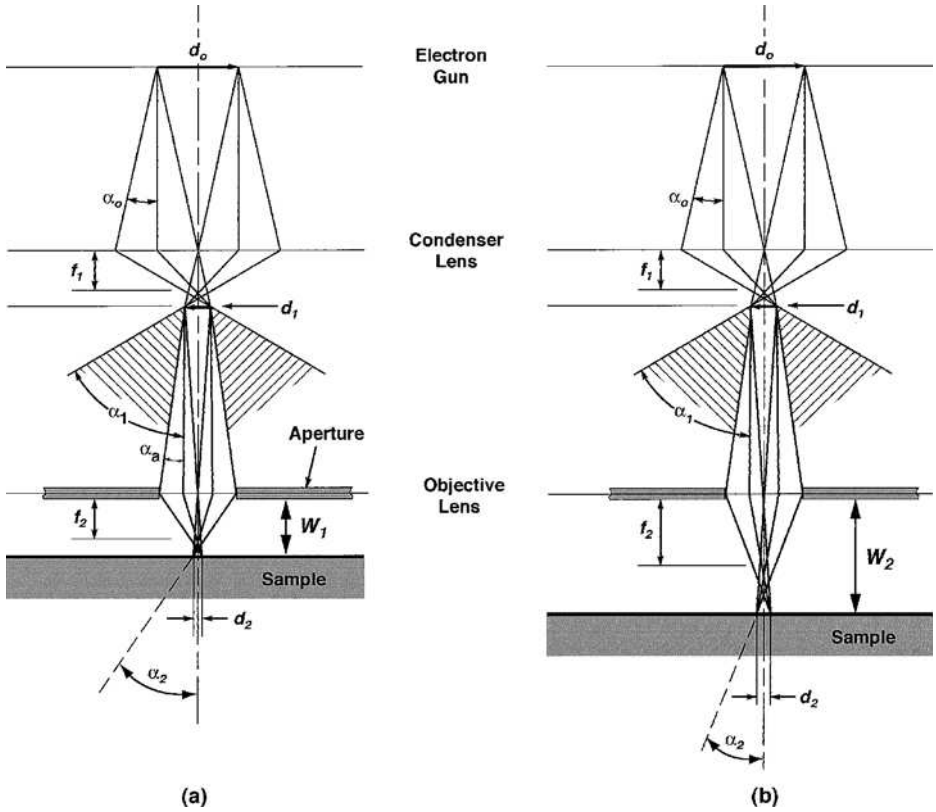


Figure 2.19. Effect of working distance in a two-lens system. (a) Small working distance, (b) large working distance.

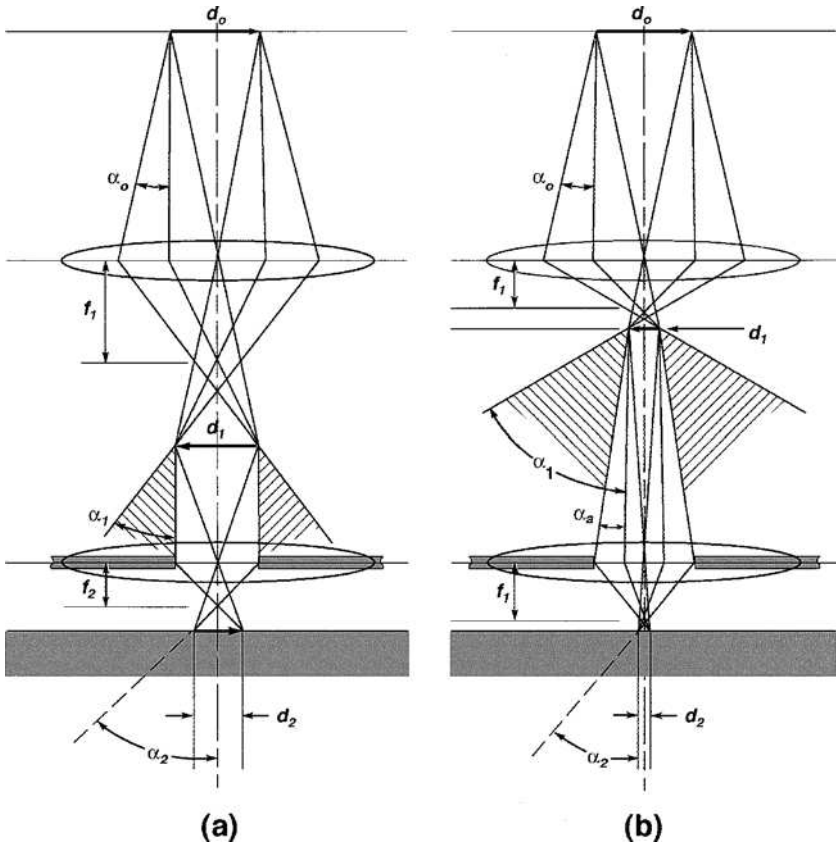


Figure 2.20. Effect of condenser lens strength in a two-lens lens system. (a) Weak condenser lens, (b) strong condenser lens.

and more current passes through the aperture to become the probe current i_p on the specimen. However, the probe size $d_p = d_2$ is relatively large. Figure 2.20b shows that, keeping the same working distance and objective lens aperture size, an increase in condenser lens strength increases the demagnification of each lens, m_1 and m_2 (i.e., reducing q_1 increases p_2; see Fig. 2.18), reducing the probe size d_p. However, more beam current will be stopped by the aperture and not go into d_p. This loss of beam current can be seen in Fig. 2.20b by the increased fraction of cross-hatched area between the two lenses compared to Fig. 2.20a. The final probe size can only be reduced at the expense of decreasing the probe current and a conscious choice between minimizing probe size or maximizing probe current must be made for each imaging situation.

2.3.4. Gaussian Probe Diameter

To fully understand how probe size varies with probe current, we need to calculate the minimum probe size and the maximum probe current. We

start with a calculation of the aberration-free Gaussian probe diameter d_G, which is the full-width at half-maximum height (FWHM) of the intensity distribution of d_G. If the emitter source size is known, then the Gaussian probe size may be calculated from the total demagnification as described above. However, to compare field-emission with thermionic sources, another, more general method is needed to estimate d_G. Because electron-optical brightness is constant throughout the electron column, we can use it to estimate d_G (if we know the brightness β) by rearranging the brightness equation, Eq. (2.1), as applied at the final probe spot:

$$d_G = \sqrt{\frac{4i_p}{\beta\pi^2\alpha_p^2}}. \tag{2.3}$$

The current in the final probe can be estimated by rearranging the brightness equation to solve for current:

$$i_p = \frac{\beta\pi^2\alpha_p^2 d_G^2}{4}. \tag{2.4}$$

If there were no aberrations in the system, it would only be necessary to increase the convergence angle α_p to increase the probe current at a constant probe diameter. However, because of the aberrations present in the electron-optical system, α_p must be kept small, and the current available for a given probe diameter is thus limited.

2.3.5. Lens Aberrations

All lenses suffer from a number of defects or aberrations in their performance. However, in electron optics, by contrast to the situation in light optics, the effects of aberrations cannot be canceled by using combinations of lenses. The only recourse therefore is to try to minimize these effects.

2.3.5.1. Spherical Aberration

Spherical aberration arises because electrons in trajectories further away from the optic axis are bent more strongly by the lens magnetic field than those rays near the axis (Fig. 2.21a). Electrons emerging from the object at point *P* along path *PA* will be focused to the Gaussian image plane, which is where the image is expected when no aberrations are present. Electrons following path *PB*, the maximum divergence allowed by the aperture of the lens, will be focused more strongly and cross the optic axis at a point closer to the lens. This results in a disk rather than a point where all rays converge. The smallest disk occurs just in front of the Gaussian plane and is often called the spherical aberration disk of least confusion. The diameter of this disk d_s can be written as

$$d_s = \frac{1}{2}C_s\alpha^3, \tag{2.5}$$

where C_s is the spherical aberration coefficient and α is the angle of the outer ray through the lens. Because the value of C_s is nearly proportional

to the lens focal length f for pinhole lenses at long working distances C_s may be 20 or 30 mm. For immersion and snorkel lenses with short focal lengths, C_s is only a few millimeters. The contribution of d_s to the final probe diameter can be reduced by limiting α with an objective lens aperture. Unfortunately, a very small aperture significantly reduces the current in the probe and introduces aperture diffraction.

2.3.5.2. Aperture Diffraction

For very small apertures, the wave nature of electrons gives rise to a circular diffraction pattern instead of a point at the Gaussian image plane. Electrons emerging from point P diffract at the edge of the small aperture and appear in the image plane as a broad "Airy disk" intensity distribution surrounded by smaller subsidiary maxima as shown in Fig. 2.21b. Following Wells (1974), we take half the diameter of the Airy disk as the diffraction contribution to the spot size d_d given by

$$d_d = \frac{0.61\lambda}{\alpha}, \tag{2.6}$$

where λ is the wavelength of the electrons and α is the beam convergence. The wavelength λ (in nanometers) for electrons of energy E_0 (in electron volts) can be calculated with only a small error as

$$\lambda = \frac{1.24}{E_0^{1/2}}. \tag{2.7}$$

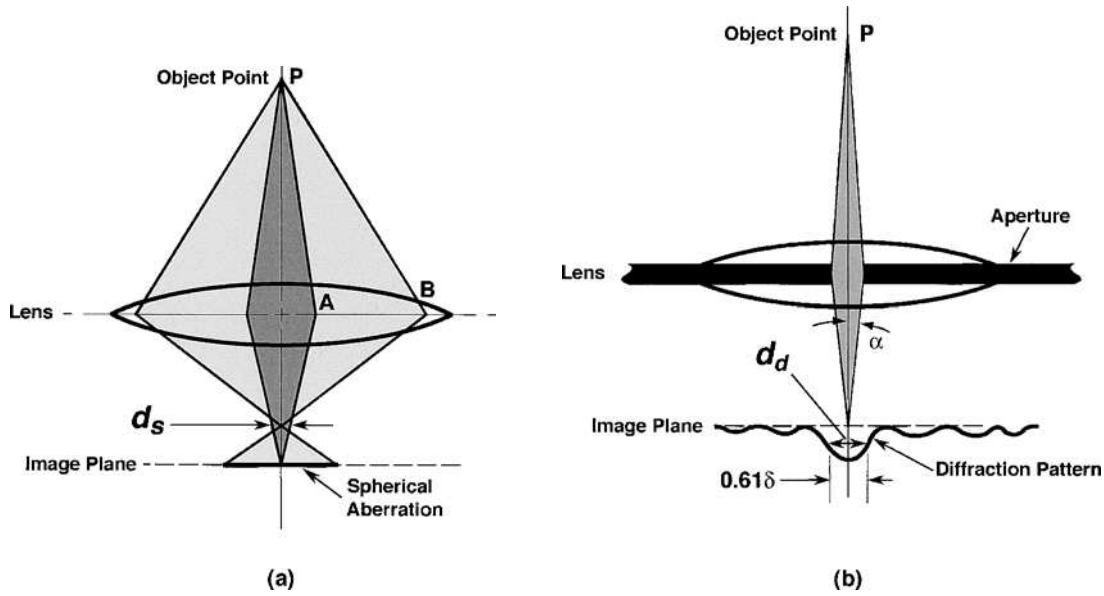


Figure 2.21. Schematic diagrams showing how (a) spherical aberration and (b) aperture diffraction in a lens cause a point object at P to blur into an enlarged spot at the Gaussian image plane. The disk of minimum confusion d_s and one-half the Airy disk d_d are used in calculations of probe size. (Adapted from Hall, 1966, and Oatley, 1972.)

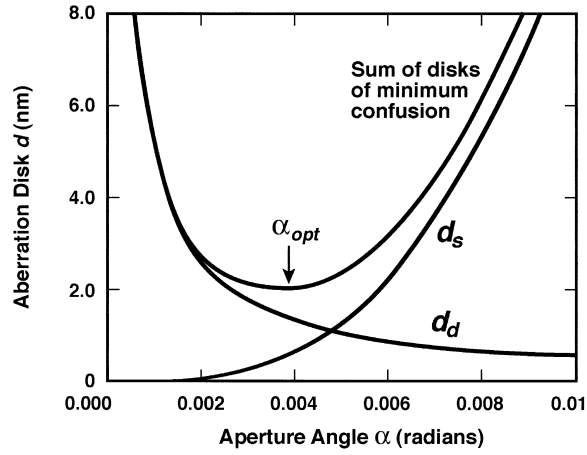


Figure 2.22. Spherical aberration disk d_s and aperture diffraction disk d_d plotted against aperture angle α. For the smallest electron probe and the best image resolution, the aperture angle is taken as α_{opt}. $C_s = 10\,mm$, $E_0 = 20\,keV$.

For aperture diffraction, the larger the value of α, the smaller will be the contribution of d_d. Thus, spherical aberration and aperture diffraction vary in opposite directions with respect to α. So an optimum aperture angle α_{opt} which produces a compromise between these two effects must be found (Fig. 2.22).

2.3.5.3. Chromatic Aberration

Electrons from point P of slightly different energies E_0 and $E_0 - \Delta E$ will be focused at different locations in the image plane (Fig. 2.23a). This again results in a disk rather than an ideal point. The diameter d_c of the disk of least confusion formed in front of the image plane is

$$d_c = C_c \alpha \left(\frac{\Delta E}{E_0} \right), \tag{2.8}$$

where C_c is the chromatic aberration coefficient, α is the convergence angle, and $\Delta E/E_0$ is the fractional variation in the electron beam energy. For example, the energy spread ΔE of electrons leaving a tungsten gun is 3 eV (Table 2.1), making $\Delta E/E_0 = 10^{-4}$ at 30 keV and 10^{-3} at 3 keV [Eq. (2.8)]. The effect of chromatic aberration will be 10 times greater at 3 keV than at 30 keV. The chromatic aberration coefficient C_c is again nearly proportional to the lens focal length f and is about the same magnitude. For most SEM objective lenses the value of C_c is similar to C_s, but it may be higher or lower than C_s depending on lens design and operating conditions. Contributions to d_c may also occur from imperfect stabilization of the accelerating voltage and the lens currents. Because modern power supplies are stable to better than 1 part in 10^6, however, these electrical instabilities should be unimportant (Oatley, 1972).

Machining errors, inhomogeneities in the polepieces, asymmetry in the lens windings, and dirty apertures all may lead to a lens that is not perfectly cylindrical, but is "astigmatic." This means that electrons diverging from a point object *P* will produce two separate line foci at right angles to each other as shown in Fig. 2.23b. The effect of astigmatism is shown in Fig. 2.24. Schematic cross sections of the electron probe are shown in Figs. 2.25a–2.25d corresponding to the set of images in Fig. 2.24. The astigmatism effect can be recognized by the stretching of image points into lines on one side of exact focus and into different lines, perpendicular to the first set, on the other side of exact focus (Figs. 2.24b and 2.24c and Figs. 2.25b and 2.25c). The stretching in the image vanishes at exact focus, but the effective probe size still may be many times the smallest possible probe, yielding a blurred image at the chosen magnification (Fig. 2.24a).

Fortunately this effect can be corrected using the stigmator, a device that applies a weak supplement magnetic field to make the lens appear

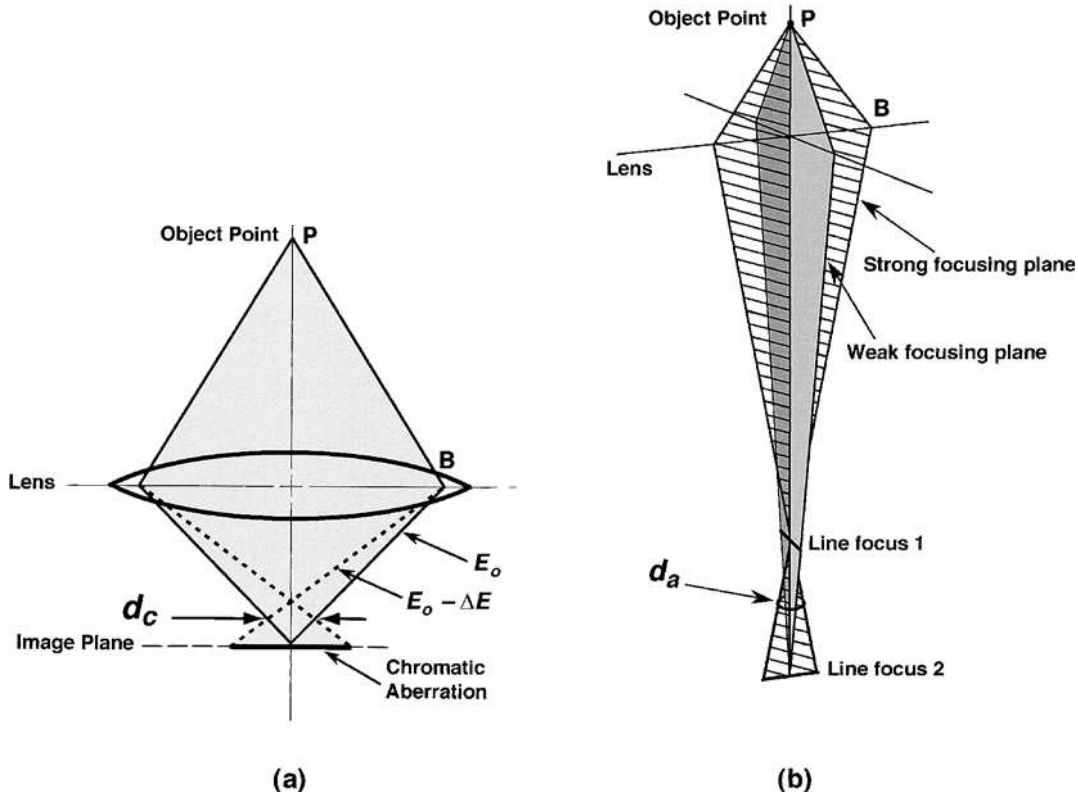


Figure 2.23. Schematic diagrams showing the origin of (a) chromatic aberration, where electrons of differing energy are focused at different locations, The chromatic disk of minimum confusion d_c is only important at low accelerating voltages. (b) A point object is focused to two line foci at the image, and the desired small focused beam can only be obtained by forcing the two line foci to coincide using the stigmator. (Adapted from Hall, 1966.)

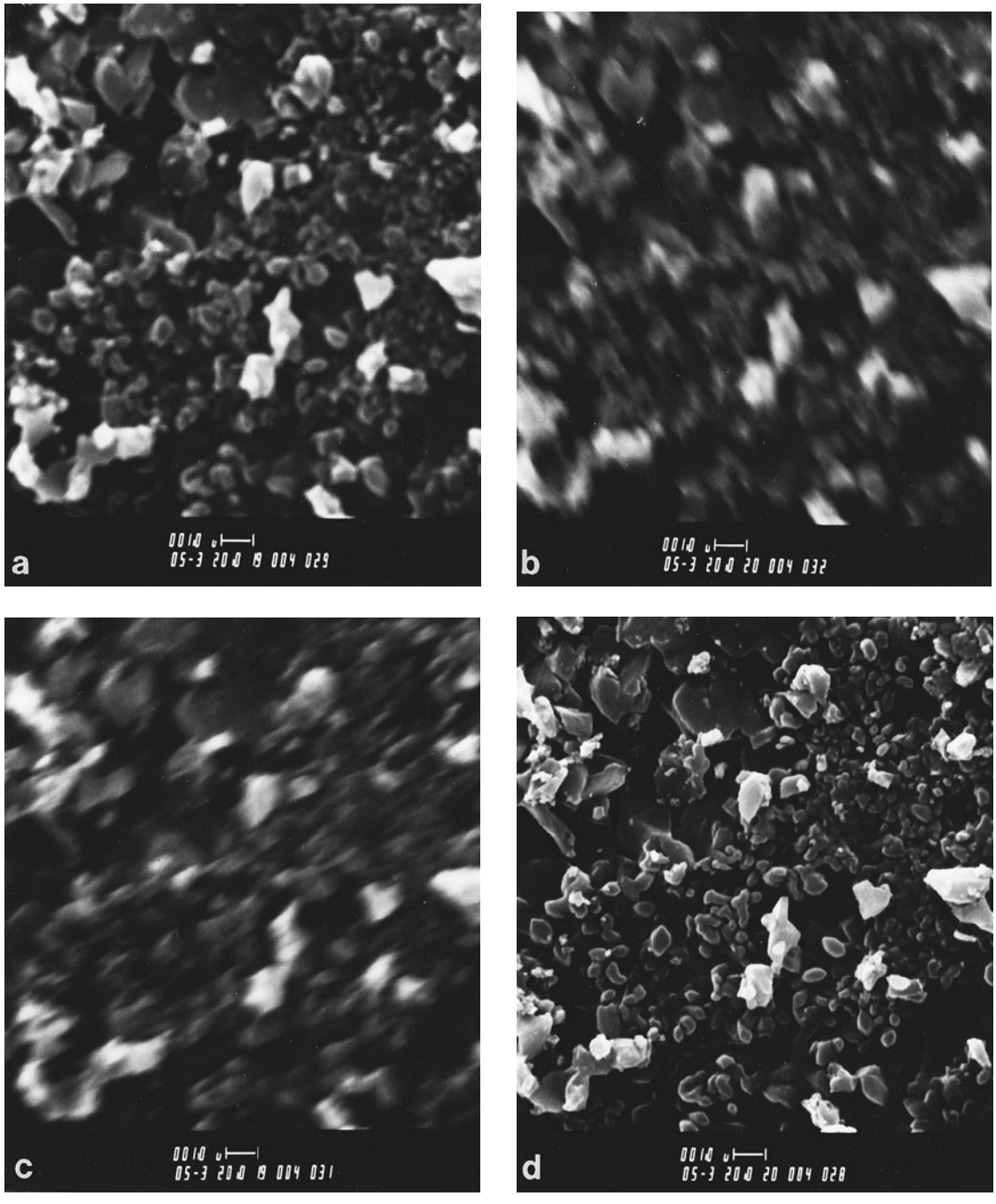
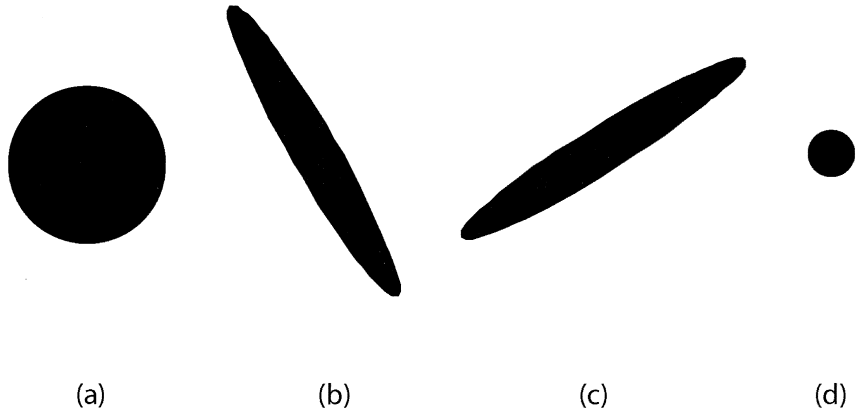


Figure 2.24. Effect of astigmatism in the probe-forming lens on an image containing fine random detail. (a) Initial situation, (b) underfocus, (c) overfocus, and (d) image corrected for astigmatism. Marker = 200 nm.

symmetric to the electron beam. The stigmator is usually an octupole of small electromagnetic coils that applies the required levels of additional field in the appropriate directions (see Fig. 2.25e). Astigmatism is corrected by alternately adjusting the *x*- and *y*-stigmator controls and the focus control while observing the image of a specimen containing fine random detail at high magnifications (>10,000×). The microscopist should avoid using a single edge or aligned edges as the test object. The astigmatism correction cycle (*x*-stigmator, focus, *y*-stigmator, focus) should be repeated until the sharpest image is obtained (Fig. 2.24d). At that point the beam cross section will be the smallest possible (Fig. 2.25d). If the stigmator cannot correct astigmatism in an image, the apertures or liner tube may need cleaning and the SEM column may need to be aligned.

2.3.5.5. Aberrations in the Objective Lens

The effects of the above aberrations are most significant in the objective lens because the amount of blurring these aberrations cause in the preceding lenses is small relative to the larger size of the beam at those lenses. Because astigmatism can be completely corrected in the final lens



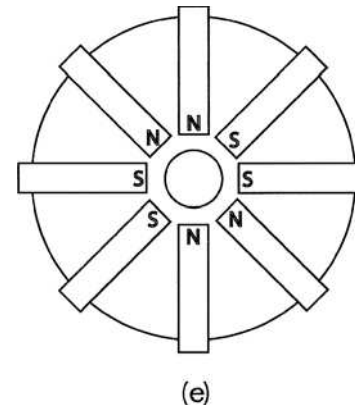


Figure 2.25. (a–d) Schematic shapes of beam cross sections corresponding to the images in Fig. 2.24. (e) Octupole stigmator with four sets of opposing magnetic poles, which can force the beam into a circular shape.

of a properly maintained SEM, only the effects of spherical aberration and aperture diffraction remain to be controlled at normal accelerating voltages of 15–30 kV. Chromatic aberration begins to seriously degrade the image at accelerating voltages under 10 kV.

2.4. Electron Probe Diameter versus Electron Probe Current

Aberrations are controlled in the SEM by the objective aperture, which limits the beam angle to α_{opt}. This optimum aperture size minimizes the effects of the major lens aberrations on the final probe size as shown in Fig. 2.22.

2.4.1. Calculation of d_{min} and i_{max}

It is of great importance to have as much current as possible in the small electron probes used for high-resolution imaging. Following Smith (1956) and Pease and Nixon (1965), it is possible to calculate the diameter d_p of the electron probe, including the effect of spherical aberration, and estimate the probe current i_p.

2.4.1.1. Minimum Probe Size

Calculations of the probe size assume that d_p is the quadrature sum of the diameters of the Gaussian probe and the various aberration disks, that is,

$$d_p = \left(d_G^2 + d_s^2 + d_d^2 + d_c^2\right)^{1/2}, \tag{2.9}$$

where d_G is the Gaussian probe size at the specimen [Eq. (2.3)], d_s is the spherical aberration disk [Eq. (2.5)], d_d is the aperture diffraction disk [Eq. (2.6)], and d_c is the chromatic aberration disk [Eq. (2.8)]. With some mathematical manipulation an optimum probe convergence angle α_{opt} can be found that produces the minimum probe size d_{min}. When the aperture size corresponding to α_{opt} is employed, this minimum probe size becomes $d_p = d_{min}$.

2.4.1.2. Minimum Probe Size at 10–30 kV

At normal accelerating voltages of 10–30 kV the relationship between probe size and probe current can be calculated at α_{opt} by considering only the spherical aberration and diffraction effects and neglecting chromatic effects. The relationship of d_p versus i_p is given by

$$d_{min} = K C_s^{1/4} \lambda^{3/4} \left(\frac{i_p}{\beta \lambda^2} + 1\right)^{3/8}, \tag{2.10}$$

where K is a constant close to unity. The value of d_{min} is given in nanometers when C_s and λ are in nanometers, and i is given in amperes (A). Equation (2.10) shows that the minimum probe size d_{min} decreases as the brightness β increases and as the electron wavelength λ and spherical aberration coefficient C_s decrease. In fact, in the limit of zero probe current, d_{min} reaches a value of approximately $C_s^{1/4}\lambda^{3/4}$, which is often regarded as a measure of the microscope's theoretical resolution.

2.4.1.3. Maximum Probe Current at 10–30 kV

Maximum probe current i_{max} is an important parameter for SEM imaging. Again neglecting chromatic effects at the optimum aperture angle α_{opt}, we find that the maximum probe current for a given d_p is given by

$$i_{max} = \frac{3\pi^2}{16}\beta\frac{d_p^{8/3}}{C_s^{2/3}}. \tag{2.11}$$

Note that when the effect of spherical aberration is included, the i_p dependence on d_p changes from d^2 [Eq. (2.4)] to nearly d^3 [Eq. (2.11)].

From Eq. (2.11) we see that there are several ways in which i_{max} can be increased. Using the condenser lens to increase the size of the electron probe will dramatically increase the beam current because i_{max} varies as $d_p^{8/3}$. At low magnifications the larger probes will not degrade image sharpness, but the increase in image quality will be significant (as discussed in Chapter 4). Second, the brightness may be increased by increasing the accelerating voltage (kV). However, the electron excitation volume beneath the surface increases as the E^5 power (Chapter 3), which influences resolution for microscopy and microanalysis applications. Third, the value of the spherical aberration coefficient can be minimized by using a short working distance W or using advanced lens designs with inherently lower C_s . For example, changing from a pinhole to a snorkel lens could reduce C_s by a factor of 10, which would increase i_{max} by a factor of 5 and reduce d_{min} by a factor of 2.

2.4.1.4. Low-Voltage Operation

Low beam energies provide a more realistic and detailed image of the surface and can often reduce the charge deposited in insulating or poorly conducting specimens, eliminating the need for metal coating. However, the performance of the electron optics at low accelerating voltages is always significantly worse than at higher energies. First, the gun brightness is reduced and so less current is available. Second, although the equations governing the effect of aberrations on the focused probe are the same as discussed above, the magnitude of the chromatic aberration and diffraction terms is now much larger than before. The effect of chromatic aberration is to remove electrons from the focused probe and to place them in a "skirt" about the probe. Although this does not have a major effect on resolution, it results in lower image contrast and visibility when sources of high energy spread (such as thermionic and Schottky guns) are used. Third,

at low energies the electrons are moving relatively slowly; and so if they are brought close together, as occurs at a focused crossover, they interfere with each other. This interaction, known as the Boersch effect (Boersch, 1954), defocuses the crossover by a factor which increases with the current density of the beam and inversely with the energy of the electron beam.

Analyzing the behavior of an electron-optical column at low energies cannot be done accurately using the quadrature addition of aberration terms discussed earlier because this no longer gives a reasonable estimate of the actual probe size and shape. More detailed analyses such as wave-optic techniques, the analytical method of Shao and Crewe (1989), or the ray-tracing method of Kenway and Cliff (1984) are therefore required.

2.4.1.5. Graphical Summary

The variation of the probe current i_p as a function of the effective probe size d_{min} [e.g., Eq. (2.10)] for a variety of imaging conditions is illustrated in Fig. 2.26. For these examples three different typical instrument configurations were considered: (1) a SEM for routine microscopy using a tungsten thermionic gun and a pinhole lens ($C_s = 20$ mm, $C_c = 10$ mm); (2) a SEM using a Schottky field emitter and a snorkel lens ($C_s = 3.2$ mm, $C_c = 2.7$ mm) designed for microanalysis; and (3) a high-resolution SEM using a cold field emitter and an immersion lens ($C_s = 2$ mm,$C_c = 1.8$ mm) optimized for high-resolution imaging. The major trends in probe size and probe current at the higher energies could be computed using the equations given earlier, but for consistency all of the curves shown here were derived using the Kenway and Cliff (1984) method. More information on the calculation of electron probe diameter and electron probe current is given in the Enhancements CD, Section E 2.5.

2.4.2. Performance in the SEM Modes

Operators should know the actual electron beam parameters for the SEM that they are operating. A chart of values for i_p, d_p, and α_p at various accelerating voltages, condenser lens strengths, and aperture sizes can be constructed, compared with Fig. 2.26, and saved for future reference. Information on the measurement of electron beam parameters is given in the Enhancements CD, Section E 2.6.

2.4.2.1. Resolution Mode

Figure 2.26a shows the probe size and probe current for the three systems at 30-kV beam voltage. The thermionic system has a minimum probe size of 10 nm (100Å), controlled by the spherical aberration of the pinhole lens. At 10 nm the probe contains in excess of 1 pA of current, which is sufficient for an acceptable photomicrograph. Because the source size of the thermionic gun is large, the probe size can be varied by changing the demagnification of the condenser lenses. As the condensers are weakened the amount of current in the probe increases by a factor of 10 without any

significant change in the actual probe diameter. Upon further weakening of the condenser lens, the demagnified image of the source becomes the dominant component in the spot size, which therefore begins to increase steadily together with a corresponding rise in beam current.

The Schottky gun and the snorkel lens, and the immersion-lens cold field emission gun (CFEG) system, by comparison, have the smallest probe diameters, between 1.2 and 1.8 nm at a current of 1 pA at 30 kV. These values reflect the much lower aberrations associated with these advanced lenses. The practical behavior of these systems is different from that of the thermionic emitter because the virtual source size of the field emitters is much smaller. On conventional SEMs the probe size and beam current are adjusted by varying the condenser lenses, but on field emission SEMs the spot size is usually maintained nearly constant and the beam current is changed by adjusting the aperture. The probe size of FEG system can be

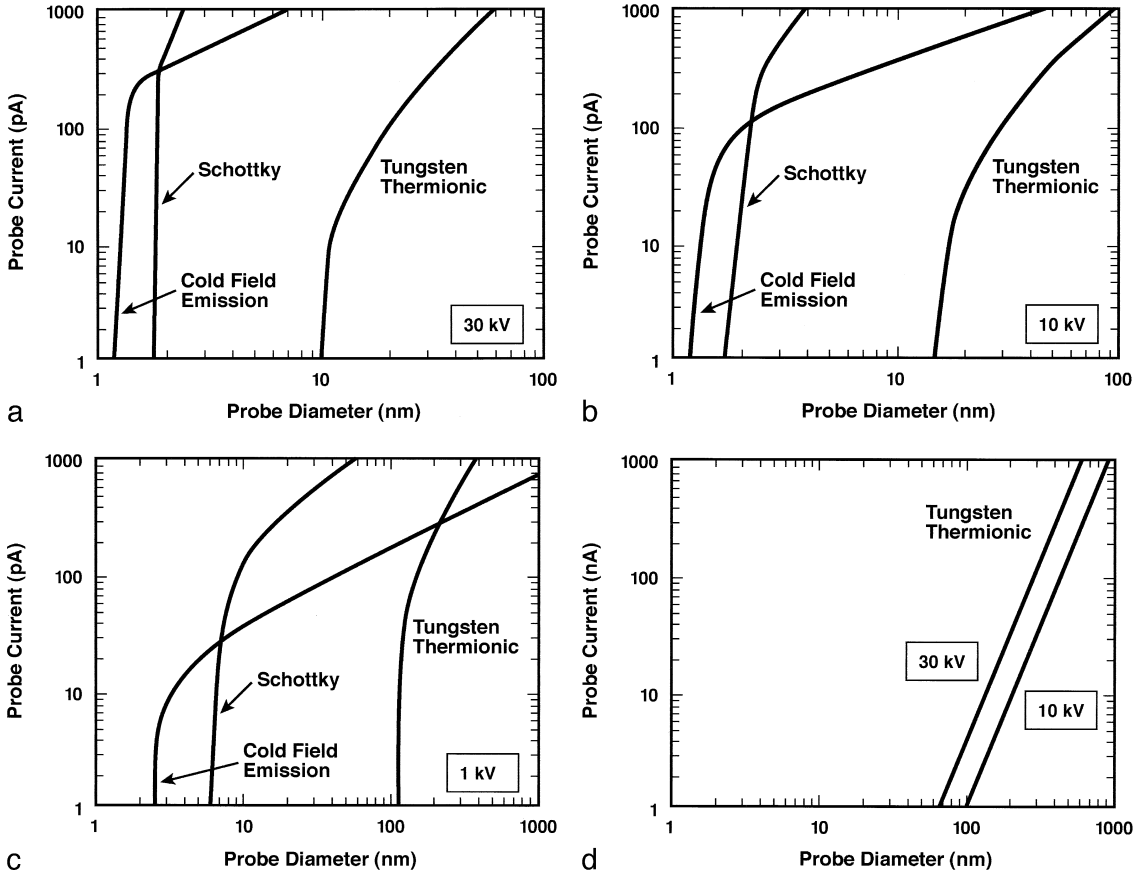


Figure 2.26. Relationship between the probe current i_p and probe diameter d_p calculated for a SEM using a tungsten thermionic source with a pinhole lens ($C_s = 20$ mm, $C_c = 10$ mm), a Schottky field emitter with a snorkel lens ($C_s = 3.2$ mm, $C_c = 2.7$ mm), and a cold field emitter with an immersion objective lens ($C_s = 2$ mm, $C_c = 1.8$ mm). (a) Imaging range at 30 kV. (b) Imaging range at 10 kV. (c) Low-voltage imaging at 1 kV. (d) Microanalysis range, higher probe currents at 10 and 30 kV.

increased by weakening the condenser excitation as before, but eventually the demagnification of the system prior to the objective lens becomes equal to unity. Because it is not efficient to magnify an electron probe since that leads to a rapid rise in spherical aberration, any further increase in probe size (and beam current) can only be achieved by increasing the convergence angle of the beam. For the Schottky system this change occurs at a probe size of about 2 nm, whereas for the cold field emitter, which has a much smaller source size, this condition is reached at about 1.4 nm. As the aperture is made larger, the beam current still continues to increase with the probe size, but at a slower rate than before. Because the Schottky emitter has a higher total emission current and comparable brightness to the cold field emission gun, it actually outperforms the CFE gun and lens for probe sizes larger than 2 nm.

Figure 2.26b plots the performance of the three systems at 10 kV. Note that the smallest probe size from the thermionic system has increased by about 50%, and by smaller amounts for the FE systems, because of the fall in brightness compared to 30 kV. Otherwise the 10-kV behavior is similar to that found at 30 kV.

2.4.2.2. High-Current Mode

Feature visibility and image quality improve with increases in probe current. However, Figs. 2.26a–2.26c show that in every case an increase in probe current has to be obtained by increasing the probe size. For the tungsten thermionic gun at 30 kV, to increase the probe current from 1 pA to 1 nA the probe diameter must be increased by about an order of magnitude to about 60 nm. This change in probe size will not, of course, affect the sharpness of images taken at low and medium magnifications because in these cases the resolution is determined by the pixel size. For SEM beam sizes of about 10 nm (100 Å), which would be adequate for good imaging up to about 10,000× magnification, the current available from the thermionic source is 10 pA or less, a value which is below the 100 pA necessary for reasonable energy-dispersive x-ray spectrometry (EDS) as discussed in Chapter 7. By comparison, the Schottky and the cold field emitters can provide more than adequate currents at probe sizes below 2 nm (20Å). Figure 2.26d demonstates that for thermionic emitters like tungsten, currents as high as 100–1000 nA can be obtained. These currents are useful for low-contrast SEM imaging and WDS trace-element and x-ray mapping (Chapter 9).

When working on bulk samples, the x-ray spatial resolution is limited to about 1 μm (Chapter 6). It would therefore be more appropriate to use electron probe diameters of 100 nm to 1 μm for optimum microanalysis. For 0.1- to 1.0 μm probes, the thermionic system can provide enough current (10^{-8} A, or 10 nA) for effective analysis not only with the EDS, but also with wavelength-dispersive systems (Chapter 6). The Schottky emitter can also provide currents of 10–15 nA, suitable for wavelength-dispersive systems, although into a much smaller probe diameter than the tungsten thermionic system. However, the cold field emitter is limited to maximum

currents of the order of 3–5 nA at 30 kV because of the low total emission current and the aberrations in the gun (Cleaver and Smith, 1973).

2.4.2.3. Depth-of-Focus Mode

To obtain a high depth of focus, α must be reduced to as small a value as possible by using a small aperture, a long working distance, or both. The small aperture will cut off much of the beam current, thus it is often necessary to intentionally increase the probe current by weakening the condenser lens. All three of these changes cause the probe size at the specimen to increase, degrading the attainable resolution. Figure 2.6b shows the effects on the depth of field in an image of changing aperture size.

For this mode the curves of probe size versus probe current are of little help except for estimating the size of the probe at a particular condenser lens setting so that the appropriate maximum magnification can be calculated. Actual probe sizes in this mode will be larger than normal because the convergence angle is smaller than optimum and the working distance is larger than optimum. Further discussion of depth of field can be found in Chapter 4.

2.4.2.4. Low-Voltage SEM

For operation at 1 keV there is a dramatic enlargement of the probe diameter for all electron sources and systems as shown in Fig. 2.26c. Chromatic aberration becomes the dominant aberration at low energies when the energy spread of the source is large. Thus for a thermionic source where ΔE is about 2.5 eV, the minimum probe size has increased to about 100 nm (1000Å). For the Schottky source with the snorkel lens, where ΔE is about 0.7 eV, the minimum probe is still only 6 nm (60Å); for the CFE source ($\Delta E = 0.3$ eV) and the immersion lens, a 2.5-nm (25-Å) spot containing 1 pA is achievable. Note, however, that for applications where high currents are required the advantage of the CFE source is quickly lost, so that the Schottky emitter becomes superior at about 7 nm (70Å), and even the thermionic system is competitive by 150 nm (1500Å). However, the size of the probe, and the current produced, is not the only consideration. For high-resolution imaging at low voltage the CFE source is greatly superior, even in the region where the Schottky emitter can produce more current, because its lower chromatic spread leads to images of much better contrast and definition. Systems employing cold field emitters also tend to suffer less from the Boersch effect because the current in the beam is smaller.

2.4.2.5. Environmental Barriers to High-Resolution Imaging

The curves of Fig. 2.26 calculate optimum probe sizes assuming that the SEM column is clean and well-aligned and that external effects such as mechanical vibration and ac magnetic field interference are negligible. In many cases, however, the ultimate resolution attained is more a function

of the room environment than of the electron optics of the microscope. Sources of vibration may be motors (over a large range of frequencies) or low-frequency oscillations (2–10 Hz) related to the natural vibrations of the building. Electromagnetic field interference (EMI) may be caused by electrical wiring (60 Hz or higher) in the vicinity of the microscope. The sensitivity of the electron beam to stray magnetic fields is greater at low accelerating voltages, and extra magnetic shielding of the column is provided in SEMs optimized for low-voltage operation. Finally, even when the microscope environment meets the criteria specified by the manufacturer, high-resolution imaging may be limited by contamination of the specimen surface with foreign molecules, either before or after insertion of the specimen into the microscope.

References

Boersch, H. (1954). *Z. Phys.* **139,** 115.
Broers, A. N. (1974). In *SEM/1974,* IIT Research Institute, Chicago, p. 9.
Broers, A. N. (1975). In *SEM/1975,* IIT Research Institute, Chicago, p. 662.
Butler, T. W. (1966). In *6th International Congress Electron Microscopy* (Kyoto) Vol. 1, p. 193.
Cleaver, J. R. A., and K. C. A. Smith (1973). In *SEM/1973/I,* SEM, Inc., AMF O'Hare, Illinois, p. 487.
Crewe, A. V., and J. Wall (1970). *Optik* **30,** 461.
Crewe, A. V., M. Isaacson and D. Johnson (1971). *Rev. Sci. Instr*. **42,** 411.
Gomer, R. (1961). *Field Emission and Field Ionization,* Harvard University Press, Cambridge, Massachusetts.
Hall, C. E. (1966). *Introduction to Electron Microscopy,* McGraw-Hill, New York.
Haine, M. E., and V. E. Cosslett (1961). *The Electron Microscope*, Spon, London.
Haine, M. E., and P. A. Einstein (1952). *Br. J. Appl. Phys,* **3,** 40.
Kenway, P. B., and G. Cliff (1984). *Inst. Phys. Conf. Ser*. **68,** 83.
Koike, H., K. Ueno and M. Suzuki (1971). In *Proceedings of the 29th Annual Meeting of the Electron Microscopy Society of America* (G. W. Bailey, ed.), Claitor's Publishing, Baton Rouge, Louisiana, p. 225.
Lafferty, J. M. (1951). *J. Appl. Phys.* **22,** 299.
Mulvey, T. (1974). In *SEM/1974*, IIT Research Institute, Chicago, Vol. 1, p. 43.
Oatley, C. W. (1972). *The Scanning Electron Microscope, Part 1, The Instrument*, Cambridge University Press, Cambridge.
Pease, R. F. W., and W. C. Nixon (1965). *J. Sci. Instr.* **42,** 281.
Shao, Z., and A. V. Crewe (1989). *Ultramicroscopy* **31,** 199.
Smith, K. C. A. (1956). Ph.D. dissertation, Cambridge University, Cambridge, England.
Swanson, L. W., M. A. Gesley, and P. R. Davis (1981). *Surf. Sci.* **107,** 263.
Troyon, M. (1987). *J. Microsc. Spectrosc. Electron*. **12,** 431.
Tuggle, D. W., J. Z. Li and L. W. Swanson (1985). *J. Microsc.* **140,** 293.
Wells, O. C. (1974). *Scanning Electron Microscopy*, McGraw-Hill, New York.

3

Electron Beam–Specimen Interactions

3.1. The Story So Far

In the introductory survey presented in Chapter 1, we learned that the SEM image is constructed by scanning a finely focused probe in a regular pattern (the scan raster) across the specimen surface. The spatial resolution achieved in this imaging process is ultimately limited by the size and shape of the focused probe that strikes the specimen. In Chapter 2 we learned how to control the critical parameters of the electron beam, energy, diameter, current, and divergence, through the use of electrical fields in the gun, magnetic fields in the lenses and stigmators, and beam-defining apertures. We saw how, depending on the type of electron source (tungsten hairpin, lanthanum hexaboride, thermal field emission, or cold field emission) and its inherent brightness (a constant dependent upon the beam energy that has been selected), it is possible to create focused beams with sizes ranging from nanometers to micrometers (three orders of magnitude) carrying currents ranging from picoamperes to microamperes (six orders of magnitude). This great flexibility in operational conditions permits the SEM microscopist/microanalyst to successfully attack a wide range of problems, provided that the proper strategy is employed. The strategy needed for selecting the proper operating conditions depends critically upon understanding (1) what happens when the beam reaches the specimen and (2) how the signals produced by the electron beam–specimen interactions are converted into images and/or spectra that convey useful information. To obtain useful information about the specimen, for example, the size, shape, composition, and certain other properties, an understanding of electron beam–specimen interactions is critical.

3.2. The Beam Enters the Specimen

Let us imagine that our task is to image and analyze a specimen of silicon. Assume that we first select a beam accelerating potential of 20 kV

from a thermal field emission source (brightness 1×10^8 A/cm^2 sr). We next choose the aperture(s), the condenser and objective lens strengths, and the stigmator strengths such that we bring to the specimen surface a focused probe with a circular cross section and a diameter of 1 nm, a divergence of 5×10^{-3} sr, and a beam current of 60 pA. The electrical and magnetic fields in the gun and lenses have determined the trajectories of the beam electrons over a distance of about 50 cm from the gun to the specimen. The electrons are traveling nearly parallel paths (varying only over a range of 5×10^{-3} sr, or $\sim$0.3 deg) and are of virtually identical energy, 20,000 $\pm$ 0.5 eV. The beam electrons are traveling through the SEM column in which the pressure has been reduced by a series of vacuum pumps to about 10^{-5} Pa ($\sim 10^{-7}$ torr). At this pressure there are so few gas molecules that only about 1 electron in 10,000 will collide with a gas atom along the 50-cm path from the electron gun to the specimen and be deflected out of the beam, whereas the rest are unaffected and strike the specimen surface in the 1-nm-diameter circle. What is the smallest amount of matter that we might possibly interrogate with such a beam? If all of the image-forming and analytical signals were immediately produced by the first layer of atoms under the 1-nm-diameter footprint of the focused beam as it strikes the surface, this would define the smallest amount of matter and therefore the best possible spatial resolution we might achieve. Unfortunately, the physics that governs the interaction of energetic electrons with atoms leads to a much different result, namely the large volume with μm dimensions

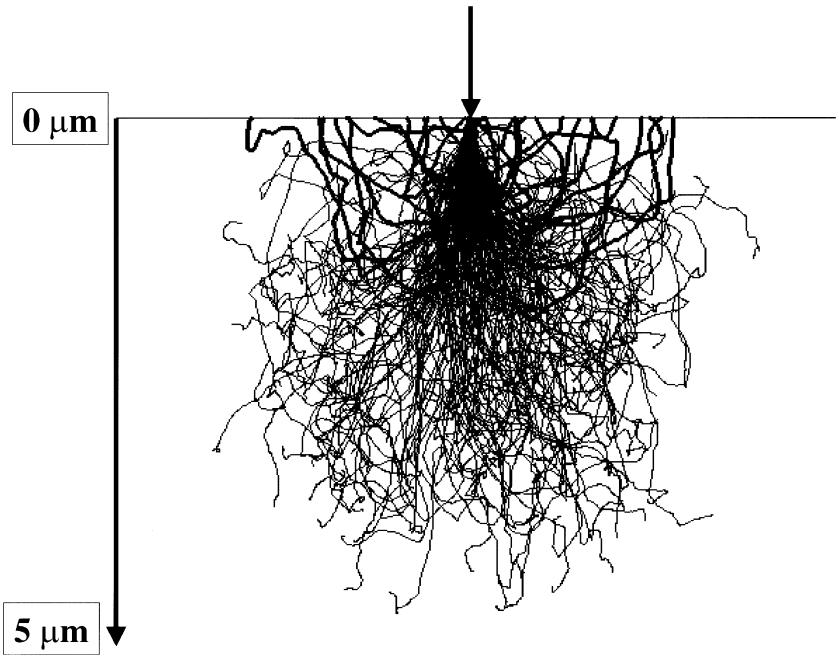


Figure 3.1. The interaction volume for a 20-keV beam striking silicon, as calculated with a Monte Carlo electron trajectory simulation (CASINO by Hovington *et al.,* 1997). The beam electron trajectories that emerge as backscattered electrons are shown as thick traces.

shown in Fig. 3.1 for a 1-nm-diameter, 20-keV beam incident on silicon. With such a small incident beam, what causes this extreme change?

As the beam electrons enter the specimen, they interact as negatively charged particles with the electrical fields of the specimen atoms. The positive charge of the protons is highly concentrated in the nucleus, whereas the negative charge of the atomic electrons is much more dispersed in a shell structure. The beam electron–specimen atom interaction can deflect the beam electrons along a new trajectory ("elastic scattering," with no kinetic energy loss), causing them to spread out laterally from the incident beam footprint. The elastic scattering can, after numerous events, actually result in beam electrons leaving the specimen (a process called "backscattering"), providing an important class of information for SEM imaging. The probability of elastic scattering increases strongly with atomic number Z, approximately as Z^2, because heavier atoms have a much stronger positive charge on the atomic nucleus, and decreases as the electron energy increases, approximately as $1/E^2$. A mathematical description of this process for an elastic scattering process at angles greater than a specified angle ϕ_0 has the form

$$Q(>\phi_0) = 1.62 \times 10^{-20}(Z^2/E^2)\cot^2(\phi_0/2) \quad (\text{events} > \phi_0)/[\text{electron (atom/cm}^2)] \tag{3.1}$$

where Q is called the cross section (cm^2) for elastic scattering (i.e., probability of elastic scattering). The distance between scattering events is known as the "mean free path," designated λ, and is calculated from the cross section and the density of atoms along the path:

$$\lambda = A/N_0\rho Q \quad (\text{cm}). \tag{3.2}$$

Simultaneously with elastic scattering, the beam electrons lose energy and transfer this energy in various ways to the specimen atoms ("inelastic scattering"), but this transfer takes place gradually, so that the beam electrons propagate through many atom layers into the specimen before losing all their energy. We shall find that inelastic scattering gives rise to useful imaging signals such as secondary electrons and analytical signals such as x-rays. Bethe (1930) described the rate of energy loss dE with distance traveled ds as

$$\frac{dE}{ds} \quad \left(\frac{\text{keV}}{\text{cm}}\right) = -2\pi e^4 N_0 \frac{Z\rho}{AE_i} \ln\left(\frac{1.166E_i}{J}\right) \tag{3.3a}$$

$$J \quad (\text{keV}) = (9.76Z + 58.5Z^{-0.19}) \times 10^{-3}, \tag{3.3b}$$

where e is the electron charge ($2\pi e^4 = 1.304 \times 10^{-19}$ for E in keV), N_0 is Avogadro's number, Z is the atomic number, ρ is the density (g/cm^3), A is the atomic weight (g/mole), E_i is the electron energy (keV) at any point in the specimen, and J is the average loss in energy per event (Berger and Seltzer, 1964). Note that the negative sign indicates the loss of energy.

Combining the constants gives

$$\frac{dE}{ds}\left(\frac{\text{keV}}{\text{cm}}\right) = -7.85 \times 10^4 \frac{Z\rho}{AE_i} \ln\left(\frac{1.166E_i}{J}\right). \tag{3.3c}$$

The behavior of energy loss as a function of beam energy for different materials is shown in Fig. 3.2a for intermediate and high beam energies, and in Fig. 3.2b for low beam energies, where a modification proposed by Joy and Luo (1989) overcomes limitations in the Bethe expression. Typical values of dE/ds at 20 keV for various pure-element targets are given in Table 3.1. For compounds or other atomic mixtures, mass-fraction-averaged materials parameters (Z, A, and ρ) are used.

How far can a beam electron travel in the specimen? By integrating the Bethe expression over the energy range from the incident value to a low threshold value (e.g., 1 keV), an estimate of the total distance an electron can travel in the specimen is obtained, the so-called "Bethe range." Note that the Bethe range is measured along the trajectory, regardless of direction changes caused by elastic scattering.

Although we can carry out useful microscopy and at least a first level of microanalysis without much knowledge of electron physics, there is a critical consequence of this scattering physics of which even a beginning student must be aware. The combined effect of these elastic and inelastic scattering processes is to distribute the beam electrons over a three-dimensional "interaction volume" with dimensions in the micrometer range, as shown in Fig. 3.1 for a 20-keV beam in silicon. We must understand the scale of the processes going on here. Our initial probe was only 1 nm in diameter (just a few atom diameters), and we might reasonably have expected to achieve image and analytical resolution on this scale. In fact, the interaction volume that results from this 1-nm-diameter, 20-keV beam on the entrance surface has linear dimensions over 1000 times greater and volume dimensions about one billion times greater than the atomic layer under the beam footprint! New and even veteran users of the SEM are often disappointed to discover that the resolution of fine detail is often much poorer than

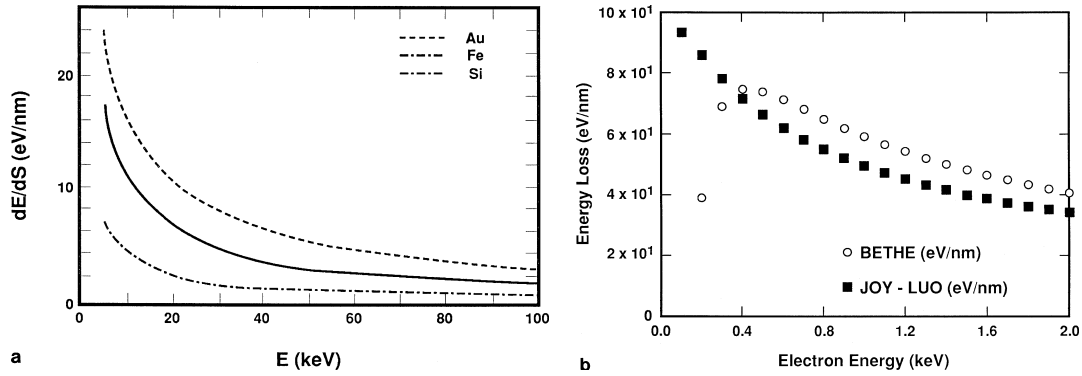


Figure 3.2. (a) Energy loss due to inelastic scattering calculated with the Bethe expression (3.3) at intermediate and high beam energies; (b) comparison of energy loss at low beam energy as calculated for Si with the Bethe expression and that proposed by Joy and Luo (1989).

Table 3.1. Energy Loss Due to Inelastic Scattering

Element	Z	A	ρ(g/cm^3)	J (keV)	dE/ds (keV/cm)	dE/ds (eV/nm)
Carbon	6	12.01	2.1	0.100	2.24×10^4	2.24
Iron	26	55.85	7.87	0.285	6.32×10^4	6.32
Silver	47	107.9	10.5	0.487	6.94×10^4	6.94
Uranium	92	238.03	18.95	0.923	9.27×10^4	9.27

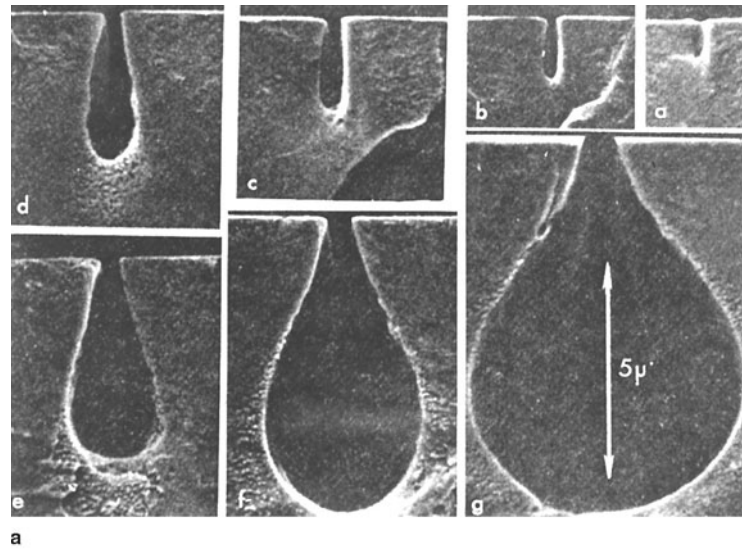
the specified resolution performance of the SEM. In this chapter we will learn that all of the imaging and analytical signals originate within this interaction volume and are therefore spread out over a substantial distance compared to the incident beam size. The interaction volume will thus have a strong influence on all of the tasks we wish to carry out with the SEM, limiting its performance both as an imaging and as an analytical tool. We need to understand all of the factors that control the size of the interaction volume and the consequent distribution of imaging and analytical signals to develop a strategy that will permit us to solve the problems at hand.

The following discussions are organized to first consider the situation for the "conventional" operational range of the SEM, which can be thought of as 10 keV and higher. This distinction is made to separately consider the "low-voltage" or "low-beam-energy" regime, which will be defined as the phenomena observed when the incident beam energy is at 5 keV or below, with the minimum as low as 100 eV or even less. In this range, some aspects of electron–specimen interactions are altered substantially. The range between 10 and 5 keV is a transition between conventional and low-beam-energy operation.

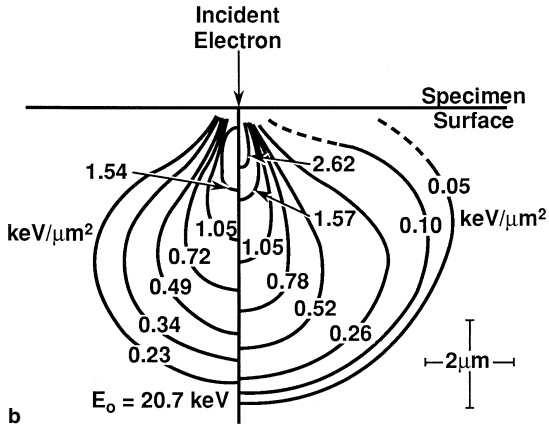
3.3. The Interaction Volume

3.3.1. Visualizing the Interaction Volume

The interaction volume, or at least its ghost, can be observed in certain plastic materials such as polymethylmethacrylate (PMMA) which undergo molecular bonding damage during electron bombardment that renders the material sensitive to etching in a suitable solvent. This phenomenon is the basis for important steps in the lithographic fabrication of semiconductor microcircuits. The same effect has been used to directly reveal the size and shape of the interaction volume (Everhart *et al.,* 1972). Figure 3.3a shows an experiment in which the contours of energy deposition by the beam within the interaction volume are revealed in a series of successively longer chemical etchings at constant electron dose. A static, 20-keV electron beam focused to less than 0.5 μm in diameter was placed for a fixed time at a series of locations on the PMMA. These constant-dose locations were then subjected to chemical etching for increasing time periods. Following etching, the PMMA was cleaved to reveal each etched volume in a vertical cross section and viewed in the SEM. The etching rate depends upon the damage, which is proportional to the deposited energy and hence to the electron dose per



a



b

Figure 3.3. (a) Etching experiment performed in polymethylmethacrylate (PMMA) for an incident beam energy of 20 keV (Everhart *et al.,* 1972) at fixed dose. (b) Interpretation of the etching experiment of (a) in terms of contours of energy deposited in the specimen (left) and as calculated with a Monte Carlo simulation (right).

unit volume, $\#e^{-}/\text{cm}^3$. The most highly damaged material etches out first in the solvent ("a"). This volume is represented by the small cylindrical shape immediately below the beam entrance surface. Etching for increasing time periods reveals contours of progressively decreasing electron energy deposition, eventually with long etching time reaching the largest structure ("g") in Fig. 3.3a. The contours of successively decreasing energy deposition are plotted numerically in Fig. 3.3b. The etched structures allow us to directly visualize electron penetration and the interaction volume in a low-atomic-number matrix similar in average atomic number to biological materials and polymers. Several points stand out: (1) Despite the fact that the incident beam diameter was well under a 1 μm in diameter, the *interaction volume at 20 keV in a low-density, low-atomic-number target has overall dimensions of several micrometers*. (2) The energy deposition rate varies rapidly throughout the interaction volume, being greatest near the beam impact

point. (3) The interaction volume for a low-density, low-atomic-number target has a distinct pear shape for the lowest energy deposition contours. The origin of this pear shape can be understood in terms of the characteristics of elastic and inelastic scattering. For this low-atomic-number matrix, elastic scattering ($\sim Z^2$) is relatively weak, so that the beam initially tends to penetrate into the target, forming the narrow neck of the pear shape. Inelastic scattering is relatively more probable in this low-atomic-number material, so that the electron energy decreases. The probability for elastic scattering increases rapidly with decreasing electron energy ($\sim 1/E^2$). With further penetration and reduction in electron energy, the cumulative effect of elastic scattering tends to cause the electron trajectories to deviate, creating the bulbous portion of the pear shape. Inelastic scattering limits the eventual range of the electrons in the target.

3.3.2. Simulating the Interaction Volume

The PMMA etching experiment gives results that are representative of low-atomic-number materials, including the important classes of biological materials and polymeric materials that are predominantly composed of carbon. There are no materials of intermediate (e.g., iron) and high (e.g., gold) atomic number for which etching can reveal the interaction volume. To explore these cases, we make use of the technique of Monte Carlo electron trajectory simulation (Berger, 1963; Shimizu and Murata, 1971; Heinrich *et al.,* 1976, Newbury *et al.,* 1986; Heinrich and Newbury, 1991). The Monte Carlo simulation technique has developed into a highly useful tool for SEM/microanalysis ranging from applications in interpreting SEM images to the x-ray microanalysis of complex structures. Because of this importance, we will briefly describe its construction. An instructional Monte Carlo simulation will be found on the CD-ROM that accompanies the text. The reader is invited to make use of the Monte Carlo simulation to provide a dynamic sense of the interactions that is lacking from the static figures of the book.

To construct a Monte Carlo electron trajectory simulation, the effects of elastic and inelastic scattering are calculated from appropriate models to determine scattering angles, distances between scattering sites ("step length"), and the rate of energy loss with distance traveled. From these parameters and equations of analytical geometry that relate the scattering angles and step length to successive electron locations, the electron trajectory can be simulated in a stepwise fashion from the location at which it enters the specimen to its final fate. Because the various physical parameters occur over a range of values (an individual elastic scattering event, for example, can vary the trajectory from 0 to 180 deg), random numbers are used with a weighting factor to produce the appropriate statistical distribution of the physical events. This use of random numbers gives rise to the name "Monte Carlo" method. As the electron trajectory is followed in a stepwise fashion, the location in the specimen (coordinates x, y, z), the electron velocity vector (v_x, v_y, v_z), and the kinetic energy ($mv^2/2$) are constantly updated. The trajectory is followed until either it loses all of its energy (to a selected threshold,

e.g., 0.5 keV) due to inelastic scattering and is captured by the specimen, or else the electron passes back through the entrance surface (for a flat, semi-infinite target) or another surface (for thin foils, particles, etc.) and escapes. These escaping beam electrons form the class known as backscattered electrons, and their spatial, angular, and energy distributions can be recorded. The production of secondary radiation, such as the generation of characteristic x-rays or secondary electrons, can be calculated with the appropriate cross sections and the step length. Extensive testing of Monte Carlo procedures with comparison to experimental values has established the accuracy of the simulations and the limits of applicability (Newbury and Myklebust, 1984). A more complete description of the Monte Carlo simulation technique can be found in several references (Newbury and Myklebust, 1979; Newbury *et al.,* 1986; Henoc and Maurice, 1991; Hovington *et al.,* 1997). In this chapter, the Monte Carlo simulation will be used as a tool to study the interaction volume, backscattered electrons, and secondary radiation products as a function of specimen and beam parameters.

3.3.3. Influence of Beam and Specimen Parameters on the Interaction Volume

3.3.3.1. Influence of Beam Energy on the Interaction Volume

The size of the interaction volume is a strong function of the energy with which the beam electrons interact with the target. The interaction volume in iron is shown as a function of beam energy for the range 10–30 keV in Fig. 3.4. The increase in size with beam energy can be understood from an examination of Eqs. (3.1) and (3.3). First, the cross section for elastic scattering has an inverse dependence on the square of the energy, $Q \sim 1/E^2$ [Eq. (3.1)]. Thus, as the beam energy increases, the electron trajectories

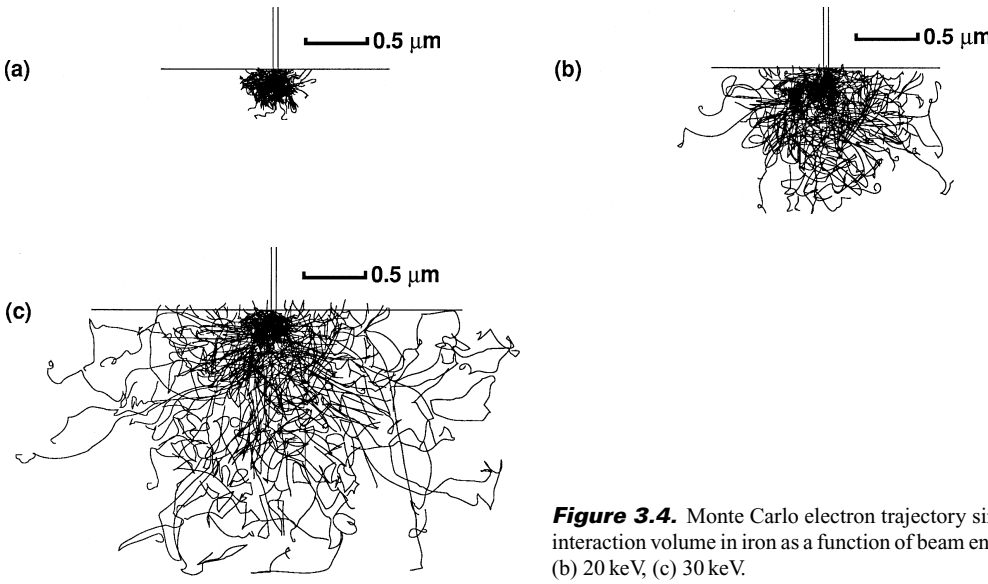


Figure 3.4. Monte Carlo electron trajectory simulations of the interaction volume in iron as a function of beam energy: (a) 10 keV, (b) 20 keV, (c) 30 keV.

near the surface become straighter and the electrons penetrate more deeply into the solid before the cumulative effects of multiple elastic scattering events cause some of the electrons to propagate back toward the surface. Second, the rate of energy loss with distance traveled, as given by the Bethe expression, is inversely related to the energy, $dE/ds \sim 1/E$ [Eq. (3.3)]. As the beam energy is increased, the electrons can penetrate to greater depths because they enter the specimen with more energy and lose it at a lower rate. The lateral dimensions of the interaction volume are seen to scale with energy in a similar fashion to the depth. The shape of the interaction volume does not change significantly with beam energy.

3.3.3.2. Influence of Atomic Number on the Interaction Volume

Monte Carlo calculations, shown in Fig. 3.5 for targets of carbon ($Z = 6$), iron ($Z = 26$), silver ($Z = 47$), and uranium ($Z = 92$) at a beam energy of 20 keV, reveal that the linear dimensions of the interaction volume

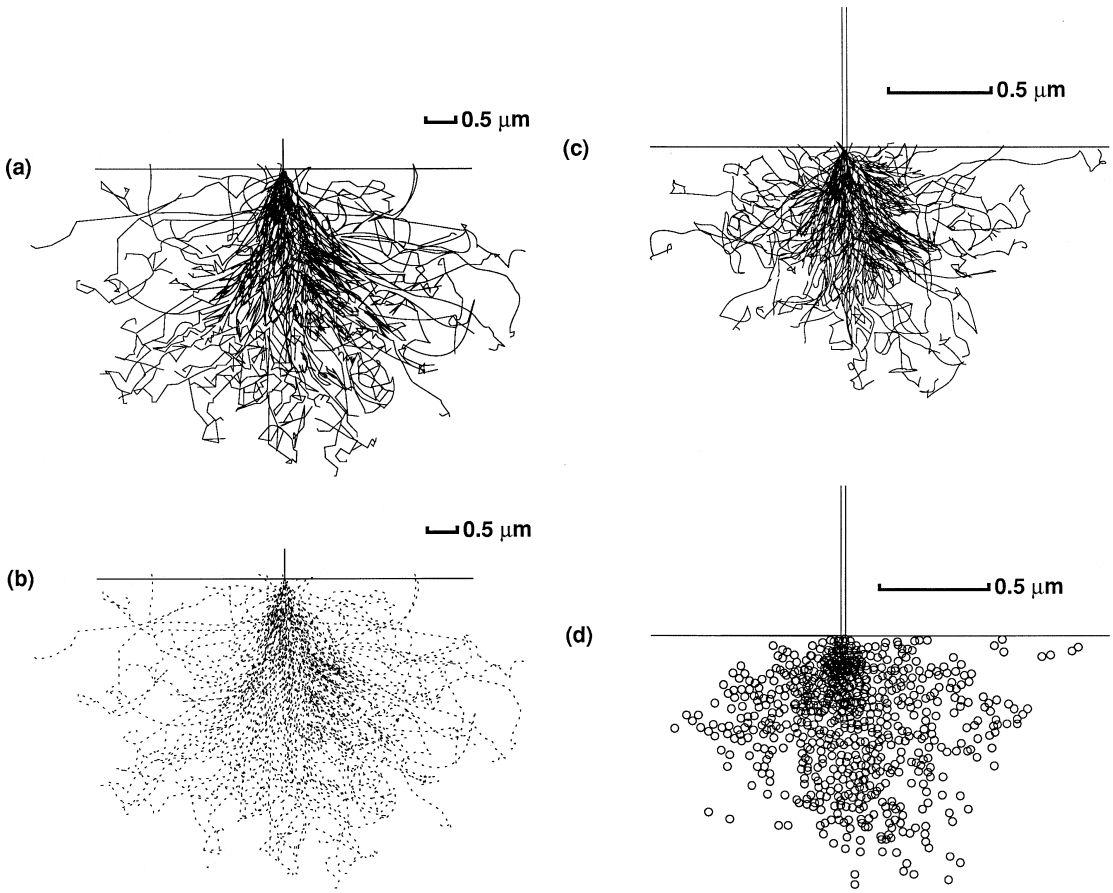


Figure 3.5. Monte Carlo electron trajectory simulations of the interaction volume and corresponding distribution of sites of inner shell ionization leading to x-ray generation at 20 keV and 0° tilt in various targets. (a) Carbon, (b) carbon *K* shell, (c) iron, (d) iron *K* shell.

decrease with increasing atomic number at a fixed beam energy. This is a direct consequence of the increase in the cross section for elastic scattering with atomic number, $Q \sim Z^2$ [Eq. (3.1)]. In targets of high atomic number, the electrons undergo more elastic scattering per unit distance and the average scattering angle is greater than for low-atomic-number targets. The electron trajectories in high-atomic-number materials thus tend to deviate out of the initial direction of travel more quickly, increasing backscattering and reducing penetration into the solid. From Table 3.1 the energy loss rate from the Bethe expression generally increases with atomic number, limiting the total range and tending to reduce the size of the interaction volume. In low-atomic-number materials, elastic scattering is less likely and the trajectories deviate less from the initial beam path, allowing for deeper penetration into the solid. Because the rate of energy loss is lower in low-Z materials, this also contributes to a larger interaction volume. The shape of the interaction volume also changes significantly as a function of atomic number. The dense region of trajectories changes from the pear shape seen in low-atomic-number materials to a more nearly hemispherical shape truncated by the plane of the surface for high-atomic-number materials.

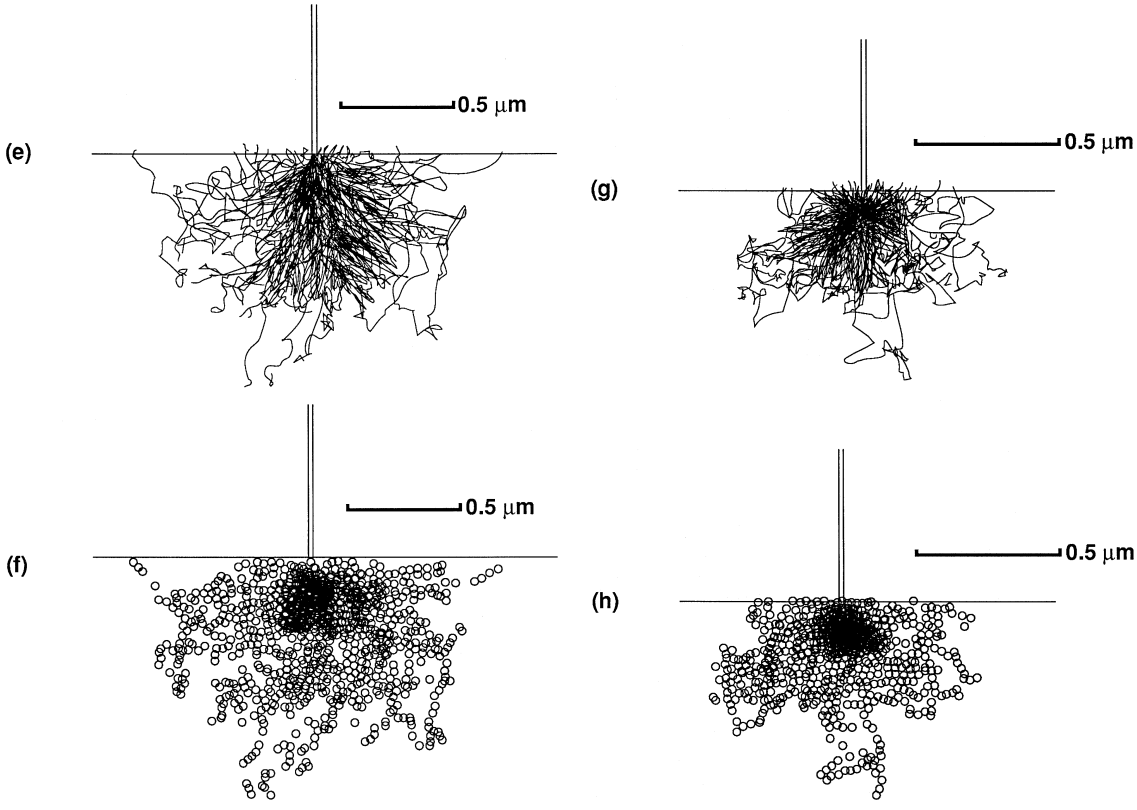


Figure 3.5. (*Continued*) (e) silver, (f) silver *L* shell, (g) uranium, (h) uranium *M* shell. Note dimension changes from C to Fe to Ag to U.

As the angle of tilt of a specimen surface increases, that is, the angle of inclination of the beam relative to the surface decreases, the interaction volume becomes smaller and asymmetric, as shown in the Monte Carlo plots of Fig. 3.6. This behavior can be understood with the aid of Fig. 3.6, which depicts the scattering cone for an elastic event. Assume that the semicone angle is equal to the most probable value of the elastic scattering angle, which is typically in the range 3°–5°. Most elastic scattering angles are so small that the electron tends to continue in the same general direction after scattering (that is, the tendency is for "forward scattering"). The electron will travel with equal probability to some point on the circumference of the base of the cone. At 0° tilt, for which the beam is perpendicular to the surface, the tendency for forward scattering causes most of the electrons to propagate down into the specimen, as shown in Fig. 3.6d. For a tilted specimen, despite the tendency for forward scattering, at least some of the beam electrons propagate nearer to the surface; indeed, even with a small scattering angle, some electrons can escape through the specimen surface after a single small-angle scattering event for a highly tilted surface. The electrons do not penetrate as deeply into the specimen at high

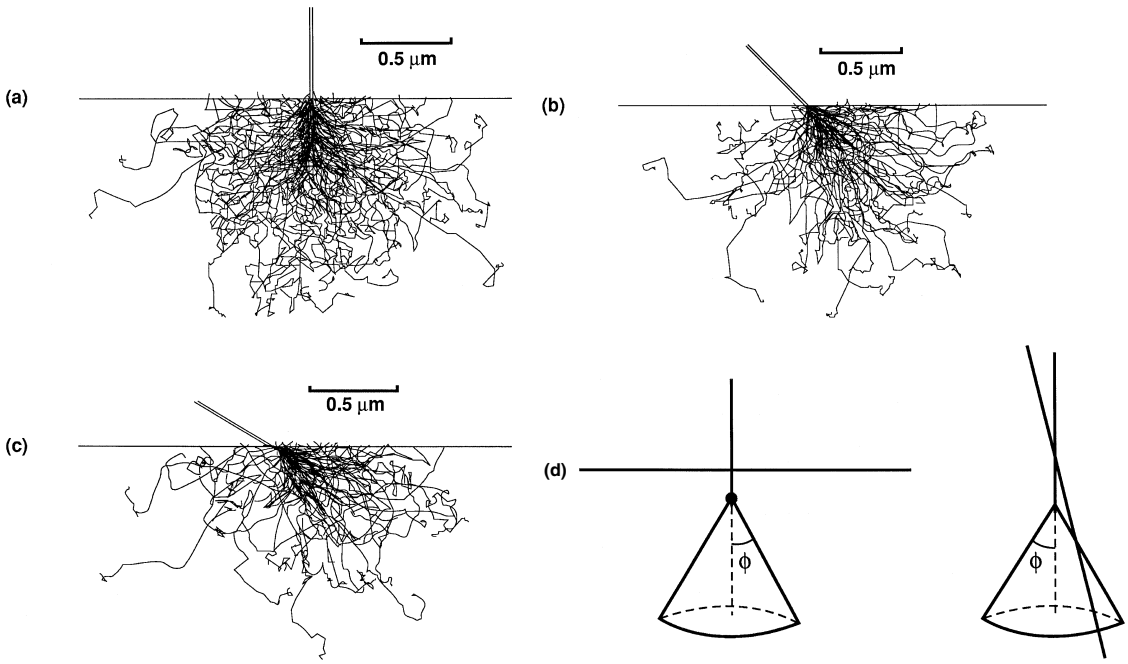


Figure 3.6. Monte Carlo electron trajectory simulations of the interaction volume in iron at $E_0 = 20\,\text{keV}$ for various tilts: (a) 0° tilt, (b) 45° tilt, (c) 60° tilt. (d) Schematic illustration of the origin of increased backscattering from tilted specimens. Consider a given average elastic scattering angle ϕ which produces the scattering cone indicated. The electron may land at any point on the base of the cone with equal probability. At normal incidence, no matter where on the base of the cone the electron lands, it tends to continue propagating into the solid. When the specimen is tilted, some of the possible locations on the base of the scattering cone actually carry the electron out of the solid immediately.

tilts, and the interaction volume thus has a reduced depth dimension. The lateral dimensions of the interaction volume from a tilted specimen present a different situation. Because of the asymmetry of the scattering situation for a tilted specimen, the interaction volume dimension perpendicular to the axis of tilt increases in the "downwind" direction from the beam impact point as compared to normal incidence, whereas the dimension parallel to the axis of tilt remains nearly the same as that for normal beam incidence.

3.3.4. Electron Range: A Simple Measure of the Interaction Volume

3.3.4.1. Introduction

From the plots presented in Figs. 3.4–3.6, it is clear that the interaction volume is a complex three-dimensional region which depends on beam energy, material parameters, and specimen tilt. Despite this complexity, there are situations when we need to describe the interaction volume with a single parameter, the so-called "electron range." Using the Monte Carlo depiction of the interaction volume shown in Fig. 3.1, consider a hemisphere constructed with a radius whose origin is the entry point of the beam into the specimen and which contains a specified fraction of the electron trajectories, for example, 90%. We could specify 100% of the trajectories, but it is clear from repeated calculations of the Monte Carlo simulation that, due to the statistical nature of the scattering processes, to capture the last 10% of the trajectories we must significantly enlarge the boundary of the interaction volume, including space where not much is happening. In experimental determinations such as the polymer etching, a similar result is found. The boundary is not really abrupt, but occurs as a gradual lowering of trajectory density. Moreover, for most purposes our measurements are relatively insensitive to the behavior of these electrons, which have lost nearly all their energy and thus constitute the boundary region. The electron range is thus properly thought of as a "gray number," that is, a simple value that is useful for rough comparisons and scaling various signal distributions. Within the electron range, the density of scattering events changes sharply with distance from the beam impact area.

Numerous descriptions of the electron range are available in the literature. With the cautions noted above, the range equation which we will use in this text is that due to Kanaya and Okayama (1972), who parameterized the electron range as

$$R_{\mathrm{KO}} \quad (\mu\mathrm{m}) = \frac{0.0276A}{Z^{0.89}\rho} E_0^{1.67}, \tag{3.4}$$

where A is the atomic weight (g/mole), Z is the atomic number, ρ is the density (g/cm^3), E_0 is the beam energy (keV), and R_{KO} is calculated in micrometers with the constant 0.0276 (to calculate R_{KO} in nanometers, the

constant becomes 27.6). The specimen is assumed to be flat, thick enough to be electron-opaque, and of sufficiently large lateral extent that there are no edges or boundaries within R_{KO} of the beam, and the incident electron beam is placed normal to the specimen surface (tilt angle $= 0^\circ$). Values calculated with Eq. (3.4) should be considered "gray numbers" and not stated beyond two significant figures because of various uncertainties.

For tilted specimens, the interaction volume becomes asymmetric, with the degree of asymmetry increasing with the tilt. As the tilt increases, the depth of penetration and therefore the z extent of the interaction volume are reduced due to the tendency for the beam electrons to travel in the forward direction, progressively closer to the entrance surface, which allows more of them to escape. The reduction in the electron range (again defined as the z boundary from the surface) as a function of tilt θ can be roughly described as

$$R(\theta) = R(0)\cos\theta, \tag{3.5}$$

where $R(0)$ is the Kanaya–Okayama range at normal beam incidence (0° tilt) from Eq. (3.4).

There are other consequences of tilting for the interaction volume. As shown in Fig. 3.5, as the tilt increases, the interaction volume becomes longer in the direction "downhill" from where the beam strikes (i.e., along the direction in the surface plane perpendicular to the tilt axis), while shortening in the "uphill" direction. Parallel to the tilt axis, the interaction volume retains about the same width as that found at normal incidence. For very high tilts above 60°, so many beam electrons escape that the interaction volume dimension parallel to the tilt axis also decreases.

3.3.4.2. The Electron Range at Low Beam Energy

Studies by Joy and Luo (1989) suggest that the Bethe expression is satisfactory over a range extending from approximately $7J$ to a maximum of approximately 50 keV. The criterion $7J$ indicates that for operation under low-beam-energy conditions with $E_0 < 5$ keV, calculations based upon the Bethe expression will result in discrepancies for targets of intermediate and high atomic number at the following energy thresholds: C (0.7 keV), Al (1.1 keV), Cu (2.2 keV), and Au (5.6 keV). Joy and Luo (1989) noted that at low beam energies the number of processes contributing to the energy loss decreases; they proposed a modified form of the Bethe energy loss relation which is appropriate at low beam energies, below $7J$, that incorporates a modified term, J^*:

$$\frac{dE}{ds} = -7.85 \times 10^4 \frac{Z\rho}{AE_i} \log_e\left(\frac{1.166E_i}{J^*}\right) \quad \left(\frac{\text{keV}}{\text{cm}}\right). \tag{3.6a}$$

The modified mean ionization potential is given by

$$J^* = \frac{J}{1 + (kJ/E)}, \tag{3.6b}$$

where J is the value given by the conventional expression, Eq. (3.3b). In this expression, k is a variable dependent on atomic number which is always close to but less than unity. The data for the variable k presented by Joy and Luo (1989) fit an equation of the form

$$k = 0.731 + 0.0688 \log_{10} Z. \tag{3.6c}$$

The Joy–Luo expressions for dE/ds can be used to calculate the distance an electron can travel in the specimen, analogous to the Bethe range, but appropriate to the low-beam-energy regime. Note that the distance s in Eqs. (3.3c) and (3.6a) is the distance *along* the trajectory. A comparison of the Bethe and Joy–Luo ranges is presented in Fig. 3.7, where it can be seen that at progressively lower beam energies, the Joy–Luo range continues to decrease while the Bethe range reaches a limit. Note that because of elastic scattering, the real trajectories deviate significantly from a straight line, especially in targets of high atomic number, so that the range describing the limit of the electron trajectories will be less than the range calculated purely on the basis of energy loss.

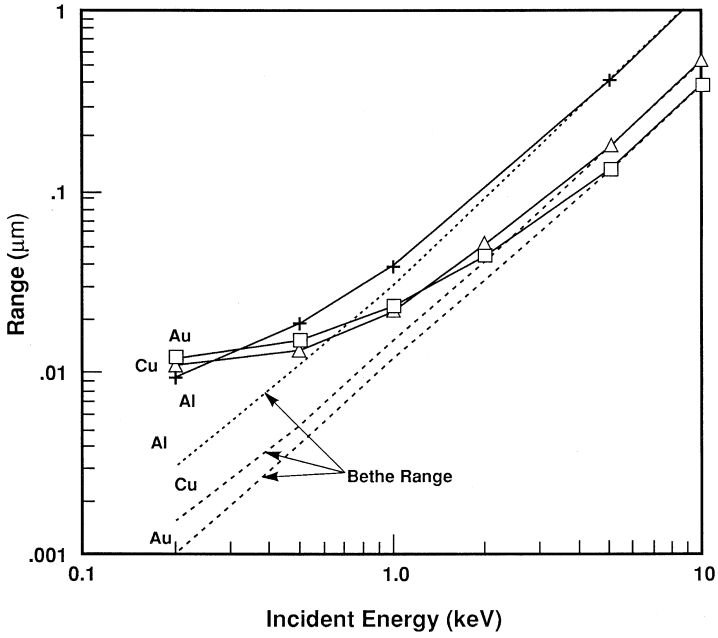


Figure 3.7. Electron range at low incident beam energy calculated by the Bethe expression as extended by Rao-Sahib and Wittry (1972) (dashed lines) and as modified by Joy and Luo (1989) (solid lines) for Al, Cu, and Au.

Backscattered and secondary electrons, the principal signals used to form images in scanning electron microscopy, are generated within the interaction volume. These signals are capable of carrying information about specimen composition, shape (topography), local fine-scale surface texture, thickness, and local inclination to the incident beam. The information presented in the next two sections forms the critical database for performing SEM image interpretation.

3.4.1. Backscattered Electrons

Backscattered electrons (BSE) are beam electrons whose trajectories have intercepted a surface, usually, but not necessarily, the entrance surface, and which thus escape the specimen. Close examination of the Monte Carlo simulations in Figs. 3.3–3.5 reveals examples of beam electron trajectories that lead to backscattering. Generally, these electrons have undergone numerous elastic scattering events to accumulate enough deviation from the incident beam path to return to the surface. If a large number of trajectories is examined, a few examples will be found in which a large-angle elastic scattering ($>90^\circ$) event will lead to a beam electron escaping as a BSE after a few events or even one event. Backscattered electrons remove a significant amount of the total energy of the primary beam, which in the absence of the backscattering effect would contribute to the production of additional secondary radiation products such as the characteristic x-rays measured in quantitative x-ray microanalysis.

Backscattering is quantified by the backscatter coefficient η, which is defined as

$$\eta = \frac{n_{\mathrm{BSE}}}{n_{\mathrm{B}}} = \frac{i_{\mathrm{BSE}}}{i_{\mathrm{B}}}, \tag{3.7}$$

where n_{B} is the number of beam electrons incident on the specimen and n_{BSE} is the number of backscattered electrons (BSE). The backscatter coefficient can also be expressed in terms of currents, where i_{B} refers to the beam current injected into the specimen and i_{BSE} to the backscattered electron current passing out of the specimen. The significant properties of BSE are surveyed in the following sections; for a more detailed survey, particularly for thin films, see Niedrig (1978).

3.4.1.1. Atomic Number Dependence of BSE

The Monte Carlo trajectory plots for different elements shown in Fig. 3.4 suggest that backscattering increases with increasing atomic number. Careful experiments have been performed in which the beam current is first accurately measured by trapping it in a Faraday cup connected to a picoammeter. The Faraday cup is made by drilling a blind hole about 3 mm in diameter into a block of conducting metal, such as brass, and fitting over

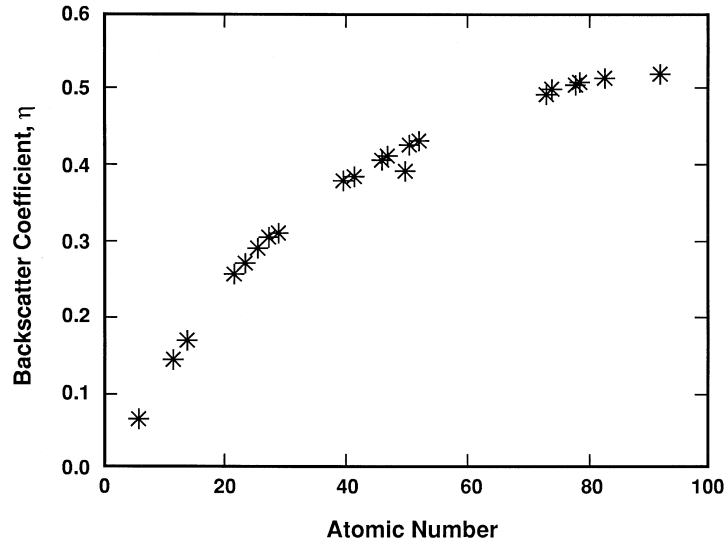


Figure 3.8. Variation of the backscattered electron coefficient η as a function of atomic number at $E_0 = 20$ keV. (Data of Heinrich, 1966.

this hole an aperture with a much smaller entrance hole, typically 100 μm in diameter or less. Once the electrons of the focused beam have passed through the aperture and scatter off the floor and walls of the hole, it is very unlikely they will pass back out of the aperture. Secondary electrons (see below) created at surfaces within the hole are also prevented from escaping and modifying the measured beam current. The entire beam current is thus absorbed by the block, and by electrically isolating it and attaching a lead to the picoammeter, the true beam current can be measured. When the beam is allowed to strike polished pure-element metal blocks similarly attached to the picoammeter, the current flowing through the meter to ground is found to be less than the beam current because a fraction of the beam has been backscattered. (Note that for maximum accuracy, it is necessary to isolate and bias the target to prevent the escape of secondary electrons.)

Figure 3.8 shows a plot of backscattering versus atomic number, η versus Z. Some important points about this plot should be noted:

1. The plot shows a general, monotonic increase in the backscattering coefficient with atomic number. Whenever a sensible relationship is found between a specimen property, such as composition, and a measurable signal in the SEM, such as BSE, the basis for a *contrast mechanism* exists. In this case, the monotonic increase of η versus Z forms the basis for *atomic number contrast* (also called compositional contrast or Z contrast).
2. The slope of η versus Z is initially steep, but decreases with increasing Z, becoming very shallow above $Z = 50$. The practical effect of this behavior is that atomic number contrast between adjacent pairs of elements is strong at low atomic number and weak at high atomic number.

The curve of η versus Z can be conveniently fit with an expression obtained by Reuter (1972):

$$\eta = -0.0254 + 0.016Z - 1.86 \times 10^{-4} Z^2 + 8.3 \times 10^{-7} Z^3. \quad (3.8)$$

This expression is useful for estimating values of η when contrast calculations must be performed (see Chapter 4).

When a target is a mixture of elements that is homogeneous on an atomic scale, for example, a solid solution, then the backscattering coefficient follows a simple rule of mixtures based on the weight (mass) concentrations C_i of the individual constituents (Heinrich, 1966):

$$\eta = \sum_i C_i \eta_i, \quad (3.9)$$

where i denotes each constituent, η_i is the pure element backscatter coefficient for element i, and the summation is taken over all constituents.

3.4.1.2. Beam Energy Dependence of BSE

The size of the interaction volume was seen in the Monte Carlo plots of Fig. 3.3 to be a strong function of the beam energy. We might reasonably expect that the backscatter coefficient would also depend strongly on beam energy. However, experimental measurements reveal that this is not the case. As shown in Fig. 3.9, there is only a small change, generally less than 10%, in the backscatter coefficient as a function of beam energy for the range 5–50 keV, which spans the conventional SEM/EPMA range. When data for specific elements are examined, the slight energy dependence of the backscatter coefficient is found to behave in a complex manner, increasing, decreasing, or remaining nearly constant, depending on the particular element (Heinrich, 1966). This unexpected insensitivity to beam energy

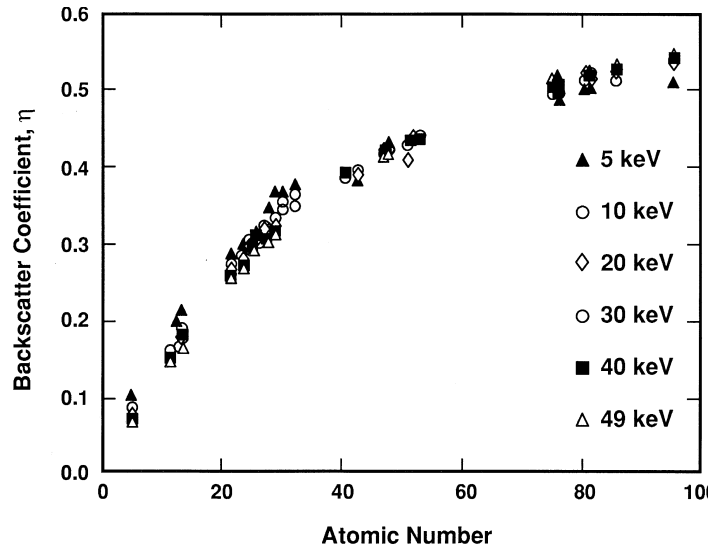


Figure 3.9. Backscattered electron coefficient as a function of atomic number plotted for a range of beam energies from 5 to 49 keV. (Data of Bishop, 1966, and Heinrich,1966).

can be understood from the following qualitative argument. Although the range increases as approximately the 1.67 power of the beam energy, the Bethe energy loss expression shows that the rate of energy loss decreases with increasing energy. Compare an average electron at the limit of the envelope of the interaction volume for an incident energy of 10 keV with an average electron at the same absolute depth in the interaction volume that started with 20-keV incident energy. In the 10-keV case, the electron loses virtually all of its energy and becomes captured by the specimen. For the 20-keV case, an electron at the same depth still has at least 10 keV of its energy remaining because it started with twice the energy and lost it at a lower rate. The electron thus has the possibility of continuing to travel and scatter, so that a significant fraction of these electrons can eventually reach the surface and escape as backscattered electrons. The backscatter coefficient is thus relatively insensitive to the beam energy.

At beam energies below 5 keV, the behavior of the backscatter coefficient is more complicated. As the beam energy is reduced toward 1 keV, the backscatter coefficients of light elements apparently increase, whereas those for heavy elements decrease. Hunger and Kuchler (1979) described an expression for backscattering, originally intended for the beam energy range 4–40 keV, which appears to work well when extended to energies as low as 1 keV (Joy, 1991):

$$\eta(Z, E) = E^m C, \tag{3.10a}$$

where

$$m = 0.1382 - (0.9211/\sqrt{Z}) \tag{3.10b}$$

and

$$C = 0.1904 - 0.2235(\log_e Z) + 0.1292(\log_e Z)^2 - 0.01491(\log_e Z)^3. \tag{3.10c}$$

There is a considerable practical difficulty in making low-beam-energy backscatter measurements under conditions appropriate to conventional SEM. The accumulation of contamination during measurement in the vacuum environment of conventional SEMs can alter the apparent backscatter coefficient. Contamination may arise from the specimen surface and/or from the residual gases of the pumping system. From Fig. 3.7, the range for 1-keV electrons is only from 0.01 μm (gold) to 0.04 μm (aluminum), which may be compared to a range of 0.4–1.3 μm for 10-keV electrons. If a layer of only a few nanometers of carbonaceous material builds up during a low-beam-energy measurement, this layer will have a large influence on the apparent backscatter coefficient. Such contamination effects may actually dominate low-beam-energy images in instruments with conventional vacuum systems when no special surface cleaning is used. The presence of carbon will reduce the apparent backscattering coefficient. If such a contamination layer extends over regions of different atomic number that would be expected to produce atomic number contrast in an SEM image, the carbon may reduce or even eliminate the contrast. Thus, the

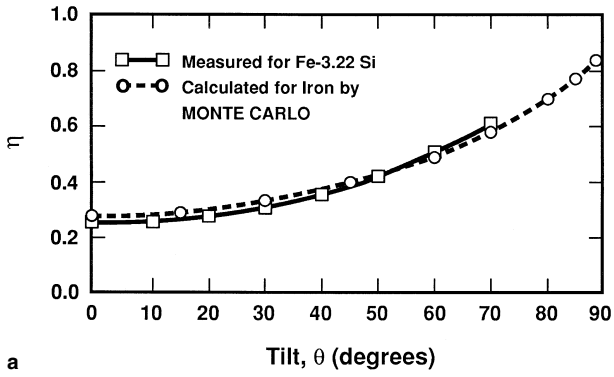


Figure 3.10. (a) Backscattered electron coefficient as a function of tilt as measured for iron and as calculated by Monte Carlo electron trajectory simulation (Newbury *et al.,* 1973).

microscopist's observations at low beam energy may be greatly influenced by unexpected surface contamination, especially in conventional vacuum instruments.

3.4.1.3. Tilt Dependence of BSE

If the backscatter coefficient is measured as a function of the tilt angle θ, which is defined as the complement of the angle between the beam and the surface plane, then a smooth, monotonic increase in backscattering with tilt is found, as demonstrated in Fig. 3.10a. The slope of η versus θ is

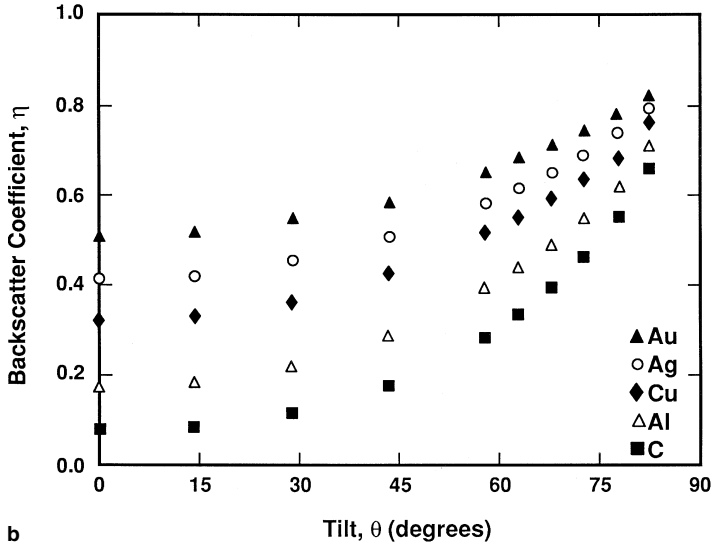


Figure 3.10. (*Continued*) (b) Backscattered electron coefficient as a function of tilt as calculated for several elements by Monte Carlo electron trajectory simulation.

initially shallow, but increases with increasing tilt. At very high tilt angles, which correspond to grazing incidence, the value of η tends toward unity. If η versus θ is plotted for a range of elements, then at high values of θ the backscatter coefficients for all elements tend to converge, as shown in Fig. 3.10b. An expression suggested by Arnal *et al.* (1969) gives the backscatter coefficient as a general function of Z and θ:

$$\eta(\theta) = 1/(1 + \cos\theta)^p, \tag{3.11}$$

where $p = 9/\sqrt{Z}$.

This behavior of η versus θ arises because of the dominant tendency of elastic scattering to be in the forward direction. That is, most elastic scattering events result in relatively small deviation angles, of the order of 5°, so the electron trajectories tend to continue in roughly the same direction after scattering as they were initially traveling. When the beam is set normal to the specimen surface, that is, $\theta = 0°$, this tendency for forward scattering means that beam electrons tend to penetrate into the target. Only by the cumulative effects of many small scattering events, and the much rarer large-angle events, do some of the electron trajectories reverse direction and travel back toward the surface to escape as backscattered electrons. However, if the tilt angle of the specimen surface is increased sufficiently, the geometry of the situation is such that, despite the tendency for forward scattering, electrons tend to travel along trajectories near the surface, as seen in Fig. 3.6. Thus, electrons can exit the tilted surface with less total angular deviation, so the backscattering coefficient increases.

The monotonic rise of η with θ forms the basis for an important component of the mechanism of topographic contrast in the SEM, by which the shape of objects is recognized. In quantitative electron probe microanalysis corrections, the enhanced loss of backscattered electrons from tilted specimens contributes to a decrease in the production of characteristic x-rays as compared to a specimen at 0° tilt.

3.4.1.4. Angular Distribution of BSE

A second important consideration is the *directionality* of the backscattered electrons. The dependence of η upon θ gives the total *number* of backscattered electrons that emerge at a particular tilt angle of the surface without regard to the trajectories which the backscattered electrons follow out of the specimen. In considering the performance of a backscattered electron detector and in order to interpret the images obtained from the BSE signal, it is necessary to understand relationship of the detector position to the BSE trajectories emitted from the specimen.

Normal Beam Incidence (0° Tilt). The angular distribution of backscattered electrons is defined relative to the normal to the surface through which the BSE emerge. Consider a specimen at 0° tilt, for which the beam is parallel to the surface normal. As shown in Fig. 3.11, an angle ϕ is defined by the vector of the surface normal **n** and a second vector **m**. If the number of backscattered electrons is measured by placing a detector with a very

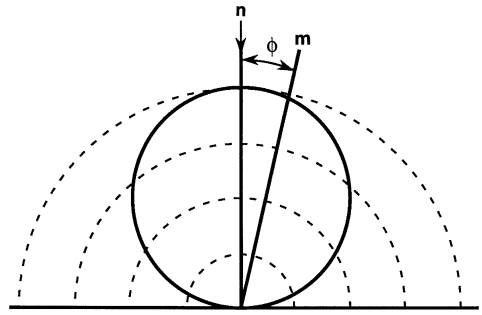


Figure 3.11. Angular distribution of backscattered electrons relative to the surface normal for a specimen surface at 0° tilt (beam perpendicular to surface).

narrow angle of view along a specific direction **m**, then the backscatter coefficient at the angle ϕ, designated $\eta(\phi)$, follows a distribution which approximates a cosine expression:

$$\eta(\phi) = \eta_n \cos\phi, \tag{3.12}$$

where $\eta_{\mathbf{n}}$ is the value measured along the normal vector **n**. In Fig. 3.11, the relative number of backscattered electrons at any angle ϕ is given by the length of the line drawn from the beam impact point to intersect the solid curve. For 0° tilt, this cosine distribution is rotationally symmetric around the surface normal, so that the same angular distribution is found in any plane that contains the surface normal.

Inspection of this cosine distribution shows that the maximum number of backscattered electrons is emitted along the surface normal, $\phi = 0°$, which means they travel back along the incident beam! As the detector is placed at larger values of ϕ away from the surface normal, the number of backscattered electrons decreases. At $\phi = 45°$, the backscattered intensity is approximately 0.707 of the intensity along $\phi = 0°$, whereas at $\phi = 60°$, the intensity has fallen to 0.50. At shallow angles just above the surface, there are virtually no backscattered electron trajectories. Thus, the angular location of a BSE detector relative to a specimen surface will have a strong influence on its collection efficiency.

Angular Distribution from Tilted Specimen Surfaces (Non-Normal Beam Incidence). When the specimen surface is tilted, the angular distribution changes from the symmetric cosine law found at normal beam incidence. As the surface is tilted, the tendency for forward elastic scattering favors backscattering from the surface in directions away from the incident beam. The angular distribution thus becomes asymmetric, with a distinct lobe in the forward scattering direction. This asymmetry is most pronounced at high tilt angles, greater than 45° tilt, as shown in Fig. 3.12a, which compares the angular distributions for 0° tilt and 80° tilt. For high tilt angles, the angular distribution resembles a highly elongated ellipse, with the long axis of the ellipse at approximately the same angle above the surface as the incident beam. Virtually no electrons are backscattered along the incident beam trajectory. At more shallow angles, less than 45° tilt, the distribution is similar to a cosine distribution with distortion in the forward direction away from the beam.

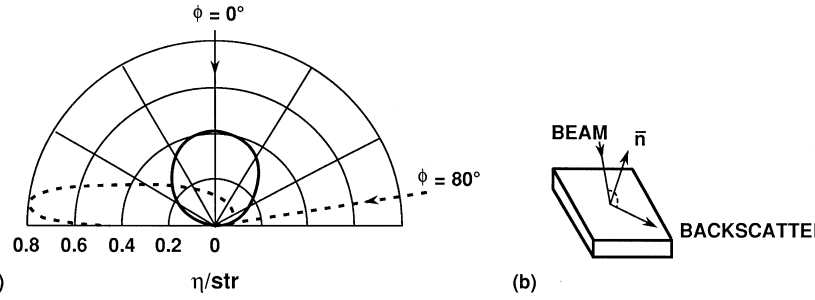


Figure 3.12. (a) Comparison of the angular distribution of backscattered electrons for sample tilts of $\theta = 0°$ and $\theta = 80°$. Note that the angular distribution is strongly peaked in the forward direction at high tilt. (b) Schematic illustration of the direction of maximum backscattering in a highly tilted specimen; the backscatter maximum tends to be in a plane defined by the incident beam vector and the surface normal.

A second consequence of tilting the specimen surface is the asymmetry of the interaction volume. This asymmetry also affects backscattering. For tilt angles greater than 45°, the angular distribution becomes progressively narrower parallel relative to the tilt axis. At very high tilt angles, for example, the 80° tilt case considered in Fig. 3.12a, the distribution is sharply peaked in the forward direction, and relatively little scattering occurs parallel to the tilt axis. That is, the distribution plotted in Fig. 3.12a is much thinner in the direction out of the plotting plane. For highly tilted surfaces, many of the beam electrons skip off the first few atom layers and exit the specimen after only a few scattering events, so that most are contained in a plane defined by the beam vector and the surface normal, Fig. 3.12b.

3.4.1.5. Energy Distribution of BSE

As the beam electrons travel within the specimen, the various processes of inelastic scattering reduce the electron energy at a rate of roughly 1–10 eV/nm, as indicated in Table 3.1. It is clear from examination of the Monte Carlo electron trajectory plots of Figs. 3.3–3.5 that individual backscattered electrons can follow trajectories which involve different distances of travel in the specimen before escaping. The energy retained by each backscattered electron depends on the history of the beam electron. Monte Carlo calculations of energy distributions of backscattered electrons, given in Fig. 3.13, show several distinct characteristics:

1. The energy distribution is a continuum extending from the incident beam energy (i.e., a small fraction of beam electrons scatter elastically with a sufficiently large angle immediately upon entering the specimen and backscatter without any significant energy loss) to essentially zero energy (i.e., a small fraction of beam electrons travel so far in the specimen that they lose virtually all of their incident energy and reach the surface just prior to being captured by the specimen upon reaching an energy equivalent to the equilibrium thermal energy of the specimen electrons).

2. The energy distribution shows two distinct regions: the uppermost region, denoted I, represents the high-energy hump of backscattered electrons that have lost less than 50% of E_0. For most targets of intermediate and high atomic number, the majority of backscattered electrons will be found in region I. Region II is the broad, gradually decreasing tail of the energy distribution, representing those beam electrons that travel progressively greater distances, losing progressively more energy within the specimen prior to backscattering.

3. The energy distribution shows a peak value, which becomes more distinct for higher atomic number targets. The peak relative energy E/E_0

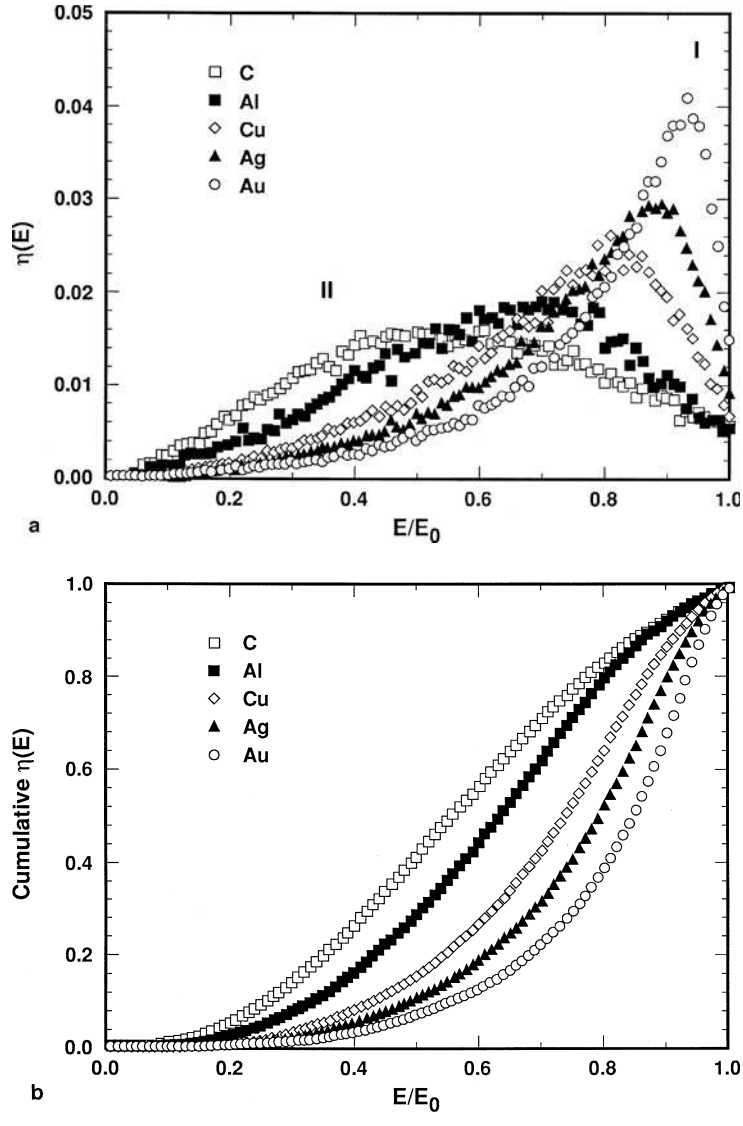


Figure 3.13. Monte Carlo electron trajectory simulations for the energy distribution of backscattered electrons emitted into 2π sr (Newbury and Myklebust, 1991): (a) $\eta(E)$ versus E; (b) cumulative $\eta(E)$ distribution.

increases from approximately 0.8 for copper to 0.95 for gold (Fig. 3.12). The distribution for a light element such as carbon is extremely wide, with only a broad maximum and no distinct peak.

4. The energy distribution of backscattered electrons also depends on the angle above the surface at which the distribution is measured. As shown in Fig. 3.14 for normal beam incidence, the peak in the energy distribution is lost at low take-off angles, that is, shallow backscatter emergence angles above the specimen surface. If the specimen surface is tilted to a high angle ($>70^\circ$) and the energy distribution is measured for forward-scattered electrons, the distribution is shifted to higher energy. Wells (1974) demonstrated that for highly tilted specimens, a significant fraction of those forward-scattered electrons have lost less than 5% of the incident energy ("low-loss electrons") and are closely confined to the beam impact footprint.

5. Cumulative integrals calculated from the backscattered electron energy distributions show that, as listed in Table 3.2, most backscattered electrons retain at least 50% of the incident beam energy, with the fraction rising sharply for intermediate- and high-atomic-number targets. Low-atomic-number scatterers produce far fewer high-energy backscattered electrons.

3.4.1.6. Lateral Spatial Distribution of BSE

Examination of the Monte Carlo electron trajectory simulations shown in Figs. 3.3–3.5 reveals that beam electrons can travel significant distances laterally from the beam impact point before escaping as backscattered electrons. Plots of the density of backscattered electrons escaping per unit area have been calculated by Monte Carlo electron trajectory simulation (Murata, 1973, 1974). As shown in Figs. 3.14a and 3.14b, the spatial distribution for normal beam incidence resembles a sombrero, with a central peak and broad tails of gradually falling density. The radial distribution

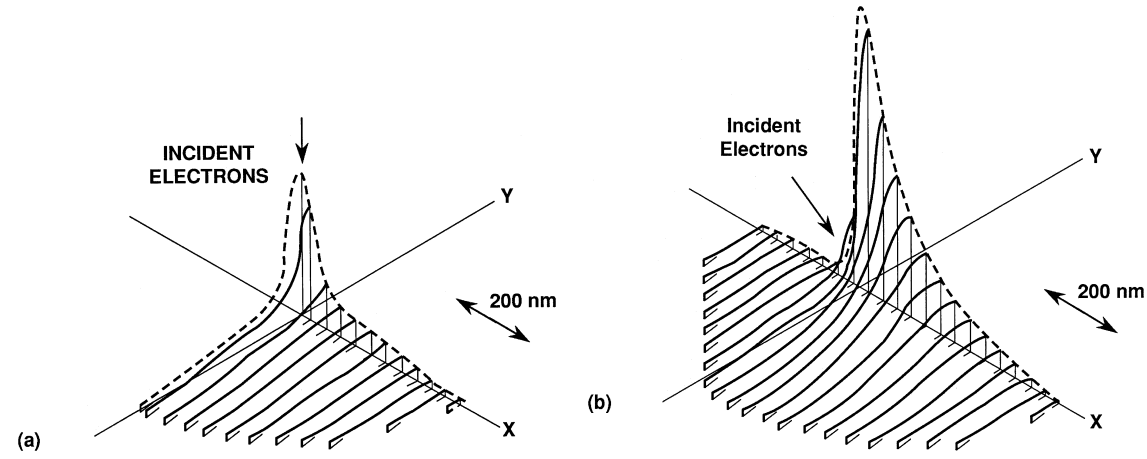


Figure 3.14. Spatial distribution of backscattered electrons from copper plotted as a histogram with the frequency axis perpendicular to the x–y surface plane: (a) Copper, $E_0 = 20$ keV, 0° tilt; (b) 45° tilt. (From Murata, 1973.)

Table 3.2. BSE Energy Distribution

Element	Cumulative fraction of BSE from 0 energy		
	0.5	0.75	0.9
Carbon	$0.55E/E_0$	$0.72E/E_0$	$0.85E/E_0$
Aluminum	0.63	0.77	0.87
Copper	0.74	0.83	0.92
Silver	0.79	0.88	0.94
Gold	0.84	0.92	0.96

function, using the Kanaya–Okayama range to provide a means of scaling the spatial distribution, is found to depend on the atomic number, as shown in Fig. 3.15 (Newbury and Myklebust, 1991). Table 3.3A gives the fraction of the Kanaya–Okayama range necessary to encompass various fractions of the total backscattering. Table 3.3B gives the constants to a power series in atomic number that describes the cumulative radial distribution.

Since the basic principle of SEM imaging is the measurement of information which is localized by focusing the beam to form a probe, a direct consequence of the lateral spreading of backscattered electrons is a decrease in the capability of the SEM to resolve fine features of the specimen on the scale of the focused probe. Examining the sombrero-shaped spatial distribution, we see that the backscattered electrons in the peak of the distribution represent the "good" portion of the backscattered electron signal, which is reasonably closely associated with the incident beam; these electrons are sensitive to fine-scale features of the specimen. The backscattered electrons in the wide brim of the sombrero are emitted so far from the beam impact area that they are insensitive to the fine features in the immediate vicinity of the beam impact, and in fact contribute to the deleterious

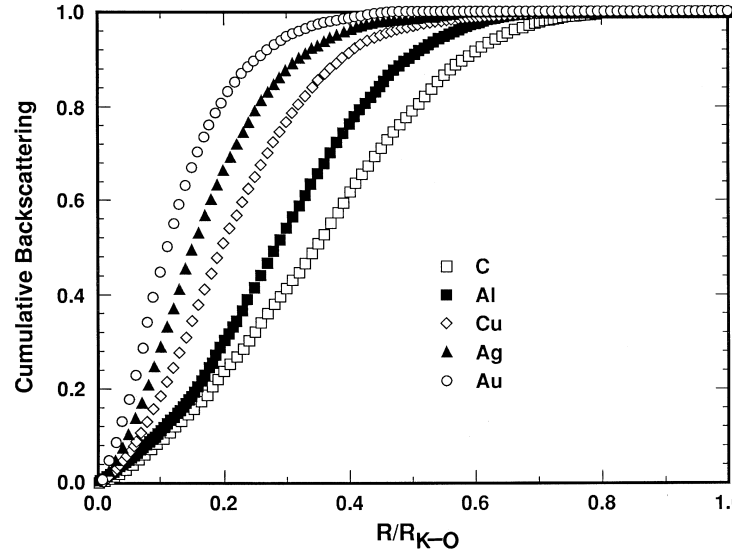


Figure 3.15. Cumulative backscattering with radial distance from beam impact, normalized to the Kanaya–Okayama range. (From Newbury and Myklebust, 1991.)

Table 3.3A. Cumulative Backscattering Radial Distribution[a]

	Distribution fraction		
Element	80%	90%	95%
C	0.502	0.575	0.63
Al	0.419	0.490	0.545
Cu	0.318	0.382	0.439
Ag	0.250	0.310	0.365
Au	0.195	0.248	0.295

[a] $E_0 = 20$ keV; normal beam incidence; lateral position given in terms of R/R_{KO}.

noise component of the backscattered electron signal. For low-atomic-number targets, the brim dominates the signal, whereas for high-atomic-number targets, the peak forms a larger fraction of the backscattered electron signal.

3.4.1.7. Sampling Depth of BSE

A second consequence of the finite distances traveled by the beam electrons within the specimen concerns the depth to which beam electrons penetrate before returning to the surface and backscattering. The SEM is sometimes thought of as a "surface" imaging tool. If we carefully examine the Monte Carlo electron trajectory simulations shown in Figs. 3.3–3.5, we can identify individual trajectories of backscattered electrons for which the beam electron actually penetrated to a significant fraction of the range before reversing its course and returning to the surface to escape as a backscattered electron. Electrons that travel along such trajectories may clearly be influenced by subsurface features of the specimen structure (e.g., inclusions of different composition, voids, etc.) and can carry information on that structure upon escaping. For example, elastic scattering depends very strongly on the atomic number of the target atom. If a low-atomic-number specimen has a small inclusion of a high-atomic-number material buried below the surface, when the probe is positioned above the inclusion, the electron trajectories that reach that inclusion will be influenced by the larger amount of elastic scattering at that location. More trajectories will be reversed because of increased scattering, which will result in a higher backscattering coefficient than would be observed from pure low-atomic-number material, depending on the depth of the feature. The backscattered electron signal can thus respond to subsurface details

Table 3.3B. Quadratic Fit ($y = M_0 + M_1 Z + M_2 Z^2$) for Cumulative Backscattering Radial Distribution (20 keV)

	M_0	M_1	M_2
80%	0.545	$-9.54E-3$	$6.49E-5$
90%	0.621	$-9.96E-3$	$6.68E-5$
95%	0.675	$-9.75E-3$	$6.30E-5$

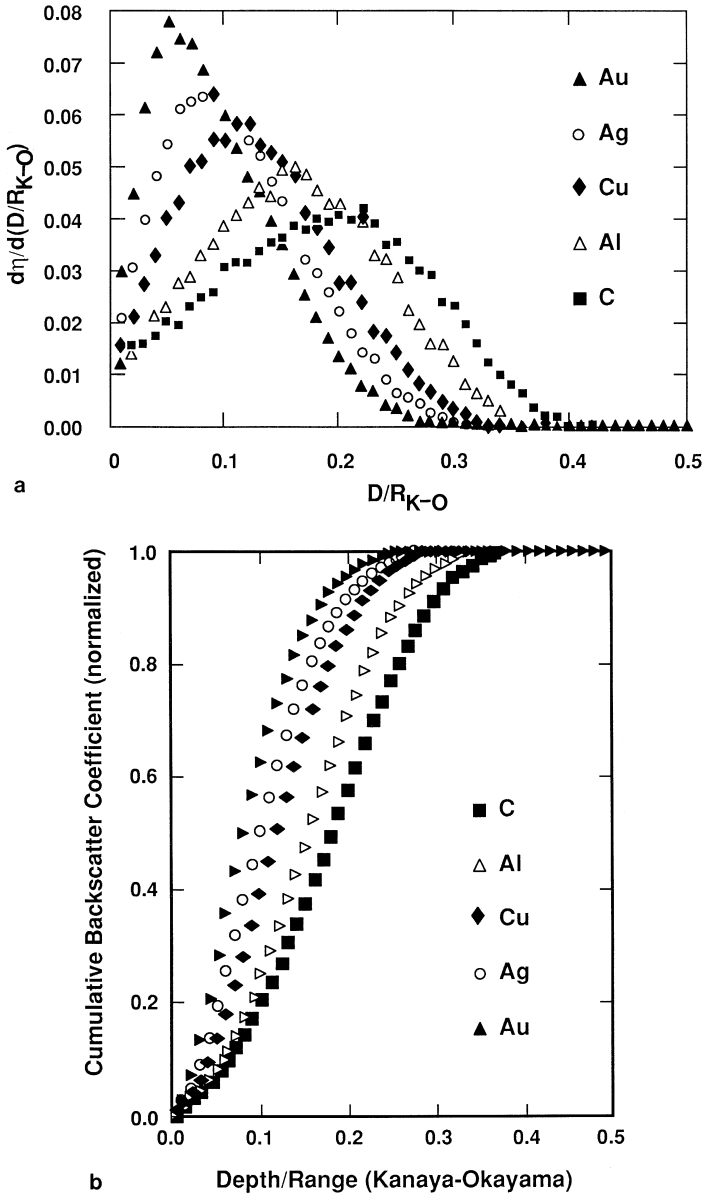


Figure 3.16. (a) Distribution of the maximum depth reached by beam electrons prior to returning to the surface to escape as backscattered electrons, as calculated by Monte Carlo electron trajectory simulation. (b) Cumulative backscattering with depth, derived from (a). $E_0 = 20$ keV. (From Newbury and Myklebust, 1991.)

of the specimen structure. The critical question is: How deep can such an inclusion be in the specimen and still be detected in a BSE image? Monte Carlo calculations are ideally suited to answering such a question because the history of each trajectory is followed and the maximum depth can be recorded. Figure 3.16 shows a Monte Carlo calculation of the frequency

Table 3.4A. Cumulative Backscattering Depth[a]

	0° Tilt			45° Tilt		
Element	63%	90%	95%	63%	90%	95%
C	0.21	0.29	0.32	0.15	0.23	0.27
Al	0.18	0.26	0.28	0.13	0.21	0.24
Cu	0.14	0.22	0.24	0.10	0.19	0.22
Ag	0.12	0.19	0.22	0.085	0.17	0.19
Au	0.10	0.17	0.19	0.075	0.15	0.17

[a] $E_0 = 20$ keV; depths given in terms of D/R_{KO}.

histogram of backscattering versus the maximum depth reached prior to backscattering (Newbury and Myklebust, 1991). Table 3.4A gives the BSE sampling depth obtained by integrating the distribution of Fig. 3.16, and Table 3.4B gives the parameters of a quadratic fit to these data. The strong influence of atomic number is readily apparent. For example, 90% of the total backscattering is obtained in $0.17\,R_{KO}$ for gold and $0.29\,R_{KO}$ for carbon.

Table 3.4A demonstrates that the sampling depth of backscattered electrons is a substantial fraction of the electron range, and the BSE signal is emphatically not a surface-sensitive signal at conventional SEM beam energies (>10 keV). Because of the general scaling of the characteristics of the interaction volume with beam energy, the data in Tables 3.4A and 3.4B are widely applicable. Note, however, that as the beam energy is decreased, the range decreases rapidly, and the backscattered electron signal becomes much more surface-sensitive at low beam energy.

3.4.2. Secondary Electrons

3.4.2.1. Definition and Origin of SE

If the energy distribution of all electrons emitted from a sample is measured over the range from E_0, the incident beam energy, down to 0 keV, a curve similar to that shown schematically in Fig. 3.17a is observed. Most of the energy distribution is dominated by the backscattered electrons, which give rise to the region denoted I, the high-energy peak, and the region II,

Table 3.4B. Quadratic Fit ($y = M_0 + M_1 Z + M_2 Z^2$) for Cumulative Backscattering Distribution with Depth (20 keV)

Fraction	M_0	M_1	M_2
0° tilt			
63%	0.226	−0.00352	2.459 E-5
90%	0.309	−0.00377	2.559 E-5
95%	0.333	−0.00374	2.469 E-5
45° tilt			
63%	0.163	−0.00263	1.935 E-5
90%	0.238	−0.00202	1.148 E-5
95%	0.280	−0.00268	1.630 E-5

the tail extending from intermediate to low energy. If this low-energy tail is extrapolated to zero energy, the yield of backscattered electrons falls smoothly to zero as expected. However, at very low energy, below 50 eV, it is found experimentally that the number of electrons emitted from the specimen increases sharply to a level much greater than the expected contribution from backscattered electrons alone. This increase in emitted electrons forms region III in Fig. 3.17a and is due to the phenomenon of secondary electron (SE) emission (Bruining, 1954; Seiler, 1984; Cailler and Ganachaud, 1990a, b). Secondary electrons are loosely bound outer shell electrons from the specimen atoms which receive sufficient kinetic energy during inelastic scattering of the beam electrons to be ejected from the atom and set into motion. The SE thus created will propagate through the solid, and some will intersect the surface and escape. Secondary electrons are defined purely on the basis of their kinetic energy; that is, all electrons emitted from the specimen with an energy less than 50 eV, an arbitrary choice, are considered as secondary electrons. Although a tiny fraction of the backscattered beam electrons are obviously included in this energy region and are thus counted as SE, their inclusion in the definition of secondary electrons introduces only a negligible effect.

The total secondary electron coefficient δ is given by

$$\delta = \frac{n_{\mathrm{SE}}}{n_{\mathrm{B}}} = \frac{i_{\mathrm{SE}}}{i_{\mathrm{B}}}, \tag{3.13}$$

where n_{SE} is the number of secondary electrons emitted from a sample bombarded by n_{B} beam electrons, and i designates the corresponding currents. In the following sections, the principal properties of SE as they impact SEM imaging will be described; for a comprehensive review of the properties of SE, see Cailler and Ganachaud (1990a, b).

3.4.2.2. SE Yield with Primary Beam Energy

The secondary electron coefficient δ generally rises as the beam energy is lowered. Table 3.5 provides experimental data for the secondary coefficients from Al and Au over a range of energy.

The behavior of the total electron emission from the sample, including both backscattering and secondary emission expressed as $\eta + \delta$, is shown

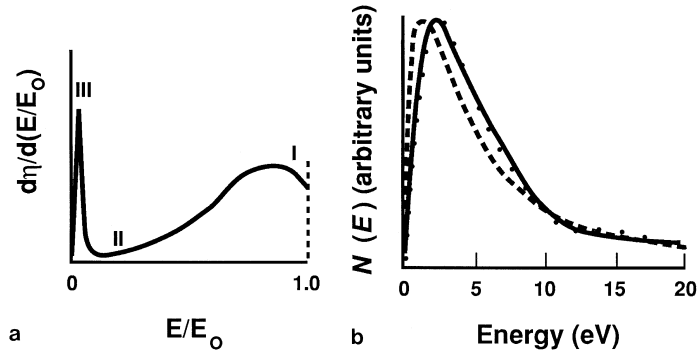


Figure 3.17. (a) Schematic representation of the complete energy distribution of electrons emitted from a target, including backscattered electrons (regions I and II) and secondary electrons (region III). Note that the width of region III is exaggerated. (b) Secondary electron energy distribution as measured (points) and as calculated (lines) with different assumptions on secondary propagation. (From Koshikawa and Shimizu, 1974.)

Table 3.5. Secondary Emission as a Function of Energy[a]

Element	5 keV	20 keV	50 keV
Al	0.4	0.1	0.05
Au	0.7	0.2	0.10

[a] After Reimer and Tollkamp (1980).

schematically as a function of energy in Fig. 3.18. As discussed above, backscattering shows a stronger variability with energy at low beam energy, but the change is usually within a factor of two of high-energy values. Most of the change in $\eta + \delta$ seen at low incident beam energy in Fig. 3.18 really reflects the behavior of δ at low beam energy. Starting with the value of $\delta = 0.1$ appropriate to a beam energy of 10 keV or higher, the value of δ begins to increase significantly as the beam energy is lowered below approximately 5 keV. This increase in δ can be understood in terms of the range of the beam electrons. As discussed below in Section 3.4.2.4, the escape depth of SE is shallow, on the order of nanometers, so all of the SE created by the beam electrons at depths greater than a few nanometers are lost. As the primary beam energy is reduced below 10 keV, the primary electron range becomes more and more shallow, so that an increasing fraction of the SE are produced near the surface and can therefore escape. In the low-beam-energy regime, the primary electron range becomes extremely shallow and most of the primary trajectory occurs within the shallow escape depth, leading to large values of δ.

As δ increases, there comes a point, denoted E_2 and called the upper (or second) crossover point, where $\eta + \delta$ reaches a value of unity. As the beam energy is lowered further, $\eta + \delta$ increases above unity, that is, more electrons are emitted from the surface as a result of backscattering and secondary emission than are supplied by the beam! As E_0 is lowered, the increase in $\eta + \delta$ eventually reaches a peak, which may be as high as 5–20 for certain materials such as insulators. As the beam energy is further

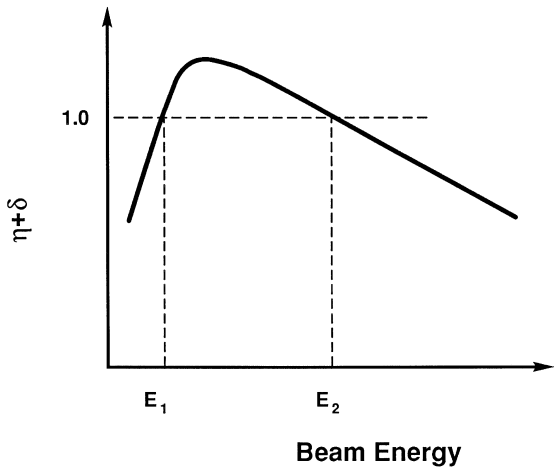


Figure 3.18. Schematic illustration of the total emitted electron coefficient ($\eta + \delta$) as a function of beam energy.

Table 3.6. Upper Crossover Energy for Various Materials (Normal Beam Incidence)

Material	E_2(keV)	Reference
Kapton	0.4	Joy (unpublished)
Electron resist	0.55–0.70	Joy(1987)
Nylon	1.18	Joy (unpublished)
5% PB7/nylon	1.40	Krause *et al.* (1987)
Acetal	1.65	Vaz (1986)
Polyvinyl chloride	1.65	Vaz (1986)
Teflon	1.82	Vaz and Krause (1986)
Glass passivation	2.0	Joy (1987)
GaAs	2.6	Joy (1987)
Quartz	3.0	Joy (1987)
Alumina	4.2	Joy (unpublished)

reduced, the value of $\eta + \delta$ decreases until the lower (or first) crossover point E_1 is reached. Below E_1, $\eta + \delta$ decreases with decreasing beam energy. E_1 is generally below 1 keV and is difficult to measure. Values of E_2 are listed for several materials in Table 3.6. For organic materials, E_2 tends to be in the range 0.5–2 keV, whereas for inorganic materials E_2 is substantially higher, 2–4 keV. As described in detail in Chapter 15, the knowledge of the E_2 point and of the behavior of $\eta + \delta$ at low beam energy forms the basis for controlling charging in SEM imaging of insulators through careful choice of the beam energy and specimen tilt.

3.4.2.3. SE Energy Distribution

One of the main characteristics of secondary electrons is the very low kinetic energy with which they escape the solid. Secondary electrons are principally produced as a result of interactions between energetic beam electrons and weakly bound conduction band electrons in metals or outer shell valence electrons in semiconductors and insulators (Streitwolf, 1959). Because of the great difference in energy between the beam electrons (keV) and the specimen electrons (eV), only a small amount of kinetic energy is likely to be transferred to the SE in a collision. Although "fast" SE with energies up to half the incident beam energy can be produced in some scattering events with more tightly bound atomic electrons, the number of these fast secondary electrons is minor compared to the low-energy "slow" SE that contribute significantly to SEM imaging. The energy distribution of the SE emitted from the specimen is narrow and peaked at very low energy, generally in the range 2–5 eV, as shown in Fig. 3.17b. The choice of 50 eV as the upper cutoff is conservative because the distribution is such that more than 90% of the SE are emitted with less than 10 eV of energy.

3.4.2.4. Range and Escape Depth of SE

An important consequence of the low kinetic energy of secondary electrons is their shallow escape depth. Secondary electrons are produced along the entirety of the beam electron trajectories within the specimen. However, secondary electrons are subject to inelastic scattering and energy

loss during their passage through the specimen. Because the SE are initially ejected with only a few electron volts of energy, they are strongly attenuated with distance traveled as a result. Moreover, when the secondary electrons reach the surface, they must overcome the surface potential barrier (work function), which requires a kinetic energy of several electron volts. As a consequence of the strong attenuation of the SE due to inelastic scattering, the probability of escape decreases exponentially with depth:

$$p \approx \exp\left(-\frac{z}{\lambda}\right), \tag{3.14}$$

where p is the probability, z is the depth below the surface where the SE is generated, and λ is the mean free path of the SE. Seiler (1967) determined that the maximum depth of emission is about 5λ, where λ is about 1 nm for metals and up to 10 nm for insulators. The mean free path is dependent on the energy of the secondary electrons, so that a range of λ actually exists over the full SE energy range (0–50 eV). For the purposes of rough estimation, the values of λ given above will suffice. The substantially greater range of SE in insulators is a direct consequence of the fact that inelastic scattering of SE takes place chiefly with conduction electrons, which are abundant in metals and greatly reduced in insulators. The escape probability of secondary electrons as a function of depth has been calculated by Koshikawa and Shimizu (1974). As shown in Fig. 3.19, there is a sharp drop off in escape probability with depth. Compared to the escape depth for BSE shown in Fig. 3.16, the escape depth for SE is about 1/100 of that for BSE for incident beam energies in the range 10–30 keV. Such a shallow escape depth seems to imply that SE collected as an imaging signal should have a shallow depth of information and be capable of imaging surface features, but the situation is much more complicated. SE are generated throughout the interaction volume of the beam electrons in the specimen, but only those generated within the escape distance from the surface carry information which can be detected by the microscopist. Such observable secondary electrons can be formed by two distinctly different processes, as shown in Fig. 3.20:

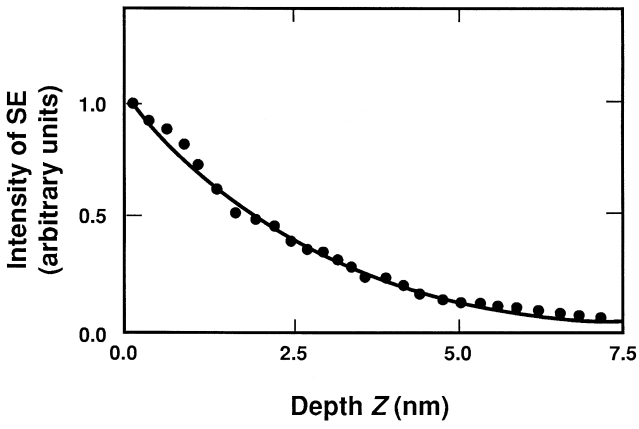


Figure 3.19. Probability of escape of secondary electrons generated at a depth Z below the surface. (From Koshikawa and Shimizu, 1974.)

1. As the incident beam electrons pass through the specimen surface, they generate observable secondaries within the 5λ escape distance of the surface. In the nomenclature of Drescher *et al.* (1970), those SE associated with the incoming beam electrons are designated SE_1. The SE_1 signal is inherently a high-resolution signal that preserves both the lateral spatial resolution of the focused beam and the shallow sampling depth of the SE.

2. As beam electrons scatter within the specimen and approach the surface to emerge as BSE, the SE generated by the exiting BSE within 5λ of the surface can also escape and are designated as SE_2 (Drescher *et al.,* 1970). Because the SE_2 signal is actually a consequence of backscattering, its characteristics as a signal carrier actually follow those of the BSE signal: changes in the SE_2 signal result from changes in the backscattered signal. Both lateral and depth distribution characteristics of BSE are found in the SE_2 signal, which is therefore inherently a low-resolution signal. Thus, the information present in the SE_1 and SE_2 signals is distinctly different.

At low beam energies, $E_0 \leq 5\,\text{keV}$, the range of the beam electrons decreases, and the sampling depth of the BSE is consequently reduced, whereas the SE escape depth is independent of primary beam energy. At a primary beam energy equal to the upper crossover energy E_2, where $\eta + \delta = 1$, the escape depth of secondary electrons and the sampling depth of backscattered electrons are approximately equal. As noted above, the increase in δ at low beam energies reflects the increased fraction of the primary beam energy that is deposited into the production of SE within the escape depth of the primary beam.

3.4.2.5. Relative Contributions of SE_1 and SE_2

Experimentalists have been able to distinguish the relative contributions of SE_1 and SE_2 to the total secondary emission by careful experiments on thin foils, where the backscattered electron component can be effectively eliminated. The total secondary electron coefficient δ_T consists

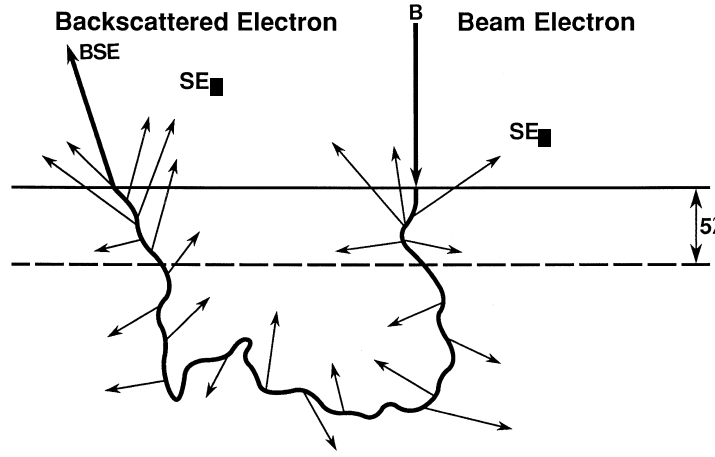


Figure 3.20. Schematic illustration of the origin of two sources of secondary electron generation in the sample. Incident beam electrons (B) generate secondary electrons (SE_1) upon entering the sample. Backscattered electrons (BSE) generate secondary electrons (SE_2) while exiting the sample. λ is the mean free path for secondary electrons.

of two components, δ_1 and δ_2, corresponding to SE_1 and SE_2:

$$\delta_T = \delta_1 + \delta_2\eta, \tag{3.15}$$

where the δ coefficients represent the contribution per incident beam electron. The backscatter coefficient η multiplies δ_2 because the flux of energetic beam electrons traveling back through the entrance surface is reduced to η. The values of δ_1 and δ_2 are not equal, which is to say that an incident beam electron and an average backscattered electron are not equally efficient in generating SE. In general, the ratio δ_2/δ_1 is of the order of 3–4 (Seiler, 1967), that is, BSE are significantly more efficient at generating SE than those same electrons were when they first entered the specimen in the incident beam! This unexpected behavior arises because of two factors:

1. Consider a layer of 5λ thickness starting at the surface; imagine that any SE produced within this layer will escape. For beam energies above 10 keV, the elastic scattering mean free path is much greater than 5λ. For a condition of 0° tilt, the incident beam electrons approach the surface at a right angle and travel through the 5λ layer without significant elastic scattering. The length of the primary electron trajectory along which production of secondaries can lead to escape is thus only 5λ. Now consider the situation for BSE. Because of the randomizing effect of elastic scattering, a large fraction of the BSE is likely to approach the surface at shallower angles and thus the path length through the 5λ escape depth will be greater by the secant of the angle to the surface. This additional path length leads to the production of more SE that can escape from the specimen.
2. Because of inelastic scattering during their passage through the specimen, the beam electrons emerging from the surface as BSE have lost energy relative to the incident beam. Electrons of lower energy are more efficient at transferring kinetic energy to the weakly bound specimen electrons to eject them as SE. Due to their lower energy, BSE have greater efficiency at generating SE as compared to the higher energy incident beam electrons.

Taking $\delta_2/\delta_1 = 3$ and using the appropriate values of δ_T and η, we can calculate the ratio of the yield of secondary electrons from both processes, SE_2/SE_1, for bulk targets. As listed in Table 3.7, the ratio of SE from the two processes is such that the SE_1 signal dominates in δ_T for light elements such as carbon and aluminum because backscattering is low. The processes are approximately equal for an intermediate-atomic-number material such as copper, whereas the SE_2 signal dominates in δ_T for gold.

Table 3.7. Ratio of SE Produced by the Beam and BSE

Element	δ_T	η	SE_2/SE_1
Carbon	0.05	0.06	0.18
Aluminum	0.1	0.16	0.48
Copper	0.1	0.30	0.9
Gold	0.2	0.50	1.5

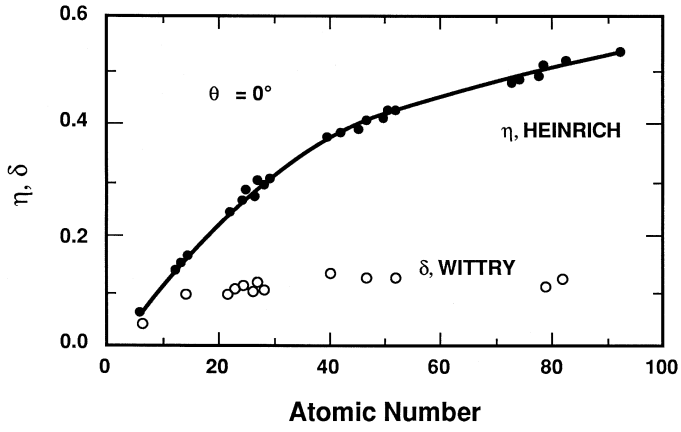


Figure 3.21. Comparison of backscattered electron coefficients and secondary electron coefficients as a function of atomic number; $E_0 = 30$ keV. (Secondary electron data of Wittry, 1966; backscattered electron data of Heinrich, 1966)

3.4.2.6. Specimen Composition Dependence of SE

Compared to the behavior of BSE, whose yield increases nearly monotonically with the atomic number of the specimen, the SE coefficient from pure-element targets is relatively insensitive to atomic number. As shown in Fig. 3.21 for a beam energy of 20 keV, δ has a value of approximately 0.1 for most elements (Wittry, 1966). Two noteworthy exceptions to this average value at 20 keV are carbon, which is anomalously low with a value of approximately 0.05, and gold, which is anomalously high with a value of 0.2 or higher, as reported by several sources (see Joy's database on the accompanying CD-ROM). These SE coefficients are representative of measurements made in the vacuum environment of a typical scanning electron microscope/electron probe x-ray microanalyzer, which operates in the pressure range 10^{-3}–10^{-6} Pa (10^{-5}–10^{-8} torr). The intense electron bombardment produced by a focused beam in such an environment often results in the almost instantaneous deposition of a layer of surface contamination. The source of this contamination is usually from cracking of hydrocarbon deposits that remain on the specimen surface after preparation. With modern vacuum systems that are well maintained and used with specimen prepumping airlocks and cooled anticontamination baffles, contamination from the residual gases that remain in the vacuum system is not usually significant. The emission of SE is very sensitive to the condition of the surface because the low kinetic energy severely limits their range. Measurements of δ made in ultrahigh vacuum conditions with *in situ* surface cleaning of the specimen suggest a much stronger compositional dependence (Seiler, 1983). However, although the values of δ plotted in Fig. 3.21 may not be representative of "true" material dependence, they are nevertheless appropriate to the actual operational environment of the conventional SEM. Examination of the δ data collected by Joy and presented in the data table in the Enhancements CD shows that wide variations in the reported δ values

occur, often by a factor of three or more among researchers, all of whom sought to do careful, reproducible measurements. This extreme variability suggests the impact that surface conditions can have on measurements of δ, and this dependence should alert the microscopist to the possibility of unexpected contrast effects in SEM images where SE form a large part of the signal [such as images prepared with the "through-the-lens" (TTL) detector often incorporated in field emission SEMs, as discussed in Chapter 4].

There are some notable exceptions to this lack of apparent compositional dependence for δ. Under certain conditions, particularly at low beam energies, δ can approach or even exceed unity for many materials that are nominally insulators. The dependence of δ upon composition can be influenced by the nature of molecular bonds, trace elements present in semiconductors, crystal orientation, conductivity, and the perfection of the crystal surface, making it difficult to formulate a predictive description of secondary emission as a function of specimen composition in these cases.

3.4.2.7. Specimen Tilt Dependence of SE

As the angle of specimen tilt θ is increased, δ increases following a relationship reasonably well described by a secant function (Kanter, 1961):

$$\delta(\theta) = \delta_0 \sec\theta, \tag{3.16}$$

where δ_0 is the value of δ found at 0° tilt (normal incidence); this behavior is shown in the plot in Fig. 3.22a. The origin of the secant function behavior can be understood with the simple argument illustrated in Fig. 3.22b. Consider that all secondary electrons created along the primary beam trajectory within a distance R_0 of the surface can escape. As described in more detail in the next section, this escape depth is very shallow, so that the primary beam electrons are effectively not scattered elastically in this distance and lose so little energy that the rate of secondary production is effectively constant. When the specimen is tilted to a value of θ, the length R of the

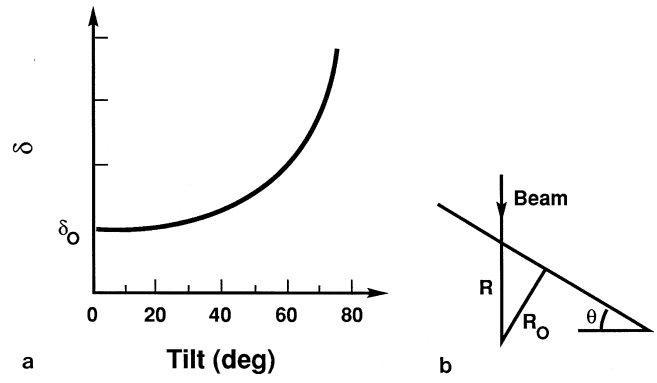


Figure 3.22. Behavior of the secondary electron coefficient δ as a function of specimen tilt θ, following a secant law. (b) Schematic illustration of the origin of the secant law behavior.

primary electron path within a distance of R_0 from the surface increases as $R = R_0 \sec\theta$. Since production and escape of secondaries are proportional to path length, the secondary electron coefficient increases following a secant relationship. This argument is not strictly complete, however, because secondary electrons can also be created by the beam electrons as they backscatter. As discussed above, backscattering also increases with tilt, which further contributes to the increase in δ at high tilt angles.

When the primary beam energy is reduced to low values of the order of E_2, the secant law behavior described by Eq. (3.16) no longer applies. The ranges of the BSE and SE are similar, so that essentially all of the SE that are generated in the interaction volume can escape. Increasing the angle of tilt does not increase the primary electron range within the escape depth, so no added secondary production is gained within the secondary escape depth.

3.4.2.8. Angular Distribution of SE

In a fashion similar to the arguments used to explain the angular distribution of BSE, when the specimen is placed at 0° tilt, it can be shown that the emission of secondary electrons follows a cosine distribution of an angle ϕ measured relative to the surface normal. When the specimen is tilted, however, the angular emission behavior of BSE and SE differ. For BSE, the angular distribution becomes asymmetric and peaked in a forward scattering direction because elastic scattering is strongly favored in the forward scattering direction. For SE the angular distribution remains a cosine relative to the local surface normal. This behavior is a result of the fact that SE are generated isotropically by the primary electrons regardless of tilt. Although the total secondary electron coefficient δ increases with increasing tilt, the angular distribution of emission remains the same because the distribution of path lengths out of the specimen follows a $1/\cos\phi$ distribution relative to the surface normal regardless of the tilt.

References

Arnal, F., P. Verdier, and P.-D. Vincinsini (1969). *C. R. Acad. Sci. Paris* **268,** 1526.

Berger, M. (1963). In *Methods in Computational Physics,* Vol. 1 (B. Adler, S. Fernback, and M. Rotenberg, eds.), Academic Press, New York, p. □□□.

Berger, M. J., and S. M. Seltzer (1964). In *Penetration of charged Particles in Matter* (U. Fano, ed.), National Academy of Sciences–National Research Council Publication 1133, Washington, D.C., p. 205.

Bethe, H. (1930). *Ann. Phys.* (Leipzig) **5,** 325.

Bishop, H. (1966). In *Proceedings of the 4th International Conference on X-ray Optics and Microanalysis* (R. Castaing, P. Deschamps, and J. Philibert, eds.), Hermann, Paris, p. 153.

Bruining, H. (1954). *Physics and Application of the Secondary Electron Process,* Pergamon, London.

Cailler, M., and J.-P. Ganachaud (1990a). *Scanning Microsc. Suppl.* **4,** 57.

Cailler, M., and J.-P. Ganachaud, (1990b). *Scanning Microsc. Suppl.* **4,** 81.

Drescher, H., L. Reimer, and H. Seidel (1970). Z. Agnew, Phys. *29*, 33.

Everhart, T. E., R. F. Herzog, M. S. Chang, and W. J. DeVore (1972). In *Proceedings of the 6th International Conference on X-ray Optics and Microanalysis* (G. Shinoda, K. Kohra, and T. Ichinokawa, eds.), University of Tokyo Press, Tokyo, p. 81.

Heinrich, K. F. J. (1966). In *Proceedings of the 4th International Conference on X-ray Optics and Microanalysis* (R. Castaing, P. Deschamps, and J. Philibert, eds.), Hermann, Paris, p. 159.

Heinrich, K. F. J., and D. E. Newbury, eds. (1991). *Electron Probe Quantitation,* Plenum Press, New York.

Heinrich, K. F. J., D. E. Newbury, and H. Yakowitz, eds. (1976). *Use of Monte Carlo Calculations in Electron Probe Microanalysis and Scanning Electron Microscopy,* National Bureau of Standards Special Publication 460, U.S. Government Printing Office, Washington, D.C.

Henoc, J., and F. Maurice (1991). In *Electron Probe Quantitation* (K. F. J. Heinrich and D. E. Newbury, eds.), Plenum Press, New York, p. 105.

Hovington, P., D. Drouin, and R. Gauvin (1997). *Scanning,* **19,** 1.

Hunger, H.-J., and L. Kuchler, (1979). *Phys. Stat. Sol. a* **56,** K45.

Joy, D. C. (1987). *J. Microsc.* **147,** 51.

Joy, D. C. (1991). *J. Microsc.* **161,** 343.

Joy, D. C., and S. Luo (1989). *Scanning* **11,** 176.

Kanaya, K., and S. Okayama (1972). *J. Phys. D. Appl. Phys.* **5,** 43.

Kanter, H. (1961). *Phys. Rev.* **121,** 677.

Koshikawa, T., and R. Shimizu (1974). *J. Phys. D Appl. Phys.* **7,** 1303.

Krause, S. J., W. W. Adams, S. Kumar, T. Reilly, and T. Suziki (1987). In *Proceedings EMSA Conference* (G. W. Bailey, ed.), San Francisco Press, San Francisco, p. 466.

Murata, K. (1973). *SEM/1973*, IIT Research Institute, Chicago, p. 268.

Murata, K. (1974). *J. Appl. Phys*. **45,** 4110.

Newbury, D. E., and R. L. Myklebust (1979). *Ultramicroscopy* **3,** 391.

Newbury, D. E., and R. L. Myklebust (1984). In *Electron Beam Interactions with Solids* D. F. Kyser, R. Shimizu, and D. E. Newbury, eds.), SEM, Inc., Chicago, p. 153.

Newbury, D. E., and R. L. Myklebust, in Microbeam Analysis (D. G. Howitt, ed.), San Francisco Press, San Francisco, p. 561.

Newbury, D. E., H. Yakowitz, and R. L. Myklebust (1973). *Appl. Phys. Lett.* **23,** 488.

Newbury, D. E., D. C. Joy, P. Echlin, C. E. Fiori, and J. I. Goldstein (1986). *Advanced Scanning Electron Microscopy and X-ray Microanalysis,* Plenum Press, New York.

Niedrig, H. (1978). *Scanning* **1,** 17.

Rao-Sahib, T. S., and D. B. Wittry (1972). In *Proceedings of the 6th International Congress on X-ray Optics and Microanalysis,* (G. Shinoda, K. Kohra, and T. Ichinokawa, eds.), University of Tokyo Press, Tokyo, p. 131.

Reimer, L., and C. Tollkamp, (1980). *Scanning* **3,** 35.

Reuter, W. (1972). In *Proc. 6th Int'l Cong. on X-ray Optics and Microanalysis* (G. Shinoda, K. Kohra, and T. Ichinokawa, eds.), Univ. Tokyo Press, Tokyo, p. 121.

Seiler, H. (1967). *Z. Angew. Phys.* **22,** 249.

Seiler, H. (1983). *J. Appl. Phys*. **54,** R1.

Seiler, H. (1984). In *Proceedings of the 1st Pfefferkorn Conference* (D. Kyser, H. Nedrig, R. Shimizu, and D. Newbury, eds.) Scanning Microscopy, Inc., Chicago, p. 33.

Shimizu, R., and K. Murata (1971). *J. Appl. Phys.* **42,** 387.

Streitwolf, H. W. (1959). *Ann. Phys.* (Leipzig) **3,** 183.

Vaz, O. W. (1986). Ph.D. Thesis, Arizona State University, Tempe, Arizona.

Vaz, O. W., and S. J. Krause (1986). In *Proceedings EMSA Conference* (G. W. Bailey, ed.), San Francisco Press, San Francisco, p. 676.

Wells, O. C. (1974). *Scanning Electron Microscopy*, McGraw-Hill, New York.

Wittry, D. B. (1966). In *Proceedings of the 4th International Conference on X-ray Optics and Microanalysis,* (R. Castaing, P. Deschamps, and J. Philibert, eds.), Hermann, Paris, p. 168.

4

Image Formation and Interpretation

4.1. The Story So Far

Scanning electron microscopy is a technique in which images form the major avenue of information to the user. A great measure of the enormous popularity of the SEM arises from the ease with which useful images can be obtained. Modern instruments incorporate many computer-controlled, automatic features that permit even a new user to rapidly obtain images that contain fine detail and features that are readily visible, even at scanning rates up to that of television (TV) display. Although such automatic "computer-aided" microscopy provides a powerful tool capable of solving many problems, there will always remain a class of problems for which the general optimized solution may not be sufficient. For these problems, the careful microscopist must be aware of the consequences of the choices for the beam parameters (Chapter 2), the range of electron–specimen interactions that sample the specimen properties (Chapter 3), and finally, the measurement of those electron signals with appropriate detectors, to be discussed in this chapter. With such a systematic approach, an advanced practice of SEM can be achieved that will significantly expand the range of application to include many difficult imaging problems.

4.2. The Basic SEM Imaging Process

This chapter will consider the formation and interpretation of SEM images. One of the most surprising aspects of scanning electron microscopy is the apparent ease with which SEM images of three-dimensional objects can be interpreted by any observer, including young children, with no prior knowledge of the SEM instrument or the arcane subject of electron–solid interactions. This aspect of the SEM is often taken for granted, and yet it is one of the most important reasons for the great utility and wide acceptance of the instrument. SEM images are routinely presented in textbooks, popular scientific articles, and even advertisements with little or no mention of

the type of microscopy employed in preparing the image, of the complex way in which the image was constructed, or with any guidance to aid the viewer's interpretation. It can safely be assumed that the reader will automatically perceive the true topographic nature of the specimen without any special knowledge concerning the origin of the image. For this to be true, the SEM imaging process must in some sense mimic the natural experience of human observers in visualizing the world around them. Such a situation is quite surprising in view of the unusual way in which the image is formed, which seems to differ greatly from normal human experience with images formed by light and viewed by the eye. In the SEM, high-energy electrons are focused into a fine beam, which is scanned across the surface of the specimen. Complex interactions of the beam electrons with the atoms of the specimen produce a wide variety of radiation products: backscattered electrons, secondary electrons, absorbed electrons, characteristic and continuum x-rays, etc. A sample of this radiation is collected by a detector, most commonly the Everhart–Thornley detector (positively biased scintillator–photomultiplier), and the collected signal is amplified and displayed on a cathode ray tube (CRT) or television screen scanned in synchronism with the scan on the specimen. Despite this complicated and unusual path to the formation of the image, the result is somehow straightforward to interpret, at least for the important class of objects that can be described as topographic (three-dimensional) in nature and are viewed at low to intermediate magnifications (up to $10{,}000\times$).

Although this "entry-level" approach to scanning electron microscopy will suffice for some problems, the experienced microscopist will frequently discover situations where it is necessary to make use of advanced concepts of image interpretation. Such situations are encountered, for example, with complex topography, especially when the specimen or feature has little familiarity; in the case of high-magnification images ($>10{,}000\times$), where questions of resolution limits and the delocalization of the imaging signals must be addressed; and with special types of specimens and contrast mechanisms, such as voltage contrast or magnetic contrast. The intellectual tools which the microscopist must bring to make a proper interpretation of an imaging problem include knowledge of electron optics, beam–specimen interactions, detector characteristics, signal quality/feature visibility relations, and signal/image processing. The first two topics, electron optics and beam–specimen interactions, were covered in detail in Chapters 2 and 3; the latter three topics will be the subject of this chapter.

Questions which constantly arise in microscopy studies include: What types of features can be seen in an image (i.e., what information about the specimen can be observed in an SEM image), and how small a region showing a particular characteristic can be detected (resolution)? These questions of image information and resolution are closely related, and in this chapter equations will be developed that permit the microscopist to make an estimate of what can be seen when specific operating conditions are chosen. SEM images are often perceived to be so spectacular that the impression arises that the technique is not subject to constraints. If a specimen contains a class of features by design and synthesis, then the SEM must be able to

image them. As a corollary, the absence of features from an image must guarantee their absence from that region of the specimen under study. In fact, the SEM, like all imaging devices, is constrained by fundamental statistical fluctuations in the signal, leading to limits of visibility. (The critical distinction between "visibility" and "resolution" will be made below.) For any choice of imaging conditions, there will always be a level of contrast below which features are not visible. Fortunately, these inevitable limitations on the performance of the SEM can be estimated numerically, and a strategy to improve visibility for a particular case can be developed. Even with a careful imaging strategy, however, it may not be possible to establish SEM visibility for some objects under any practical imaging conditions within the operating parameters of a particular SEM. Recognizing that such limitations exist and understanding how to optimize instrument parameters to achieve the best performance is critical to advanced use of the SEM.

Digital imaging has now been established as the dominant form of SEM image storage and manipulation. Indeed, for many users, SEM images may never appear in the form of photographs, but rather images will be recorded electronically, stored as a computer file, examined, modified, and measured on a computer display, and assembled into text reports or presentation graphics, and only in the final presentation to a client may a "hard copy" of the document be produced. Some reference will still be made to analog systems because SEMs often have a useful life of 20 years or more, so that readers of this book will likely use equipment spanning full-analog to full-digital control.

4.2.1. Scanning Action

The detailed description of the electron optics of the SEM, as given in Chapter 2, can be briefly summarized by considering the degree of control which the microscopist has in choosing the parameters that characterize the beam as focused to form a probe at the specimen plane: probe current i_B (pA to μA); probe diameter d, which ranges from a minimum of 1 nm in the highest resolution instruments (typically limited to 5 nm in conventional instruments) to a maximum of 1 μm; and probe divergence α (10^{-4}–10^{-2} rad). Of course, these three parameters cannot be selected independently because they are interrelated, as described by the brightness equation. The brightness equation gives the theoretical limit to the beam performance; lens defects (aberrations) result in beams which are either larger than expected for a given beam current and divergence or else lower in current for a specified diameter and divergence.

The SEM imaging process is illustrated schematically in Fig. 4.1. The electron beam, defined by the parameters d, i_B, and α, exits the electron column, enters the specimen chamber, and strikes the specimen at a single location on the optic axis of the column. The beam electrons interact elastically and inelastically with the specimen, forming the limiting interaction volume from which the various types of radiation emerge, including backscattered, secondary, and absorbed electrons, characteristic and bremsstrahlung x-rays, cathodoluminescence radiation (long-wavelength

photons in the ultraviolet, visible, and infrared ranges) in some materials (compounds, minerals, semiconductors, etc.), and finally, in certain semiconductor materials and structures, beam-induced currents. By recording the magnitude of these signals with suitable detectors, a measure of certain properties of the specimen, such as topography, composition, etc., can be made at the single location where the electron beam strikes. To study more than a single location and eventually construct an image, the beam must be moved from place to place by means of a scanning system, as illustrated in Fig. 4.1. This "scanning action" is usually accomplished by energizing electromagnetic coils arranged in sets consisting of two pairs, one pair each for deflection in the x and y directions. (The convention will be followed that an xyz coordinate system will be established with the z direction parallel to the optic axis of the microscope and the x and y axes defining a plane perpendicular to the beam.) A typical double-deflection scanning system, as shown in Fig. 4.1, has two sets of electromagnetic scan coils, located in the bore of the final (objective) lens. As controlled by the scan generator, the upper coils act to drive the beam off-axis, and the lower coils deflect the beam to cross the optic axis again, with this second crossing of the optic axis taking place in the final (beam-defining) aperture. This system has the advantage that by placing the scan coils within the lens, the region below is kept open and the specimen can be placed relatively close to the lens at a short working distance. A short working distance means the objective lens will be relatively more highly excited, which minimizes some of the lens aberrations, improving the beam performance. By locating the beam-defining aperture at the second crossover (pivot point), large scan angles necessary for low magnifications can be obtained without cutting off the field of view on the aperture (Oatley, 1972).

Scanning action is produced by altering the strength of the current in the scan coils as a function of time, so that the beam is moved through a sequence of positions on the specimen (e.g., locations 1, 2, 3, 4, etc., in Fig. 4.1) and the detector(s) samples the electron–specimen interaction at a defined sequence of points. In an analog scanning system, the beam is moved continuously, with a rapid scan along the x axis (the line scan)

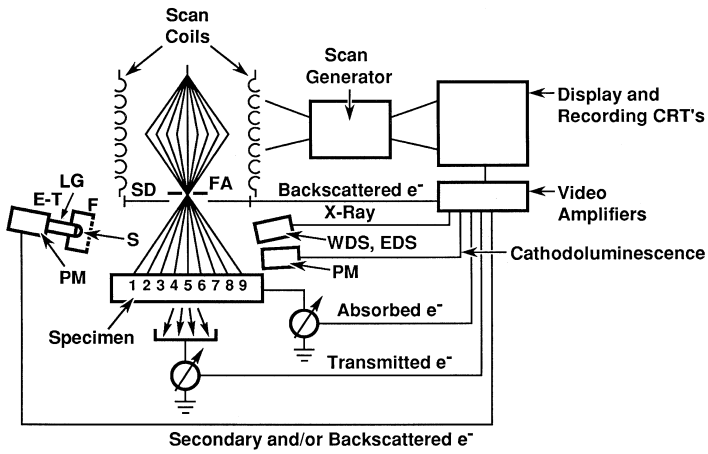


Figure 4.1. Schematic illustration of the scanning system of the SEM. Abbreviations: FA, final aperture; SD, solid-state backscattered electron detector; EDS, energy-dispersive x-ray spectrometer; WDS, wavelength-dispersive x-ray spectrometer; CRT, cathode ray tube; E–T, Everhart–Thornley secondary/backscattered electron detector, consisting of F, Faraday cage; S, scintillator; LG, light guide; and PM, photomultiplier. Successive beam positions are indicated by the numbered rays of a scanning sequence.

and a slow scan, typically at 1/500 of the line rate, at right angles along the y axis (the frame scan), so that a good approximation to an orthogonal scan is produced. In digital scanning systems, only discrete beam locations are allowed. The beam is addressed to a particular location (x, y) in a matrix, remains there for a fixed time (the dwell time), and is then rapidly moved to the next point. The beam current may actually be turned off during the movement between pixels by an electrostatic blanking system placed just below the gun. The time to shift the beam between points is negligible compared to the dwell time. For display, the image is constructed on a cathode ray tube (CRT), which is scanned in synchronism with the scan on the specimen, controlled by the same scan generator. The signal derived from one of the detectors is amplified and used to control the brightness of the CRT ("intensity modulation"), often with some form of signal processing applied to enhance the visibility of the features of interest, a topic to be covered in detail below. At the resolution at which a human observer usually examines or photographs the CRT image, there is no effective difference between images prepared with analog and high-density (e.g., 1024 × 1024, or higher) digital scans; both types of images appear continuous in nature. An added benefit of the digital scan is that the numerical address of the beam location is accurately and reproducibly known, and therefore the information on the electron interaction can be encoded in the form (x, y, I), where x and y give the address and I is the intensity. Such encoding provides the basis for digital image processing, to be discussed later in this chapter.

4.2.2. Image Construction (Mapping)

The information flow from the scanning electron microscope consists of the scan location in x–y space and a corresponding set of intensities from the set of detectors (backscattered electron, secondary electron, transmitted electron, specimen current, x-ray, cathodoluminescence, etc.) that monitor the beam–specimen interactions, all of which can be monitored in parallel. This information can be displayed to the observer in two principal ways: line scans and area scans.

4.2.2.1. Line Scans

In the line scan mode, the beam is scanned along a single vector on the specimen, e.g., in the x or the y direction. The same scan generator signal is used to drive the horizontal scan of the CRT. The resulting synchronous line scan on the specimen and the CRT produces a one-to-one correspondence between a series of points in the "specimen space" and on the CRT or "display space." In such a line scan displayed on a CRT, the horizontal position is related to distance along a particular line on the specimen. The effective magnification factor M between the specimen space and the CRT space is given by the ratio of the lengths of the scans:

$$M = L_{\text{CRT}}/L_{\text{specimen}}, \tag{4.1}$$

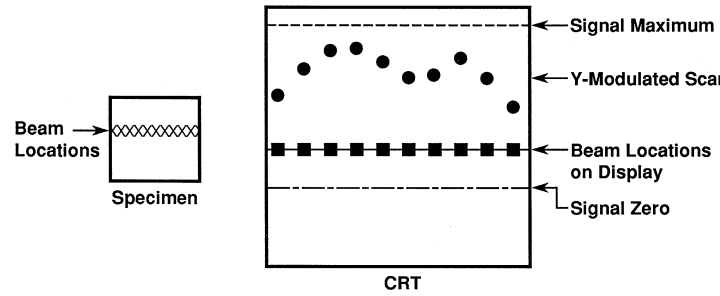


Figure 4.2. SEM line scan display: The beam is scanned along a locus of points on the specimen, and the beam on the CRT is scanned in a similar locus along the horizontal (x axis). A signal measured from one of the detectors (backscattered electrons, x-rays, etc.) is used to adjust the vertical position (y axis). For quantitative measurements, the trace corresponding to the locus of points should be recorded, along with traces corresponding to the signal zero and maximum values.

where L denotes the length of the scans. The numerical value of the magnification reported on the alphanumeric display typically refers to the final image format recorded on the SEM photographic system. The magnification value is thus appropriate to the size of the printed photographic image (often a 10×12.5-cm format), and it may underestimate the true magnification of the large viewing CRT by a large factor, often two or three times. On the other hand, the calibrated scale bar, usually expressed in micrometers or nanometers, refers to a specific distance in the scanned image. This value scales appropriately with any reproduction of the image with any device, including projection on a large screen, and is thus the meaningful dimensional quantity.

The intensity measured during the line scan by one of the detectors, for example, the signal from a backscattered electron detector or from an x-ray spectrometer, can be used to adjust the y deflection ("y modulation") of the CRT, which produces a trace such as that illustrated in Fig. 4.2. For example, if the signal is the characteristic x-ray intensity, the y deflection of the CRT will be proportional to the amount of a particular element present at each location. Although line scans are the simplest type of scanning action, they are almost never used alone and are typically superimposed on the more familiar two-dimensional SEM image. Line scans are extremely useful for diagnostic work, where the signal profile across a feature is needed. Line scans can be used to display small signal changes that can easily be detected as offsets in y modulation, but which would be difficult to discern in a conventional intensity-modulated area image. In recording a line scan, as shown in Fig. 4.2, it is important to record (1) the scan locus (i.e., the location in the area raster at which the line scan is taken), (2) the y-modulated scan of the signal, (3) the signal zero level, and (4) the signal maximum. This information is critical if any quantitative use is to be made of the line scan data.

4.2.2.2. Image (Area) Scanning

To form the SEM image with which we are most familiar, the beam is scanned on the specimen in an x–y pattern while the CRT is scanned

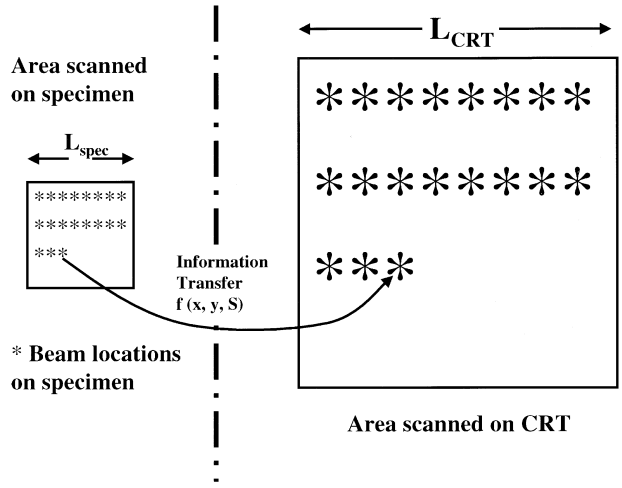


Figure 4.3. The principle of image display by area scanning. A correspondence is established between a set of locations on the specimen and a set on the CRT. Magnification = $L_{CRT}/L_{specimen}$.

in the same x–y pattern, as illustrated in Fig. 4.3. Again, a one-to-one correspondence is established between the set of beam locations on the specimen and the points on the CRT, and the linear magnification of the image is given by Eq. (4.1). To display the beam–specimen interaction information, the signal intensity S derived from a detector is used to adjust the brightness of the spot on the CRT, a process referred to as "intensity modulation," shown schematically in Fig. 4.4. Typically, an SEM contains one or more slow-scan CRT visual displays, a photographic CRT, and a television-rate display, or all of these displays may be replaced by a single computer monitor on a fully digital instrument. The television rate display operates with a rapid decay phosphor that can respond to scan speeds of

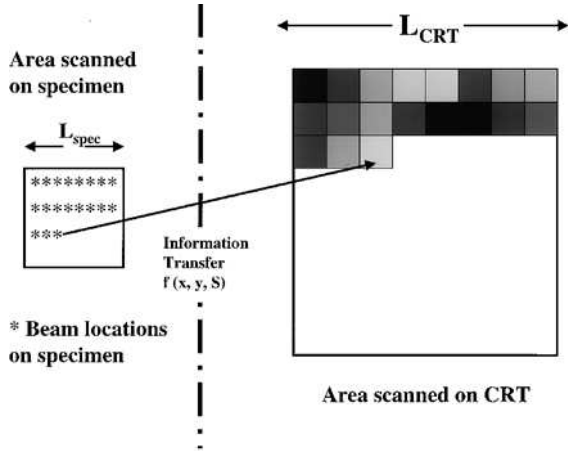


Figure 4.4. The principle of intensity modulation used to display the magnitude of the signal produced by electron–specimen interaction at the locations scanned in Fig. 4.3. Black represents low intensity; gray, intermediate intensity; white, high intensity.

1/30 s per frame, which the observer perceives as "flicker-free." The visual display CRT contains a phosphor with a relatively long persistence (usually 1 s or more) to permit the observer to view the image with frame scan speeds in the range 0.5–5 s. A high-resolution (small spot size) CRT with a short persistence phosphor is used for photographic recording, almost invariably with rapid-development film.

The creation of an SEM image consists in constructing an intensity map in the analog or digital domain. Unlike an optical microscope or transmission electron microscope, no true image exists in the SEM. In a true image, there are actual ray paths connecting points on the specimen to corresponding points in the image as displayed on a screen or detected by the eye or film. It is not possible to place a sheet of film anywhere in the SEM chamber and record an image, as it is in a light-optical microscope or a transmission electron microscope. In the SEM, image formation is produced by a mapping operation that collects information from the specimen space and passes the information to the display space. That such an abstract process of creating an image can produce a class of images of topographic objects which are readily understandable is a considerable surprise, fortunately a pleasant and extremely useful one.

The information contained in the image will convey the true shape of the object if the specimen and CRT scans are synchronous and are constructed to maintain the geometric relationship of any arbitrarily chosen set of points on the specimen and on the CRT. When this condition is satisfied, as shown in Fig. 4.5, a triangle on the specimen remains a triangle of the same shape on the CRT. The influence of projection distortion for three-dimensional objects and various scan defects on the shape of objects in SEM images will be described below. The nature of the intensity variations that produce the gray-level shading of various objects in the image will be discussed below in Section 4.4.

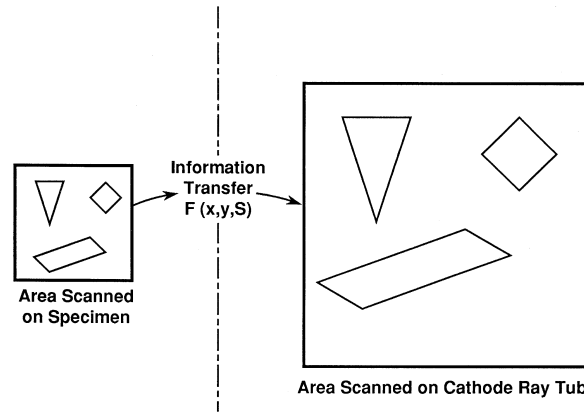


Figure 4.5. Shape correspondence between the scanned field on the specimen and the display on the CRT. In a perfect scan system, the shapes in the plane of the scan in the microscope are transferred without distortion to the CRT display.

A digital image consists of a numerical array (x, y, S) stored in a computer memory, where each entry consists of three values, two for the position and one for the signal intensity. To create a digital image, the x–y pattern of scanned locations is made to correspond to a matrix array in a computer memory or in a framestore (a special dedicated computer memory board consisting of an array of memory registers adequate to record one or more complete images). Each scan position is generated as a digital address and converted with a digital-to-analog converter (DAC) into a voltage to drive the scan circuit. The number of discrete beam positions, the picture elements or "pixels," is specified as the digital resolution, for example, 512×512, 1024×1024, etc. The pattern does not have to be square, so that rectangular or other scan shapes can be generated, but it is important to maintain symmetric (i.e., square) pixels to avoid introducing a distortion into the digital image. When the beam has been addressed to a location on the specimen, the analog signal intensity is measured by the detector(s) (SE, BSE, x-rays, etc.), integrating for the length of the pixel dwell. To ensure proper matching of the measured signal to the processing chain, the analog signal amplifier is first adjusted with a separate oscilloscope or an equivalent y-modulated scan on the CRT to bring the range of the signal over the entire frame scan into the input acceptance range of the analog-to-digital converter (ADC). The voltage signal produced by the signal amplifier is digitized to a value S for the computer and stored as a discrete numerical value in the corresponding register (x, y, S). Typically, the intensity is digitized into a range of 8 bits, which gives $2^n = 2^8 = 256$ discrete levels.

With only 256 levels in the digital representation of the signal, it is only possible to recognize changes at the level of $1/256 = 0.4\%$ with subsequent digital functions, given that the input signal spans the entire range of the DAC. Generally, to avoid the possibility of saturating the response of the DAC, the input signal range is adjusted more conservatively so that the signal does not reach the 0 or 255 digital levels. Note that once the digitization has been performed, it is not possible to recover information not recorded in the original digitization step. For many applications, it is desirable to provide more discrete digital levels for subsequent processing, so the original signal can be digitized to 12 bits (4096 levels), 16 bits (65,536 levels), or even higher. This additional signal resolution is gained by digitizing for longer integration times or sampling repeatedly at each pixel and averaging. Generally, we do not wish to record all images at the maximum digital resolution available. The penalty for recording large digital images is the need to provide mass storage to save the images. A $1024 \times 1024 \times 256$-level image requires over 1 megabyte (1 MB) of mass storage, whereas a $2048 \times 2048 \times 256$-level image requires 4 MB. The choice of the digital x–y resolution should be made to slightly oversample the image in terms of the picture element (see Section 4.2.4), and the digital resolution of the intensity depends on the need for subsequent contrast manipulation.

The digital image is viewed by converting the numerical values stored in memory back into an analog signal (digital-to-analog conversion) for

display on a CRT. The density and speed of computer memory are such that large image arrays (512 × 512, or larger) can be stored digitally and read out in "real time," producing a "flicker-free" image. Such digitally stored images have many advantages. One special advantage previously mentioned is the possibility of minimizing the radiation dose necessary to view the specimen by viewing the digitally stored image until a fresh image field is needed. To continuously view an image with an analog SEM system, the beam must be constantly scanned on the specimen in order to refresh the display. The information in each scan is discarded unless recorded by film or videotape. Radiation damage or beam-induced contamination generally scales with the total dose, so that analog imaging with repeated scans exposes the specimen to possible degradation. Digital images are stored in memory, and can be repeatedly read out to the analog display for inspection of that particular frame. Thus, the beam can be scanned for a single frame on the specimen to accumulate data in a memory and then "blanked" or deflected off the specimen, generally into a Faraday cup for measurement of the current as a stability check, while the digital memory is repeatedly read to refresh a CRT display. As described in detail below, image processing functions can be applied to this stored image to enhance the visibility of objects of interest. Once the displayed image has been satisfactorily adjusted, the resulting image can be transmitted to the photographic recording CRT or other form of "photoquality" hard copy output device.

4.2.3. Magnification

Changing magnification in the SEM image is accomplished by adjusting the length of the scan on the specimen for a constant length of scan on the CRT, as described by Eq. (4.1), which gives the linear magnification of the image. There are several important points about the SEM magnification process:

1. *Numerical value of magnification:* Because the maximum CRT scan length is fixed to the full dimension L of the tube, for example, 10 cm, an increase in magnification is obtained by *reducing* the length l of the scan on the specimen. Table 4.1 gives the size of the area sampled on the specimen as a function of magnification.

Table 4.1. Area Sampled as a Function of Magnification

Magnification[a]	Area on sample
10×	$(1\text{ cm})^2 = 100\text{ mm}^2$
100×	$(1\text{ mm})^2 = 1\text{ mm}^2$
1,000×	$(100\ \mu\text{m})^2 = 0.01\text{ mm}^2$
10,000×	$(10\ \mu\text{m})^2 = 10^{-4}\text{ mm}^2$
100,000×	$(1\ \mu\text{m})^2 = 10^{-6}\text{ mm}^2$
1,000,000×	$(100\text{ nm})^2 = 10^{-8}\text{ mm}^2$

[a] Assumes magnification is relative to a CRT screen that measures 10 cm ×10 cm.

2. *Area sampled:* The capability of the SEM to obtain useful images at high magnification sometimes leads to a "cult of magnification," which is a tendency to represent a specimen by a limited number of high-magnification images. The practical importance of Table 4.1 becomes clear when we consider the problem of surveying a specimen to determine its significant features. Note that in changing the magnification from 100× to 10,000×, the area sampled by the image decreases by four orders of magnitude. Although higher magnifications are often necessary to solve a problem, when only one or two such images are taken, the microscopist samples only a few square micrometers. This sampled area may represent only a few parts per million or parts per billion of the possible specimen area when the specimen has overall dimensions of square millimeters to square centimeters. Such a small sample may not be sufficient to obtain an adequate impression of the specimen characteristics. For survey work, a combination of both low-magnification and high-magnification imaging should be used, as shown in Fig. 4.6, and an adequate number of areas must be recorded to gain a valid description of the specimen.

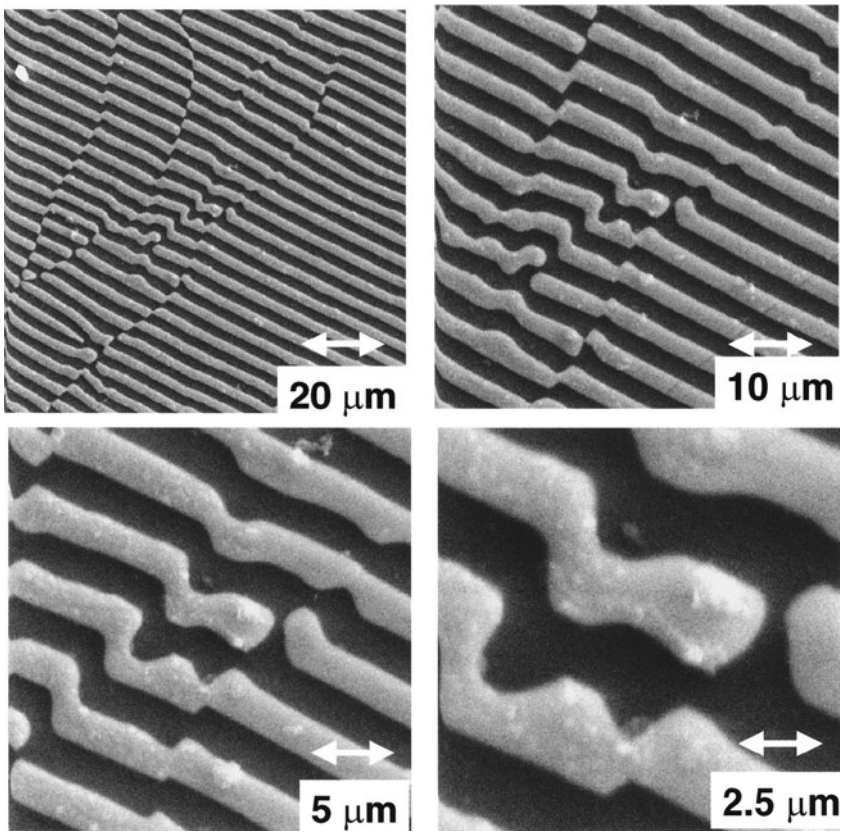


Figure 4.6. Magnification series of a polished cross section of aluminum–copper eutectic alloy, illustrating the rapid surveying capability of the SEM. The images are recorded at constant objective lens strength and working distance. Note that focus is maintained throughout the series, and there is no image rotation.

3. *Zoom capability:* Magnification in the SEM depends only on the excitation of the scan coils and not on the excitation of the objective lens, which determines the focus of the beam. Thus, once the objective lens is adjusted in strength to focus the image at high magnification, lower magnifications of the same region remain in focus as the scan strength is increased to scan a larger area. This "zoom magnification" feature is very useful for rapid surveying of the specimen, as shown for a polished Al-Cu eutectic alloy in Fig. 4.6.

4. *Lack of image rotation:* Figure 4.6 also demonstrates that for operation at a fixed working distance with constant objective lens strength, the image does not rotate as the magnification is changed. As discussed in Chapter 2, a relative rotation of the image will occur if the working distance, that is, the polepiece-to-specimen distance, is changed, and the objective lens strength must be altered to focus the beam at the new working distance. Image rotation with a change in working distance is illustrated in Fig. 4.7.

5. *Absolute value of the magnification:* Can the magnification printed on the SEM screen as a numerical value or the scale marker be trusted? If accurate measurements are to be made, the magnification should be verified by means of an external standard. Calibrated gratings with known spacings are available from several vendors of microscopy supplies. "Traceability" to a national standards organization is provided by materials such as Standard Reference Material (SRM) 484, available from the National Institute of Standards and Technology (NIST, 2002; complete information available on the World Wide Web at www.nist.gov). SRM 484 is a "stage micrometer" consisting of electrodeposited layers (nominal spacings 0.5, 1, 2, 5, 10, 30, 50 μm) of nickel and gold. This SRM permits image magnification calibration to an accuracy of 5% at the micrometer scale. The accuracy limitation arises because of the difficulty in exactly and consistently locating the edges of the calibration structures. Ideally these edges should be atomically flat, but the finite beam diameter and the effects of scattering broaden the apparent edge. In general, the error becomes greater as the spacing becomes smaller because the uncertainty in the two edge positions used to define a spacing becomes an increasingly greater fraction of the spacing as this value is reduced. Extension of the stage micrometer approach to smaller structures is the subject of on-going research, which has led to new dimensional reference standards such as NIST SRM 8091, created by means of semiconductor device lithography techniques (Vladar *et al.,* 2001).

4.2.4. Picture Element (Pixel) Size

The "picture element" ("pixel") is related to scanning action and magnification and is a concept critical to the interpretation of SEM images. The picture element can be understood with the aid of Fig. 4.3. As described above, the scan consists of a series of discrete locations to which the beam is addressed. The picture element is the size of the area *on the specimen* from which information is collected and transferred to the CRT image or computer memory. Usually the picture element is considered as a

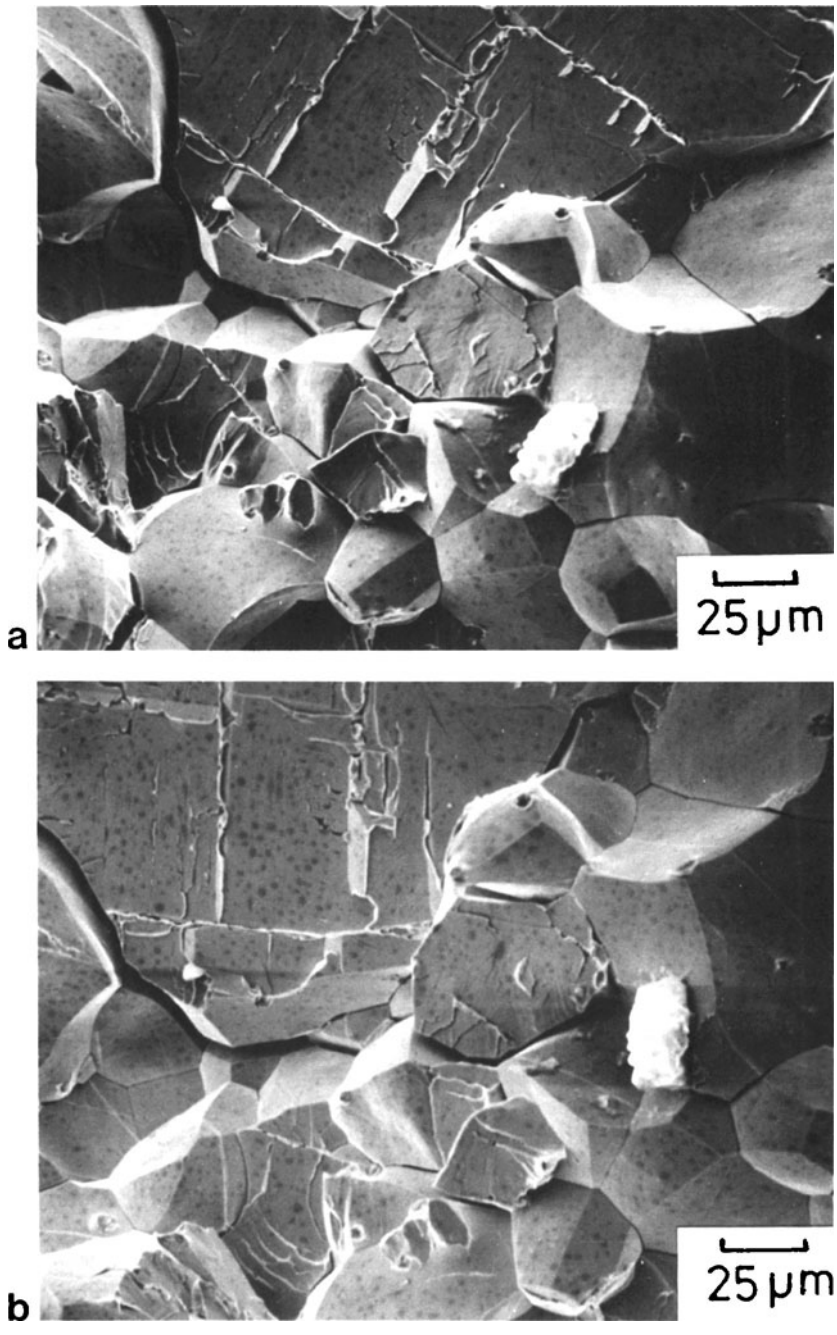


Figure 4.7. Images of a fracture surface at constant magnification, but with different working distances. The change in lens excitation causes a pronounced image rotation.

Table 4.2. Size of the Picture Element as a Function of Magnification.[a]

Magnification	Edge dimension of picture element
10×	10 μm
100×	1 μm
1,000×	0.1 μm (100 nm)
10,000×	0.01 μm (10 nm)
100,000×	0.001 μm (1 nm)

[a] Assumptions: 1000 pixel ×1000 pixel scan matrix, 10 cm ×10 cm display CRT.

circle or square, described by a linear measure of diameter or edge length. Considering the rectilinear scan to be divided into equal-sized square boxes filling all of the scan frame, the linear dimension D_{PE} of the picture element is given by the expression

$$D_{\mathrm{PE}} = L_{\mathrm{spec}}/N_{\mathrm{PE}}, \tag{4.2}$$

where L_{spec} is the length of the scan on the specimen from Eq. (4.1) and N_{PE} is the number of discrete locations (picture elements) along the scan line (often listed as the number of lines per frame on the SEM scan control module). Table 4.2 lists the length of the picture element edge as a function of magnification for the particular case of a high-resolution (1024 × 1024) image.

For a given choice of magnification, images are considered to be in sharpest focus if the signal that is measured when the beam is addressed to a given picture element comes only from that picture element. From Table 4.2, the picture element dimension rapidly decreases as the magnification increases. At high magnification, overlap of adjacent picture elements must eventually occur. To determine the magnification at which pixel overlap begins, both the incident probe diameter and the spreading (delocalization) of the signal in the interaction volume must be considered. To determine the limiting magnification for a given situation of beam size, energy, and specimen composition, we must calculate the signal-producing area from information on beam–specimen interactions provided in Chapter 3 and compare it to the pixel size.

How can the signal "leak out" to overlap adjacent picture elements? First, the beam diameter at the specimen plane is controlled by the strength of the final lens. Consider a flat specimen placed at a right angle to the optic axis (i.e., 0° tilt). As the lens strength is adjusted, the crossover moves along the optic axis (z = optic axis) and the effective beam size at the specimen is varied. On the CRT display, we observe this focusing operation as a maximization of fine detail when the beam is focused at the specimen plane containing the feature of interest. A second and frequently more important source of "signal leakage" occurs as a result of electron scattering and the finite size of the interaction volume from which the imaging signals emanate. How do we estimate the size of the signal-producing area from these two contributions? Consider the case of backscattered electrons. The dimension of the signal-producing area contains contributions from the

finite size of the beam as well as the projected surface area of backscatter emission due to electron scattering. Table 3.4B provides a description for the radius of the emission area (normalized to the Kanaya–Okayama range) that contains various percentages of the backscattered electrons. The diameter of the effective signal-producing area is found by adding in quadrature the beam diameter d_B and the diameter of the signal-producing area d_{BSE}:

$$d_{eff} = \left(d_B^2 + d_{BSE}^2\right)^{1/2}. \tag{4.3}$$

Consider a 10-keV beam focused to 50 nm which is incident on a specimen of aluminum and a specimen of gold; the specimens have flat surfaces set normal to the incident beam. From Table 3.4B, the diameter of the region emitting 90% of the backscattered electrons is $0.98R_{KO}$ or 1.3 μm (1300 nm) for aluminum and $0.50R_{KO}$ or 0.13 μm (130 nm) for gold at 10 keV. From Eq. (4.3), the total lateral spreading of the backscatter signal considering both the incident beam and the BSE spread due to the interaction volume is

For aluminum $\quad d_{eff} = 1300 \text{ nm} = 1.3\ \mu\text{m}$

For gold $\quad d_{eff} = 140 \text{ nm} = 0.14\ \mu\text{m}.$

(Note that because of the addition in quadrature and the large difference in the beam size and interaction volume for aluminum, the beam size makes no significant contribution to the spreading for aluminum. For the gold, the beam size makes only a small contribution and beam scattering is the controlling factor.) When these values of d_{eff} are compared to the pixel size from Table 4.2, we see that even at a low magnification of 100×, the adjacent picture elements are slightly overlapped for aluminum, whereas for gold an overlap situation is encountered above 700×. Thus, at surprisingly low magnifications, adjacent pixels begin to overlap. Figure 4.8 illustrates the effect of interaction volume on pixel size. The consequences of this situation will be considered in detail later.

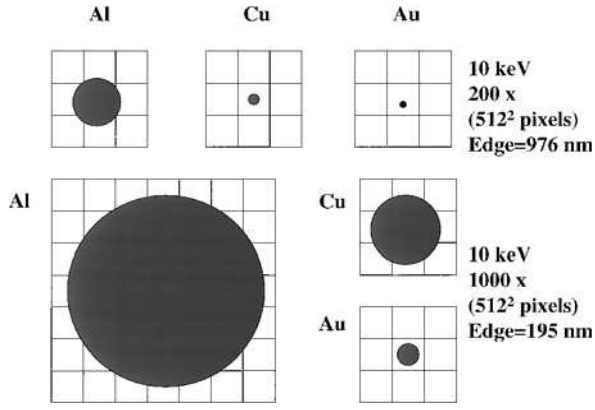


Figure 4.8. Beam-BSE footprint relative to pixel size for various materials at two different magnifications, 200× (pixel edge 976 nm) and 1000× (pixel edge 195 nm). Assumptions: 512 × 512 scan (10 × 10 cm CRT display); $E_0 = 10$ keV; incident beam size = 50 nm.

4.2.5. Low-Magnification Operation

Many images recorded with the SEM are prepared with modest magnifications, less than 1000×. For the FEG SEM, the focused beam is of the order of a few nanometers, so for the low-magnification range the beam is always much smaller than the pixel size. What is the proper strategy for choosing beam parameters for low-magnification operation with a thermionic gun SEM (tungsten or LaB_6 gun), where the microscopist has control over the beam current covering five orders of magnitude (~10 pA to 1 μA)? Should we, for example, choose to operate with a very finely focused beam in the 10-nm range or less, which is appropriate for high-magnification operation? We know from the discussions in Chapter 3 that for a thermionic source SEM, a severe penalty in beam current must be paid if we choose to focus a very small probe. In arguments to be developed later in this chapter (Section 4.5), we shall learn that the beam current is a critical parameter in establishing the visibility of an object: the more current in the beam, the lower is the contrast level that can be seen. From the exercise presented above in Section 4.2.4 for estimating the size of the effective signal sampling region, we found that, at least for the specific situation described, the beam size could be a small contributor compared to the overall spread of the BSE signal. We could therefore safely increase the size of the focused probe to 100 nm or more by weakening the first condenser lens, which has the direct effect of increasing the current in the beam approximately as the square of the beam diameter. Increasing the beam diameter from 10 nm to 100 nm would increase the beam current by a factor of 100 (an estimate in the absence of aberrations), while causing no increase in d_{eff} for aluminum and only increasing d_{eff} for gold from 140 nm to 170 nm. The key point in a strategy for low-magnification operation is therefore to use a large beam size that does not significantly degrade the spatial resolution, but does carry sufficient beam current (achieved with a weakly excited first condenser lens and a large final aperture) to produce images with a high signal quality.

4.2.6. Depth of Field (Focus)

The large depth of focus that can be obtained is one of the most striking aspects of SEM images. Depth of field, also referred to as "depth of focus," rather than high resolution is often the more valuable feature to users. How can we optimize the depth of field? The picture element concept can be employed to understand more fully the concept of depth of field and to develop an imaging strategy for depth-of-field microscopy. As explained in Chapter 2 and illustrated in Fig. 4.9, the action of focusing the beam to a crossover ("plane of optimum focus" in Fig. 4.9) inevitably produces angular divergence of the ray package that constitutes the focused beam. This divergence causes the beam to broaden above and below the plane of optimum focus. Consider a rough specimen such that there are features at various distances from the final lens, that is, the features have different values of the z coordinate. With the objective lens strength selected to bring the beam to a focus at the plane indicated, different beam sizes will strike

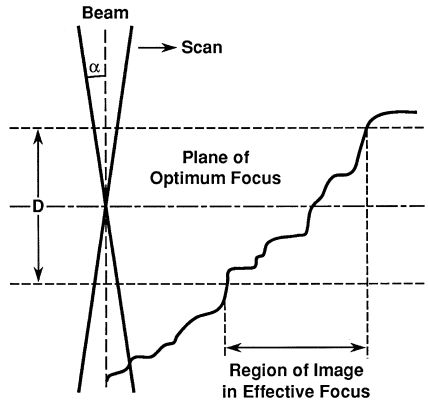


Figure 4.9. Schematic illustration of the depth of focus (field) in an SEM image. Any portion of the specimen that lies within the range of Z (along the optic axis) defined by the planes located at $\pm D/2$ from the plane of optimum focus will appear to be in focus.

specimen features, depending on their z position relative to the plane of optimum focus. For a sufficiently rough specimen, it is clear from Fig. 4.9 that the beam intersecting some features will become so large that those features of the specimen will not appear in focus. We must understand what factors control the depth of field so as to optimize the range of the image that appears in focus.

To calculate the depth of field, we must know at what distance above and below the plane of optimum focus the beam has broadened to a noticeable size. The depth-of-field criterion then depends on reaching a condition of overlapping adjacent pixels. The geometric argument in Fig. 4.9 indicates that, to a first approximation, the vertical distance $D/2$ required to broaden the beam of minimum radius r_0 (in the plane of optimum focus) to a radius r that causes noticeable defocusing is given by

$$\tan\alpha = r/(D/2), \tag{4.4a}$$

where α is the beam divergence, as defined by the semicone angle. For small angles in radian measure, $\alpha \approx \tan\alpha$, so that

$$\alpha \approx r/(D/2) \tag{4.4b}$$

$$D \approx 2r/\alpha. \tag{4.4c}$$

On a high-resolution CRT display where the spot size $= 0.1$ mm ($100\,\mu$m), most observers will find that defocusing becomes detectable when two pixels are fully overlapped, that is, when the condition of Fig. 4.9 and Eq. (4.4c) has been reached with $r = 1$ pixel. The pixel size referred to the specimen space is given by $0.1\text{ mm}/M$, where M is the magnification. Substituting $r = 0.1\text{ mm}/M$ into Eq. (4.4c) gives a practical expression for the depth of field:

$$D \approx 2(0.1\text{ mm}/M)/\alpha = 0.2\text{ mm}/\alpha M. \tag{4.4d}$$

Equation (4.4d) indicates that to increase the depth of focus D, the microscopist can choose to reduce either the magnification M or the divergence

Table 4.3. Depth of Field

Magnification	Image width (μm)[a]	Depth of Field (μm)[b]			
		$\alpha = 2$ mrad	$\alpha = 5$ mrad	$\alpha = 10$ mrad	$\alpha = 30$ mrad
10×	10,000	10,000	4000	2000	667
50×	2,000	2,000	800	400	133
100×	1,000	1,000	400	200	67
500×	200	200	80	40	13
1,000×	100	100	40	20	6.7
10,000×	10	10	4	2	0.67
100,000×	1	1	0.4	0.2	0.067

[a] For a 10-cm CRT display.

[b] $\alpha = 2$ mrad: $R_{AP} = 100\,\mu$m, $D_W = 25$ mm; $\alpha = 5$ mrad: $R_{AP} = 100\,\mu$m, $D_W = 10$ mm; d $= 10$ mrad: $R_{AP} = 200\,\mu$m, $D_W = 10$ mm; $\alpha = 30$ mrad: $R_{AP} = 600\,\mu$m, $D_W = 10$ mm.

α. Changing the magnification is not generally an option because the magnification is chosen to fill the image with the specimen features of interest. This leaves the divergence as the adjustable parameter that must be decreased to obtain greater depth of field. The divergence is adjusted by the selection of the final beam-defining aperture. For an aperture with radius R_{Ap} located at the scan crossover in the exit of the objective lens and working distance D_W we find

$$\alpha = R_{Ap}/D_W. \tag{4.5}$$

Decreasing R_{Ap} and increasing D_W both increase the depth of field. (In an SEM with a virtual aperture, the final divergence-defining aperture is located above the final lens. For similar depth of field, such a virtual aperture is smaller in radius than the equivalent real aperture in a position below the lens by a factor equal to the demagnification of the objective.)

A typical set of final aperture sizes, as specified by the diameter, might consist of 100-μm, 200-μm, and 600-μm apertures, whereas the typical working distance would be approximately 10 mm, with a possible increase to 50 mm or more in some instruments, depending on the sample stage motions. The depth of field calculated with Eqs. (4.4d) and (4.5) is given in Table 4.3 for several combinations of the operating parameters. Examination of Table 4.3 reveals that it is possible to choose operating parameters (R_{Ap}, smallest aperture, D_W, longest working distance: e.g., 100-μm aperture at 25-mm working distance) to achieve a condition in which all objects are in focus within a volume of space whose z extent along the optic axis is as great as the image width. (Note that at high magnifications, some of these aperture selections may be impractical due to degradation of the beam by aberrations when the optimum aperture size is not used.)

Figure 4.10 gives examples of the appearance of a rough object, a fracture surface, under different conditions of depth of field. The initial image pair (Figs. 4.10a and 4.10b) shows a choice of depth of field that is

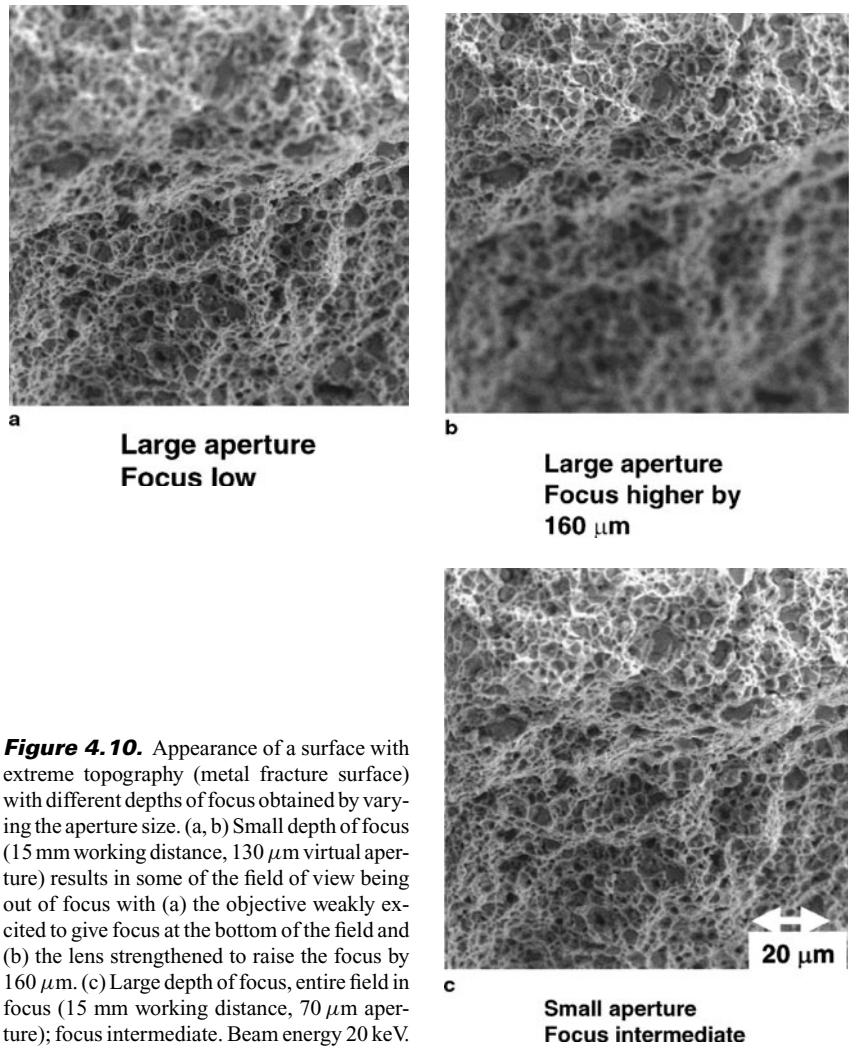


Figure 4.10. Appearance of a surface with extreme topography (metal fracture surface) with different depths of focus obtained by varying the aperture size. (a, b) Small depth of focus (15 mm working distance, 130 μm virtual aperture) results in some of the field of view being out of focus with (a) the objective weakly excited to give focus at the bottom of the field and (b) the lens strengthened to raise the focus by 160 μm. (c) Large depth of focus, entire field in focus (15 mm working distance, 70 μm aperture); focus intermediate. Beam energy 20 keV.

inadequate to bring the entire field of view into focus. Indeed, by changing the lens strength we can focus on the features near the lens (lens strong) or the features further away (lens weak), but both areas cannot be viewed in focus at the same time. Figure 4.10c shows the same field of view with a smaller divergence, bringing the entire field of view into focus.

Choosing the SEM operating parameters to optimize the depth of field is of special importance when stereomicroscopy is performed. Stereomicroscopy, discussed in detail in Chapter 5, permits visualization of the three-dimensional nature of an object. The perceived stereo effect is generally much stronger when the entire field of view of an object appears in focus. Conversely, when only a single image of a three-dimensional object

is viewed, it is sometimes difficult to judge the roughness if the image is prepared with high depth of focus. In Figures 4.10a and 4.10b, which were prepared with poor depth of focus, it is actually easier to judge which part of the specimen is nearest to the polepiece by adjusting the strength of the lens to determine the upper and the lower regions of the sample (along the optic axis), whereas in Fig. 4.10c, prepared with a high depth of focus, this information is lost.

4.2.7. Image Distortion

4.2.7.1. Projection Distortion: Gnomonic Projection

Consider the simplest case of SEM imaging, that of a flat object placed perpendicular to the optic axis. The image is constructed by a scan that diverges from a rocking point, as shown in Fig. 4.1. Are there any distortions inherent in such a scanning system? For a scan generated by rocking about a point, the SEM image is geometrically equivalent to a projection obtained by extending the beam vector to intersect a plane perpendicular to the optic axis of the instrument. The nature of this projection is shown schematically in Fig. 4.11. (Note that this drawing must not be construed to imply that an actual image exists in the SEM anywhere below the specimen; this is simply the equivalent geometric construction.) The scan can be considered to originate as a series of ray vectors with their point of origin in the final aperture. The geometric equivalent of the image, which is actually mapped on the CRT, is the intersection of these vectors extended to meet a plane perpendicular to the beam axis. The position of the scan S_x on the specimen is given by

$$S_x = \text{WD} \ \tan\phi, \tag{4.6}$$

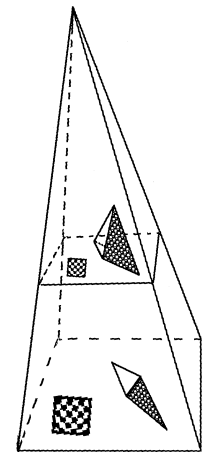


Figure 4.11. Construction of an SEM image as a projection onto a plane perpendicular to the beam axis. (Note: This is only the geometric equivalent of the imaging process; no such image actually exists anywhere within the SEM chamber.)

where WD is the working distance and ϕ is the angle of the scan from the central optic axis. At high magnifications, the total scan angle is small. For example, at 1000× for a 10-mm working distance, the semicone angle ϕ of the scan is 5 mrad (0.29°). For such small angles, $\phi \sim \tan \phi$, and for constant increments of ϕ the displacements S are linearly related, resulting in negligible distortion. To lower the magnification, the scan angle must be increased. At low magnifications, the angle of the scan cone can be substantial. For example, at 10×, the scanned field is 10 mm wide (for a 100-mm-wide display), and if the working distance is 10 mm, the semicone angle of the scan is 26.5°. For such large angles the relation between ϕ and $\tan \phi$ is no longer linear, resulting in a nonlinear "projection distortion." This type of projection is equivalent in map making to a gnomonic projection. The magnitude of the distortion of the gnomonic projection increases with distance from the optic axis (centerline) of the system. As the magnification is reduced and the scan excursion increases, the position of the scan deviates from linearity. For the 10×, 10-mm-working-distance example, the scan position at 25° from the optic axis is 6.5% greater than that predicted by linear extrapolation of the length of the scan at 5° from the optic axis. These nonlinear effects must be considered when accurate measurements are to be made.

At higher magnifications (>500×) the total scan cone angle is reduced to 0.5°or less, and the beam changes angle only slightly across the scan field. The distortion caused by the gnomonic projection can be neglected because the beam is effectively perpendicular to the projection plane at every point in the scan field and the projection is effectively parallel at all points.

For SEMs dedicated to critical dimension measurements and to electron beam fabrication, the scan system is modified to correct for gnomonic projection distortion as well as other secondary scan distortions.

4.2.7.2. Projection Distortion: Image Foreshortening

When we form an SEM image of a rough specimen, the resulting image construction is a two-dimensional representation of a three-dimensional object. Just as the creation of a two-dimensional map of a spherical world contains distortions, the SEM image of a rough surface must inevitably incorporate distortions.

Flat Object, Tilted. Tilting a planar object represents the simplest deviation from the flat object placed normal to the beam. When the specimen is a flat surface set normal to the beam and a square raster is scanned synchronously on the specimen and the CRT, the magnification will be identical along any direction in the image; the shape of objects in the field will be reproduced accurately in the SEM image, as depicted schematically in Fig. 4.5. An equivalent SEM image example of real objects is shown in Fig. 4.12b, which shows a copper grid with square openings and spherical polystyrene latex particles. As expected, the grid, which lies in the plane perpendicular to the beam, is seen with openings reproduced as squares.

The sphere is a three-dimensional object, which is projected onto the plane perpendicular to the beam as a circle.

Now consider the consequences of tilting this planar object. When the planar specimen is tilted around an axis parallel to the image horizontal, the projection situation changes, as illustrated in Fig. 4.12a. The projection of the scan line on the projection plane intercepts the expected length of specimen *parallel* to the tilt axis, but *perpendicular* to the tilt axis,

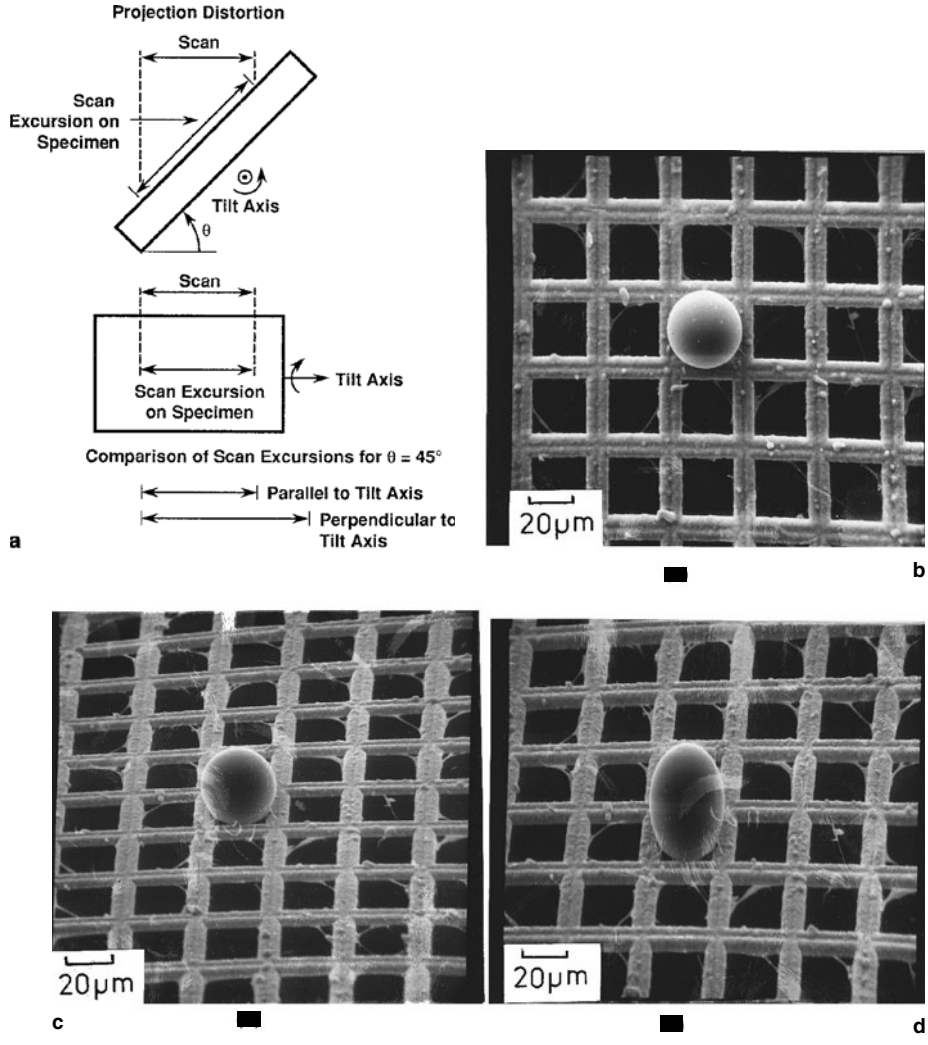


Figure 4.12. Schematic illustration of projection distortion (image foreshortening) that results from the tilt of a specimen for a tilt angle of 45°. The scan has the correct length parallel to the tilt axis (horizontal in the displayed image), but is lengthened perpendicular to the tilt axis (vertical in the displayed image), thus reducing the magnification. (b) Latex sphere on a copper grid, specimen plane normal to the beam (0° tilt). Note the square grid openings and the circular form of the sphere. (c) Grid–sphere specimen tilted to 55°; note the change in the shape of the grid openings, which are now rectangles, whereas the spherical particle still appears as a circle. (d) Grid–sphere specimen tilted to 55°, tilt correction applied, note the restoration of the square grid openings, but the extreme distortion of the sphere into an ovoid!

and a greater length of the specimen is projected into the same length of scan. The magnification is therefore *lower* perpendicular to the tilt axis (more distance on the specimen is mapped) compared to parallel to the tilt axis. This effect is referred to as "foreshortening." The length of the scan excursion perpendicular to the tilt axis $S_\perp$ in terms of the scan length parallel to the tilt axis $S_\parallel$ is given by

$$S_\perp = S_\parallel / \cos\theta = S_\parallel \sec\theta, \tag{4.7}$$

where θ is the tilt angle. Because $S_\perp$ is always greater than $S_\parallel$ for any angle $\theta > 0$, the magnification along a line perpendicular to the tilt axis will always be smaller than that along a line parallel to the tilt axis. This effect can be easily seen for the grid/spherical particle specimen, as shown in Fig. 4.12c, where the specimen is tilted to 55° around a tilt axis parallel to the horizontal grid bars. Note the change in the shape of the grid openings, which have become rectangles, with a shorter length (a consequence of lower magnification) for those grid bars that lie perpendicular to the tilt axis. Note that in this image, the spherical particle still appears as a circular feature without any apparent distortion.

Consider a specimen consisting of randomly tilted, intersecting flat surfaces or facets imaged at high magnification where the scan angle is small so that the rays are nearly parallel to the optic axis, as shown in Fig. 4.13. The projection of these surfaces to form the image will produce

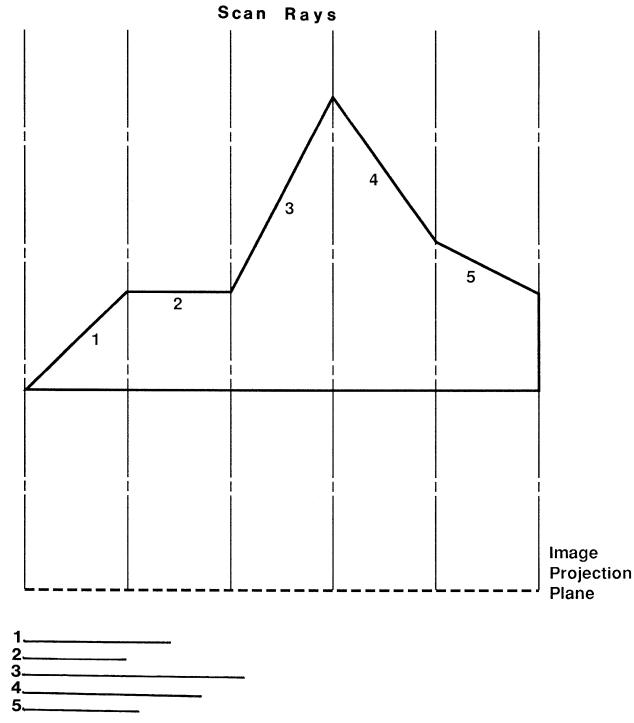


Figure 4.13. Projection distortion of facets with different inclinations to the beam. In this example, the magnification is sufficiently high that the scan angle is negligible and the scan rays are nearly parallel to the optic axis. Although the numbered facets have different lengths, their images in projection all have the same apparent length.

a final image in which the effective magnification will vary with the local tilt angle according to Eq. (4.7). The more highly tilted a facet is, the more foreshortened will be its image. Although each of the numbered facets in Fig. 4.13 is of different length, in the SEM image each will appear to have the *same* length.

Consider what happens when we attempt to use the "scale marker" digitally generated in the SEM image to measure features in the image. The scale bar is only accurate for a measurement between any two arbitrarily selected points within the facet that is normal (i.e., at equivalent 0° tilt) to the beam. For any arbitrarily tilted facet, accurate measurements can only be made when the points selected in the facet lie along a line perpendicular to the optic axis, but this line cannot be determined by simple inspection of the image. For any two points that form a line that tilts at an angle θ to the horizontal, the error will be proportional to the secant of θ. For example, if a facet is tilted 60° from the horizontal, then measurements of the length of features running on that surface will be *underestimated* by a factor as high as 2 (sec 60° = 2) compared to measurements within a facet at 0° tilt in the same image. Unfortunately, it is not possible to obtain accurate information on the local tilt from a single SEM image, so calculating a correction is not feasible. For such situations, the technique of stereomicroscopy, described in Chapter 5, must be applied to calculate true (x, y, z) coordinates for the points of interest to achieve accurate measurements.

Corrections for Tilted Flat Surfaces. Tilt Correction. If the tilt angle of a plane is accurately known, the effect of foreshortening can be calculated from Eq. (4.7) and a quantitative correction can be applied to the distance measured in the image. *Note that such a correction only applies for a flat specimen placed at a known tilt angle and only when the tilt axis is parallel to the image horizontal.* Alternatively, the correction can be applied directly to the image by electronically reducing the strength of the scan in the direction perpendicular to the tilt axis by the amount predicted by Eq. (4.7). The scan generator control for this operation is usually called "tilt correction." It must be emphasized that electronic tilt correction is only valid for the special class of specimens that have a flat planar surface tilted to a known angle. Tilt correction is an invalid operation to apply to any randomly shaped three-dimensional specimen, and applying the "tilt correction" by choosing some nominal tilt angle will in fact introduce unnecessary and misleading distortions into the image! For SEM imaging of all specimens except a flat, planar surface set at a known tilt angle, the tilt correction should be turned off or set to 0°.

To appreciate the magnitude of this "tilt correction" distortion, consider the images in Figs. 4.12c and 4.12d. In Fig. 4.12c, the specimen is tilted, as evidenced by the foreshortening of the square openings of the grid to form rectangles, whereas the projection of the three-dimensional sphere is still a circle, as expected. When tilt correction is applied, Fig. 4.12d, the grid openings are restored to their proper square shape because the grid is a flat object lying in the plane of tilt. After tilt correction has been

applied, however, the image of the sphere becomes highly distorted. The sphere is a three-dimensional object, the projection of which is a circle, but the normal projection of the SEM image has been disrupted by the modification of the scan strength in orthogonal directions through tilt correction, with the undesirable result that the sphere now looks like an egg. *Tilt correction is an invalid operation for any three-dimensional object*. As part of a personal quality assurance plan for microscopy, the microscopist should locate the tilt correction function on the SEM and learn how to turn it off! Checking the status of tilt correction is especially important in a multiuser environment because the previous user may have activated tilt correction.

Dynamic Focus Correction. An additional feature which is found on some SEMs is that of "dynamic focus," which should not be confused with "tilt correction." When a flat surface is highly tilted and is of such a large size that some portion of it extends out of the depth of field of the SEM image, dynamic focus can restore focus to these areas. Dynamic focusing is achieved by adjusting the strength of the final objective lens as a function of the scan position perpendicular to the tilt axis so as to bring the plane of optimum focus to coincide with the surface at all working distances in the image field. The lens is strengthened when the scan is at the top of the field and progressively weakened as the scan proceeds down the slope of the tilted object, thus keeping the specimen in optimum focus at all points in the image. Dynamic focus depends on satisfying a simple angular relationship between the scan position in the SEM image and the working distance. This relationship can be established only if the specimen is flat and tilted to a known angle with the tilt axis parallel to the horizontal scan line of the SEM image. *The same restrictions apply to dynamic focus as applied to tilt correction, namely, that it is only vald for a planar sample placed at a known tilt angle. It is invalid to apply this dynamic focus correction to the image of any arbitrary three-dimensional object.* Again, the microscopist must learn where the SEM control for "dynamic focus" is located and learn how to turn it off!

4.2.7.3. Scan Distortion: Pathological Defects

The SEM image may contain unexpected distortions due to pathological defects in the scan. Such distortions can arise from improper adjustment of the scan, typically because the two orthogonal scans are produced by different circuits which may not be adjusted to exactly the same value of deflection. The scan pattern may also become nonlinear, particularly at the edges of the image, even at high magnifications where the nonlinearity of the gnomonic projection is not significant. Scan distortions may also increase with the scan speed, particularly at television rates, due to hysteresis in the scan coils.

An example of distortion that can occur at the edge of the image field is illustrated in Fig. 4.14, which shows the change in the appearance of

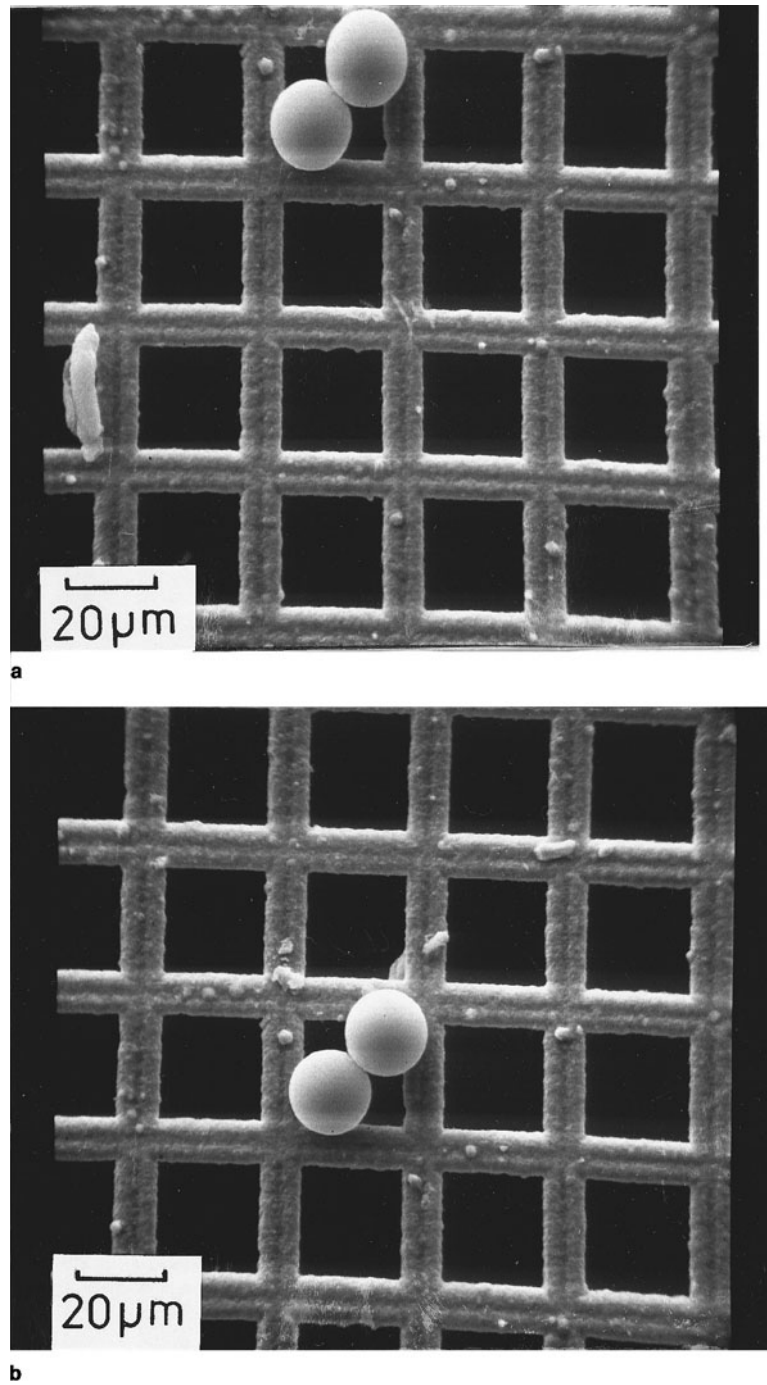


Figure 4.14. Illustration of image distortion that can occur at the edge of the image field. (a) Note the pronounced distortion of the image of the squares at the center of the field into rectangles near the top; a sphere located at the top of the field of view appears to have an elliptical shape. (b) When the stage is translated to move the spheres to the center of the field, the expected round shape is obtained, and the local grid openings adjacent to the spheres are squares again.

a spherical object as it is translated from the edge to the center of the field. Such scan distortions may not be noticed when images are prepared of objects that are irregular and lack any points of obvious reference. To determine the existence of scan distortions, it is necessary to make careful measurements on known regular objects. Carefully fabricated grids make excellent test specimens. Another good choice is polystyrene spheres. NIST Standard Reference Materials 1960 (9.89 μm) and 1961 (29.64 μm) consist of polystyrene spheres fabricated in microgravity in earth orbit aboard the NASA Space Shuttle (NIST 2002). These spheres exhibit an extremely high degree of sphericity and uniformity of diameter. As such, the spheres can serve as measures of both the absolute magnification and the uniformity of the scan field.

4.2.7.4. Moiré Effects

Although the recorded SEM image appears to be continuous to the human eye, it is composed of a grid of picture points with a spatial periodicity. We are thus always looking at the specimen through a grating, although the image appears continuous to the unaided eye. Viewing a typical specimen with irregular features through this grating introduces no anomalous effects. However, when the specimen itself has features with a periodic structure, such as a grid, the superposition of the periodicities of the specimen and the image can lead to the formation of moire fringes. Moiré fringes are the interference patterns between gratings of similar period. When the specimen has a periodic structure, moiré patterns may appear in SEM images and should not be confused with real features of the specimen. Figure 4.15 shows the appearance of moiré fringes in low-magnification images of a copper grid. Moiré fringes can be identified because the pattern changes rapidly with specimen rotation/translation relative to the fixed scan.

4.3. Detectors

4.3.1. Introduction

An appropriate detector must be employed to collect and convert the radiation of interest that leaves the specimen into an electrical signal for manipulation and display by the signal processing electronics to create an SEM image. Considering electron imaging, we know from Chapter 3 that backscattered electrons (BSE) and secondary electrons (SE) leave the specimen with drastically different properties in terms of energy, fraction relative to the beam, and directionality of emission. Each specific type of radiation potentially carries information on different aspects of the specimen characteristics. Moreover, this information can be contained in the number of BSE and SE, their energy distributions, and their emission directions, or a combination of all three factors.

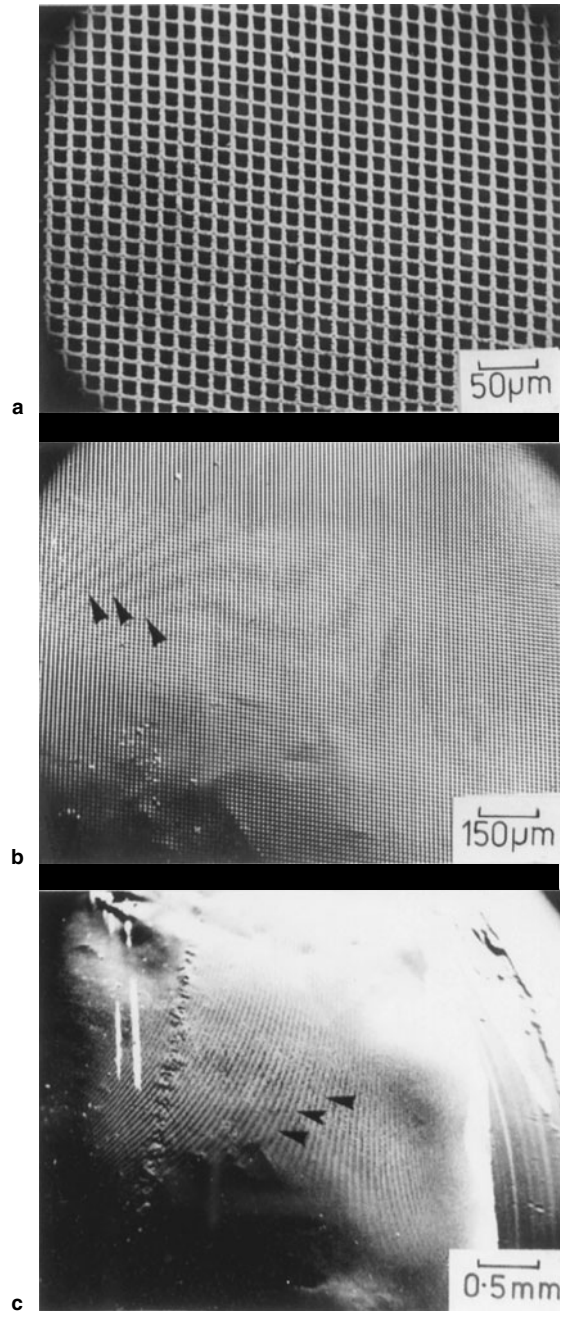


Figure 4.15. SEM images of a grid showing the appearance of moiré fringes as a function of magnification. The moiré fringes appear as the magnification is decreased and the scanning raster periodicity approaches the grid periodicity. The dark lines are moiré fringes and are artefacts; these features do not actually exist on the specimen. (a) Regular grid pattern at high magnification. (b, c) Lower magnifications showing moiré fringes (arrows). The fringes appear as wavy lines due to misalignment between the specimen pattern and the scanning pattern.

When considering a detector, we need to ask four questions. The answers to the first two of these questions are illustrated in Fig. 4.16:

1. Where is the detector relative to the beam and the specimen? The position of the detector is described by the take-off angle ψ, which is defined as the angle from the surface of the specimen to the line connecting a point on the specimen surface to the center of the detector face.

2. How large is the detector? The size of the detector is described by the solid angle Ω, which is the ratio of the area of the face of the detector to the square of the radial distance to the beam impact point, $\Omega = A/r^2$. The unit of measure of solid angle is the steradian (sr); a sphere represents a solid angle of 4π sr (i.e., $4\pi r^2/r^2$).

3. How efficient is the detector for converting radiation which strikes the collection area into a useful signal? Often, the detector response will change for different radiation energies, for example, the detector efficiency may not be uniform over the broad energy range of an incoming signal such as backscattered electrons.

4. What is the bandwidth of the detector/amplification system? Bandwidth refers to the range of signal frequencies that the detector system as a whole can process. Since the SEM image is scanned, the signal appears as a function of time, and the spatial detail in the image is converted into the domain of time. For example, as the beam is scanned across fine-scale features, the signal changes rapidly. Fine-scale spatial features are converted to high-frequency signals. In comparison, the general shading of coarse features varies much more slowly with the scan, so that these features are converted to low-frequency signals. A detector/amplification system typically has a high-frequency cutoff beyond which information is lost.

4.3.2. Electron Detectors

From Chapter 3, the electrons that escape the specimen under bombardment by keV electrons fall into two classes with widely differing properties:

1. Backscattered electrons are beam electrons which escape the specimen as a result of multiple elastic scattering and have an energy distribution $0 \leq E_{BSE} \leq E_0$, with the energy distribution peaked in the range (0.7–0.9) E_0 for intermediate- and high-atomic-number targets.

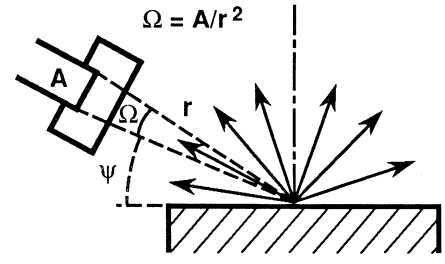


Figure 4.16. General characteristics of detectors. The position of the detector relative to the beam and the specimen is described by the take-off angle ψ measured relative to the local surface plane. The size of the detector is given by the solid angle Ω, which is equal to the area of the detector divided by the square of the radial distance from the detector face to the beam impact point on the specimen, $\Omega = A/r^2$.

2. Secondary electrons are specimen electrons given a small amount of kinetic energy by inelastic collisions with beam electrons and are emitted with energies in the range $0 \leq E_{SE} \leq 50$ eV, with a most probable energy of 3–5 eV, and 90% of the total between 0 and 10 eV.

These widely differing energy characteristics of BSE and SE present a considerable challenge to the design of a detector that can make use of both signals. Alternatively, the differences permit the design of detectors that are selective for one of the signals and thus selective for specific specimen properties.

4.3.2.1. Everhart–Thornley Detector

The electron detector most commonly used in scanning electron microscopy is the combined secondary/backscattered electron detector developed by Everhart and Thornley (1960), based upon a sequence of earlier developments by the Cambridge group under Sir Charles Oatley (Oatley, 1972). The Everhart–Thornley (E–T) detector is so popular that it is extremely rare to find a conventional SEM without one. The development of the E–T detector provided the first efficient use of the rich secondary/backscattered electron signal with a large solid angle of collection, high amplifier gain, low noise, and robust, low-maintenance performance. The rise of the SEM as a tool of broad use to the scientific community is due in considerable measure to the performance and utility of this detector. Because of its efficient collection of secondary electrons, the E–T detector is often mistakenly considered only a secondary electron detector. As the following discussion will demonstrate, backscattered electrons form an important and often dominant part of the E–T signal.

The E–T detector, illustrated in Fig. 4.17, operates in the following manner. When an energetic electron (~10 keV energy) strikes the scintillator material, light is emitted. (The scintillator may be a doped plastic or glass target, or a crystalline compound such as CaF_2 doped with europium; for a review of scintillators, see Pawley, 1974.) The light is conducted by total internal reflection in a light guide (a solid plastic or glass rod) to a

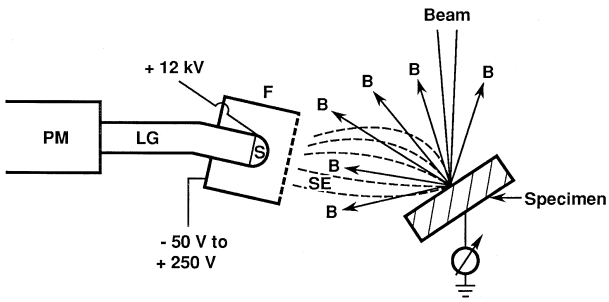


Figure 4.17. Schematic diagram of the Everhart–Thornley detector: B, backscattered electron trajectories; SE, secondary electron trajectories; F, Faraday cage (bias range −50 V to +250 V); S, scintillator with thin metallic coating; high bias (+12 kV) supply to the scintillator coating; LG, light guide; PM, photomultiplier.

photomultiplier. Because the signal is now in the form of light, the signal can pass through a quartz glass window, which forms a vacuum barrier, to the first electrode of a photomultiplier. At this photocathode, the photons are converted back into electrons, and the electrons are accelerated onto the successive electrodes of the photomultiplier, producing an ever-increasing cascade of electrons until the final collector is reached. The typical gain of the photomultiplier is 10^5–10^6, and is adjustable by selecting the voltage on the electrodes. The photomultiplication process provides high gain with little noise degradation and high bandwidth. The response is sufficiently fast to operate at television scan rates.

A large fraction of the backscattered electrons that originate from incident beams with energies from 10 to 30 keV carry sufficient energy to directly excite the scintillator, even in the absence of postspecimen acceleration. By applying a large positive potential (+10 kV to +12 kV) to a thin metal coating on the face of the scintillator, the low-energy secondary electrons (as well as lower energy BSE, such as those generated from low-energy incident beams) will be accelerated to a sufficient energy to generate light in the scintillator. To protect the beam from unwanted deflection and distortion by this large potential in the specimen chamber, the scintillator is surrounded by a Faraday cage, which is insulated from the scintillator bias. To collect the low-energy secondary electrons with higher geometric efficiency than simply collecting the fraction defined by the line-of-sight solid angle, a separate bias potential is applied to the Faraday cage, typically selectable in the range −50 V to +250 V. Note that this range from negative to positive provides the possibility of completely rejecting secondary electrons (−50 V) or efficiently collecting secondary electrons (+250 V). Let us consider the action of the detector for two bias conditions.

Everhart–Thornley Detector, Negative Bias. When the E–T detector is biased negatively, only backscattered electrons are detected. All secondary electrons are rejected, including those that are emitted from the specimen in the direction of the E–T detector within the line-of-sight solid angle for direct geometric collection. The take-off angle and solid angle of collection for the E–T detector for the direct collection of backscattered electrons are illustrated in Fig. 4.16. Those high-energy backscattered electrons that leave the specimen with trajectories directly toward the face of the scintillator are collected (that is, within the "line-of-sight" cone of solid angle); all other backscattered electrons emitted from the specimen are lost. For a specimen at 0° tilt (normal beam incidence), the E–T detector usually views the specimen from a take-off angle of approximately 30°. For a 1-cm-diameter scintillator placed at a distance of 4 cm from the beam impact point on the specimen, the solid angle of the E–T detector is 0.05 sr. As a fraction of the 2π-sr hemispherical space above a flat specimen, this solid angle gives a geometric efficiency of only 0.8%. The collection efficiency also depends on the angular distribution of the emitted signal. Consider the cosine angular distribution over which the backscattered electrons are emitted at normal beam incidence. The portion of the distribution that is collected at

different take-off angles is illustrated in Fig. 4.18a. Because the cosine angular distribution for 0° tilt favors BSE trajectories near the surface normal, when the E–T detector is placed at a low detector take-off angle, the fraction of the BSE collected is even less than that given by the simple geometric efficiency. The negatively biased E–T detector is thus a highly directional, asymmetrically placed, low-geometric-efficiency detector for BSEs.

The rigid nature of the light pipe of the E–T detector means that it occupies a fixed position in the specimen chamber. The take-off angle and collection situation thus also depend on the specimen tilt. When the specimen is highly tilted, the angular distribution of BSE becomes skewed in the forward direction, increasing the fraction collected by the E–T detector. Figure 4.18b shows how the collection situation changes with detector position. Comparing Figs. 4.18a and 4.18b, we see that the detector position which is most favored for collection at a 0° specimen tilt is least favored at a high specimen tilt.

The energy response of the scintillator of the E–T detector depends on the exact nature of the material used, but it generally increases with increasing electron energy up to a saturation point. There will also be a threshold energy (typically a few keV) below which the detector will produce no response from BSE unless postspecimen acceleration is applied.

Everhart–Thornley Detector, Positive Bias. The positively biased E–T detector behaves in a markedly different manner. The direct effect of the positive bias is to attract secondary electrons (and low-energy BSEs) to enter the Faraday cage for subsequent acceleration by the bias on the scintillator. In addition to those secondaries emitted from the specimen into the solid angle of collection of the E–T detector (identical to the solid angle for BSE), the action of the attractive positive bias is to cause a deviation of the trajectories of SE emitted from the specimen over a much wider range of solid angle into the detector, as shown in Fig. 4.19. Calculations of the trajectories of SE under the positive field of the E–T detector reveal that SE can be collected even if there is not direct line of sight from the specimen to the detector. From a flat surface, the collection efficiency can approach 100%.

The positive bias on the Faraday cage does not affect the directly collected high-energy BSE which are emitted from the specimen into the solid angle of the detector. These BSE are collected regardless of the cage bias. The positive bias does, however, have an unexpected, strong effect on the *indirect* collection of BSE, as illustrated in Fig. 4.20. The vast majority of BSE follow trajectories which miss the small direct collection solid angle of

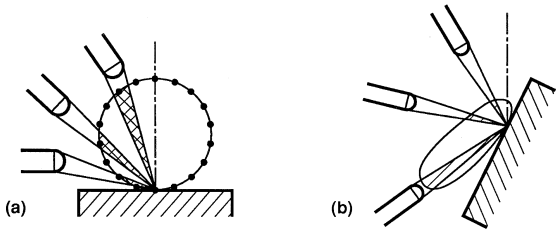


Figure 4.18. (a) Relative collection of backscattered electrons emitted in a cosine distribution by negatively biased E–T detectors placed at various take-off angles. (b) Relative collection of backscattered electrons emitted from a highly tilted surface.

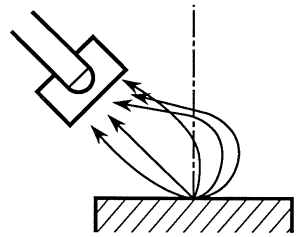


Figure 4.19. Schematic illustration of deflection of trajectories of secondary electrons of various energies by positive potential on the Faraday cage.

the E–T detector. The BSE following these trajectories strike the polepiece and specimen chamber walls, where they cause the emission of secondary electrons, designated as the SE_3 component. The BSE may actually bounce repeatedly in the chamber before losing all energy, and at each collision, more SE_3 are released. These SE_3 are collected with high efficiency by the positive bias on the E–T detector. Although apparently part of the SE signal, this BSE-to-SE conversion signal component actually represents the detection of the remote BSE signal. This apparent contradiction can be understood by considering that the magnitude of the SE_3 signal scales with changes in the BSE signal that leaves the specimen; for example, if the composition increases in average atomic number from one beam location to the next, the BSE signal will increase in all directions, and consequently, so will the SE_3 signal. Some idea of the efficiency of collection for these indirect BSE events is given by the measurements of Peters (1984). As listed in Table 4.4, for a material such as gold with a high backscatter coefficient, over 60% of the total SE signal of a positively biased E–T detector arises from indirect collection of the remote BSE signal component.

Although the negatively biased E–T detector is highly directional in its acceptance of BSE, the positively biased E–T detector accepts both this direct component as well as the indirect component, which is effectively

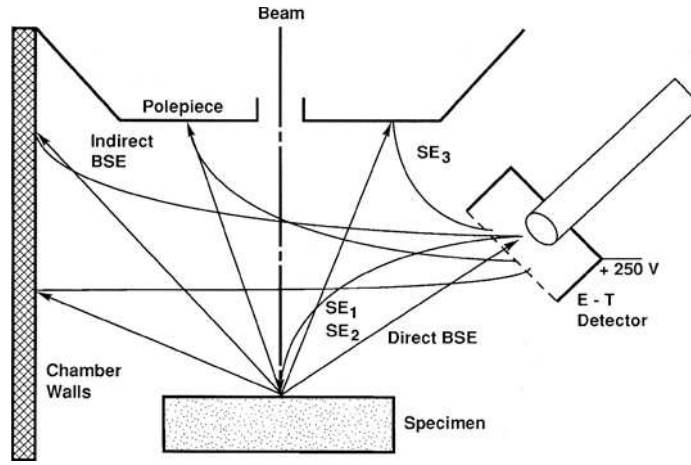


Figure 4.20. Schematic illustration of the indirect collection of backscattered electrons by a positively biased E–T detector. The BSE strike the chamber walls, where they create secondary electrons. These SE are collected by the E–T detector with high efficiency. Although nominally a contribution to the SE signal, these SE actually represent the BSE signal component because their number must rise and fall according to the behavior of the BSE.

Table 4.4. Collection of Secondary Electron Signals from a Gold Specimen with the Everhart–Thornley Detector[a]

Signal	Source	*Percentage* of total
SE_1	Beam-produced SE	9
SE_2	BSE-produced SE	28
SE_3	Chamber wall BSE-produced SE	61
SE_4	Beam-produced SE at final aperture	2

[a] From Peters (1984).

collected over a much larger solid angle approaching 2π sr. The positively biased E–T detector must thus be thought of as a *combined secondary and backscattered* electron detector. Because the BSE coefficient is typically larger than the SE coefficient in the conventional beam energy range ($\geq$10 keV), in many situations the BSE signal will be the most significant component of the total signal, especially for intermediate- and high-atomic-number specimens. Thus, the E–T detector is not really a "secondary electron detector," to which it is commonly referred in the literature, but nearly always has a strong BSE component in its signal. Recognition of this fact is critical to proper image interpretation.

Another advantage of the positively biased E–T detector is its performance at low beam energy. The positive Faraday cage bias results in efficient collection of direct SE and BSE as well as remote SE (equivalent to BSE). Because the potential applied to the scintillator can accelerate even low-energy SE to the energy needed to excite light emission from the scintillator, the positively biased E–T detector is an excellent detector for low-beam-energy microscopy. The E–T detector is capable of working with incident beam energies of 1 keV or even less.

4.3.2.2. ''Through-the-Lens'' (TTL) Detector

The high-performance field-emission-gun scanning electron microscope (FEG SEM) equipped with a "snorkel"-type lens described in Chapter 2 produces a strong objective lens magnetic field that is projected into the specimen chamber to reach the specimen plane. This markedly contrasts with the "pinhole" lens of the conventional SEM in which the lens magnetic field is contained within the bore of the lens so that the specimen resides in a field-free region. One major consequence of the strong magnetic field is to trap with high efficiency those SE emitted from the specimen, that is, the types SE_1 and SE_2, especially at short working distance, as shown in Fig. 4.21. The SE spiral up along the magnetic field lines and pass up through the lens bore. By placing a scintillator with a high bias (+10 kV) on the surface electrode above the objective lens, a "through-the-lens" (TTL) secondary electron detector is created (Fig. 4.21). The TTL detector virtually eliminates the SE_3 component produced remotely by the BSE, because the vast majority of these SE_3 are created well off the optic axis by the BSE colliding with the bottom of the lens and chamber walls. The TTL also discriminates against the BSE, because the magnetic field

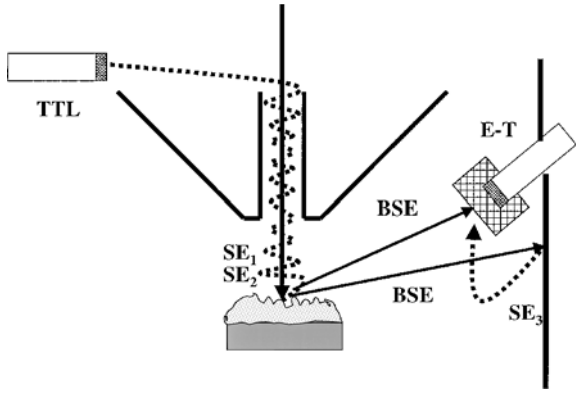


Figure 4.21. "Through-the-lens" (TTL) detector for SE used in high-performance field emission SEMs. The magnetic field of the objective lens projects into the sample chamber and the SE (types SE_1 and SE_2) are trapped by this field and follow spiral trajectories around the lines of magnetic flux up through the lens. A scintillator–light guide–photomultiplier detector above the lens collects the emerging SE and the high potential on the scintillator raises their kinetic energy sufficiently to cause light emission. Note that secondary electrons of type SE_3 are excluded by the TTL because they are formed too far off-axis to be collected by the on-axis magnetic field. The strong magnetic field of the snorkel lens captures most of the SE_1 and SE_2, greatly reducing the number available for capture by the more remote E–T detector. BSE and SE_3 collection by the E–T detector is relatively unaffected.

does not act to collect off-axis BSE, and those on-axis BSE that are emitted directly into the solid angle of the lens bore tend to travel in straight lines and are too energetic for collection by the potential on the scintillator above the lens. Although the TTL acts as a nearly pure SE detector, it must be noted that the SE_2 component of the SE signal collected by the TTL is actually controlled by the BSE leaving the specimen surface and thus the TTL is still sensitive to contrast effects arising from the BSE component. Nevertheless, the TTL signal has a significantly reduced BSE component compared to the positively biased Everhart–Thornley detector because of the nearly complete elimination of the SE_3 signal component.

4.3.2.3. Dedicated Backscattered Electron Detectors

There are usually 2–10 times more backscattered electrons emitted from the sample than secondary electrons. Backscattered electrons carry much useful information on specimen composition, topography, crystallinity, etc. Because of the large differences in energy between backscattered and secondary electrons, it is relatively easy to develop detectors that are only sensitive to high-energy backscattered electrons. Dedicated backscattered electron detectors are frequently included in the basic specification of an SEM.

The negatively biased E–T detector provides a signal which is composed exclusively of backscattered electrons, but it has a very low geometric efficiency, is asymmetrically placed on one side of the specimen, and is highly directional in its collection. Dedicated backscattered electron detectors are designed to greatly increase the solid angle of collection.

Passive Scintillator Backscattered Electron Detectors. Passive scintillator backscattered electron detectors operate on the principle that, with an incident beam energy of 10 keV or more, the majority of the backscattered electrons carry sufficient energy to excite the scintillator even in the absence of postspecimen acceleration. Without such "active" acceleration, the secondary electrons have no effect on the scintillator, so the signal from an unbiased or passive scintillator detector will consist only of contributions from backscattered electrons. The elimination of biasing also has the benefit that the detector potential will not disturb the beam, so the detector can be placed close to the specimen for more efficient collection. The scintillator still must be coated with a conducting layer (~10 nm of aluminum) to avoid charging. Wells (1974a) and Robinson (1975) described passive scintillator detectors with large collection angles. Figure 4.22a shows a large, passive scintillator placed in proximity to a tilted specimen; the detector take-off angle has been optimized to permit collection of forward-scattered electrons from a tilted surface (Wells, 1974a).

By making the scintillator of the same material as the light guide, designs that collect over much larger solid angles are possible (Robinson, 1975). Figure 4.22b shows a combined scintillator/light guide placed above the specimen. A hole drilled through the material permits access for the beam. Such a detector placed close to the specimen occupies nearly 2π sr of solid angle. For a specimen placed at normal beam incidence, only electrons scattered out of specimen at low angles to the surface are lost, as well as a small fraction at very high take-off angles near the incident beam which are scattered into the beam access hole. In this configuration, the detector surrounds the specimen nearly symmetrically, so that the signal is integrated in all directions, nearly eliminating sensitivity to trajectory effects. Note that

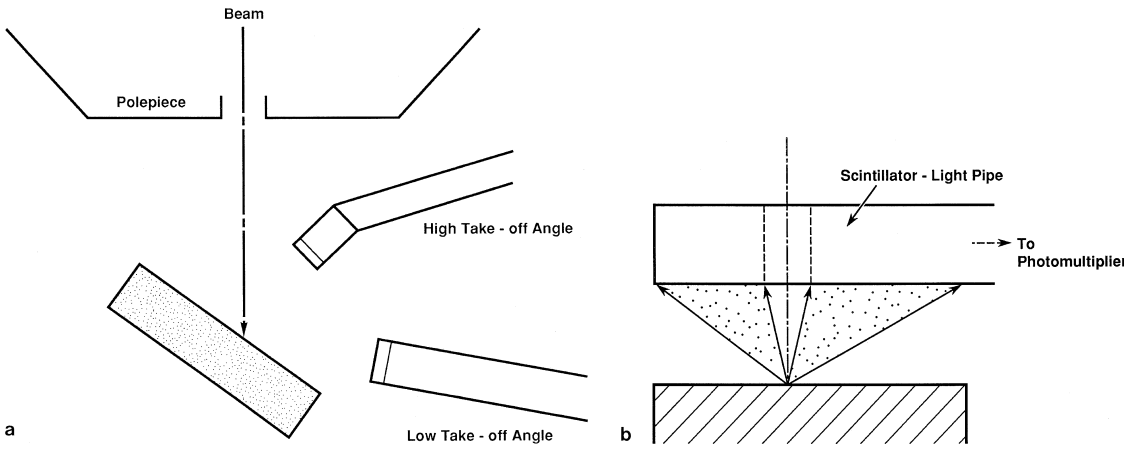


Figure 4.22. Scintillator backscattered electron detectors. (a) Large scintillator–light guide placed in proximity to the specimen, but asymmetrically (after Wells, 1974a). (b) Combined scintillator–light guide designed as an inverted bowl placed symmetrically above the specimen and collecting a solid angle of nearly 2π sr (Robinson, 1975).

if the specimen is tilted or has strong topography, backscattered electrons are directed away from such a detector placed above the specimen, the efficiency is significantly decreased, and sensitivity to trajectory effects develops.

BSE-to-SE Conversion Detector. The backscattered-to-secondary electron (BSE-to-SE) conversion detector operates on the same principle that provides the sensitivity of the positively biased E–T detector for indirect collection of backscattered electrons through collection of the remotely generated secondary electrons. The BSE-to-SE conversion detector (Moll *et al.,* 1978; Reimer and Volbert, 1979) seeks to optimize this conversion while excluding the secondary electrons produced at the surface of the specimen by the incoming beam electrons and the exiting backscattered electrons. As shown in Fig. 4.23, the separation of the two classes of secondaries is accomplished by surrounding the specimen with a grid biased to −50 V. Secondary electrons are repelled and cannot pass this grid, whereas the high-energy backscattered electrons pass through the grid openings and are not significantly affected by the grid potential. To improve the yield of secondary electrons produced by the backscattered electrons, the converter target can be covered with a thin layer of a material such as MgO, which has an extremely high secondary yield. The converter target is placed on the objective lens to intercept a large fraction of backscattered electrons. Once produced, these remote secondary electrons are then collected by the conventional E–T detector operated with a positive bias on the Faraday cage.

By collecting backscattered electrons striking over a large fraction of the polepiece and chamber walls, the BSE-to-SE conversion detector achieves a large solid angle of collection and greatly reduced sensitivity to trajectory effects. Because the yield of secondary electrons generally increases as the energy of the exciting electron decreases, the BSE-to-SE conversion detector is unique in that its gain per backscattered electron actually increases at low incident beam energy. The BSE-to-SE detector is therefore especially useful for low-voltage microscopy.

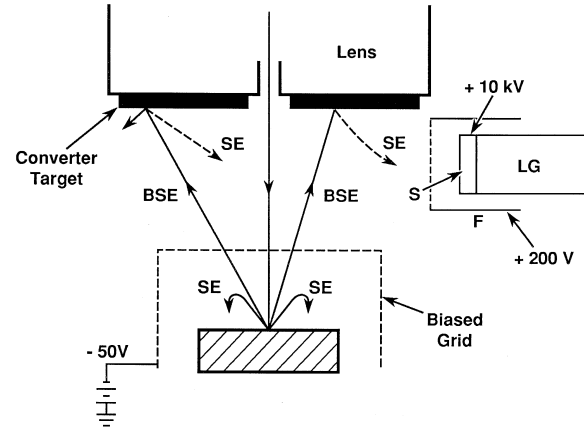


Figure 4.23. Schematic diagram of a backscatter electron-to-secondary electron (BSE-to-SE) conversion detector.

Solid State Diode Detector. The solid state diode detector operates on the principle of electron–hole production induced in a semiconductor by energetic electrons (Kimoto and Hashimoto, 1966; Gedcke *et al.,* 1978). The electronic structure of a semiconductor consists of an empty conduction band separated by a band gap of forbidden energy states from a filled valence band. When energetic backscattered electrons enter and scatter inelastically in the semiconductor, electrons are promoted to the conduction band, leaving an absence of an electron, or "hole," in the valence band. For silicon, the energy required to create an electron–hole pair (averaged over all energy loss processes) is about 3.6 eV, so when a single 10-keV backscattered electron strikes the detector, about 2800 electron–hole pairs are produced. If not separated, the electron and hole will recombine. As shown in Fig. 4.24, by applying an external bias across thin surface metal

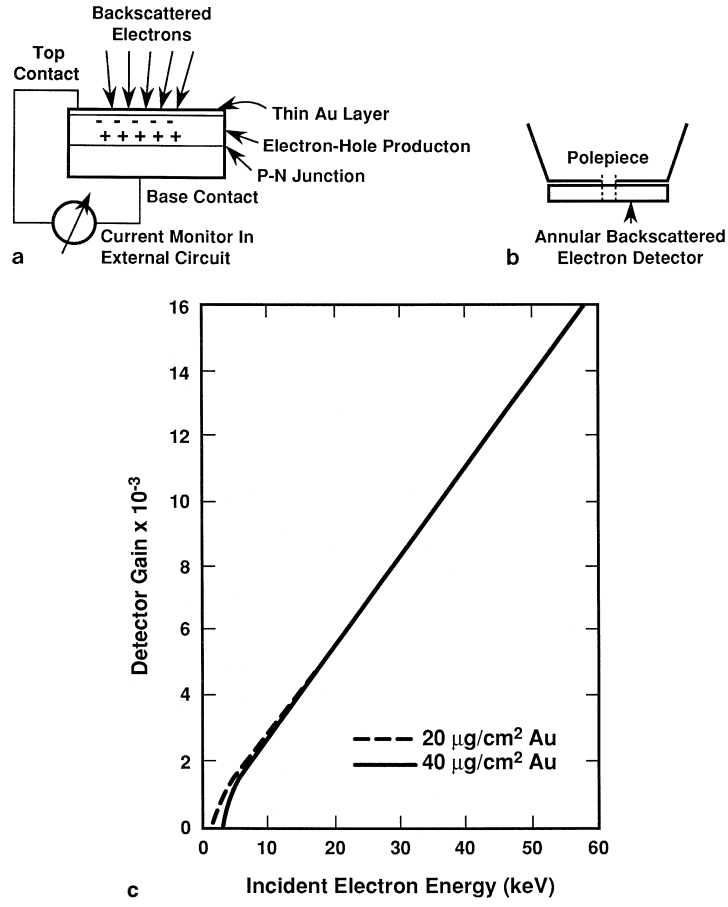


Figure 4.24. (a) Schematic diagram of a solid state backscattered electron showing separation of electron–hole pairs by an applied external potential. (b) Typical installation of a solid state detector on the polepiece of the objective lens. (c) Response of a solid state detector with respect to electron energy (from Gedcke *et al.,* 1978); note threshold, nonlinear region, and linear response region.

electrodes or by using the internal self-bias of certain electronic device structures, the free electron and the hole will move in opposite directions, separating the charges and preventing recombination. The charge collected on the external electrodes is used as the input to a current amplifier. The initial energy to electron–hole pair conversion gives a gain of about a factor of 1000, so the current amplification requirement is reduced.

A number of important features of solid state detectors can be noted:

1. The solid state detector has the form of a flat, thin wafer (typically several millimeters thick), which can be obtained in a variety of sizes, from small squares to large annular detectors. Annular detectors or arrays of smaller discrete detectors are typically placed on the polepiece of the objective lens, as shown as in Fig. 4.24.

2. The thinness of the detector permits it to be placed in close proximity to the specimen, which combined with the large area possible provides a large solid angle for high geometric efficiency.

3. Multiple detector arrays can be readily created by using separated solid state chips so that the discrete signals can be viewed separately or combined. The advantages of detector arrays will be described below.

4. The detector is sensitive to high-energy backscattered electrons only and not to secondary electrons. The detector is also affected by x-rays, but these are rare compared to backscattered electrons. Visible light can also affect the detector, such as cathodoluminescence from the specimen. Light from artificial sources, such as inspection lights or optical microscope illumination, must be eliminated to avoid interference with the solid state detector. Fluorescent room light entering through an SEM observation port window can produce interference with a 60-Hz period.

5. A typical energy response for a solid state detector is shown in Fig. 4.24c. (Such a response curve can be readily measured by placing the detector directly under the beam and using a Faraday cup to quantify the beam current to normalize the results.) Above a threshold, the response of the detector is linear with energy. The energy threshold is a result of the loss of energy suffered by the energetic electrons during penetration of the surface electrode ($\sim$10–20 nm) and silicon dead layer ($\sim$100–200 nm) to reach the active volume of the detector. Typically, the threshold is in the range 2–5 keV. As shown in Fig. 4.24c, the position of the threshold depends on the thickness of the front surface electrode. This threshold has both positive and negative impacts on performance. On the positive side, the loss of the low-energy backscattered electrons actually improves contrast and resolution in most situations because the low-energy backscattered electrons are likely to have undergone more elastic and inelastic scattering and are therefore likely to emerge far from the beam impact point and with little specific information on specimen characteristics. On the negative side, the threshold of several keV obviously eliminates the use of the solid state detector for low-voltage microscopy.

6. Through the mechanism of electron–hole production, the solid state detector acts to boost the signal by about three orders of magnitude prior to the current amplification stage. A current amplifier is required, preferably

of the operational amplifier type; such an amplifier can also be used to amplify the direct specimen current signal (Fiori *et al.,* 1974). For a detailed discussion of specimen current, see the Enhancements CD, Chapter 4.

7. Because of the capacitance of the solid state device, the bandwidth is relatively narrow, which compromises the performance at TV scan rates. High capacitance produces a long time constant, which has the effect that the detector signal does not accurately follow a large, rapid change in signal, such as occurs across a sharp edge, resulting in apparent blurring at TV rates. When the scan speed is reduced to a slow visual scan, the available detector bandwidth becomes adequate and the image quality improves. Photographic scan recording is sufficiently slow that no bandwidth limitations are encountered. "Premium" solid state detector designs can overcome this limitation, but even for these detectors, the high-speed performance is usually noticeably poorer when compared to the E–T detector.

Channel Plate Detector. The interest in low-voltage microscopy has stimulated the development of electron detectors that can operate with high performance for electron energies in the range of a few keV down to 100 eV (Joy, 1991). The channel plate electron multiplier detector, illustrated in Fig. 4.25, operates on the principle of backscattered-to-secondary electron conversion, but in this case the conversion occurs within the detector. The channel plate consists of at least one array of short (millimeter length) tilted glass capillaries with an accelerating potential applied at the exit face relative to a grounded entrance face. When an energetic electron strikes the wall at the entrance of a capillary, it liberates secondary electrons, which are accelerated down the capillary in response to the applied potential. Each capillary acts as a miniature electron multiplier, so that every time the electrons strike the walls, more secondary electrons are emitted, building into a cascade. To increase the gain, two channel plates can be stacked to produce a chevron. Because the secondary electron yield actually increases as the energy of the incoming electron decreases into the low-keV range, the channel plate detector is ideal for low-voltage operation.

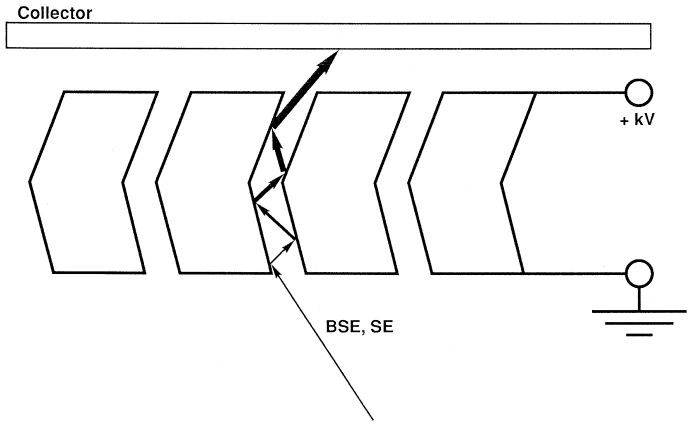


Figure 4.25. Principle of operation of a channel plate electron detector. Electrons that strike the walls at the entrance of a hole create secondary electrons, which are accelerated by the potential across the plate, producing a cascade of current.

Note that without postspecimen acceleration, the channel plate detector is not sensitive to secondary electrons emitted from the specimen because they will not have enough energy to initiate the secondary emission at the detector entrance.

4.4. The Roles of the Specimen and Detector in Contrast Formation

An SEM image is created by recording as an intensity level on a CRT/film/computer memory the product(s) of the interaction of the electron beam with the specimen at a series of points selected by the location of the beam. The task of image interpretation is to relate such gray-level images to the characteristics of the specimen so as to derive useful information. Two critical concepts form the basis of image interpretation: (1) the origin of contrast mechanisms and (2) the relationship of the size of the beam interaction volume/signal-emitting region to the picture element size in the recording. The understanding of contrast formation in the SEM has developed progressively throughout the history of the field and is the subject of a large literature (e.g., Oatley *et al.,* 1965; Oatley, 1972; Wells, 1974a, b; Goldstein *et al.,* 1975; Peters, 1982, 1984; Joy, 1991).

The following discussion considers the interpretation of image contrast for the case of low-magnification imaging, $M < 1000\times$. For these low magnifications, the beam size and the spread of the signal due to the finite size of the interaction volume are such that the signal for each pixel, is generated only from that pixel, so neighboring pixels are essentially independent. The overlap of signals emitted from adjacent pixels is a minimum. We should recognize that the selection of $M < 1000\times$ for the low-magnification regime of imaging is a "gray" number. Depending on the composition of the specimen, the spread of the beam can vary by more than a factor of five. Moreover, the shape of the surface distribution function is also important. Higher atomic number targets produce more sharply peaked distributions of backscattered electrons. Many of the tools for image interpretation discussed below will remain valid into the intermediate magnification range ($1000\times$ to $10{,}000\times$), depending on the exact circumstances of specimen composition and topography and the beam energy.

4.4.1. Contrast

Contrast is defined according to the equation

$$C = (S_2 - S_1)/S_2, \qquad S_2 > S_1, \tag{4.8}$$

where S_1 and S_2 represent the signals detected at any two arbitrarily chosen points of interest in the scan raster that defines the imaged field. Generally we are interested in a feature and its immediate surroundings, so that S_2 would be selected within the feature and S_1 in the surrounding background. By this definition, C is always positive and is restricted to the range

$0 \leq C \leq 1$. The case $C = 0$ is the trivial case for which the signal is the same (except for statistical fluctuations) at the two points, whereas $C = 1$ is the case, for example, in which an object is suspended over a hole from which no signal emerges, so that $S_1 = 0$.

The concept of contrast and its numerical meaning constitute one of the important basic factors in scanning electron microscopy. Contrast is a measure of the real information in the signal. Contrast is related to the properties of the specimen (shape, composition, etc.) that we wish to determine. As discussed in detail later in this chapter, establishing the *visibility* of features is a critical issue in properly operating the SEM. The criterion for visibility will be shown to be tied directly to the numerical value of the contrast.

The signals in Eq. (4.8) are those that arrive at the detector prior to any amplification in the signal processing chain that eventually produces the displayed/recorded image. In discussing contrast, we must consider the specimen and detector as a complete system. Thus, the contrast carried in the signal leaving the specimen is a result of events within the specimen (e.g., scattering from different kinds or densities of atoms) or in its immediate vicinity (e.g., by electric or magnetic fields just above its surface). This contrast can be subsequently modified by the response of the detector; for example, the detector may be more sensitive to high-energy BSE than to low-energy BSE. The signal reaching the detector represents the maximum amount of information available for the particular set of operating conditions employed. Subsequent amplification and signal processing, described in following sections, can only serve to control the way in which the information is displayed. The information in the signal cannot be increased after the detector. All subsequent processing, although useful for improving the match of the image display to the characteristics of the observer's visual system, cannot recover information that is not already present in the signal as it leaves the detector.

Contrast can be influenced by a complex mix of the characteristics of the beam – specimen interaction, the properties of the specimen, the nature of the signal carriers, and the position, size, and response of the detector. We can identify three different ways that contrast can be measured (*contrast components*):

1. *Number component.* The number component refers to contrast that arises simply as a result of different numbers of electrons leaving the specimen at different beam locations in response to changes in the specimen characteristics at those locations.
2. *Trajectory component.* The trajectory component refers to contrast effects resulting from the paths the electrons travel after leaving the specimen.
3. *Energy component.* The energy component arises when the contrast is carried by a certain portion of the backscattered electron energy distribution. Typically, the high-energy backscattered electrons are the most useful for imaging contrast mechanisms.

4.4.2.1. Introduction

The simplest contrast mechanism, compositional contrast, arises from differences in the local chemical composition within an object under study. In analogy to our experience with light, this mechanism conveys the type of information we normally obtain optically with color, that is, differences in color are often the result of differences in composition. Compositional contrast is principally a BSE number contrast mechanism, with a component of energy dependence.

4.4.2.2. Compositional Contrast with Backscattered Electrons

To eliminate the complexity that arises when a specimen has both compositional differences and an irregular surface (topography), we shall begin by considering a specimen, as shown in Fig. 4.26a, that consists of a flat, smooth block of an amorphous (i.e., noncrystalline) pure element. This flat specimen is sufficiently thick compared to the electron range so as to be effectively infinite in thickness to the beam electrons for the incident energy. In forming an SEM image of this object, a sufficiently high magnification is chosen so that the area scanned is much smaller than the lateral area of the specimen, ensuring that the beam and interaction volume never approach the edge of the specimen. Under such conditions, the backscattered, secondary, and absorbed electron signals emitted from all beam locations in the scan will not change, except for statistical fluctuations (an extremely important factor to be considered subsequently under image quality). From Eq. (4.8), there is no contrast formed, and therefore no information is obtained about the specimen. Now consider a slightly more complicated specimen formed by placing two such amorphous regions adjacent, but with differing compositions and atomic numbers, and separated by an atomically sharp interface, as shown in Fig. 4.26b. Why might the signals differ from the two regions? If the atomic numbers are such that $Z_2 > Z_1$, then based upon our knowledge of electron–beam specimen interactions, we can predict with a high degree of confidence that the behavior of backscattering is such that $\eta_2 > \eta_1$. The secondary

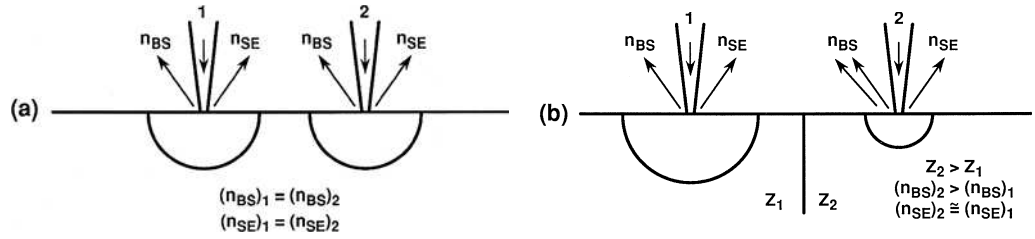


Figure 4.26. Schematic illustration of electron currents emitted from a sample which is (a) the same composition at two beam locations and (b) different in composition. n_{BS} is the number of backscattered electrons and n_{SE} is the number of secondary electrons.

electron coefficients δ_2 and δ_1 from the two regions may also differ, but in a less predictable manner. Assume that, to at least a first approximation, the secondary coefficients are similar in the two materials. Considering then only the backscattered electrons collected for beam locations sufficiently far from the interface so that the interaction volume is not affected by the interface region, contrast should exist between the two regions. This contrast mechanism arising from the specimen composition influence on backscattering is referred to as "atomic number," "compositional," or "material" contrast. An example of this contrast mechanism is shown in Fig. 4.27, which depicts a multiphase aluminum–nickel alloy, with the BSE detected with (1) the negatively biased Everhart–Thornley detector and (2) a large solid state detector. The major features are similar in both images, with three regions recognizable from their distinctly different gray levels. There are some differences in the fine details of the images, which will be considered further under topographic contrast.

Atomic number contrast has the following characteristics:

1. Because of the nearly monotonic increase of η with atomic number Z (or weight-fraction-averaged Z for compound targets), regions of high average atomic number will appear bright relative to regions of low atomic number. Thus, the increasing gray levels in Fig. 4.27 can thus be interpreted as regions of different composition with increasing average atomic number as the image gets brighter.

2. The magnitude of atomic number contrast can be predicted. If we make use of a detector that is sensitive to the number of backscattered electrons, for example, a negatively biased Everhart–Thornley detector, a solid state diode, or a passive scintillator, then the signal from the detector will be proportional to the backscatter coefficient, neglecting any energy response characteristics of the detector:

$$S_{\text{detector}} = \varepsilon_{\text{BS}}\, \eta, \tag{4.9}$$

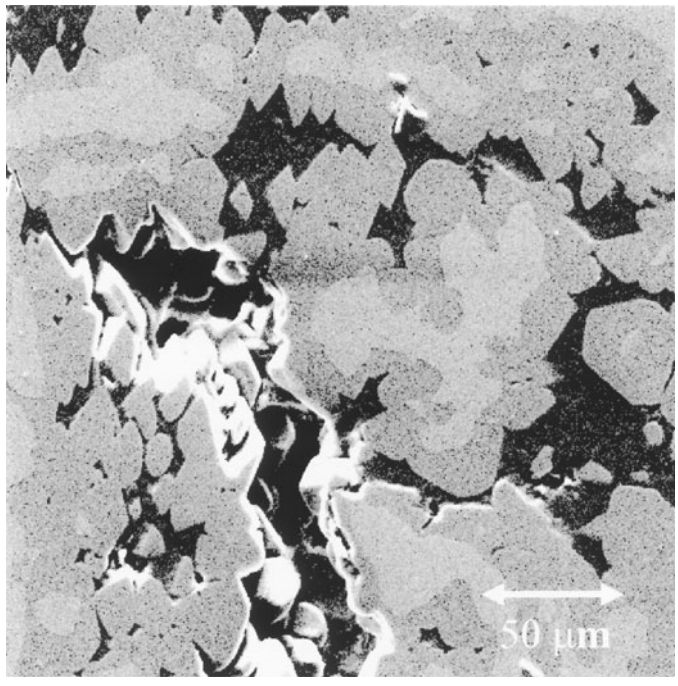
where ε_{BS} is the efficiency with which the backscattered electrons are detected. The efficiency involves both the geometric effects, that is, the angular size of the detector, and the energy response of the detector. From Eq. (4.9), the contrast can be calculated as

$$C_Z = (S_2 - S_1)/S_2 = (\varepsilon_{\text{BS}}\, \eta_2 - \varepsilon_{\text{BS}}\, \eta_1)/\varepsilon_{\text{BS}}\, \eta_2. \tag{4.10a}$$

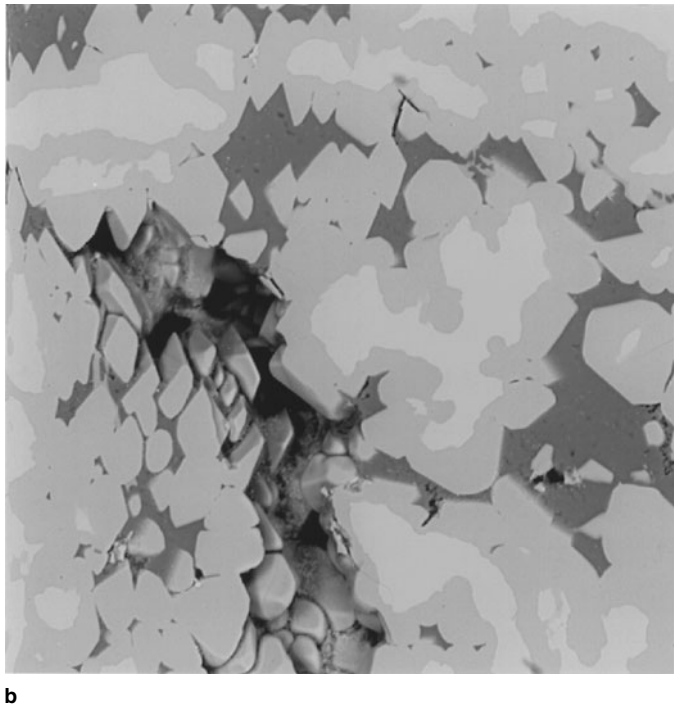
If ε_{BS} can be considered constant for the two different compositions (a reasonable approximation for elements of similar atomic number; when widely different atomic numbers are compared, the BSE energy distributions will be quite different, so ε_{BS} will vary), then Eq. (4.10a) simplifies to

$$C_Z = (\eta_2 - \eta_1)/\eta_2. \tag{4.10b}$$

Equation (4.10b) is useful for estimating atomic number contrast, but the result should only be thought of as an approximate or "gray" number. For a more accurate calculation of C_Z, the detector response per electron ε_{BS} must be known as a function of atomic number to use the more rigorous



a



b

Figure 4.27. Atomic number (compositional contrast) observed in an aluminum–nickel alloy (Raney nickel). (a) BSE image from an Everhart–Thornley detector, negatively biased; note strong contrast from topography associated with the cavity. (b) BSE image from a solid state detector (signal from both detector halves added); note improved visibility of the alloy phases.

equation (4.10a). Most BSE detectors have the property that their efficiency depends on the energy of the BSE. For example, a solid state diode has a response which, above a threshold, increases proportionally to the energy of the backscattered electrons. If the two different materials produce different BSE energy distributions, then ε_{BS} will not be a constant. Thus, if element 2 has a high atomic number, for example, gold, and element 1 has a low atomic number, for example, aluminum, the greater fraction of high-energy backscattered electrons from the high-atomic-number target has the effect of making e_{BS} greater for gold than for aluminum. Generally, ε_{BS} is an increasing function of Z. For an accurate calculation of C_Z, the energy distribution effect on the contrast can be calculated by convolving the energy distribution from Fig. 3.13 with the energy response of the backscatter detector, Fig. 4.24c. In an extreme case, for example, gold versus carbon, the contrast may be increased by as much as a factor of two by the energy dependence of the detector compared to the simple value predicted by Eq. (4.10b).

Equations (4.8)–(4.10) introduce the important concept of a *contrast calculation*. To make a contrast calculation, we require a mathematical description, or else systematic experimental data, that shows how some property of the specimen, for example, composition, systematically modifies a measurable signal such as backscattering. A contrast calculation is the first step to placing scanning electron microscopy on a quantitative basis and to developing an *imaging strategy*.

3. The contrast calculated from Eq. (4.10) and an expression fitted to η versus Z is listed in Table 4.5 for several combinations of elemental pairs. Elements separated by one unit of atomic number produce low contrast, for example, Al and Si yield a contrast of only 0.067 (6.7%). For elemental pairs widely separated in atomic number the contrast is much larger, for example, Al and Au produce a contrast of 0.69 (69%).

4. When pairs of elements separated by one unit of atomic number are considered, the predicted contrast decreases as the atomic number increases, as shown in Table 4.5 and Fig. 4.28. This behavior occurs because the slope of the η versus Z curve decreases with increasing Z. The decrease in contrast with Z is quite pronounced, dropping from 0.14 (14%) for B–C,

Table 4.5. Atomic Number Contrast ($E_0 = 20$ keV) between Pure Elements

Z_A	Z_B	η_A	η_B	C
13 (Al)	14 (Si)	0.153	0.164	0.067
13 (Al)	26 (Fe)	0.153	0.279	0.451
13 (Al)	79 (Au)	0.153	0.487	0.686
Adjacent element pairs				
5 (B)	6 (C)	0.055	0.067	0.14
13 (Al)	14 (Si)	0.153	0.164	0.067
26 (Fe)	27 (Co)	0.279	0.287	0.028
41 (Nb)	42 (Mo)	0.375	0.379	0.013
57 (La)	58 (Ce)	0.436	0.439	0.0068
78 (Pt)	79 (Au)	0.485	0.487	0.0041

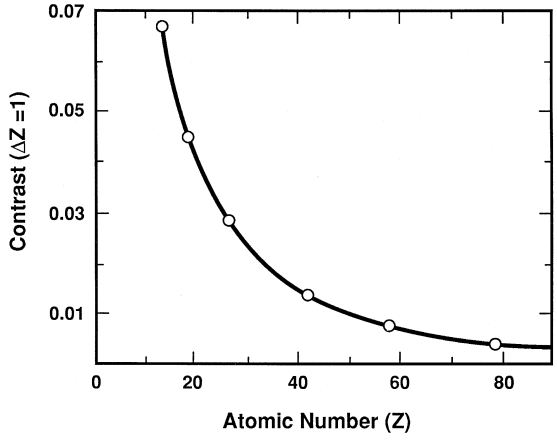


Figure 4.28. Atomic number contrast from adjacent pairs of elements ($\Delta Z = 1$) as predicted by a fit to the η versus Z curve.

to 0.067 (6.7%) for Al–Si, to 0.028 (2.8%) for Fe–Co, to 0.0041 (0.41%) for Au–Pt. As discussed later in this chapter, contrast above 10% is relatively easy to image in the SEM, whereas imaging contrast in the range 1–10% requires a careful strategy, and a contrast of $<1\%$ requires extreme measures for successful imaging. As noted in Chapter 3, the curve of η versus Z is in fact not perfectly monotonic, and in some cases, the backscatter coefficient may actually decrease with a unit increase in Z, resulting in an unexpected reversal in the atomic number contrast.

5. Specimen tilt influences atomic number contrast. As shown in Fig. 3.10b, the backscatter coefficients from highly tilted specimens tend toward the same value of unity at grazing incidence, so the atomic number contrast between any two elements decreases as tilt increases.

6. The directionality of backscattering also has an influence on atomic number contrast. For normal beam incidence, backscattering follows a cosine distribution, so the most favorable detector placement to maximize atomic number contrast is at a high take-off angle directly above the specimen. A backscatter detector placed at a low take-off angle will intercept a smaller fraction of the backscattered electrons; these electrons tend to be those that have lost the most energy and are less sensitive to atomic number effects. At high tilt angles backscattering is strongly peaked in the forward direction, so that the most favorable detector location for atomic number contrast is below the beam impact point on the specimen. For a highly tilted specimen, the usual location of the solid state backscattered electron detector on the polepiece of the final lens is not favorable for atomic number contrast.

4.4.3. Topographic Contrast

Probably the most frequent application of the SEM is the visualization of the topography (shape, size, surface texture) of three-dimensional

objects. Topographic contrast has a complex origin with strong number and trajectory components of both BSE and SE.

4.4.3.1. Origins of Topographic Contrast

Topographic contrast includes all of those effects by which the morphology (shape) of the specimen can be imaged. As discussed in detail below and in Chapter 3, topographic contrast arises because the number and trajectories of backscattered electrons and the number of secondary electrons depend on the angle of incidence between the beam and the specimen surface. Except for the lowest magnifications ($<100\times$; at $100\times$ the scan cone angle is only 5.6° across the full 1-mm scan field), the incident beam can be thought of as effectively parallel over the scanned field. The angle of incidence between the beam and the local surface will vary only because of the local inclination of the specimen. At each point the beam strikes, the numbers of backscattered and secondary electrons measured by the Everhart–Thornley detector give direct information on the inclination of the local specimen surface and its orientation to the E–T detector. Surprisingly, any observer, including one without special training or even a child, can interpret such topographic images and understand the true shape of the object without any prior knowledge of electron interactions or detector characteristics. In this section we will examine the topographic image formation process in detail to understand this convenient, but unexpected property of SEM images.

From the information presented in Chapter 3, the following effects can be expected to contribute to the formation of topographic contrast:

1. The backscatter coefficient increases as a monotonic function of the specimen tilt. The more highly inclined the local surface is to the incident beam, the higher is the backscatter coefficient. This effect produces a *number component* contribution to topographic contrast in backscattered electrons.

2. The angular distribution of backscattered electrons is strongly dependent on the local surface tilt. At normal beam incidence (local tilt angle $\theta = 0^\circ$), the angular distribution approximately follows a cosine function relative to the local surface normal; this cosine distribution is rotationally symmetric around the local surface normal and the incident beam. As the local tilt angle increases, the angular distribution gradually deviates from the cosine law and becomes distorted with the peak in the distribution occurring away from the incident beam. For $\theta > 50^\circ$ the angular distribution becomes highly asymmetric, with the peak in the forward scattering direction. The maximum likelihood for scattering is in a plane defined by the beam vector and the local surface normal vector. This directionality of backscattering from tilted surfaces contributes a *trajectory component* to the backscattered electron signal.

3. The secondary electron coefficient varies with the specimen tilt angle in a monotonic fashion, varying approximately as a secant function. Tilted surfaces thus yield more secondary electrons than a surface normal to the beam. This effect introduces a *number component* to topographic contrast in the secondary electron signal. The angular distribution of secondary

electron emission does not vary significantly with tilt angle, so trajectory effects are small.

The topographic contrast that is actually observed depends on the exact mix of backscattered and secondary electrons detected, which in turn depends on the detector used and its placement relative to the specimen.

4.4.3.2. Topographic Contrast with the Everhart–Thornley Detector

The Everhart–Thornley detector is most commonly used for imaging topographic contrast, but depending on the application, a dedicated backscattered electron detector may be used as an additional source of information. To understand how the number and trajectory components combine in the signal from the Everhart–Thornley detector to produce the topographic images with which we are familiar, the image of a specimen that presents only topographic contrast will be constructed in a stepwise fashion. The specimen is a fracture surface from a piece of high-purity iron, so that there is no contribution from atomic number contrast, only surface morphology. Such a contrast component separation experiment can be performed because the E–T detector permits us to distinguish the contribution of a portion of the backscattered electrons while excluding secondary electrons.

Backscattered Electron Component. We shall first consider the appearance of the fracture surface obtained with the portion of the backscattered electron signal collected by the E–T detector along the line of sight to the specimen. This BSE signal is selected by placing a negative bias (usually in the range 0 V to −50 V) on the Faraday cage which surrounds the scintillator so as to exclude *all* secondary electrons, as shown schematically in Fig. 4.29a. Operated under these conditions, three properties of the E–T detector affect the observed contrast:

1. The detector is located on one side of the specimen, so it has an anisotropic view of the specimen.
2. The solid angle of collection of the detector is small, so that only a small fraction of the backscattered electrons can be collected.
3. The detector is at a low take-off angle relative to the horizontal plane (i.e., the detector line of sight is at a high angle relative to the beam).

These three factors combine to give the detector a highly directional view of the specimen. The actual appearance of the fracture surface under this condition is shown in Fig. 4.29b and is characterized by bright regions from which very high signals are collected and dark regions from which essentially no signal is collected. As a result, the contrast in the image is harsh, with few intermediate gray levels represented. This appearance can be explained as follows with reference to the schematic diagram in Fig. 4.29a. Although all surfaces struck by the primary electron beam produce backscattered electrons, only those surfaces that face the scintillator will

direct at least some of their backscattered electrons toward it, resulting in a detectable signal. Secondary electrons, although produced at all surfaces struck by the primary beam, are completely rejected by the negative bias on the E–T detector cage. Because backscattering from tilted surfaces is increasingly directional with tilt angle, any surface tilted away from the detector will send very few backscattered electrons toward the detector, resulting in a low signal, so that the surface appears dark. The signals are so low that the only information conveyed from these surfaces is the fact that the tilt must be *away* from the E–T detector. Shadowing effects created by details of the local topography are also found. In Fig. 4.29a, a portion of face d is screened from the detector by the ridge formed by surfaces b and c.

As observers of scanning electron microscope images, we unconsciously tend to interpret images based on our most familiar imaging situation, namely light interacting with everyday objects. Oatley *et al*. (1965) demonstrated that a highly effective analogy to our experience with everyday light illumination can be constructed for the SEM imaging process of topographic contrast from rough, three-dimensional objects in images

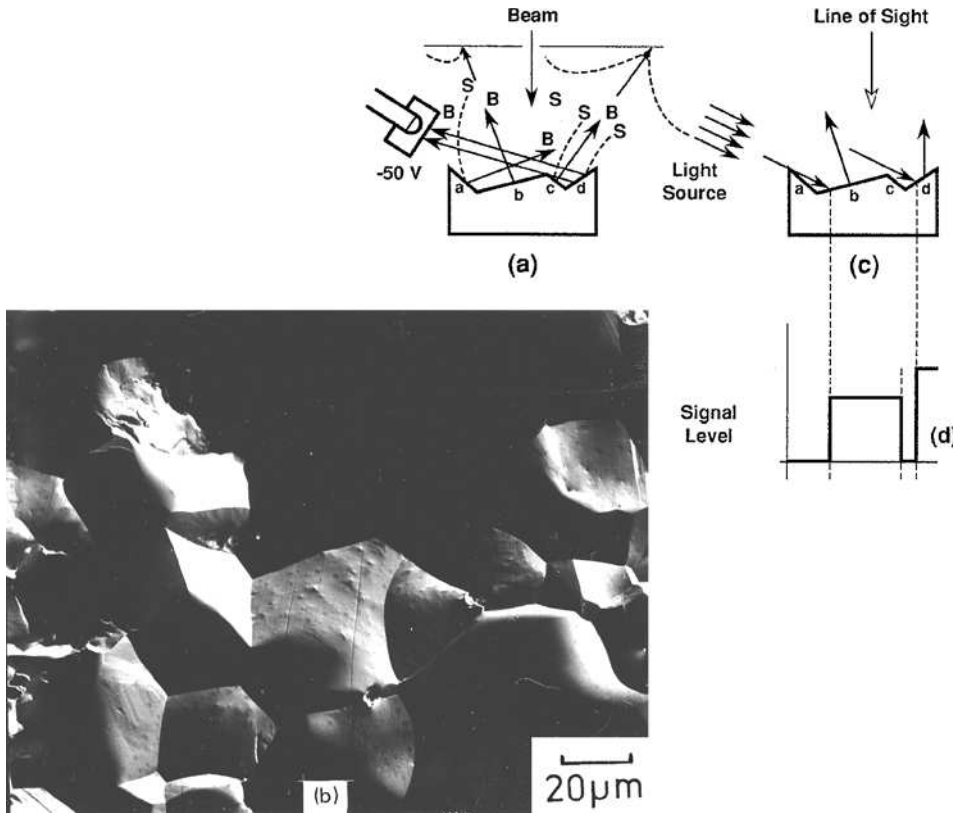


Figure 4.29. (a) Electron collection from randomly oriented surfaces. E–T detector, biased negatively. Solid trajectories: backscattered electrons; broken trajectories: secondary electrons. (b) Image of an iron fracture surface with this detector condition; the detector is located at the top of the image. (c) Equivalent optical illumination and position of the line of sight to produce a similar image. (d) Schematic of the signal received as a function of scan position.

prepared with the E–T detector. The light-optical analogy to Figs. 4.29a and 4.29b is illustrated in Fig. 4.29c. If we wanted to view the rough object with light and produce the same sense of contrast as seen in the SEM image, it would be necessary to place a directional light source such as a flashlight at the position of the E–T detector and place our eye as detector looking down on the specimen, taking the position of the electron beam. Just as in Fig. 4.29a, the surfaces tilted toward the flashlight appear bright, and those tilted away appear dark when viewed from above. Shadowing effects from local topography are also found; ridges which occur between the detector and object shadow the object. The light strength received by the viewer along the line of sight is shown in Fig. 4.29d.

Secondary + Backscattered Electron Component of Topographic Contrast. Let us now consider the case where the Faraday cage is biased positively (usually in the range +50 V to +300 V), as illustrated in Fig. 4.30a, so that secondary electrons are collected. This is the usual mode of operation for the E–T detector (and for many microscopists the ONLY

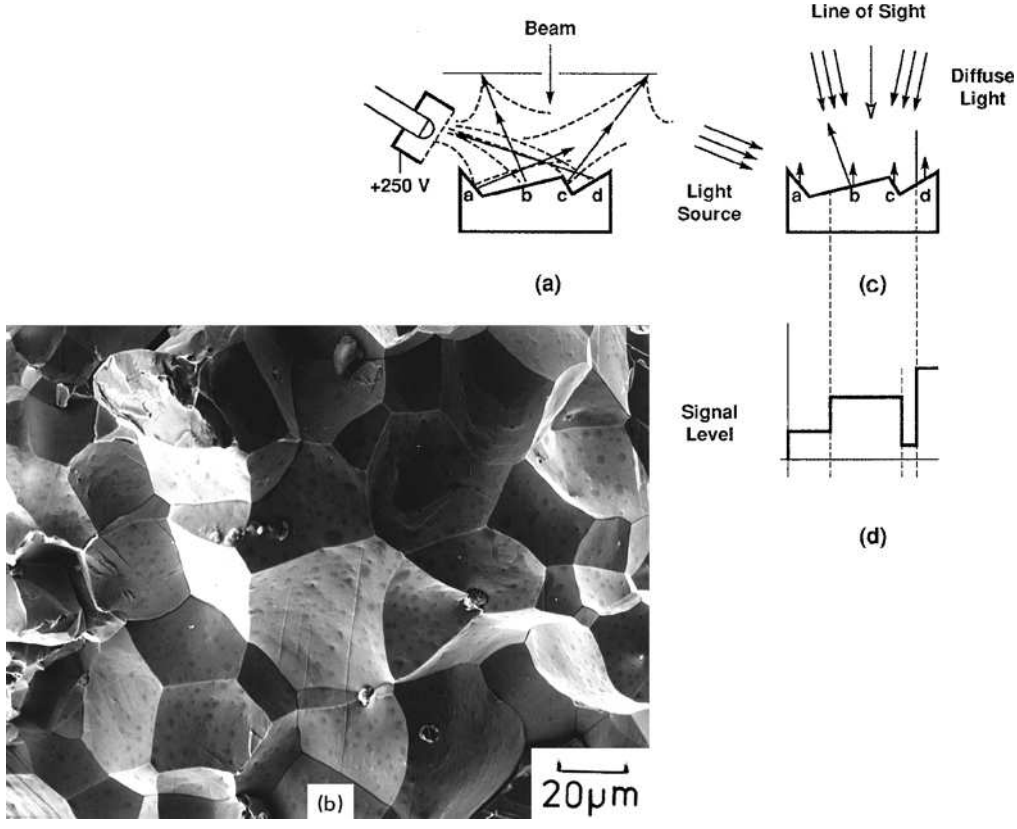


Figure 4.30. (a) Electron collection from randomly oriented surfaces with the E–T detector biased positively. Solid trajectories: backscattered electrons; broken trajectories: secondary electrons. (b) Image of an iron fracture surface with this detector condition; the detector is located at the top of the image; same field as Fig. 4.29b; $E_0 = 15$ keV. (c) Equivalent optical illumination and position of the line of sight to produce a similar image. (d) Schematic of the signal received as a function of scan position.

way they have operated it). The signal is now a complex mixture of three components:

1. *Direct backscattered electrons:* The schematic diagram of Fig. 4.30a has been drawn to indicate that the backscattered electrons directly collected or lost in the negative-bias case of Fig. 4.30a behave in virtually the same way when positive bias is applied to the Faraday cage. A small positive bias (+200 V to +400 V compared to the typical incident beam energy of 10 keV or more) does not significantly deflect the trajectories of high-energy backscattered electrons to alter their *direct* collection or to perturb the trajectories of BSE following paths away from the E–T detector.

2. *Secondary electrons:* The small positive bias profoundly alters the collection of secondary electrons. Secondary electrons are emitted from all surfaces struck by the primary beam, with the number increasing sharply as the secant of the local tilt angle. The contrast in the secondary electron signal produced by this number effect can be estimated as follows. The secondary electron coefficient δ for a surface with tilt θ is given by

$$\delta = \delta_0 \sec\theta, \tag{4.11}$$

where $\delta_0$is the secondary coefficient at 0° tilt. The difference in the SE signal $d\delta$ between two surfaces with a difference in tilt angle $d\theta$ is found by differentiating Eq. (4.11):

$$d\delta = \delta_0 \sec\theta \tan\theta \; \delta\theta. \tag{4.12}$$

The contrast C is found by forming the ratio of Eqs. (4.12) to (4.11):

$$C \sim d\delta/\delta = \delta_0 \sec\theta \tan\theta \; d\theta/\delta_0 \sec\theta = \tan\theta \; d\theta. \tag{4.13}$$

In addition to the number component of contrast described by Eqs. (4.11)–(4.13), there is also a strong trajectory effect on secondary electron collection. Secondary electrons are emitted from all surfaces with a cosine distribution relative to the local surface normal. However, because secondary electrons are low in energy (the majority are emitted with less than 10 eV), even the modest positive collection field on the Faraday cage of the E–T detector can cause wide deflections of the secondary electron trajectories. This results in high SE collection efficiency from surfaces that face the detector and at least partial collection from most surfaces that are tilted away from the detector. Some secondary electrons are collected from all surfaces that the primary beam strikes.

3. *Indirect backscattered electrons:* In Figs. 4.29a and 4.30a, the backscattered electrons that are not intercepted by the detector for direct collection strike the polepiece of the final lens and the walls of the specimen chamber. These backscattered electrons retain sufficient energy to cause secondary electrons (designated SE_3) to be emitted wherever BSE strike interior surfaces of the microscope. Since most BSE strike these surfaces at acute angles, they are even more efficient at generating SE_3. The BSE may re-emerge and continue traveling to strike other chamber surfaces, producing still more SE_3 secondary electrons. All of this BSE scattering and SE_3 production takes place within the dwell time of the

beam on the pixel to which it has been addressed. The positive bias on the E–T detector results in such an efficient secondary electron "sponge" that a significant fraction of these remotely produced SE_3 from the chamber walls are collected as part of the total signal. What information do these remotely generated SE_3 contribute to the signal? Surprisingly, these SE_3 from BSE conversion do not represent true secondary electron information, like that provided by the SE_1 component emitted where the beam enters the surface. These remote conversion SE_3 actually represent *indirect* collection of BSE. This correspondence can be best understood by considering what causes the SE_3 signal to rise and fall. The remote SE_3 signal rises and falls purely with the number of BSE emitted from the sample, and thus despite being low-energy secondary electons, the SE_{III} signal carries BSE information, not SE_I information. Peters (1984) has shown that the remote BSE-to-SE conversion can form a large fraction of the total E–T detector signal, particularly from targets of intermediate and high atomic number.

The image obtained with the positively biased E–T detector is shown in Fig. 4.30b, which depicts exactly the same area as Fig. 4.29b. With positive bias on the E–T detector, the harsh trajectory component of backscattered electron topographic contrast is reduced. SE and BSE are collected with at least some efficiency no matter how well screened a feature is from direct sight to the detector due to the mechanisms of direct and indirect collection of remote secondary electrons. Note that those features that faced the detector and appeared bright in Fig. 4.29b remain the brightest features in Fig. 4.30b because of the unchanged situation for the direct collection of backscattered electrons. Features facing the detector appear as highlights in the image. The equivalent lighting situation for the positively biased E–T detector is illustrated in Fig. 4.30c, showing the same situation as Fig. 4.29c, but augmented with a general diffuse illumination. Now from the point of view of an observer looking down on the scene, all surfaces are illuminated, and those tilted toward the directional source of light are highlighted.

4.4.3.3. Light-Optical Analogy

There exists a strong correspondence of the SEM imaging process with a positively biased E–T detector to the normal human vision process for scenes with illumination from above (e.g., the sun, artificial lighting), as shown in Fig. 4.31. The remarkable fact that SEM images of rough, three-dimensional objects can be readily interpreted, even by viewers with no prior knowledge of the nature of the SEM image formation process, flows from this strong "light-optical analogy." As observers of our natural environment, we are used to viewing features lighted from above by a highly directional source, the sun, supplemented by diffuse illumination due to light scattering in the atmosphere. Figure 4.30c constructs the light-optical analogy as equivalent to the SEM imaging situation. To properly interpret the SEM image, we must match the components of the two systems (SEM and human vision) that have similar characteristics. Thus, the electron

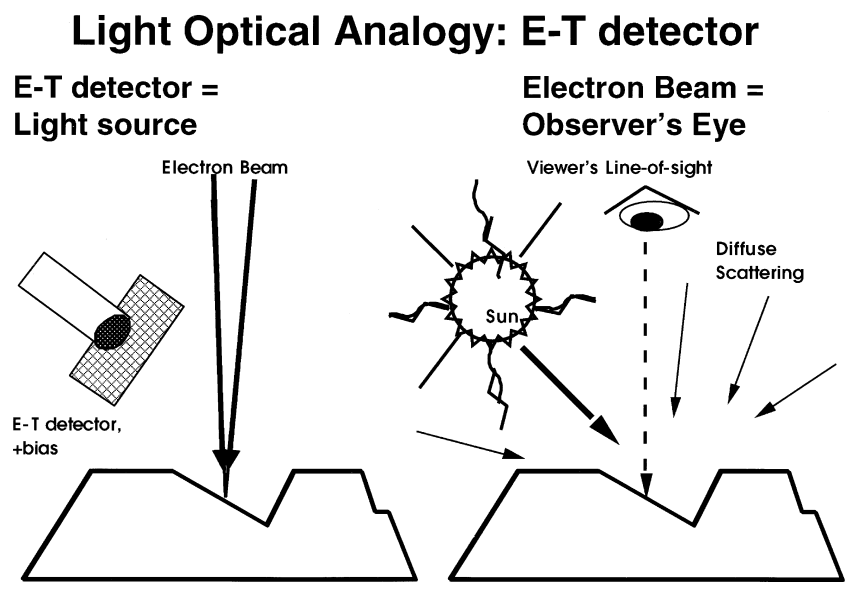


Figure 4.31. Light-optical analogy. The eye of the viewer is equivalent to the beam in the SEM, and the SEM detector is equivalent to the light source in the conventional light illumination.

beam is a highly directional component because only those positions of the specimen that can be directly struck by the incident electron beam are accessible in the final SEM image. The human eye is highly directional. We cannot see the person standing behind a tree; there must be a direct line of sight to any feature for it to be visible. To establish the light-optical analogy, we must place the viewer at the position of the electron beam, that is, viewing the specimen from above as if looking through the final aperture with our eyes ("Be the beam"). The positively biased E–T detector has a highly directional signal component, the direct BSE component, and a diffuse component, the true SE and remote BSE collected as the SE_3 component. The E–T detector thus has the characteristics of the sun in our everyday illumination situation, with direct BSE equivalent to direct sunbeams and the SE_3/remoteBSE equivalent to atmospherically scattered indirect light. To interpret the SEM image, we therefore must imagine that the apparent source of direct illumination seems to come from the E–T detector. The image appears as if we placed a flashlight at the position of the E–T detector in the specimen chamber and we view that specimen by looking along the electron beam. To complete the recipe for the light-optical analogy, because the everyday illumination is from above (sun high in the sky), the E–T detector must be located at the top of our SEM frame, with the apparent light source shining down on the specimen. In summary, the SEM image made with a positively biased Everhart–Thornley detector is to be interpreted as if the viewer looks along the electron beam and the illumination of the scene comes from the electron detector, which should be oriented to be at the top of the SEM image.

The light-optical analogy is a powerful effect, virtually automatic in its influence on an observer, but to ensure proper interpretation of the topography of surfaces, it is vital that we always establish this critical condition of top lighting in our SEM images. When the lighting source comes from an unexpected direction, for example, below or from the side, our ability to recognize features is compromised. (Consider the disorienting view of a human face illuminated with a flashlight from below in a dark room!) We are so conditioned to expect top lighting that failure to provide top lighting can cause the apparent sense of the topography to invert, particularly where the topography is shallow and there are no clues from experience or prior knowledge of the specimen. The appearance of topography with top lighting (correct) and illuminated from below (incorrect) is shown in Fig. 4.32. The correct sense of topography can be lost if the apparent illumination comes from anywhere but the top of the image when the topography does not contain strong visual clues as to the nature of its shape. For example, an image of an insect will appear to have the correct shape no matter what the illumination direction because we inherently understand the insect to have a shape that generally consists of convex surfaces.

It is important to realize that the apparent lighting of an SEM image may not be correct in an arbitrarily selected SEM. (In a multiuser facility, always assume the worst and check!) The apparent lighting is controlled by the position of the E–T detector relative to the scanned field. The E–T detector is rigidly fixed in the specimen chamber. To establish the proper illumination for the light-optical analogy to function, the scan must be oriented to place the apparent position of the E–T detector at the top of the image, both on the visual CRT and on the final photographs. The SEM usually has an electronic scan rotation feature that permits the microscopist to orient the image on the screen to any position through a rotation of 360°. Scan rotation changes the orientation of the scan field, but not the orientation of the specimen relative to the E–T detector (mechanical stage rotation will change the specimen orientation relative to the E–T detector). This scan rotation can lead to very misleading images because the apparent position of the E–T detector will travel around the image field as the scan is rotated. If the apparent position of the E–T detector is placed at the bottom of the image field, the apparent illumination comes from the bottom, and the sense of topography may appear inverted. To avoid this error, the following procedures can be used to establish the proper setting for scan rotation.

Method 1. If spherical particles are available (a small ball bearing will suffice), or any other object with strong topography whose sense is known (e.g., the test feature is known to be a hole or a protuberance), then obtain an SEM image of the ball with the Faraday cage of the E–T detector biased negatively. The strong directional illumination from direct BSE will reveal the direction of the E–T detector, as shown in Fig. 4.33a. If the position of the E–T detector is not at the top of the display, adjust the scan rotation control (not the mechanical stage rotation) to place the effective detector position at the top center of the image, as illustrated in Fig. 4.33b. Mark this important position on the control knob! Note that the correct

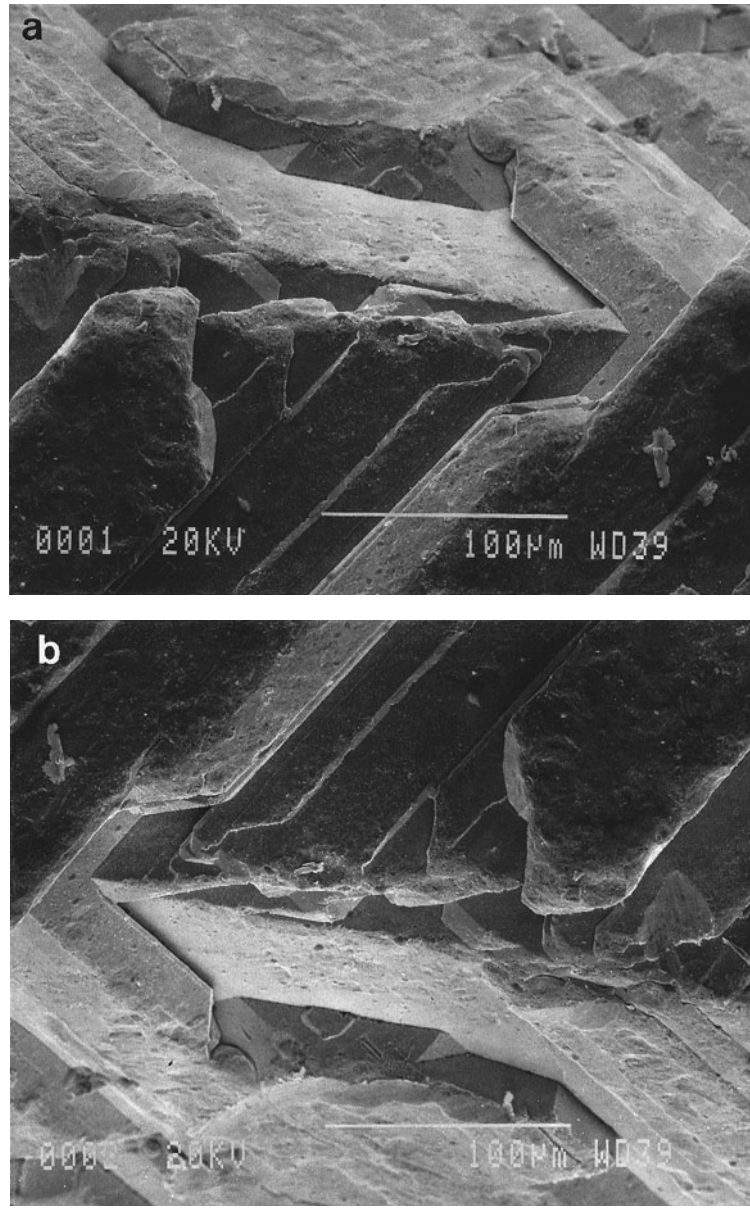
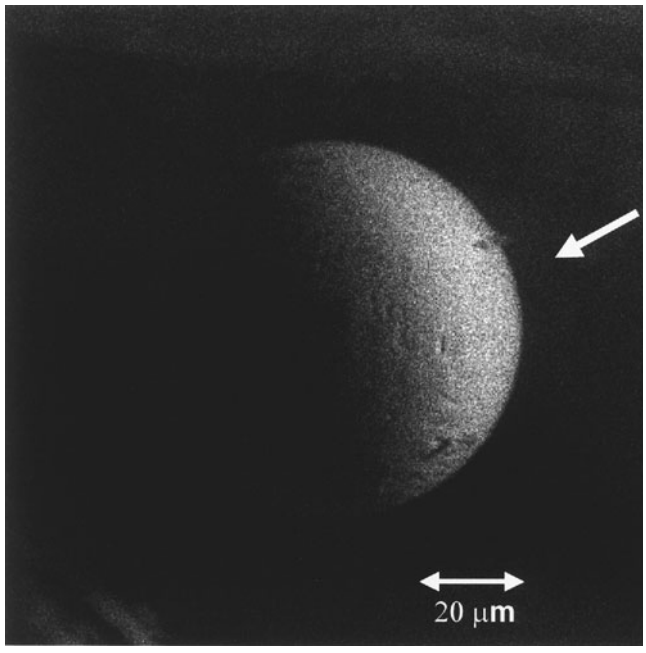
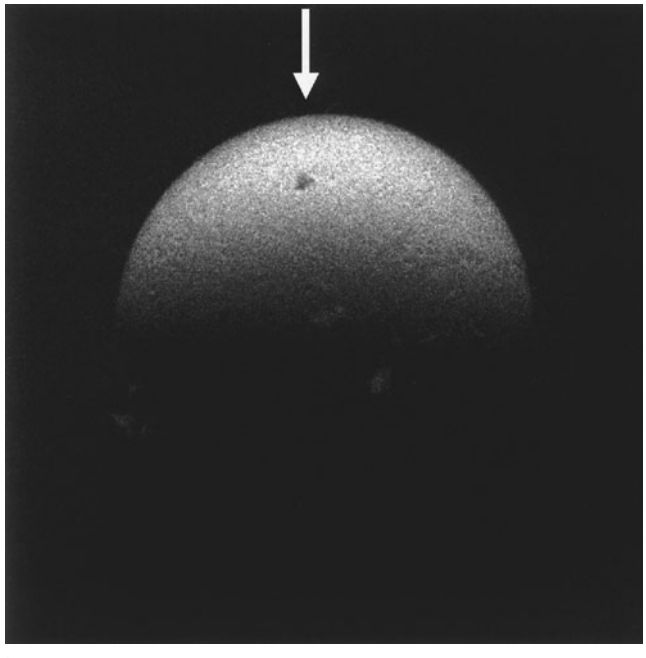


Figure 4.32. Comparison of topographic (positively biased E–T detector) images taken (a) with the proper sense of lighting by placing the E–T detector at the top of the field of view and (b) improper sense of lighting by placing the E–T detector at the bottom of the field.

position of the scan rotation may not correspond to an obvious angular value such as 0° or 180° because the dial on the knob may never have been set properly, it may have slipped on the shaft, or the E–T detector itself may have been moved to a different port on the SEM. Note also that the image of the ball taken with the positively biased E–T detector, Fig. 4.33c has such soft shadowing that it is not very useful for locating the detector position.

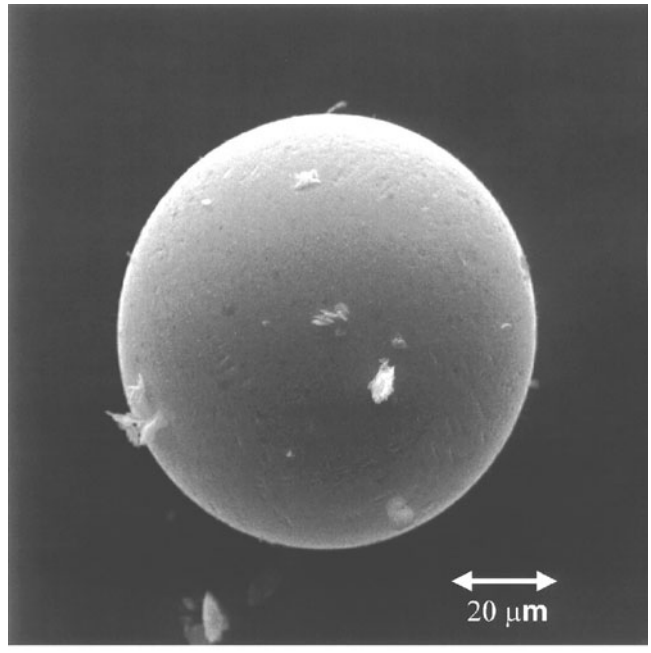


a



b

Figure 4.33. Use of a spherical object (particle of metal alloy) to locate the source of apparent illumination with an E–T detector. (a) E–T biased negatively to detect only BSE; the image of a spherical object (particle of metal alloy) shows the apparent illumination source from the right. (b) Rotation of the image by 90° counterclockwise places apparent the light source at the top of the field. (c) Same particle with positively biased E–T detector.



c

Figure 4.33. (*Continued*)

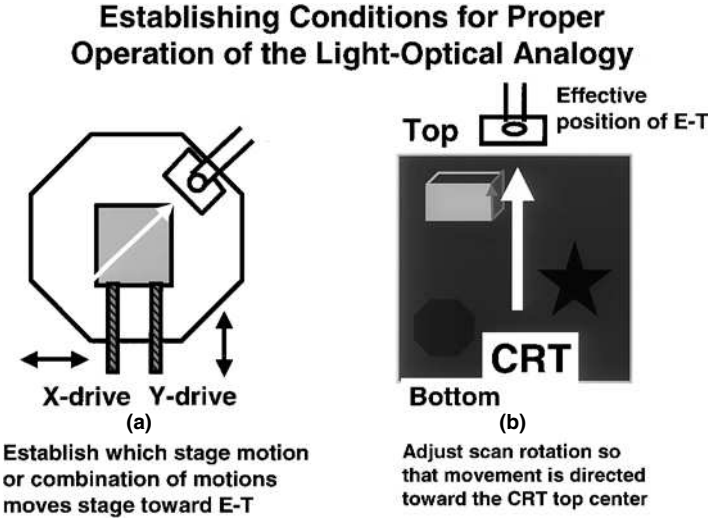


Figure 4.34. (a) Determination of which mechanical motions drive the specimen stage toward the E–T detector. (b) SEM imaging operation with this stage motion engaged; the image should be seen to be moving toward top center when the image orientation is correct.

Method 2. If it is not possible to set the E–T Faraday cage to negative bias to obtain the strongly shadowed view, then this second method should be used. The microscopist should first determine in what way the mechanical translation of the specimen stage moves the specimen toward the E–T detector. That is, does the x or y drive (or a combination of x and y) move the stage toward the E–T detector, as shown in Fig. 4.34a? With this knowledge, operate the SEM in the normal imaging fashion with the E–T detector biased positively. Then operate the mechanical drive(s) to move the specimen toward the E–T detector, and observe the apparent motion on the CRT screen, Fig. 4.34b. The apparent motion should be from the bottom to the top of the frame. If not, the scan rotation should be adjusted to produce this result. This will ensure that top lighting is obtained. Again, mark the position on the scan rotation dial that corresponds to establishing this condition of top lighting.

Establishing a top lighting condition will produce a situation where the sense of the topography can be properly interpreted in most cases. Occasionally images will be encountered where the sense of the topography is not obvious, even with top lighting, perhaps because of the shallowness of the relief or some local peculiarity in the mix of secondary and backscattered electrons that is collected. An example of this effect is shown in Fig. 4.35a, where the feature may be a mound or a dimple; we simply cannot tell without an indication of the direction of the illumination. Only at a slightly lower magnification, when a spherical particle is also included in the field of view, as shown in Fig. 4.35b, is the direction of illumination established, and the feature can be recognized as a dimple. In such cases where the topography seems ambiguous, the microscopist should try lowering the magnification to view a larger area. Also, tilting and/or rotating the specimen to view that particular field from a different angle may be useful.

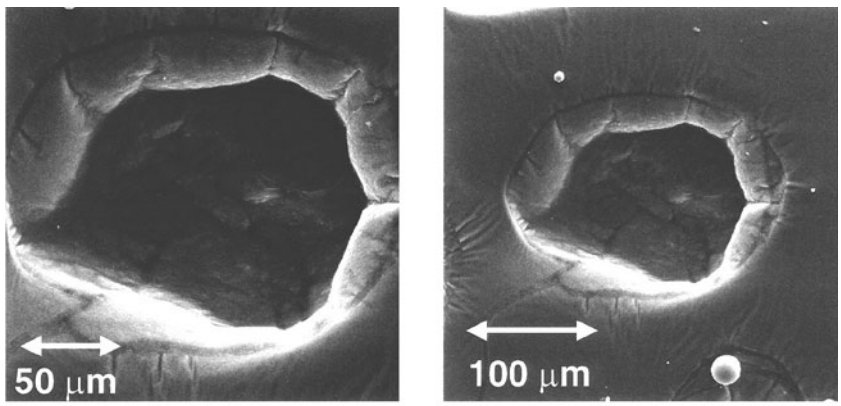


Figure 4.35. (a) Example of an image of a structure with ambiguous topography; is it a dimple or a bulge? (b) Lower magnification image of the same area, where a spherical particle is also included. The illumination of the sphere aids interpretation of the ambiguous object, which can thus be seen to be a dimple.

Finally, the preparation of stereo pairs may be needed to properly assess the nature of the topography.

4.4.3.4. Interpreting Topographic Contrast with Other Detectors

A typical arrangement of detectors surrounding a topographic specimen is shown in Fig. 4.36. The appearance of a rough surface, especially the apparent illumination, varies significantly with the type, solid angle, and take-off angle of the detector used to collect the signal. We shall consider the image obtained with the positively biased E–T detector, properly oriented with one of the tests described above, to be the "reference" image because this detector satisfies the light-optical analogy, which permits the most natural interpretation.

General Rule. When dealing with *any* electron detector that is external to the specimen (i.e., excluding images prepared with specimen or absorbed current) the general rule for interpreting the illumination of an SEM image is that the detector appears to be the source of light that illuminates the image and the viewer's line of sight is along the electron beam.

Thus, to properly interpret images made with detectors other than the E–T detector, we must consider where the detector is located relative to the specimen and electron beam, how large it is, and its symmetry or asymmetry relative to the electron beam (i.e., the eye of the viewer).

BSE. Solid State. Consider a dedicated backscattered electron detector having a large solid angle and placed at a high take-off angle, as shown in Fig. 4.36. The detector consists of two segments: A, located at the top of the image field, and B, located at the bottom. The reference image of a rough surface prepared with the positively biased E–T detector is shown in Fig. 4.37a, and the corresponding negatively biased E–T detector in

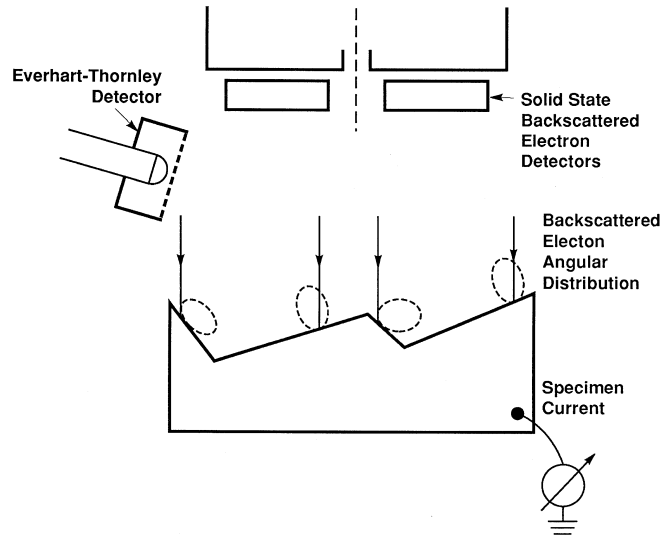
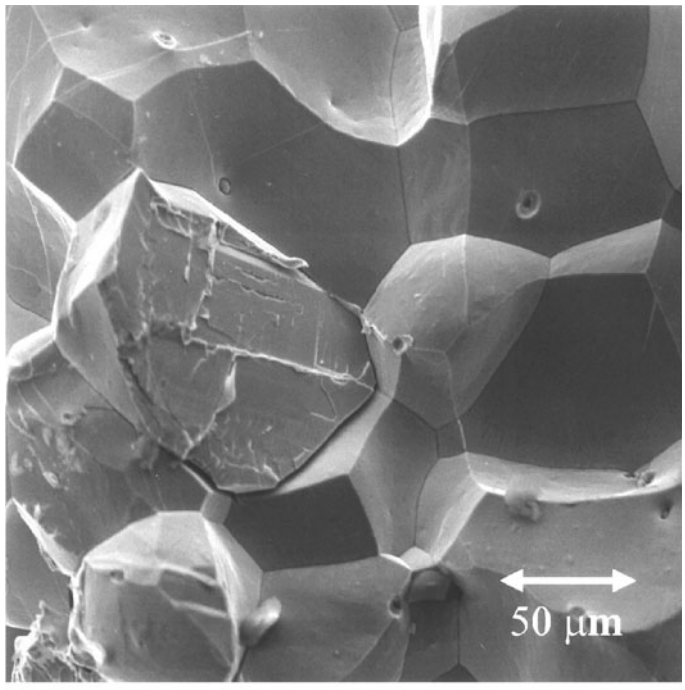
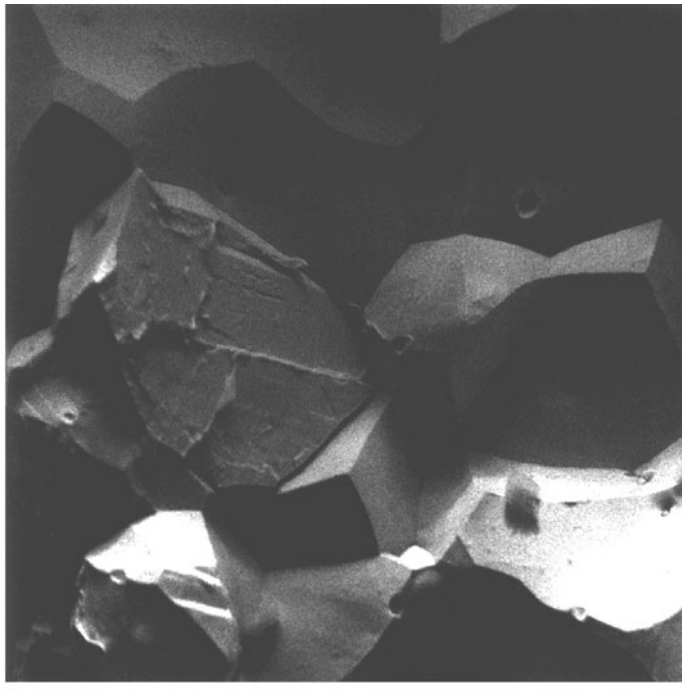


Figure 4.36. Typical arrangement of detectors in an SEM chamber relative to a rough surface with random facets.

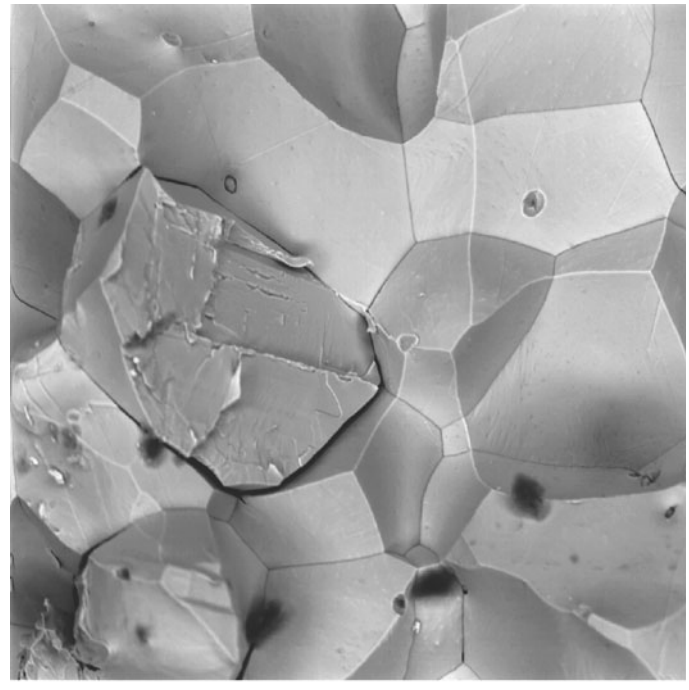


a

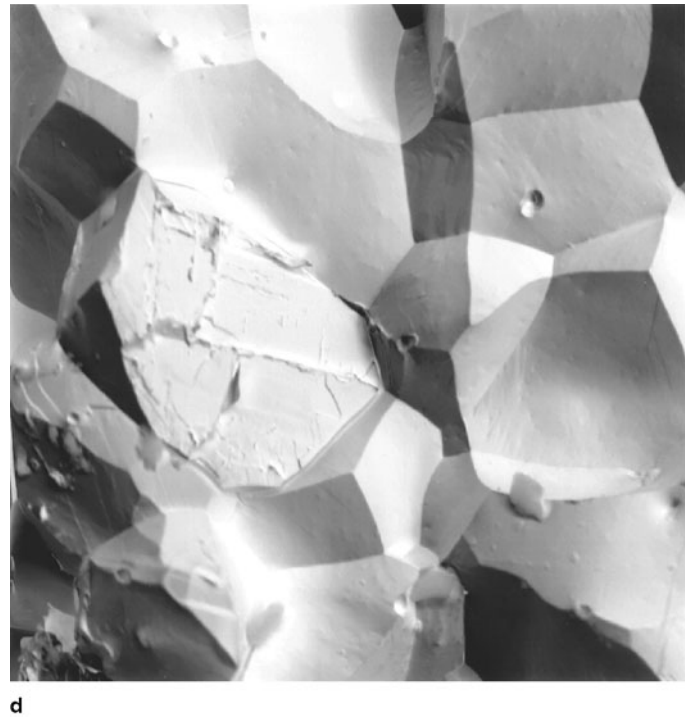


b

Figure 4.37. Appearance of a rough surface as imaged with various detectors: (a) Positively biased E–T detector; (b) negatively biased E–T detector; (c) solid state detector, A + B mode; (d) solid state detector, A − B mode; (e) solid state detector, B − A mode; (f) passive scintillator detector, with a central ring and a large shaft placed to the left horizontal of the scan field.

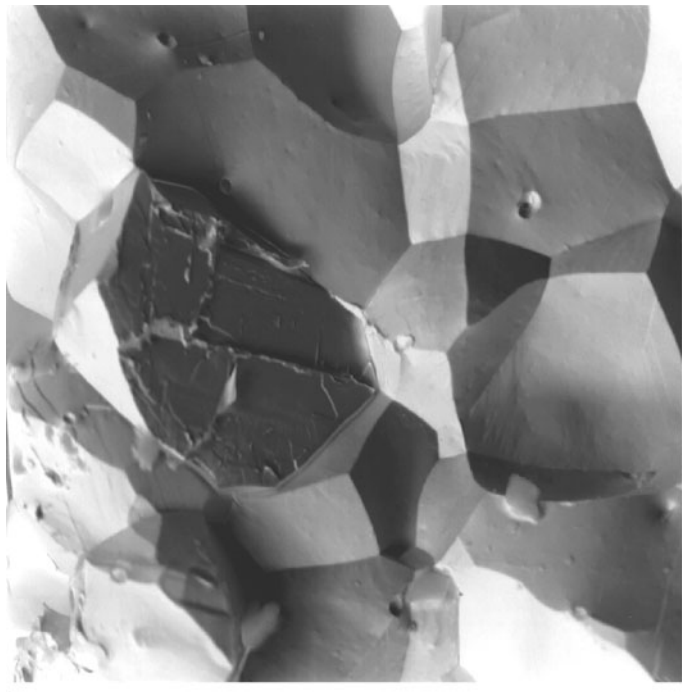


c

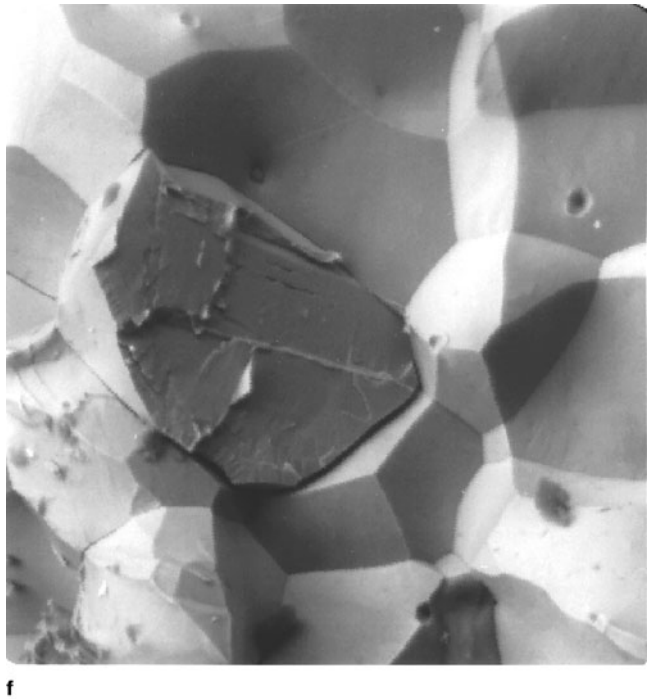


d

Figure 4.37. (*Continued*)



e



f

Figure 4.37. (*Continued)*

Fig. 4.37b. Fig. 4.37c shows the appearance of the same field of view, but prepared with the sum of the signals from the solid state detectors A + B. The A + B image is distinctly different in many respects from the conventional positively biased E–T image. With a large-solid-angle detector placed near the beam at a high take-off angle, the apparent illumination is along the line of sight of the viewer, so that the surface appears as if we were looking along the beam of a flashlight. A signal is obtained preferentially from surfaces that are nearly perpendicular to the beam, whereas highly tilted surfaces appear dark because electrons backscattered from these surfaces are directed away from the detector. Thus, a large A + B BSE detector placed above a specimen whose general plane is perpendicular to the beam produces topographic contrast which is nearly the *Reverse* of the positively biased ET detector, so that fine and even coarse features may give the illusion of the wrong topography, that is, holes appear to be dimples. The A + B BSE topographic contrast is also somewhat reduced in magnitude because some electrons emitted from tilted surfaces and going in opposite directions are still likely to be collected because of the large solid angle of the detector. Another striking difference is the greater sensitivity to the small dimple-like structures in the A + B image and their dark appearance. Thus, the conventional light-optical analogy of the E–T detector is not fulfilled by the solid state BSE detector in the sum mode. For specimens with strong and easily recognized topography such as an insect, this may not cause any problems for the viewer. However, for specimens with modest to slight topography of uncertain sense, the viewer's understanding of the sense of that topography may be compromised.

When the image is formed as the difference between the two detectors A − B as shown in Fig. 4.37d, the appearance of the topography again changes and becomes similar to the conventional positively biased E–T image as far as the general shading of the facets is concerned. Note, however, that the fine-scale dimples nearly disappear in the difference image. If the order of the difference operation is reversed, B − A, as shown in Fig. 4.37e, the sense of the topography is seen to invert. Obviously only one of these difference images can be correct. It is important to establish the correct order for taking the difference. This is best done on a specimen with simple, unambiguous, easily understood topography, such as a flat metallic surface with sharp scratches. Taking the negatively biased E–T detector image of such a scratched surface, Fig. 4.38a, as a reference, one can compare the A − B (Fig. 4.38c) and B − A (Fig. 4.38d) images. The scratches will appear as indentations in the correct image and as ridges in the incorrect image.

Passive Scintillator. The passive scintillator BSE detector typically is characterized by a very large solid angle, a high take-off angle (for a specimen placed normal to the electron beam), and symmetric placement of the detector around the electron beam. It thus resembles the large solid state BSE detector operated in the sum mode in the sense of the illumination that it produces, very similar to Fig. 4.37c. The proximity of the passive scintillator BSE detector to the electron beam means that images

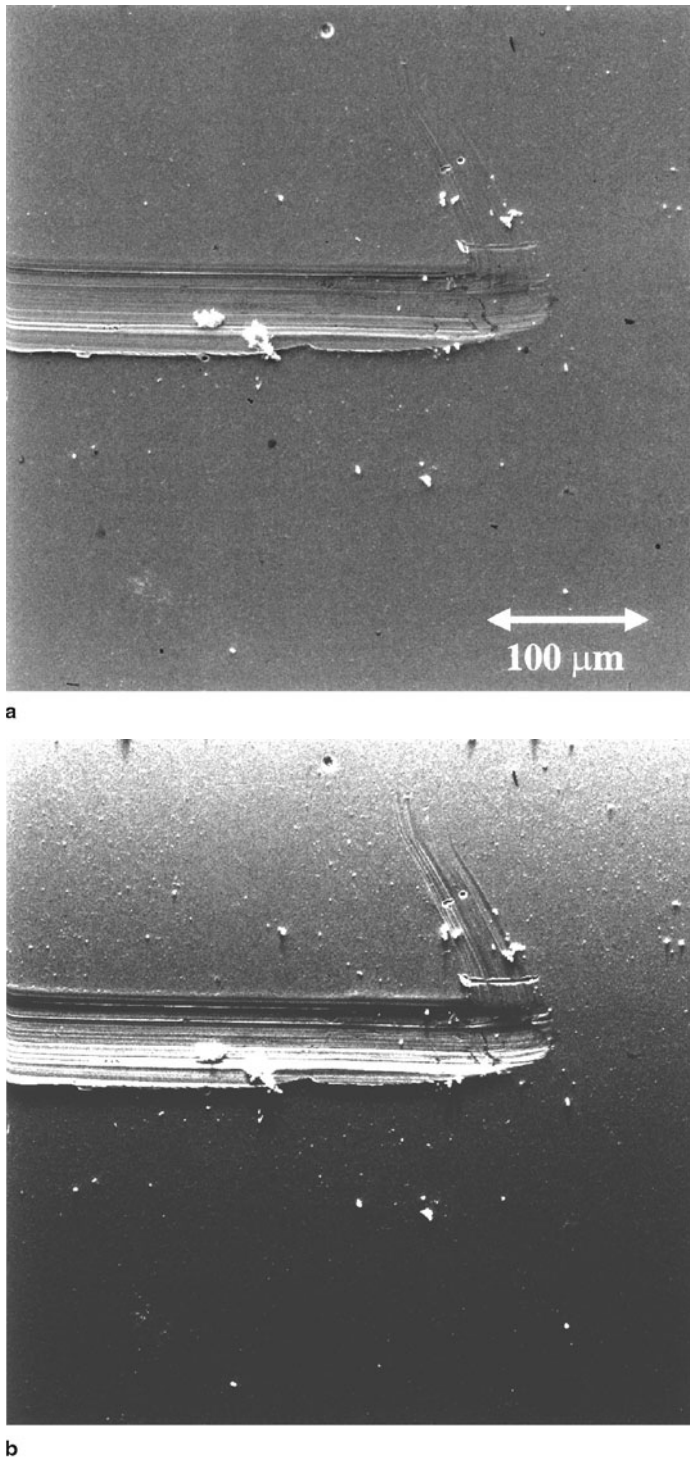
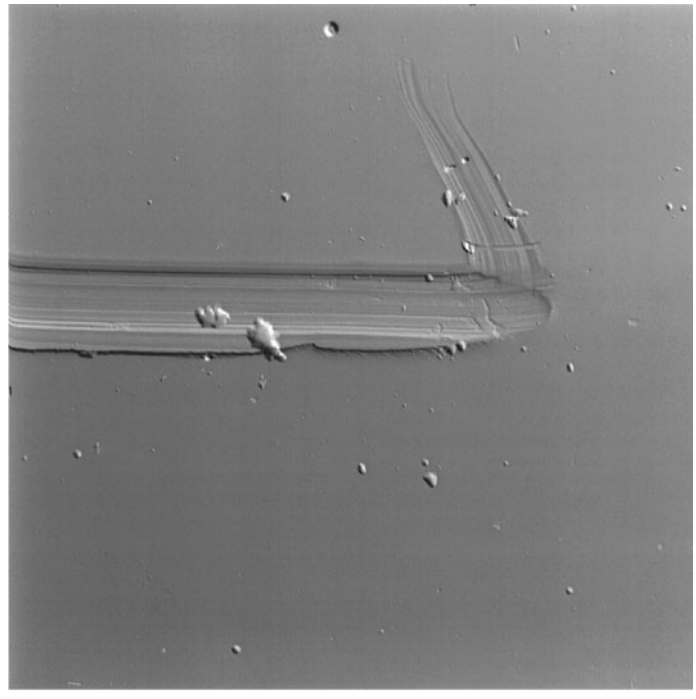
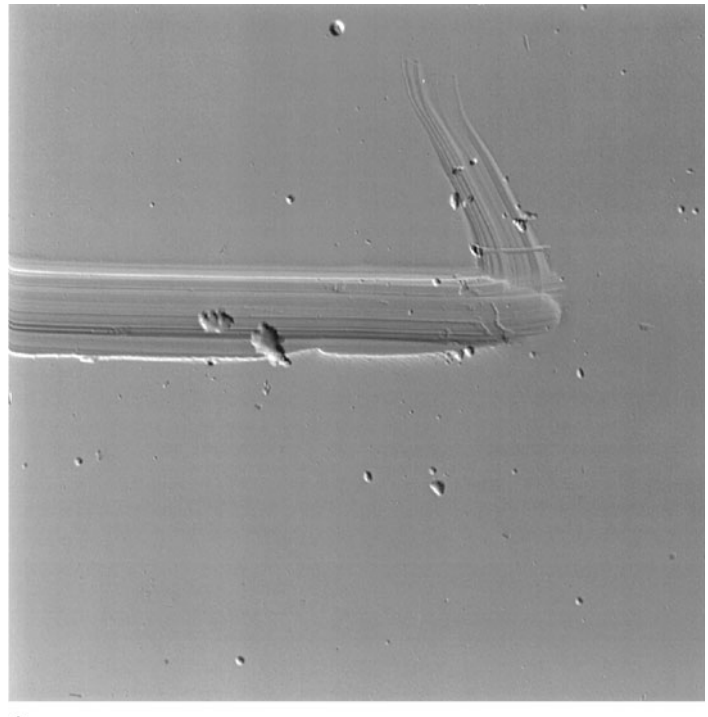


Figure 4.38. Use of a metal surface with simple scratches to assess the proper order of the detector difference operation: (a) Negatively biased E–T detector, placed at the top center of the field; (b) positively-biased E–T detector; (c) A–B; (d) B–A. Note differences in the apparent topography.



c



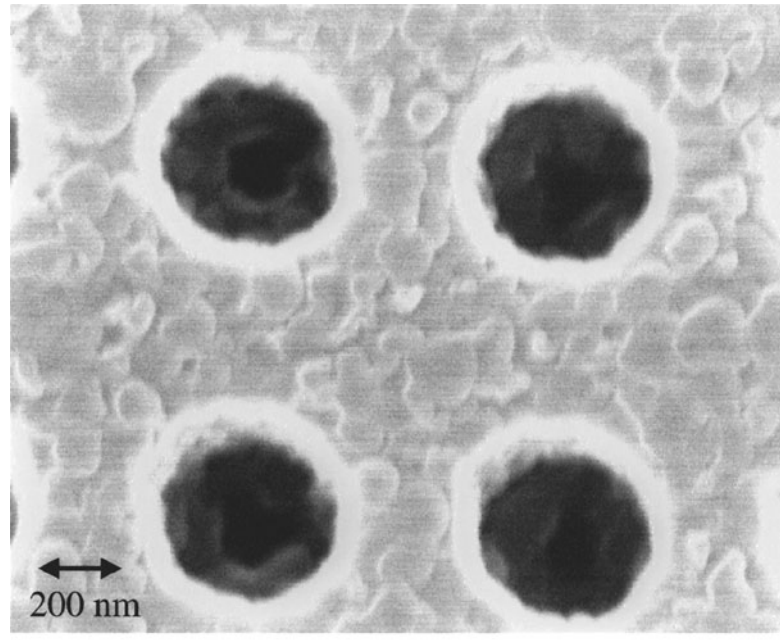
d

Figure 4.38. (*Continued*)

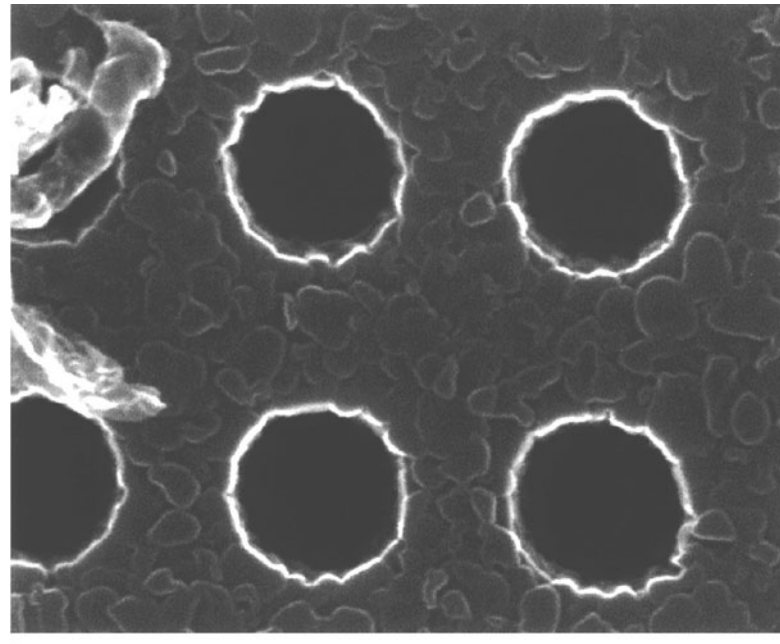
will appear to be illuminated nearly along the observer's line of sight. To provide more sensitivity to topography, the passive scintillator detector can be made asymmetric, with a central ring and a long shaft to the side leading to the photomultiplier. Because the detector serves as the apparent source of illumination, the enhanced collection along the shaft of the passive scintillator increases the apparent illumination from that side. The image of the rough surface then shows stronger sensitivity to topography, as shown in Fig. 4.37f. Note that in this case, the apparent sense of illumination is from the side of the image, rather than the top, and so to better satisfy the light-optical analogy, the image should be rotated by 90°. When interpreting topography with the passive scintillator, the microscopist must carefully regard the detector's special characteristics.

Through-the-Lens SE Detector. The TTL SE detector of the FEG SEM instruments collects signal along the same axis as the beam, so that illumination appears to occur exactly along the observer's line of sight. As a result of this unusual illumination, it is possible to look down blind holes and view features at the bottom. Any feature that the beam can reach has a good chance of being viewed with the TTL signal. An example is shown in Fig. 4.39a, where details in the bottom of blind holes in silicon are visible, whereas similar holes viewed with an E–T appear completely black, Fig. 4.39b. Because the collection of SE depends on the magnetic field of the lens penetrating into the hole, the TTL SE image is limited in the depth to which it can reach. The effect can be very useful for the study of blind holes in engineered structures. A second aspect of the TTL detector is its surface sensitivity because the TTL signal is relatively enriched in the SE_1 signal, which is surface-sensitive, compared to the SE_2 and SE_3 signals, which are created by BSE events and have the depth sensitivity of BSE. This effect can be seen in Fig. 4.39c, a TTL image of an aluminum alloy with Mg–Zn intermetallic precipitates and additional features due to thin surface films, compared to Fig. 4.39d, the corresponding BSE image, which shows the precipitates only.

Separation of Contrast Components. When multiple detectors are available in the same instrument, the microscopist can examine images made with each detector separately or in combination. Such studies can be very useful in elucidating the nature of the specimen, when a single detector image might not give adequate information. The added information that can be deduced by making use of the detector illumination paradigm is illustrated in Fig. 4.40 for a field of view of particles on a nearly featureless substrate. Three detectors are available, as shown in Fig. 4.40a: an ET detector (positive bias) located in the upper right corner of the image, and two annular solid state BSE detectors, one located nearly on the optic axis and the second placed around this at a greater angle from the axis. The image prepared with the ET(+) detector, Fig. 4.40b, shows the particles with asymmetric edge illumination, with their brightest edges in the upper right, toward the ET detector. The inner on-axis BSE

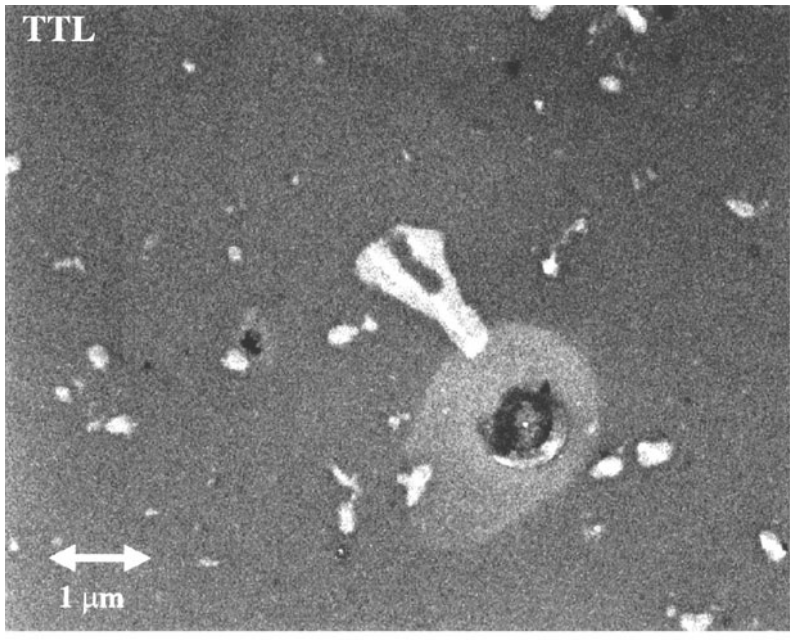


a

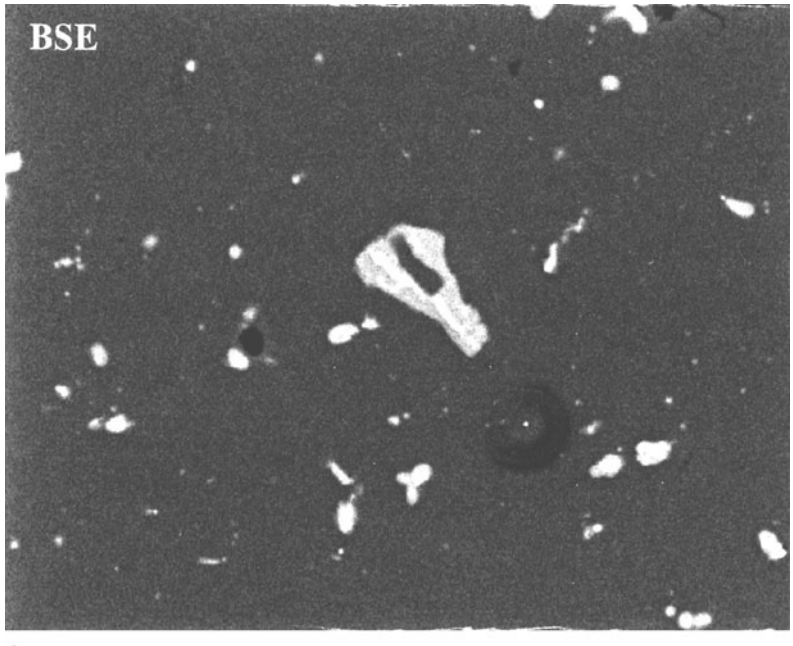


b

Figure 4.39. Comparison of TTL detector with other detectors: (a) TTL SE image of blind holes in a silicon wafer; note details at the bottom of holes due to apparent illumination along the beam. (b) E–T detector of the same specimen at a different location; note that the holes give no information because of an off-axis apparent illumination source. (c) TTL SE image of aluminum with precipitates of Mg–Zn intermetallic, showing precipitates and surface deposits. (d) BSE detector image of the same field of view showing precipitates, but without any evidence of surface deposits. (Al–Mg–Zn sample courtesy of Raynald Gauvin; images courtesy of John Small.)



c



d

Figure 4.39. (*Continued*)

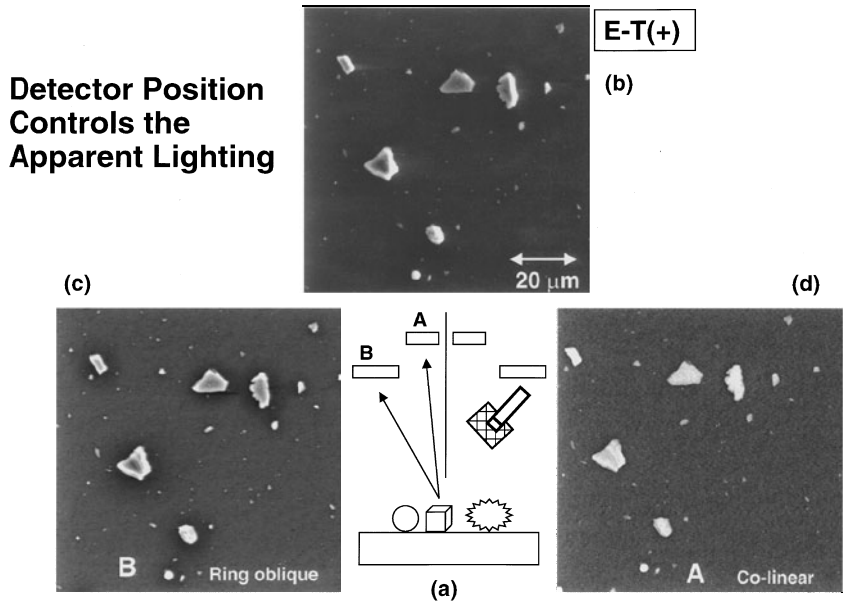


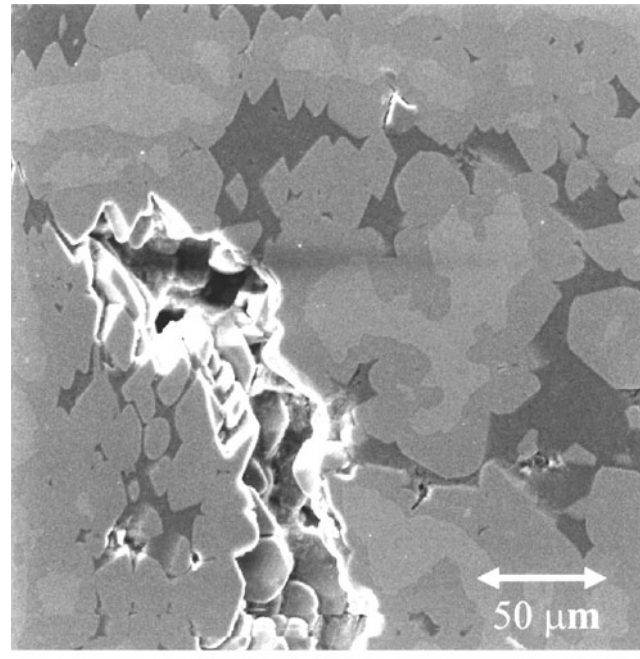
Figure 4.40. Appearance of a field of particles with three detectors, (a) Schematic. (b) Image with positively biased E–T detector, placed in the upper right corner of the scan field; (c) image with a central A + B annular solid state detector; (d) image with a larger bore, A + B annular solid state detector.

detector gives an image, Fig. 4.40c, in which the top of each particle is strongly and evenly illuminated, an image which is useful for measuring the projected area of the particles. The outer annular detector shows the outline of the particles, but compared to the ET(+) image, the outline is nearly uniform because this detector is positioned to act like a ring of oblique illumination.

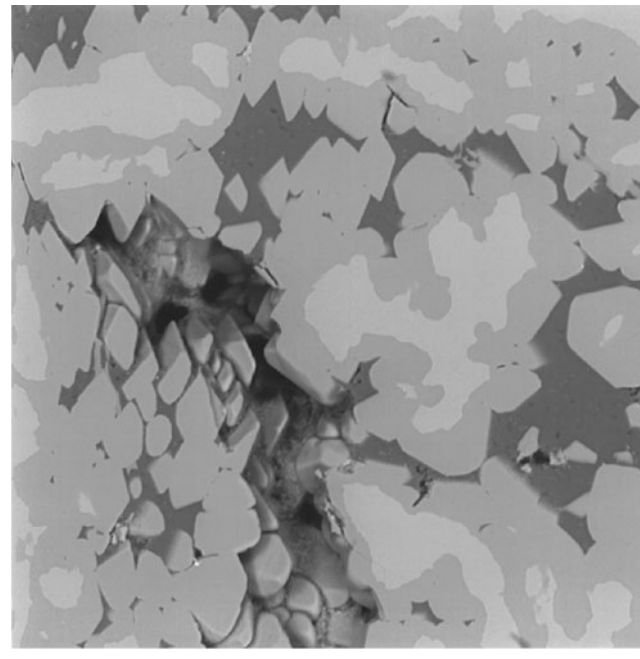
When the character of the specimen is such that multiple contrast mechanisms can operate, the resulting SEM image will contain information from each source. The visibility of each contrast component in the final image depends upon the relative strength of the signal modulation that each causes. A strong contrast mechanism will dominate weak mechanisms in relative visibility. Because contrast mechanisms have different characteristics in terms of their number and trajectory components, separation of competing contrast mechanisms is sometimes possible through the use of multiple detectors. The classic case is the separation of atomic number and topographic contrast (Kimoto and Hashimoto, 1966). Atomic number (compositional) contrast is almost exclusively a number mechanism, whereas topographic contrast has both a number component and a very strong topographic component. By using a carefully positioned BSE detector array operated in the sum and difference modes, separation of the number and trajectory components is possible. In the sum mode, number effects dominate because the detector has such a large collection angle that

no matter which direction electrons leave the specimen, they are collected by the detector, eliminating trajectory effects. In the difference mode, trajectory effects dominate because each detector views the specimen from a different location, and therefore those features of the specimen that face a particular detector dominate the signal from that detector. When differences are taken between the signals from various detectors, the illumination appears to come from the detector signal with the positive multiplier (e.g., A − B is accomplished by scaling B with a negative multiplier and adding). Because each detector in a difference signal receives the same number components of contrast from the specimen, these number effects are eliminated in the difference image. An example of sum and difference mode imaging is shown in Fig. 4.41. The specimen is Raney nickel, which contains multiple phases that produce atomic number contrast and surface defects that produce topographic contrast. To provide a point of familiar reference, the specimen is viewed in Fig. 4.41a with a positively biased E–T detector, showing atomic number, topographic, and surface film contrast. In Fig. 4.41b, the summation signal from both halves of a solid state BSE detector is shown, and the atomic number contrast dominates the image, whereas the topography is nearly completely suppressed. In Fig. 4.41d the difference mode A − B for a detector at the top of the field of view minus a detector at the bottom of the field of view is shown. Topographic contrast dominates, the atomic number contrast is nearly lost, and the apparent illumination is from the top. The sensitivity to topography is not uniform. Linear features that run parallel to a line connecting the detector components produce very low contrast, whereas features running perpendicular to that line are seen in high contrast. The topographic contrast also appears much stronger than that which we obtain from the conventional E–T detector.

There is also evidence that artifacts can appear in A − B difference images at phase boundaries. The artifacts appear as narrow linear features at boundaries between phases (i.e., regions of different chemical composition). The linear features are often interpreted as steplike changes in elevation, and yet independent measurement reveals that the surface is actually flat and featureless. This boundary artifact arises from the lateral spread of the beam that arises from scattering, which means that although the phase boundary is atomically sharp, the BSE signal does change with the abruptness of atomic dimensions as the beam is scanned across the boundary. Consider the scattering situation shown schematically in Fig. 4.42. The beam approaches the boundary from the lower atomic number (i.e., weaker scattering) phase. For beam electrons scattered into paths with a vector component parallel to the surface, entering the higher Z material is more difficult because of its stronger scattering, so the boundary tends to act as a mirror which scatters these electrons back into the weaker scattering phase. This creates asymmetry in the trajectories of the electrons that create atomic number contrast, but just near the boundary. When the beam is at least an electron range unit away from the boundary and completely immersed in either phase, the atomic number scattering is “normal,” with

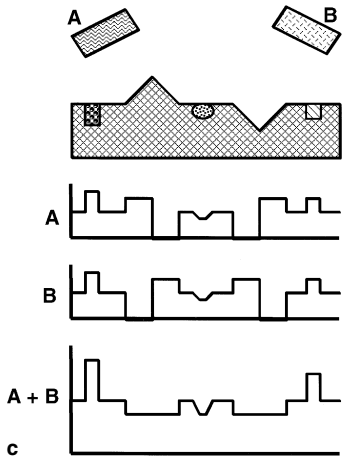


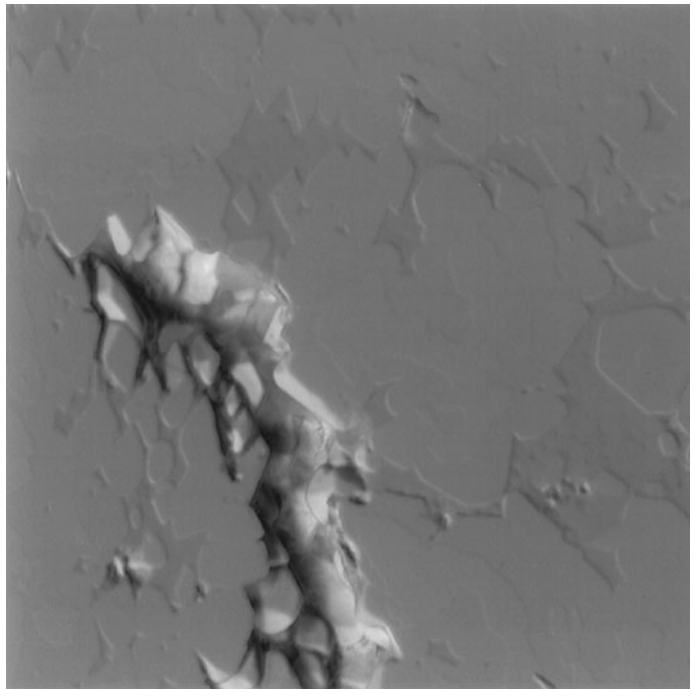
a



b

Figure 4.41. Separation of contrast components. The specimen is Raney nickel, which contains multiple phases that produce atomic number contrast and surface defects that produce topographic contrast. (a) The specimen is viewed with a positively biased E–T detector, showing both topographic contrast in the pit and atomic number contrast on the flat surface. (b) Two-segment solid state BSE detector, sum mode: atomic number contrast dominates the image and the topography of the pit is not strongly evident.

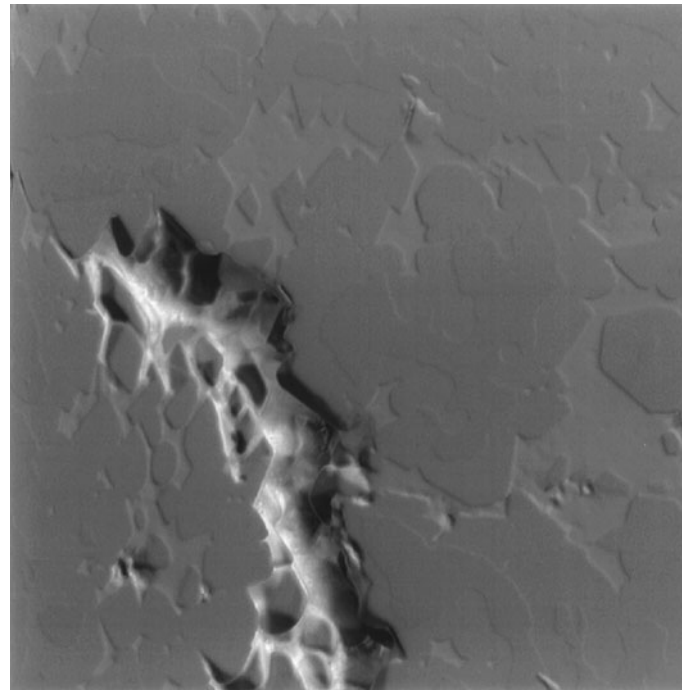




d

Figure 4.41. (*Continued*) (c) Schematic showing detector and signals for the sum mode. (d) Difference mode, A − B; topographic contrast dominates, with apparent top illumination, so that the pit is the strongest feature in the image, whereas the atomic number contrast is almost completely lost.

the usual cosine distribution (if the specimen surface is normal to the beam) and no particular trajectory component. Because the A − B detector difference mode is so sensitive to trajectory effects, the small trajectory asymmetry introduced by the boundary is readily detected, but it must not be confused with real topography.



e

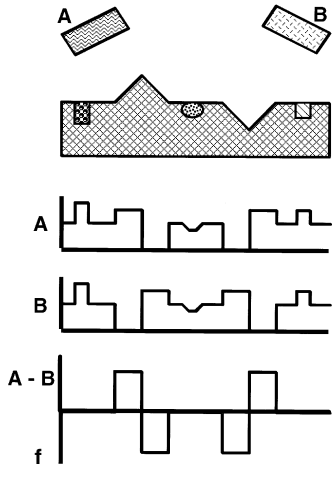


Figure 4.41. (*Continued*) (e) Difference mode, B – A. (f) Schematic showing detectors and signals for the difference mode.

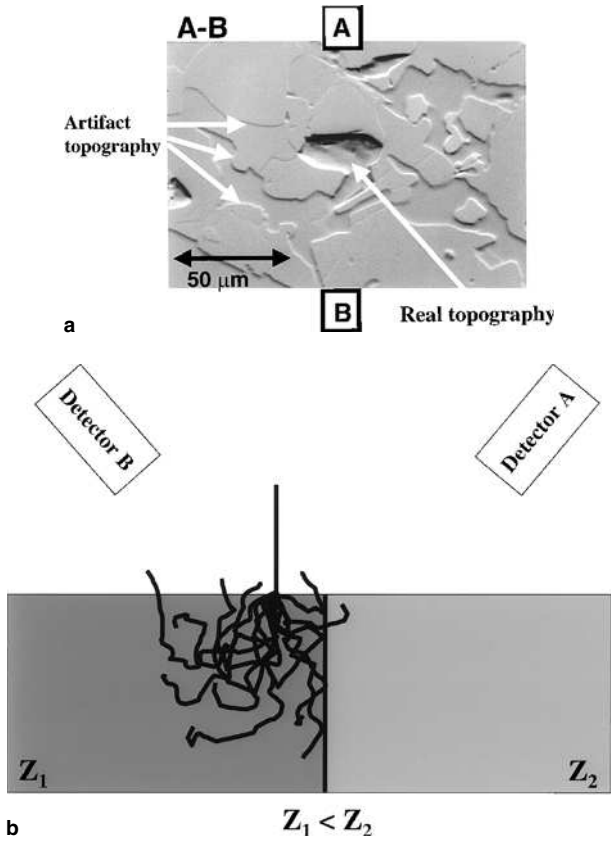


Figure 4.42. (a) Example of artifact contrast in the A–B mode at the phase boundaries of a highly polished, flat surface. (b) Origin of image artifacts at phase boundaries: apparent shadowing due to the development of asymmetry in the BSE due to increased reflection at the boundary.

4.5. Image Quality

Assuming that a specimen has features (such as compositional differences or topography) that can be imaged with the SEM, it is next necessary to ask what conditions must be satisfied to produce a final image on the display screen that conveys this contrast information to the observer. These conditions can be separated into two categories:

1. Ensuring that the contrast information potentially available in the signal is not degraded or obliterated by the presence of random fluctuations ("noise")
2. Applying the techniques of signal processing to render the contrast information in the amplified signal visible to the observer

One of the ways of studying the nature of a specimen in the SEM is to make a line scan across it, that is, plot the signal strength as a function of

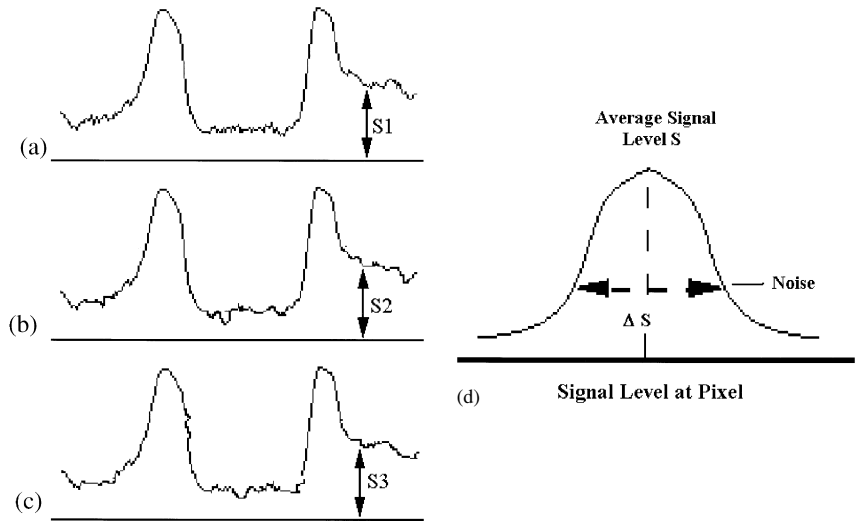


Figure 4.43. Plot of signal as a function of position for three sequential line scans across the same features in an SEM image. Although the same major features are visible in each trace, the intensity S_1 measured at a given pixel in (a) is different from the corresponding intensity S_2 at the same pixel in (b) and S_3 in (c). (d) A histogram of the signal level at a given pixel for a large number of scans that would allow the average signal level S at that pixel to be determined as a function of time for a single pixel. The uncertainty ΔS of this measurement represents the noise in the signal.

the position of the beam. If we do this once, then we might get a profile of the kind shown in Fig. 4.43a which clearly shows the presence of several large and small features. If we now repeat this process, leaving all of the conditions the same as before, then a second line scan profile would be obtained like that shown in Fig. 4.43b. This profile is clearly very similar to that in Fig. 4.43a, but it is not identical because the signal intensity at corresponding pixels is not the same in the two cases. If a third line profile were to be taken, Fig. 4.43c, then again we would find that although the main features of the profile are again evident, on a pixel-by-pixel basis there is a difference in the signal. Repeating this process for a very large number of line scans would never result in the production of a pair of identical profiles.

If the signal intensity at a given pixel were recorded for all of these scans and the values were to be plotted as a histogram, then the result would appear like that shown in Fig. 4.43d. The histogram would be in a Gaussian (bell-shaped) form. The position of the peak represents the average value of the signal S at this pixel position, but the Gaussian shape means that this average value is associated with some degree of uncertainty. The magnitude of this uncertainty is quantified by calculating the standard deviation of the signal from the histogram, and it is this variability that is usually described as being "noise." This is not a felicitous choice of name because noise is usually considered as being something external to the signal of interest and so, at least in principle, of being something which

can be eliminated if proper precautions are taken. In this case, however, if the mean pixel signal level has a value n (e.g., the number of signal carriers such as BSE or SE), then the expected standard deviation (noise) is found to be $n^{1/2}$. This means that as the signal level increases, the noise also increases. The reason for this is that the noise is inherently a part of the signal. Although we think that we are irradiating the sample with a fixed beam current and, for example, measuring the resultant secondary electron signal, the actual number of electrons hitting the specimen per unit time fluctuates randomly because of the way in which the gun emits electrons, and the yield of secondary or backscattered electrons is also variable because each electron trajectory through the specimen is different from every other one.

Since the signal and the noise are coupled in this way, the quality of the signal is best defined by the signal-to-noise ratio S/N:

$$\frac{S}{N} = \frac{n}{n^{1/2}} = n^{1/2}. \tag{4.14}$$

Because either noise and/or genuine contrast effects can change the signal level, the important issue now is to determine under what conditions real signal changes can be distinguished from those due to random noise. Rose (1948) made an extensive study of the ability of observers to detect the contrast between objects of different size and the background in a scanned TV image in the presence of various levels of noise. He found that the average observer could distinguish small features in the presence of noise provided that the change in signal ΔS exceeded the noise N by a factor of at least five:

$$\Delta S > 5N \qquad \text{(Rose criterion)} \tag{4.15}$$

The Rose criterion is not a fundamental physical law [although more recent work by Bright *et al.* (1998) has confirmed its general validity], but only the measure of the typical performance of an average observer. However, it provides a useful basis for deciding under what conditions the detail in a SEM image is meaningful. Specifically, we want to know what is the smallest (or threshold) beam current that will permit a given contrast level ($C = \Delta S/S$) to be discerned. The noise can be considered in terms of the number of signal events *n:*

$$\Delta S > 5n^{1/2}. \tag{4.16}$$

Equation (4.16) can be rewritten in terms of the contrast by dividing through by the signal S:

$$\frac{\Delta S}{S} = C > \frac{5n^{1/2}}{S} = \frac{5n^{1/2}}{n} \tag{4.17}$$

$$C > \frac{5}{n^{1/2}} \tag{4.18}$$

$$n > \left(\frac{5}{C}\right)^2. \tag{4.19}$$

Equation (4.19) indicates that in order to observe a specific level of contrast C, a minimum number of electrons, given by $(5/C)^2$, must be collected at each pixel position. This number of electrons can be converted into a current I_S by writing

$$I_S = \frac{ne}{\tau}, \tag{4.20}$$

where τ is the dwell time per pixel and e is the electron charge (1.6×10^{-19} C). Substituting Eq. (4.19) into (4.20) gives

$$I_S = \frac{25e}{C^2\tau}. \tag{4.21}$$

The signal current I_S differs from the beam current I_B by the efficiency of signal collection ε, a factor which depends on the yield of signal carriers by the beam–specimen interaction (e.g., δ for secondary and η for backscattered electrons), and on the collection efficiency, of the detector (the detector quantum efficiency, DQE; Joy *et al.,* 1996). For example for secondary electrons we could write

$$I_S = I_B\delta\,(\mathrm{DQE})\,. \tag{4.22}$$

Combining Eqs. (4.21) and (4.22) gives

$$I_B > \frac{25(1.6 \times 10^{-19}\,\mathrm{C})}{\delta(\mathrm{DQE})\tau}. \tag{4.23}$$

The pixel dwell time τ can be replaced by the time t_f taken to scan one full frame of the image from the relation

$$\tau = \frac{t_f}{n_{\mathrm{pixels}}}, \tag{4.24}$$

where n_{pixels} is the number of pixels in the image. Substituting Eq. (4.24) into (4.23) gives

$$I_B > \frac{(4 \times 10^{-18})\,n_{\mathrm{pixels}}}{\delta\,(\mathrm{DQE})\,C^2 t_f} \quad \text{(amperes)}. \tag{4.25}$$

For a high-quality image there are typically about 1 megapixels (10^6) in an image, so Eq. (4.25) can be restated as

$$I_B > \frac{4 \times 10^{-12}}{\delta\,(\mathrm{DQE})\,C^2 t_f} \quad \text{(amperes)} \tag{4.26}$$

Equation (4.26), referred to as the *threshold equation,* defines the minimum beam current, the *threshold current,* necessary to observe a certain level of contrast C under specified operating conditions (Oatley 1972). Alternatively, for a specified level of beam current and operating parameters Eq. (4.26) predicts the minimum contrast, or *threshold contrast,* which could be observed in the image. Objects in the field of view below this level of contrast cannot be distinguished from the noise of the random background fluctuations.

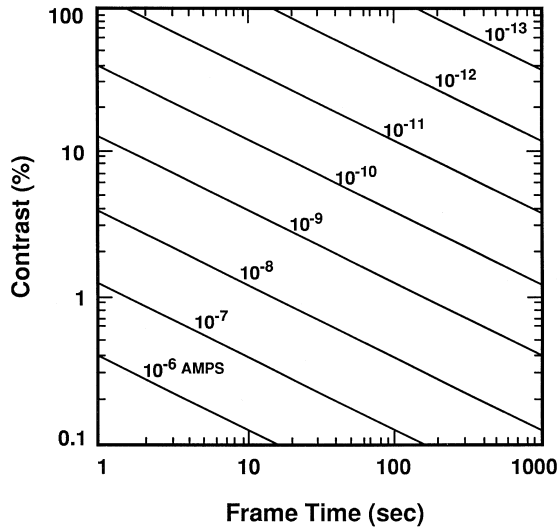


Figure 4.44. Graphical representation of the relationship of parameters of contrast, frame time, and beam current in the threshold equation. Assumptions: The signal production/collection efficiency $\varepsilon = 0.25$ and the scan is 1000×1000 picture elements.

A useful way to understand the relationships of the parameters in the threshold equation is via the graphical plot shown in Fig. 4.44, derived from Eq. (4.26) with the assumption that that the overall signal generation efficiency is 0.25, that is, on average it takes four incident electrons striking the specimen to produce one signal carrying electron (secondary or backscattered) that reaches the detector. This is a reasonable assumption for most materials in SE mode, and for heavier materials imaged using BSE mode. From Fig. 4.44 it can be seen that to image a contrast level of 10% ($C = 0.1$) with a frame time of 100 s, a beam current in excess of 10^{-11} A (10 pA) must be supplied. If the contrast were only 0.5% ($C = 0.005$), then a current of 10^{-8} A (10 nA) would be required. Conversely, for a specified beam current there will be levels of contrast which will not be detectable for given scan times. For example, if a beam current of 10^{-10} A is used for a 100-s frame time, only objects producing more than 5% contrast will be visible. Once we know the current needed to image a given contrast level we can calculate the probe size obtainable with this current, using the equations in Chapter 2. Table 4.6 shows the approximate probe sizes that can image various contrast levels under the conditions specified in Fig. 4.44. It can be seen that poor contrast levels and short recording times produce a severe penalty on the resolution available.

The Rose criterion is actually a conservative estimate of the threshold condition. For an object that forms a large fraction of the field of view of the image or has extended linear detail (such as an edge), the ability of the human brain to process and combine information from many related pixels actually relaxes the strict visibility criterion (Bright *et al.,* 1998). In general, however, the proper procedure is always to ensure that the beam current and scan times chosen are adequate, as computed from the Rose criterion, in order to guarantee visibility of the feature of interest.

Table 4.6. Gaussian Probe Size for Imaging Various Contrast Levels[a]

Contrast	d_{min} (nm)
1.0 (100%)	2.3
0.5 (50%)	4.6
0.25 (25%)	9.1
0.10 (10%)	23
0.05 (5%)	46
0.025 (2.5%)	91
0.01 (1%)	230
0.005 (0.5%)	460
0.0025 (0.25%)	910
0.001 (0.1%)	2300

[a] Tungsten filament, 20 keV; $\varepsilon = 0.25$; 100-s frame; 1000×1000 pixels.

4.6. Image Processing for the Display of Contrast Information

The time-serial nature of the SEM signal permits real-time application of the processing operations to modify the signal actually displayed on the viewing or recording screens of the SEM. With proper signal processing we can readily observe features of very low contrast, provided, of course, that the threshold current criterion has been satisfied. Signal processing enables the microscopist to create an image which is comfortable to view in the illumination conditions of the SEM laboratory. Further, by using digital storage on the SEM, advanced processing techniques which can greatly enhance the utility, presentation, and aesthetic appeal of the image become available. Proper application of both real-time analog and digital processing greatly increases the power of the SEM.

4.6.1. The Signal Chain

The signal passes from the detector to the display screens through a video amplifier. If the beam were stopped on a particular pixel and the amplification of this system was slowly increased, then the corresponding pixel on the display screen would become brighter until ultimately the signal reached a value known as *peak white*. This is the maximum output signal from the amplifier and corresponds to the highest possible brightness on the screen. Similarly, if the amplifier gain were to be reduced, then the screen would become darker until, at the amplifier output known as *black level,* the signal would cease falling and the screen would be at its darkest possible level. The difference between peak white and the black level defines the *dynamic range* of the signal that we can view on the SEM display. When properly adjusted, the time-varying signal should span as much as possible of the dynamic range because this produces the most visually pleasing and useful image (Figs. 4.45 and 4.46). Under no conditions must the amplifiers be driven so as to force the signal to the peak white or to the black level

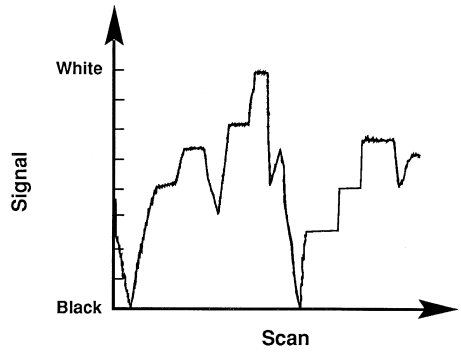


Figure 4.45. Schematic illustration of a line trace display of a signal with characteristics that would produce a "good" image to an observer. The image will be perceived as filling out the full dynamic range of the display, from near full black to near full white, with no saturation.

because this will result in *clipping* as the signal reaches one of the limits of the dynamic range. This condition is readily visible on the display screen by the ugly distortion that it causes, or from the flattening of the peaks on the line profile display.

The properties of the video amplifier are usually set by two controls; one, labeled *brightness,* varies the standing dc level of the amplifier output; the other, labeled *contrast,* varies the gain of the amplifier. By adjusting

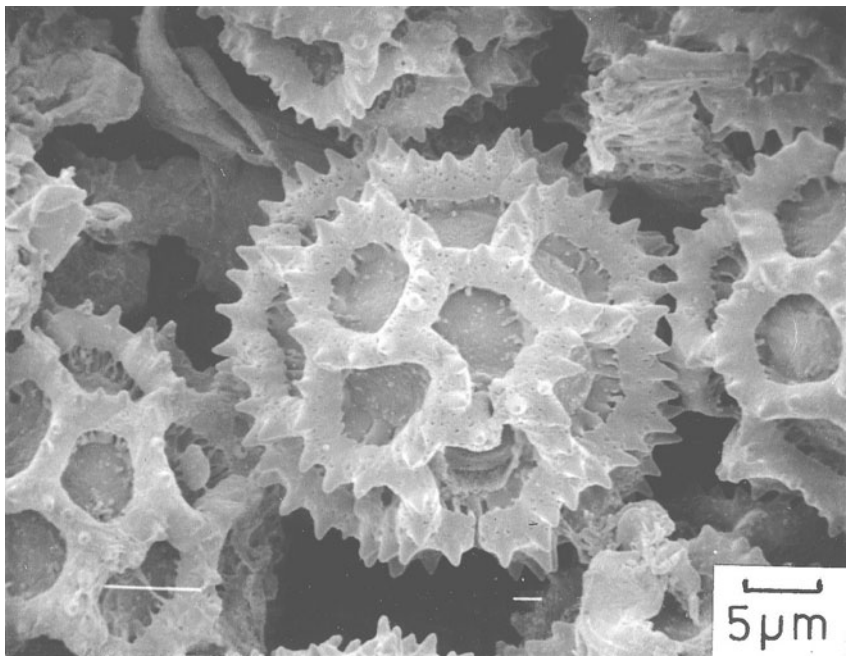


Figure 4.46. Example of linear amplification applied to a sample which produces high natural contrast by the topographic contrast mechanism. Detector: positively biased E–T detector.

these two controls the signal can be set to fill, but not to clip, the full dynamic range of the system. These controls should not be confused with the brightness and contrast controls fitted to each of the microscope display screens, as these only adjust the display properties of the screen to which they refer. Display brightness and contrast are set up before turning the electron beam on by adjusting the controls on each screen to be used so that the scan raster is just visible in the ambient lighting of the SEM room. All subsequent adjustments are then made with the amplifier controls.

4.6.2. The Visibility Problem

If the conditions imposed by the threshold equation are satisfied, and if the specimen contains the features of interest, then the SEM signal should be able to display this information on the screen. However, the contrast produced may not be visible to the operator on the display because of the limitations of our eyes. For most observers the minimum useful image contrast level is about 5%, a value which is actually greater than that found from SEM contrast modes such as atomic number contrast between most adjacent pairs of elements, or magnetic or channeling contrast. Some adjustment of the signal is therefore required to bring the image up to a viewable level and it is here that the brightness and contrast controls of the video amplifier are employed to great benefit.

The procedure, usually called *differential amplification,* is illustrated in Fig. 4.47. The signal profile (Fig. 4.47a) is seen to be safely lying between

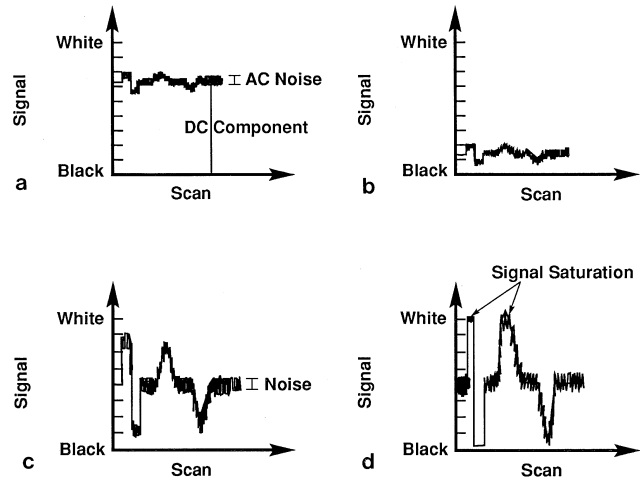


Figure 4.47. Schematic illustration of the steps of applying differential amplification to enhance weak contrast in an image. (a) Line trace of a specimen which produces low natural contrast. The modulations of the signal (contrast) due to specimen features are confined to one or two gray levels and are poorly visible. (b) A constant (dc) level is subtracted from the signal at every point. (c) The difference signal is amplified, expanding the range of gray levels over which the specimen contrast is displayed. (d) Excessive application of differential amplification leads to signal saturation.

the peak white and the peak black levels, but the variation in the signal caused by the contrast events of interest is small. The visibility of the contrast effects cannot be increased just by amplifying the signal because the waveform will then be clipped at the peak white level. An examination of the signal profile shows that it consists of two components, a fixed (dc) signal and a time (i.e., scan position) -varying (ac) signal. The dc signal, which is unvarying, carries no information and so can be discarded. This is done by using the brightness control on the amplifier because this controls the dc level. As shown in Fig. 4.47b, the brightness control is adjusted until the dc component is removed and the ac, information-carrying, portion of the signal is lying just above the black level. On the SEM image screen, the image would now appear to be very dark. The amplifier contrast control is now used (Fig. 4.47c) to increase the signal amplitude so that the waveform fills as much of the dynamic range as possible. The weak contrast of the original image is now readily visible (Fig. 4.48). This procedure can be used to view contrast levels as low as 0.1%, provided that the signal-to-noise ratio is adequately high.

Although differential amplification is a vital operation on the SEM, it must be used with care. If the contrast control is set too high (Fig. 4.47d), then clipping may occur at signal peaks, leading to a badly distorted image. Further, when the contrast is enhanced the visibility of noise in the image is also increased, and excessive contrast adjustment will result in a micrograph which has a grainy or speckled appearance. When this situation is encountered it is necessary either to reduce the amount of contrast expansion, which risks rendering the desired contrast invisible, or to slow down the scan and/or to increase the beam current in order to improve the signal-to-noise ratio.

With the control over the display of contrast provided by differential amplification it might be thought that the limitations of the threshold

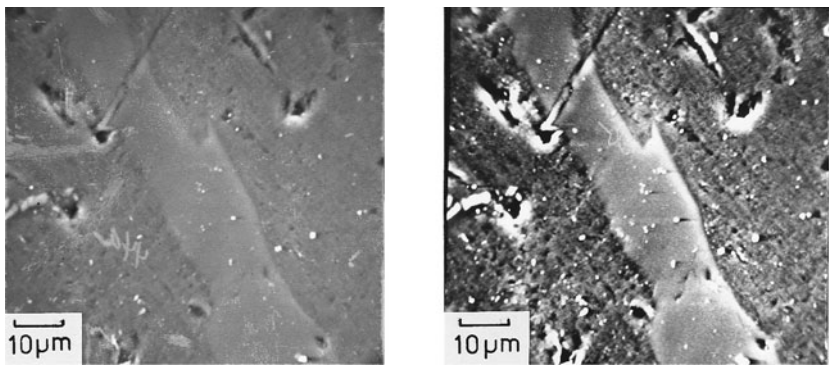


Figure 4.48. Weak atomic number contrast in an aluminum–silicon eutectic enhanced by differential amplification. (a) Natural contrast approximately 7%, displayed with linear amplification. The brighter silicon band is barely visible against the darker, surrounding aluminum. (b) Same image as (a), but with differential amplification applied. Note the bright areas of saturation around some of the voids in the specimen.

equation have been transcended. This is not correct. The critical parameter in the threshold equation is the contrast, which enters as a squared term and therefore has a strong effect on the threshold current condition. The contrast referred to in the threshold equation is the natural contrast from the sample, that is, the differences in signal received at the detector when the beam is scanned to various locations on the specimen, not the value obtained by subsequently processing the image. Although we can increase the display contrast, we do not improve the signal-to-noise ratio at the same time, and so the threshold current situation is not altered.

4.6.3. Analog and Digital Image Processing

Differential amplification is the most important and the most basic image processing operation and, as such, it is available on every SEM of whatever age or origin. However, other signal processing tools are also offered on most instruments. On older machines these options will be in the form of *analog* processing. In analog processing the signal voltage (which is the "analog" of the input from the specimen) is manipulated in a variety of ways to attempt to solve or ameliorate a variety of problems such as excessive signal contrast or the need to enhance a selected contrast range. Although often quite effective, these procedures have the drawback that they irreversibly modify and destroy the original signal. In general, analog signals also have the fundamental limitation that they can only be stored by recording them onto photographic film, which then makes it impossible to further enhance or correct the image. For these reasons all new SEMs now carry out their signal processing and storage operation in a *digital* mode. In a digital image the brightness variations have been transformed to strings of numbers and so the image can readily be stored on a computer. This image storage method has the merit that any modifications to the image can then be made on a copy, rather than on the original, allowing much more safety and flexibility in processing.

All digital images are considered to consist of red, green, and blue (R, G, B) components because any color can be synthesized by appropriately mixing these primary colors. Any desired color can therefore be coded using three numbers representing the fraction of R, G, and B needed to get the desired shade. For example, yellow is produced by equal amounts of red and green, so yellow would be coded as (100%, 100%, 0). For convenience the percentage is instead represented as an 8-bit binary number (i.e., from 0 to 255), so yellow would actually be stored as (255, 255, 0). Thus the color of each pixel is represented by three numbers. Because each number is 1 byte in size, each pixel represents 3 bytes, and because 1 byte is 8 bits in length, the representation of each pixel is therefore 24 bits deep. More than 16 million (actually 16,777,216) colors can be coded in this way. A 1024×1024 pixel color image requires 3,145,728 bytes (3 MB) of storage.

In the SEM the image is always encoded as a gray scale between black and white, and because of this we can save memory. Pure gray tones are

always composed of equal amounts of the RGB primary colors, so we need only store one of the primary color values for each pixel in a gray-scale image. Only 256 shades of gray can be coded in this way, but even this number easily exceeds the actual number that an average observer can discern (typically 100–140).

In all modern SEMs the first step is to temporarily store the image in this digital format in a live, electronic, framestore. The contents of the framestore are used to produce the display for the operator, and the framestore is also often used for some real-time image processing. In one mode the framestore accumulates several incoming image frames, acquired at TV rates or at slower scan speeds, and *integrates* them to form a final composite image (Fig. 4.49). Frame integration enhances the signal-to-noise ratio because signal detail will be constant and will add, whereas noise is random and so variations in successive frames will tend to cancel themselves out. Using this approach, images acquired at TV scan rates and at relatively small beam currents can be displayed with acceptable signal-to-noise ratios. Alternatively the framestore may be used as a *recursive filter*. In this mode the image displayed is the running average of the last 5 or 10 previous frames, the information in each pixel being weighted so that the most recent frame contributes the highest fraction of the final result. This also produces a significant enhancement in signal-to-noise ratio while still maintaining an almost real-time live image.

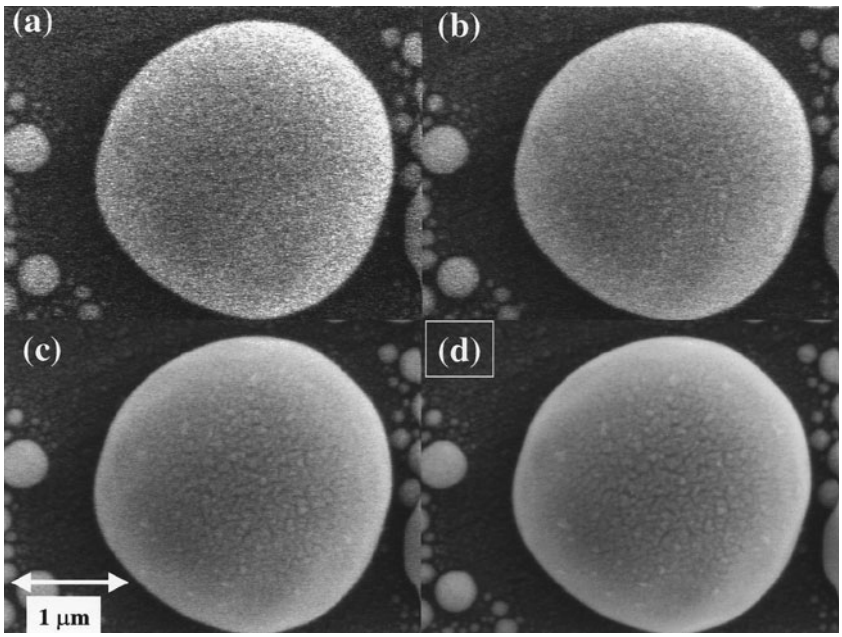


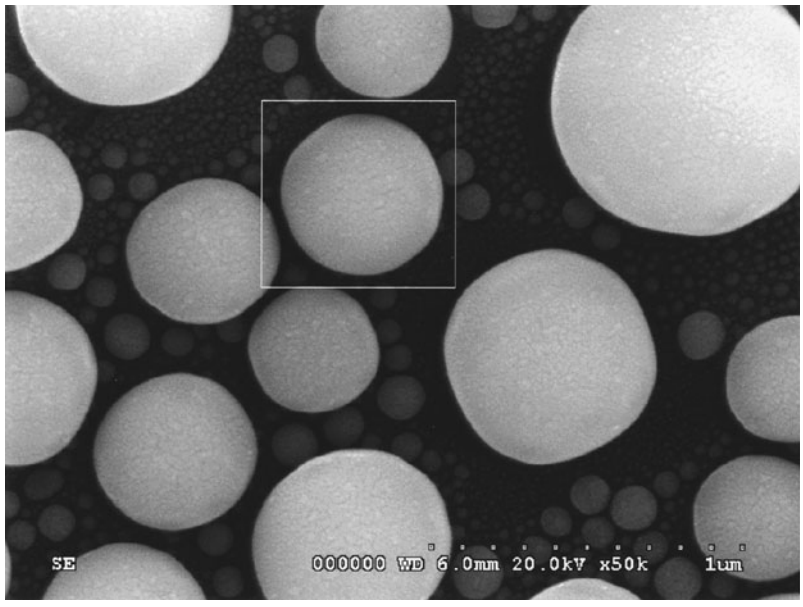
Figure 4.49. The use of a digital framestore as an image integrator. (a) One frame of the input signal, (b) the sum of 2 frames, (c) the sum of 8 frames, and (d) the sum of 64 frames. Note the significant reduction in noise and the increase in the visibility of detail.

Although digital image processing can be performed on-line, it is usually more efficient and practical to do it off-line because in this way the processing can be carried out without tying up valuable time on the SEM itself. It is thus necessary to store the contents of the framestore into an archival digital format so that it can be processed when convenient. Several image file formats are in common use for this purpose, the four most popular being bitmap (.BMP), graphics interchange format (.GIF), tagged image file format (.TIF), and joint photographic experts group (.JPG) format.

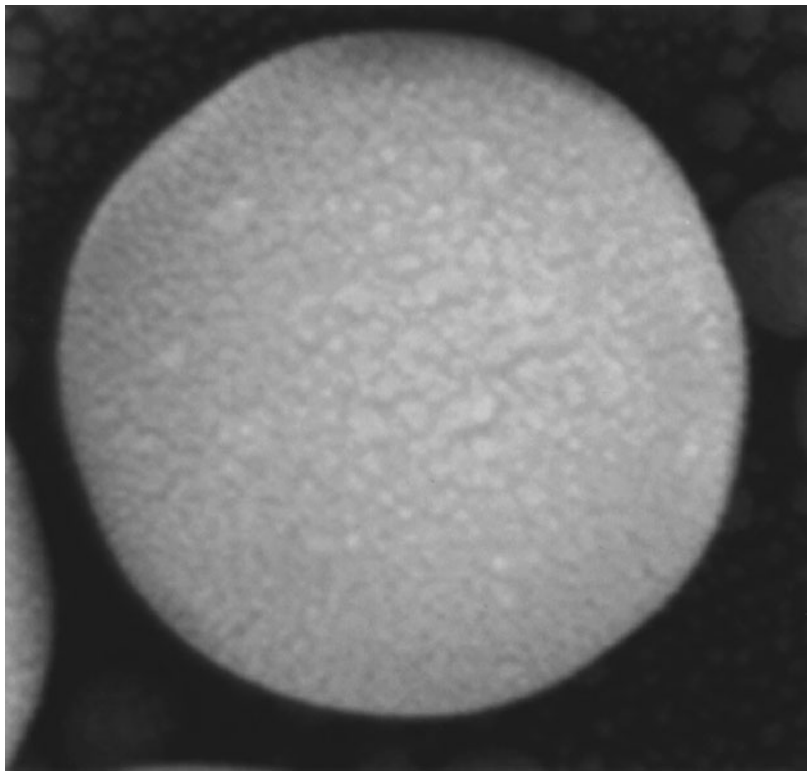
The bitmap simply stores the pixel matrix values directly, there is no data compression and hence no loss of quality. The bitmap is therefore the simplest file format and the easiest to use, but it gives huge files. A typical 1024 × 1024 SEM micrograph will require about 1 MB of storage. The GIF format compresses the data from the image before storage by looking for repetition in the data stream (for example, successive pixels all of the same value). This procedure works well for drawings, but less effectively for micrographs. The advantages of the GIF format are that it provides significant compression (a typical 1024 × 1024 micrograph occupies just about 50 kB) and there is no loss of quality. But GIF is limited to 8-bit black-and-white (or color) images and it is a proprietary format belonging to Compuserve, which means that software to produce it requires a license fee from the user. TIF files carry not just the image data, but also many "tags" which contain other data such as printer settings, aspect ratio, calibration, etc. They have the advantage that they can handle 24-bit images and offer moderate compression: a typical micrograph is about 650 kB in size. The complexity of the TIF format, however, and the fact that it is frequently updated, can lead to compatibility problems with older software. The JPG format uses an aggressive data compression algorithm relying on perceptual coding techniques. This provides very good, and adjustable, data compression (the typical micrograph is about 100 kB) and the ability to handle 24-bit images. There is a loss of image quality at the higher compressions, however, and artifacts are generated by some kinds of image detail. An example of the loss of image fidelity at high jpeg (.JPG) compression is shown in Fig. 4.50. Figure 4.50a shows an image stored with the TIF format. The same image stored with a highly compressed jpeg yields a very similar image when both formats are printed on a photographic-quality printer. When the image is magnified by a factor of four, however, the difference in image fidelity becomes more obvious, as shown in Figs. 4.50b and 4.50c. The expanded jpeg shows coarse pixels at the edge of the sphere, and the fine details on the sphere itself are beginning to break up because of the level of the pixelization. Because the cost of data storage continues to fall, the bitmap format is the most reliable recommendation.

4.6.4. Basic Digital Image Processing

Once the SEM micrograph is stored in a suitable format it can then be processed either on the computer that runs the SEM or on a separate

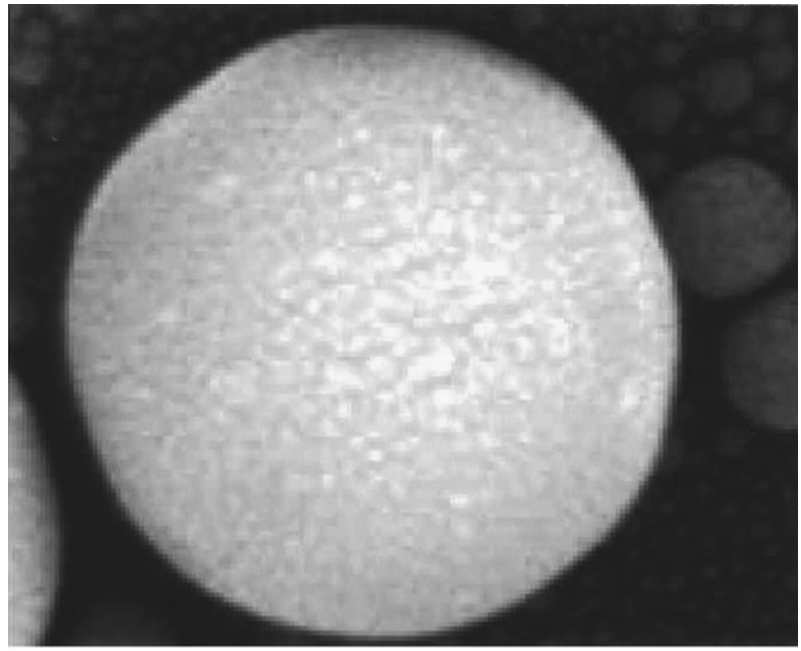


a



b

Figure 4.50. An image of tin spheres stored as (a) a .TIF file format occupying 306 kB; (b) a portion of the TIF image selected and enlarged by a factor of four; (c) the same area selected and enlarged by a factor of four from an original jpeg (.JPG) file occupying 42 kB. Note the degradation of image quality and the obvious pixelation in (c) compared to (b).



c

Figure 4.50. (*Continued*)

personal computer. A wide range of commercial, as well as free, software is now available for image processing and a vast variety of options suitable for all purposes is available. Digital image processing can be divided into two categories, *digital image enhancement,* which seeks to render information in the image more readily visible to the operator, and *digital image interpretation,* which provides automated measurement and analysis of features, such as particles, in the field of view.

The properties of the digital image are conveniently summarized by the image histogram (Fig. 4.51), which plots the number of pixels associated with each gray level (0–255). In the ideal image every possible gray level will be represented by some fraction of the pixels so that the dynamic range is fully occupied. If the histogram shows that this is not being achieved, then the contrast and brightness controls should be adjusted as discussed above until the histogram fills the dynamic range. This should be done even though most systems provide some means of "histogram equalization" to redistribute the intensity variation across the entire range. If the original recorded signal only covers a few of the available gray levels, then histogram equalization will spread these out to span the dynamic range, but will not add any new levels. As a result the image will look posterized (i.e., having discrete blocks of each brightness level) instead of varying smoothly. Digital imaging processing cannot be used to excuse the operator from the responsibility of acquiring an acceptable image in the first place.

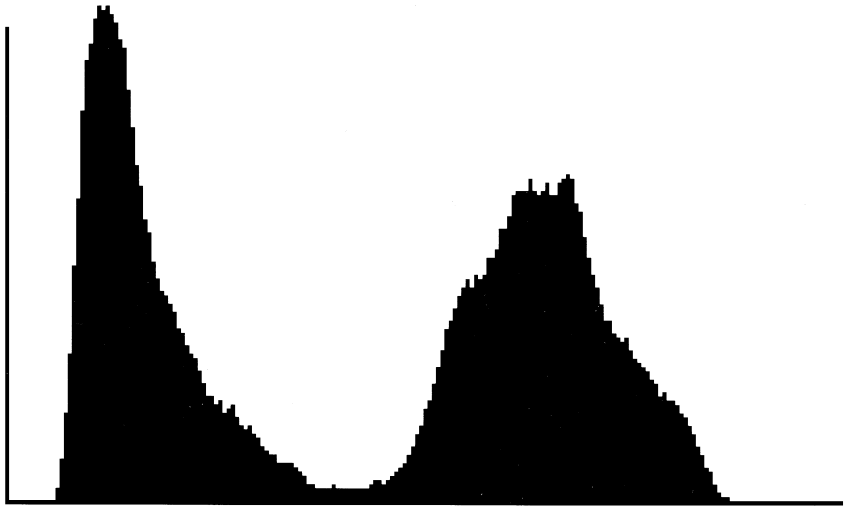


Figure 4.51. Histogram of Fig. 4.50 showing that the signal fills most of the available dynamic range and has two main, well-separated bands of contrast.

4.6.4.1. Digital Image Enhancement

Digital processing can mimic any of the conventional analog image processing operations as well as provide entirely new tools for use. The tasks performed are carried out in one of two distinct ways, either on the basis of on pixel at a time or by considering a group of pixels at a time. In both cases the algorithm works on an archived and protected copy of the original image and produces as output a new image containing the modified data. Many digital processing tools are presented to the user in the form of a graphical interface, so that the mathematics presented below is hidden, but the user can visually inspect the relationship between the input signal and the output signal as various operations are selected.

Single-pixel operations consist of simple mathematical transformations on the data. For example, if the pixel $P_{x,y}$ at position x, y has an original value $I_{x,y}$, then the intensity $I'_{x,y}$ in the corresponding pixel $P'_{x,y}$ of the output image can be written as

$$I'_{x,y} = A(I_{x,y} + B), \tag{4.27}$$

where A and B are constants. This operation has the same effect on the digital image as the differential amplification method had on the analog image, with the parameter A being the contrast control and B representing the brightness control. Other popular variants include operations of the form

$$I'_{x,y} = F(I_{x,y}), \tag{4.28}$$

where the function F can include raising to some exponent (which will expand or contract selected portions of the dynamic range) or taking a logarithm (which compresses the overall dynamic range).

The most powerful options for digital image processing involve treating pixels not individually, but as a group or "kernel". The assumption is that the intensity $I(x, y)$ at the central pixel x, y will be related in some way to the intensity of the pixels that are its neighbors, both because the beam interaction volume is large enough to spread over many pixels and because in nature most things have a degree of continuity and so a given set of characteristics will typically be shared by several pixels. An example of a kernel is

$$\begin{array}{lll} I(x-1, y-1) & I(x, y-1) & I(x+1, y-1) \\ I(x-1, y) & I(x, y) & I(x+1, y) \\ I(x-1, y+1) & I(x, y+1) & I(x+1, y+1). \end{array} \tag{4.29}$$

It is usual to consider a 3×3 kernel of pixels centered around the target pixel, although larger kernels (5×5 or even 7×7) are sometimes used. The general equation for kernel processing relates the new value $I'(x, y)$ to the original value $I(x, y)$ by the formula

$$\begin{aligned} I'(x, y) = {} & A \cdot I(x-1, y-1) + B \cdot I(x, y-1) \\ & + C \cdot I(x+1, y-1) + D \cdot I(x-1, y) \\ & + E \cdot I(x, y) + F \cdot I(x+1, y) \\ & + G \cdot I(x-1, y+1) + H \cdot I(x, y+1) \\ & + J \cdot I(x+1, y+1), \end{aligned} \tag{4.30}$$

where the coefficients A through J are constants. It is customary to divide the result of the operation at each pixel by the sum of the coefficients so as to keep the pixel intensity within an 8-bit span.

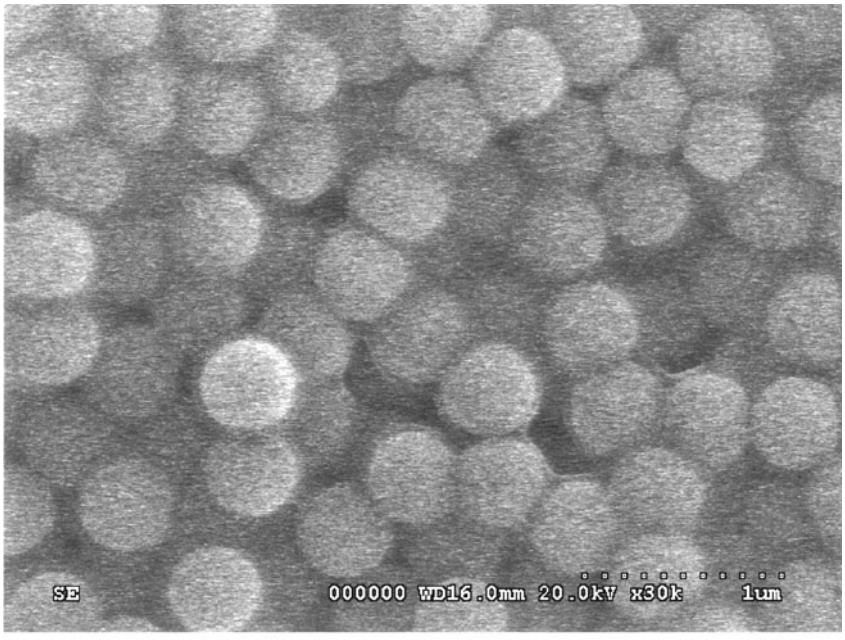
The kernel operator is conveniently represented by the shorthand form

$$\begin{array}{lll} A & B & C \\ D & E & F \\ G & H & J \end{array} \tag{4.31}$$

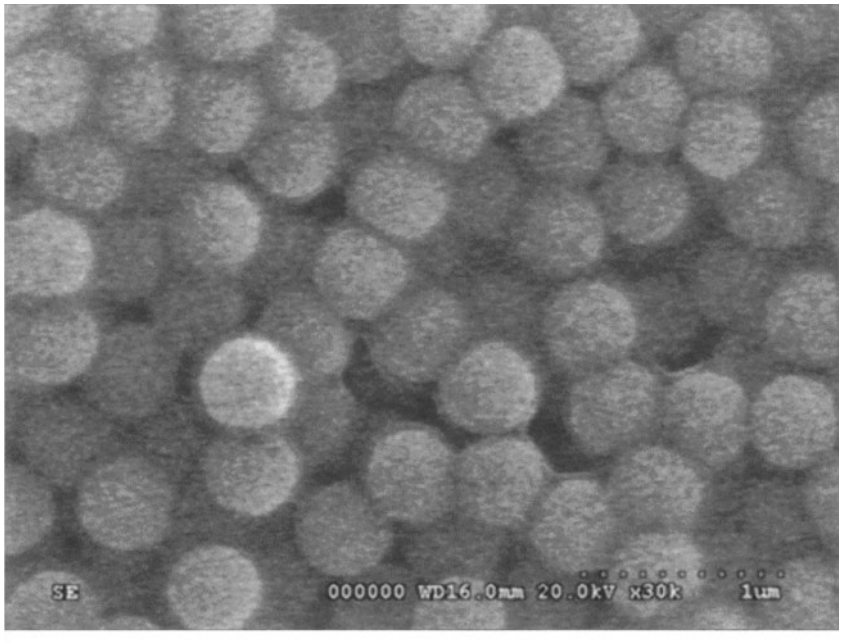
where the array of coefficients is assumed to have the functional effect shown in Eq. (4.30). The computation is performed on a copy of the original image, sequentially selecting the target pixel x, y, then applying the kernel operation equation, and constructing the new image pixel $I'(x,y)$. For a typical 1024×1024 pixel image this requires nearly 10 million arithmetic operations, so a fast computer is mandatory. Note that kernel procedures cannot be correctly applied at the edges or the corners of an image. For example, when using a 3×3 kernel the first pixel that can be processed is two pixels in and two pixels down from the corner. Thus the application of a kernel operator to an image reduces its effective size by at least one row of pixels, at the top, bottom, and on both sides, and repeated applications can have a significant effect on the image size.

Figures 4.52a and 4.52b show the effect of the "smoothing" kernel

$$\begin{array}{lll} 1 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 1 \end{array} \tag{4.32}$$

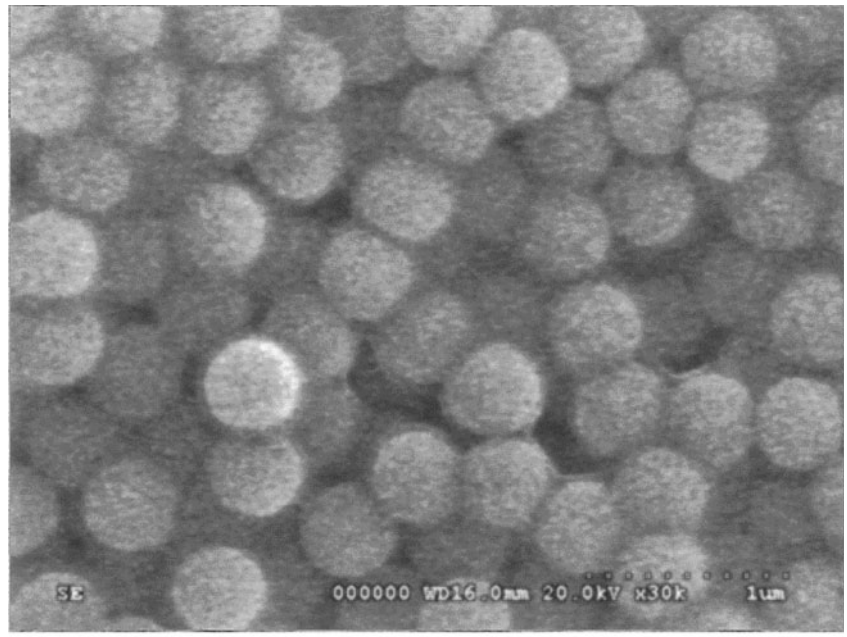


a



b

Figure 4.52. Comparison of (a) the original, noisy image and (b) the same image after processing with the simple smoothing kernel. Note the reduction in noise, but also the slight loss in image sharpness. (c) The image of (a) after processing with a 3×3 Gaussian smoothing kernel. Note the reduction in noise, but more preservation of image detail than that shown by the simple smoothing kernel.



c

Figure 4.52. (*Continued*)

on a noisy image. This kernel computes the center-weighted average of the pixel intensities around the target pixel. All the neighbor pixels are treated equally, but the center, target, pixel is weighted by a factor of four. Assuming that all this group of pixels all equally affected by random noise, then summing in this way will reduce the visibility of the noise by a factor of about three.

An alternative smoothing kernel has the form

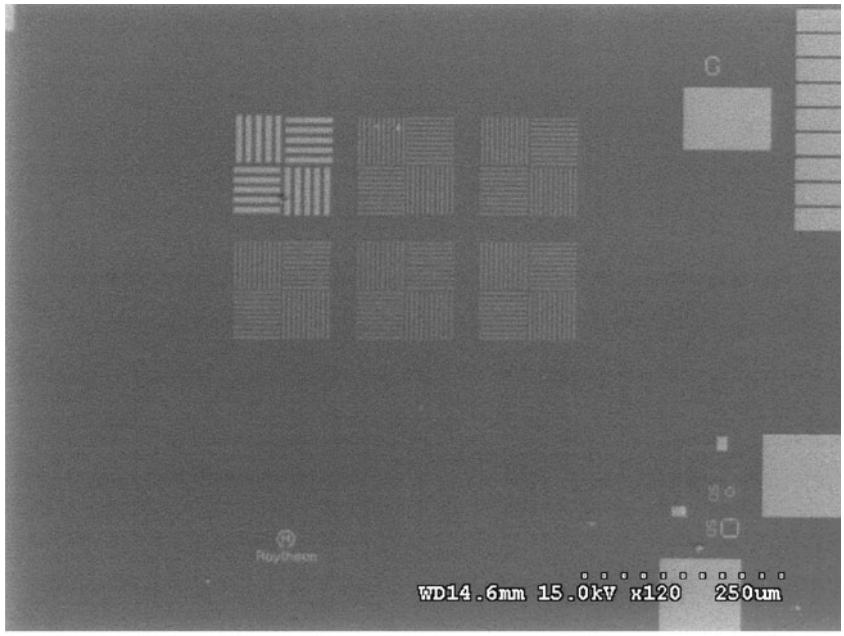
$$\begin{matrix} 1 & 4 & 1 \\ 4 & 12 & 4 \\ 1 & 4 & 1 \end{matrix} \tag{4.33}$$

which approximates the weighting of a Gaussian function with a standard deviation of about 0.4 (Fig. 4.52c). Either of these kernels is useful for enhancing the appearance of a noisy image, although the reduction in noise will inevitably be accompanied by some slight blurring of details in the micrograph.

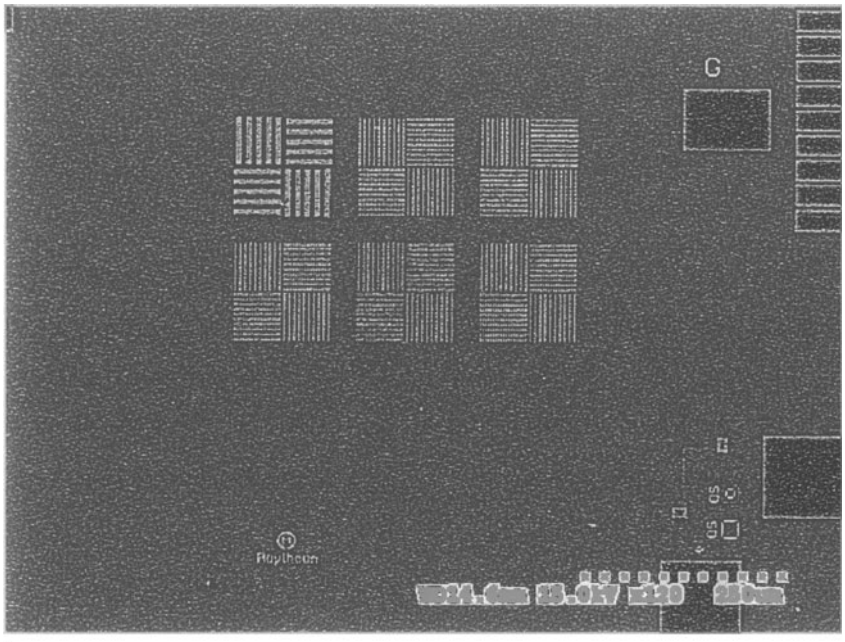
The kernels

$$\begin{matrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{matrix} \quad \text{and} \quad \begin{matrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{matrix} \tag{4.34}$$

when used together form a "sharpening Laplacian" operator, which has the effect of making detail in an image crisper and more visible (Fig. 4.53).



a



b

Figure 4.53. (a) Original image as recorded and (b) the same image after processing with a sharpening Laplacian kernel. Note the increased definition and enhancement of edges and contrast in the processed image.

The first kernel extracts and enhances edge detail in the image, and the application of the second kernel then adds these data back on top of a copy of the unprocessed image so as to maintain the original gray-level information.

It must be remembered that whereas "smoothing" and "sharpening" look and sound as if they are complementary events, in reality the sequences of operations Smooth + Sharpen and Sharpen + Smooth do not result in the same end product, and neither combination will restore the image to its original unprocessed state, an excellent example of the need to preserve the original image data in a protected archive before starting any manipulation of the data.

A rich variety of kernels providing many sophisticated and powerful options is available as part of image processing programs, and the simplicity with which new kernels can be written also encourages experimentation. Kernel processing offers greater flexibility and better results, with a lower incidence of artifacts than can be achieved using analog or single-pixel digital methods. When artifacts do occur with kernel processing procedures, the artifacts generally occur at the scale of the processing kernel, a good argument for keeping the processing kernel small.

4.6.4.2. Digital Image Measurements

Once an image is captured in digital form, then a wide range of measurements on features in the image field, such as counting objects, or finding the area, center of mass, perimeter length, aspect ratio, etc., of particles, can be performed automatically and at high speed. The essential first step in every case is to convert from a gray level to binary image, which is done by thresholding the picture. Note that in this context, "thresholding" is in no way related to the threshold equation described previously. The image processing term "threshold" refers to a user-selected gray-level value. Every pixel above the selected threshold level is painted white, every pixel below the threshold level is painted black. The threshold is selected so as to separate the features of interest from their background. Because any SEM micrograph will contain some random noise, this separation will not be perfect. The next step therefore is to remove any noise speckle by a set of logical operators called erosion (which identifies and removes single black pixels in white areas) and dilation (which identifies and paints black any individual white pixels in black areas). Finally, a specialized suite of kernel operations replaces each white area by a continuous boundary, and then measures and computes the actual parameters requested by the user. In this way hundreds of items in a field of view can be counted and characterized in about 1 s with little or no operator intervention.

References

Bright, D. S., D. E. Newbury, and E. B. Steel (1998). *J. Microsc.* **189,** 25.

Everhart, T. E., and R. F. M. Thornley (1960). *J. Sci. Instr.* **37,** 246.

Fiori, C. E., H. Yakowitz, and D. E. Newbury (1974). In *SEM/1974,* IIT Research Institute, Chicago, p. 167.

Gedcke, D. A., J. B. Ayers and P. B. DeNee (1978). In *SEM/1978/1,* SEM Inc., AMF O'Hare, Illinois, p. 581.

Goldstein, J. I., H. Yakowitz, D. E. Newbury, E. Lifshin, J. Colby, and J. Coleman (1975). *Practical Scanning Electron Microscopy,* Plenum Press, New York.

Joy, D. C. (1991). *J. Microsc.* **161,** 343.

Joy, D. C., C. S. Joy, and R. D. Bunn (1996). Scanning 18, 181.

Kimoto, S., and H. Hashimoto (1966). In *The Electron Microprobe,* Wiley, New York, p. 480.

Moll, S. H., F. Healy, B. Sullivan, and W. Johnson (1978). In *SEM/1978/1,* SEM, Inc., AMF O'Hare, Illinois, p. 303.

National Institute of Standards and Technology (2002). *Standard Reference Materials Catalog,* NIST, Gaithersburg, Maryland. Available at www.nist.gov.

Oatley, C. W. (1972). *The Scanning Electron Microscope, Part 1, The Instrument,* Cambridge University Press, Cambridge.

Oatley, C. W., W. C. Nixon, and R. F. W. Pease (1965). In *Advances in Electronics and Electron Physics,* Academic Press, New York, p. 181.

Pawley, J. B. (1974). In *SEM/1974,* IIT Research Institute, Chicago, p. 27.

Peters, K.-R. (1982). In *SEM/1982/IV,* SEM, Inc., AMF O'Hare, Illinois, p. 1359.

Peters, K.-R. (1984). In *Electron Beam Interactions with Solids for Microscopy, Microanalysis, and Microlithography* (D. F. Kyser, H. Niedrig, D. E. Newbury, and R. Shimizu, eds.), SEM, Inc., AMF O'Hare, Illinois, p. 363.

Reimer, L., and B. Volbert (1979). *Scanning* **2,** 238.

Robinson, V. N. E. (1975). In *SEM/1975,* IIT Research Institute, Chicago, p. 51.

Rose, A. (1948). In *Advances in Electronics* (A. Marton, ed.), Academic Press, New York, p. 131.

Vladar, A. E., M. T. Postek, N. Zhang, R. Larrabee, and S. Jones (2001). Reference Material 8091: New Scanning Electron Microscope Sharpness Standard, In *Proceedings SPIE* 4344–104.

Wells, O. C. (1974a). *Scanning Electron Microscopy,* McGraw-Hill, New York.

Wells, O. C. (1974b). In *SEM/1974,* IIT Research Institute, Chicago, p. 1.

5

Special Topics in Scanning Electron Microscopy

5.1. High-Resolution Imaging

5.1.1. The Resolution Problem

From the basic description in Chapter 4, the SEM image formation process can be summarized as a geometric mapping of information collected when the beam is sequentially addressed to an x–y pattern of specific locations on the specimen. When we are interested in studying the fine-scale details of a specimen, we must understand the factors that influence SEM image resolution. We can define the limit of resolution as the minimum spacing at which two features of the specimen can be recognized as distinct and separate. Such a definition may seem straightforward, but actually applying it to a real situation becomes complicated because we must consider issues beyond the obvious problem of adjusting the beam diameter to the scale of the features of interest. The visibility of a feature must be established before we can consider any issues concerning the spatial scale. For a feature to be visible above the surrounding general background, we must first satisfy the conditions contained within the threshold equation (4.26). For a specified beam current, pixel dwell time, and detector efficiency, the threshold equation defines the threshold contrast, the minimum level of contrast ($C = \Delta S/S_{\max}$) that the feature must produce relative to the background to be visible in an image presented to the viewer with appropriate image processing.

What controls the localization of the signal to the feature of interest to permit the development of this contrast? The beam size upon entering the specimen sets a lower limit on the signal localization and the resolution that can be potentially achieved. However, as discussed in Chapter 3, scattering of the beam electrons creates an interaction volume that determines the spatial characteristics of the BSE and SE imaging signals. Depending on specimen composition and incident beam energy, BSE and SE signals suffer considerable delocalization from the beam entrance footprint. In Fig. 4.8, for the low-magnification regime ($\leq 1000\times$), the combined beam and BSE

(90%) signal delocalization footprint is shown in terms of pixel overlap. For a beam energy of 10 keV and even with a beam only 10 nm in diameter, adjacent pixels are already significantly overlapped at a magnification of 1000× for a low-atomic-number matrix such as carbon or aluminum. Moreover, for intermediate atomic numbers, it is clear that pixel overlap will begin to occur at a magnification of approximately 2000×. Only for the high-atomic-number materials such as gold is overlap avoided until at least the beginning of the high-magnification range (>10,000×). These are actually very modest magnifications. How can we achieve useful high-resolution imaging at finer spatial scales?

The signal footprints shown in Fig. 4.8 represent the full range of backscattered electrons. When we consider the details of the BSE and SE signals, shown schematically in Fig. 5.1, it quickly becomes clear that not all BSE and SE have the same spatial properties. We can recognize different classes of BSE and SE. The secondary electrons generated as the beam enters the specimen, designated SE_1, are localized within a few nanometers of the impact sites of the incident high-energy electrons and can respond to local fine-scale features. The SE_1 component represents a high-resolution signal. Indeed, SE_1 are the only true secondary electrons. As the beam electrons scatter and propagate within the specimen, some emerge as backscattered electrons. If we energy select only BSE that retain a large fraction of the beam energy, for example, 98%, we can define a "low-loss" component BSE_1 whose range is restricted to the immediate vicinity of the incident beam and which constitutes another high-resolution-capable signal.

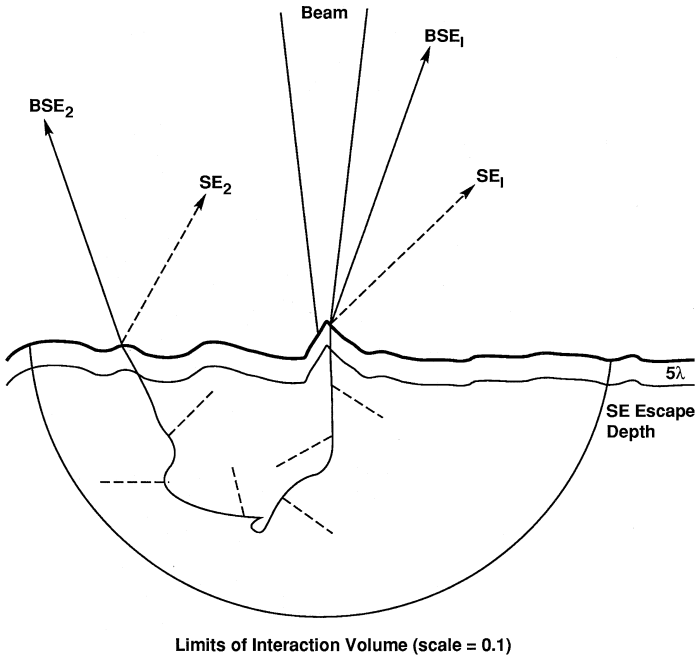


Figure 5.1. Backscattered and secondary electron signals emitted with a finely focused high-resolution beam.

Backscattered electrons create secondary electrons while exiting the surface. Any SE generated by the exiting BSE_1 will clearly also contribute to the high-resolution SE_1. But BSE that lose progressively more energy will travel further from the beam impact point and become delocalized, unable to retain information about the specimen in the beam impact area. The secondary electrons created by these BSE are designated SE_2. Although these SE_2 are secondary electrons, their number depends upon the amount of backscattering, so the information they carry is actually the same as the delocalized backscattered electrons. Finally, SE_3 secondaries are formed where the backscattered electrons strike the objective lens and chamber walls. SE_3 also scale in number with backscattering and again represent a delocalized BSE process, not true secondary electron information.

To achieve useful high resolution, we can seek to separate the high-resolution signals from the low-resolution signals. Alternatively, we can seek operating conditions that create similar ranges for all of the signals.

5.1.2. Achieving High Resolution at High Beam Energy

Some separation of the SE_1/BSE_1 and SE_2/BSE_2 signals can be achieved by choosing an appropriate ratio between the image field of view and pixel size and the beam interaction volume. As shown in Fig. 5.2, for a 30-keV probe focused to 1 nm on a silicon target, the SE_1/BSE_1 component will have a spatial width of about 2 nm, but the range of the primary electrons, $R_{KO} = 9.5\ \mu$m, is so great that the BSE_2 and SE_2 signals emerge over an area approximately 10,000 times larger in radius. If the sample is scanned at a magnification less than 10,000× (equivalent to a 10 μm × 10 μm scan field with 10-nm pixels for 1024 × 1024 digitization), then the field of view is many micrometers across and larger than the electron range. As the beam moves from pixel to pixel, the SE_2 component will change in response to features of the field of view on the size scale of one pixel and larger. The SE_1/BSE_1 signal will also change, but because this component is coming from a region which is much smaller than the pixel size, the contributions

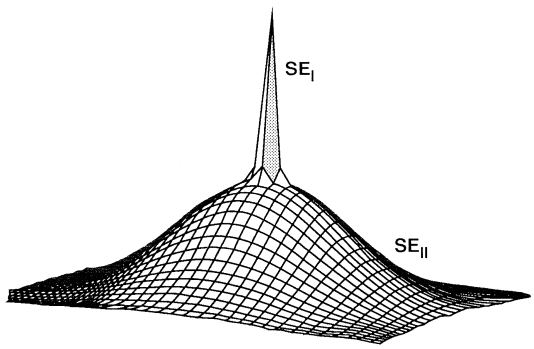


Figure 5.2. Secondary electron emission for a 1-nm, 30-keV probe focused onto a silicon target. The SE_1 has a FWHM of 2 nm and the SE_2 component has a FWHM of approximately 10 μm.

from each pixel will not form a continuous signal, but will be discrete and random like noise. However, at very high-magnification, $>100{,}000\times$ (equivalent to a 1×1-μm scan field with 1-nm pixels for 1024×1024 digitization), the field of view is much smaller than the diameter of the interaction volume. As the beam is scanned within this area, the effective change in the SE_2 distribution is negligible, and so the absolute magnitude of the SE_2 component remains relatively constant. Although the SE_2 component has not been removed from the signal, by forcing it to be relatively constant its contribution to the detail in the image is minimized. However, the random noise contributed by the SE_2 component is significant, and contributes to resolution limitations imposed by the signal-to-noise ratio. The changes in the total signal are almost all due to the SE_1 component, and since the pixel size is now comparable with the SE_1 generation volume, a signal waveform that preserves the high-resolution information in the image is produced. An example is shown in Fig. 5.3 of a high-resolution image prepared with a field emission SEM operated at 35 keV, showing the surface of the magnetic storage medium of a computer hard disk.

When low-atomic-number targets, for example, the typical matrix of biological or polymer samples, are to be imaged, the low secondary electron yield typical of such compositions creates limitations to high-resolution imaging which arise from problems of poor signal-to-noise ratio. The signal-to-noise ratio can be improved by the application of an ultrathin (3–5 nm) metal coating. Such layers are formed from individual metal crystals 2–3 nm in size. Because this size is comparable with the escape depth for secondary electrons, such layers exhibit a very high secondary electron yield. Gold–palladium, platinum, and tungsten are the preferred choices for such coatings because they can readily be deposited in a low-cost sputter coater. For magnifications lower than about $100{,}000\times$ the grains are sufficiently fine to be effectively invisible, but at higher magnifications the individual particles are discernible and so may produce artifacts in the

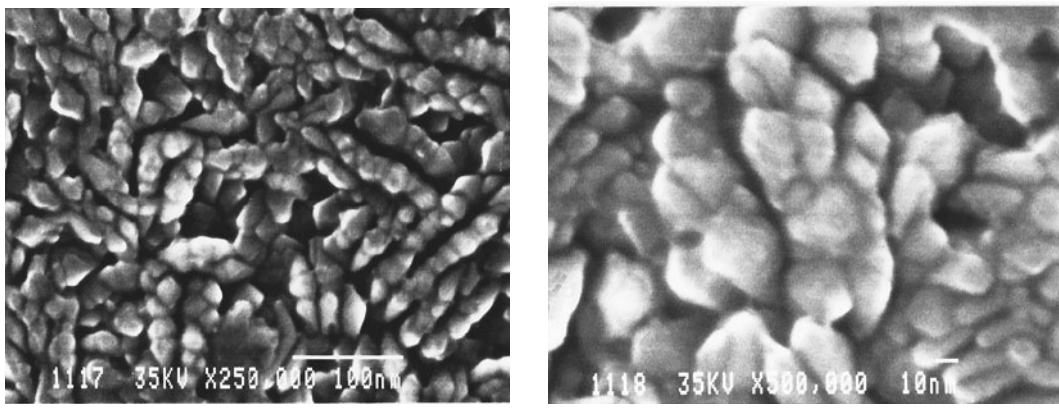


Figure 5.3. High-resolution secondary electron images of the surface of the magnetic storage of a computer hard disk. Images recorded at 35 keV on a field emission SEM with signal collection from an E–T detector. Two magnifications of the same area are shown.

image or even mask real surface detail. On the other hand, the ultra-fine-grain coating makes accurate focus and astigmatism correction very easy and many users find this to be a sufficiently valuable asset to outweigh the disadvantages.

To avoid artifacts arising from such structure in the coatings, advanced thin-film deposition techniques such as ion beam deposition and alternative choices of coating metals, such as titanium, vanadium, and particularly chromium, must be used (Peters, 1982, 1985). Under controlled conditions, continuous films exhibiting no visible structure and as thin as 1 nm can be deposited. These ultrathin coatings are less effective at enhancing the signal-to-noise ratio than a particulate coating, but they play an important role in contrast formation and image resolution at high-magnification. As noted above, the SE_1 signal comes from a region of the sample a few nanometers in size as a result of diffusion and delocalization. The actual spread, or mean free path ζ, of the emitted signal varies with material and can be as small as 1 nm for a heavy metal. For materials such as carbon, polymers, biological tissue, or even silicon, ζ is of the order of 5 nm. This diffusion of the SE_1 produces the familiar bright white line along the edge of every feature in a high-magnification SE image because when the incident beam is close enough to an edge then SE can escape through both the top and the side surfaces of the sample (rather than just through the top) leading to about a doubling of the signal (Fig. 5.4) over a region equal to the mean free path in size. If a feature becomes so small that it has edges which are separated by a distance smaller than the length of the SE mean free path ζ, then the bright lines from the two edges will overlap, neither edge will be visible, and hence the feature is no longer resolved because its size and shape are indeterminate. However, if the surface is covered with a continuous layer of metal, then all of the SE_1 generation is localized in, and comes only from, the film. As shown in Fig. 5.5, the magnitude of the SE_1 signal emitted varies monotonically with the thickness of the metal layer

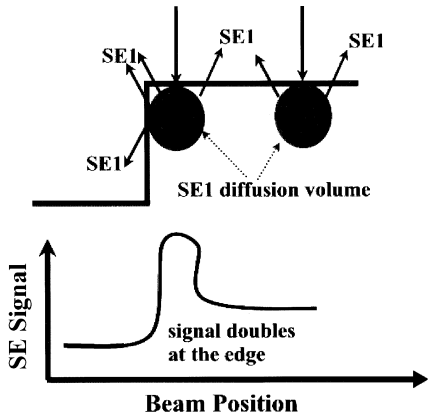


Figure 5.4. Origin of the edge bright line as a result of the finite size ζ of the SE_1 diffusion volume; as the boundary is approached, the surface area encountered by the interaction volume increases by at least a factor of two.

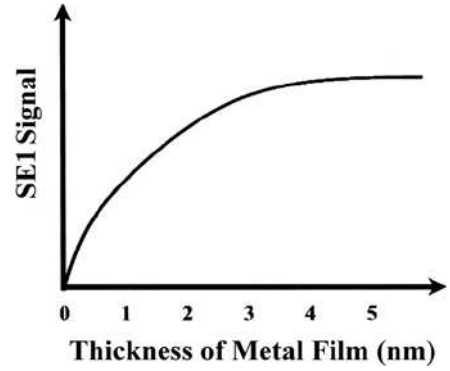


Figure 5.5. Variation of the SE_1 signal as a function of thickness from a thin film of chromium.

until this thickness typically exceeds about three to five times the value of ζ. If a film of thickness, say, 2ζ is deposited onto a surface which has topography, then as the beam scans the projected thickness of the metal film the intensity will change point to point depending on the geometry of the surface, leading to corresponding signal variations. The signal response is illustrated in Fig. 5.6, in which SE_1 profiles from a very small object, such as a carbon nanotube, are compared uncoated and coated. The uncoated sample is not resolved because its diameter is less than the ζ value for the material, and so it simply appears as a bright blur with no definite edges. After coating with a thin metal film of small ζ, however, the rapid change

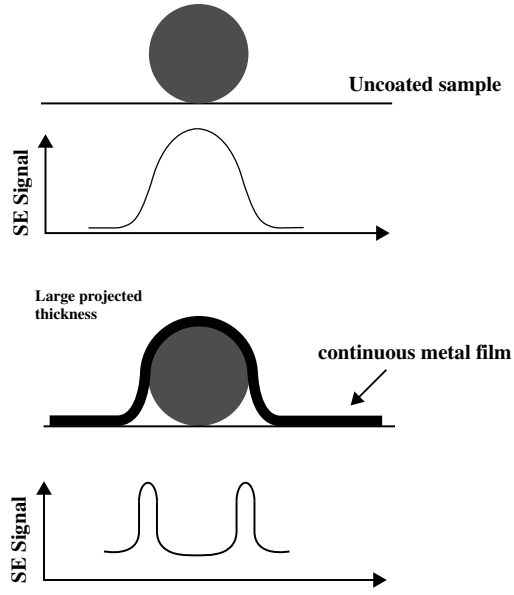


Figure 5.6. Comparison of SE signal profiles from a low-atomic-number, nanometer-scale object, and from the same object, but coated with an ultrathin continuous metal film.

in projected metal thickness that occurs as the beam is scanned over the object leads to the appearance of sharp maxima in the signal at the edges. The size and shape of the feature are now clearly resolved, and the effective resolution of the image is essentially determined by the thickness of the coating because this defines the width of the SE_1 maxima at the edge. This technique is now in wide use and makes it possible on state-of-the-art SEMs to image biological samples and other low-Z materials with a spatial resolution close to 1 nm.

5.1.3. High-Resolution Imaging at Low Voltage

An alternative approach to achieving high-resolution microscopy is to operate at low beam energies in the range from 5 keV down to below 500 eV (Pawley, 1984; Joy 1984). The dimensions of the interaction volume decrease rapidly with beam energy, approximately as the 1.7 power. Lowering E_0 from 20 keV to 2 keV results in a decrease in the range by a factor of 50 (i.e., $10^{1.7}$), so that the FWHM dimensions of the SE_2 and BSE_2 components begin to approach that for SE_1. This is a desirable situation because now all of the emitted electrons carry information at high spatial resolution. There is no need to separate the weak SE_1 signal from the strong, but low-resolution SE_2 background. Consequently the imaging process is more efficient and the contrast is higher. This also means that BSE imaging with a resolution similar to that for SE images is possible. For example, Fig. 5.7 shows a BSE image of sputtered tin spheres on an aluminum

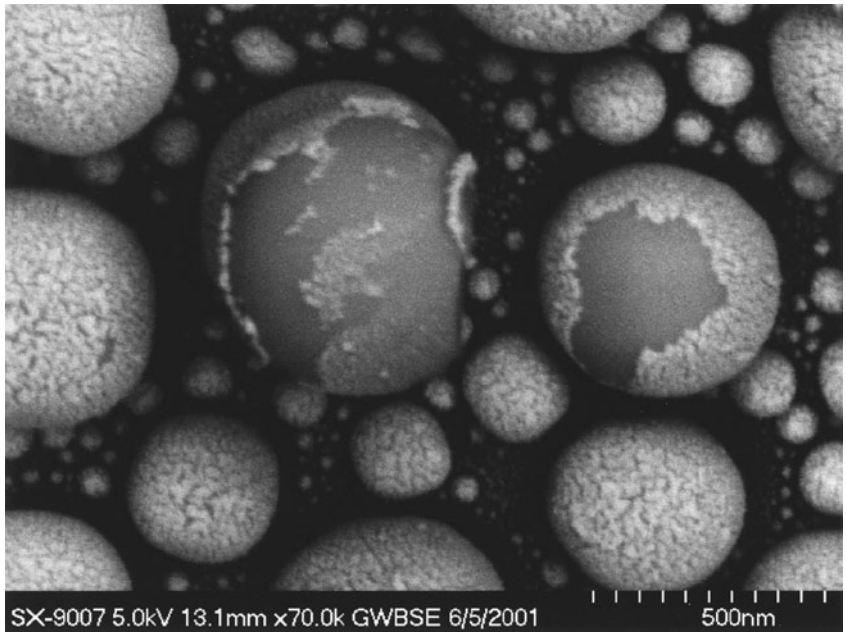


Figure 5.7. Backscattered electron image of sputtered tin spheres on an aluminum substrate recorded at 5 keV using a microchannel plate detector.

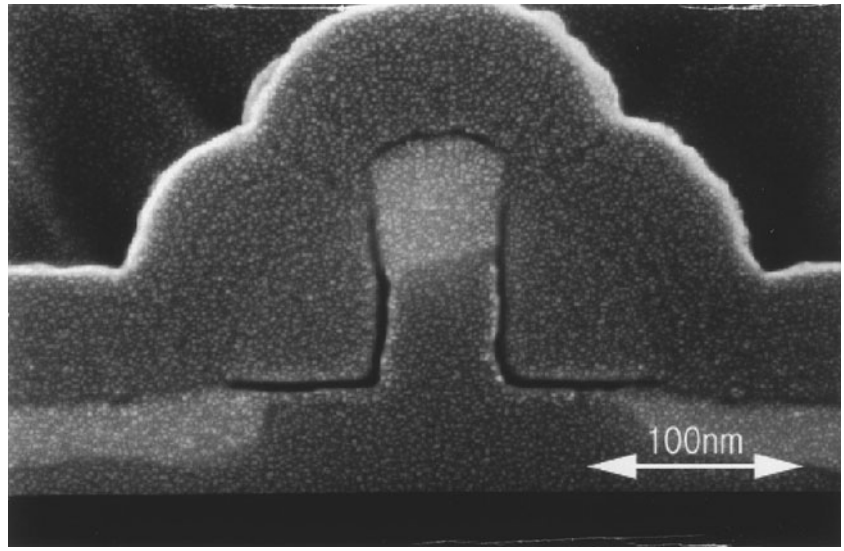


Figure 5.8. Cross-sectioned silicon semiconductor device coated with a 2-nm-thick layer of platinum metal, beam energy = 9 keV. Individual metal particles down to 1 nm diameter are clearly visible in addition to contrast from the doped semiconductor and topographic contrast. (Courtesy of Bryan Tracy of AMD.)

substrate recorded at 5 keV using a microchannel plate detector. The image, although a little noisy because of the reduced efficiency of the BSE detector at low energies, has a resolution below 10 nm and provides excellent atomic number contrast between the oxide film and the metal spheres. The use of BSE imaging at low energies is often advantageous because this signal is much less sensitive to the effects of charging, allowing good images to be obtained even from samples that are uncoated.

The secondary electron signal rises at low beam energies at a rate which offsets the fall in incident beam current that results from the inevitable decrease in gun brightness. The signal-to-noise ratio of the SE image therefore remains good down to energies below 500 eV, which is a distinct benefit for satisfying the threshold equation. Finally, the situation greatly improves for the intermediate- and high-atomic-number specimens because the spatial range of signals decreases to an even greater extent for strongly scattering materials. Figure 5.8 shows a cross-sectioned silicon semiconductor device coated with a 2-nm-thick layer of platinum metal. In this image, recorded at 9 keV, individual metal particles of a diameter down to 1 nm are clearly visible in addition to contrast from the doped semiconductor and topographic contrast. The absence of edge brightness effects and charging artifacts also adds to the clarity of the image.

There are, however, some drawbacks to operation at low beam energy (Pawley, 1984). The source brightness falls proportionally to the beam energy, with the result that at 500 eV even a field emission gun has only about the same brightness as a basic tungsten hairpin emitter at 20 keV.

The available beam current may thus fall below the threshold current for any useful contrast, although for secondary electron signals this decrease is usually more than offset by the corresponding improvement in the SE yield. The electron-optical performance is also degraded by the chromatic aberration which results from the energy spread of the gun. This does not greatly increase the actual probe size, but it does degrade the profile of the beam by producing a broad skirt of intensity about the central focused probe, reducing the image contrast. Low-energy beams are also much more susceptible to deflection/degradation by stray electromagnetic fields and so the use of short working distances is mandatory. A cold field emission source on a well-shielded, high-performance SEM equipped with an immersion or a snorkel lens is the ideal answer to these problems, but other configurations can still be useful, although great care must be taken to achieve reproducible, high-resolution performance.

Contamination is a serious limitation to low-voltage microscopy because the range of the signals is so short that the contamination may dominate the image as it builds up. Careful sample handling followed by cleaning, preferably by means of a plasma rather than a solvent, is therefore necessary to ensure useful images. The rate of contamination build-up can be reduced by minimizing the electron dose to the specimen with the following procedures: (1) avoiding high magnification except when essential; (2) focusing and stigmating the image in an area remote from the one that is to be used to record a micrograph; and (3) never using spot or reduced-area raster modes. Because contamination grows as a result of the field-enhanced mobility of hydrocarbons across the sample surface, the rate of growth can often be drastically reduced by a very thin metal coating because this eliminates fields from any charged region. Other useful techniques include mildly baking the specimen with a heat lamp in the specimen airlock during the pump-down cycle to drive off hydrocarbons, and/or ultraviolet irradiation to "crack" remaining hydrocarbons in place to minimize their mobility to the beam impact site.

5.2. STEM-in-SEM: High Resolution for the Special Case of Thin Specimens

When the dimensions of the object of interest approach and then become less than that of the interaction volume, progressively more of the electron scattering occurs from the substrate and progressively less from the object. This situation is frequently encountered in the case of particles. The examination of submicrometer-sized particles with the SEM often leads to disappointing results when a conventional sample mounting approach is used. Particles are collected from air samples deposited directly on polycarbonate filters or are extracted from suspensions in liquids by filtering. A typical preparation involves attaching the filter onto a bulk substrate or transferring the particles to a bulk carbon substrate and applying a conductive coating. Such an approach is generally satisfactory

for supermicrometer-sized particles because most of the scattering of the beam electrons takes place in the particle. For submicrometer particles, particularly those composed of low- to intermediate-atomic-number elements, the primary beam will penetrate through the particle and into the substrate. Although the substrate is usually chosen to have a low atomic number and density, the extraneous scattering from the substrate acts as noise and thus affects the signal quality, raising the necessary threshold current and diminishing the image quality. Eventually a particle size is reached where the substrate scattering dominates the signal, and the particle contrast drops below the threshold contrast for the beam current available in a small probe. The transition situation is especially difficult. Situations are often encountered where a satisfactory photograph of a small particle can be prepared with a slow scan or a long accumulation digital image, but as soon as the scan speed is increased to a rapid visual or TV rate to locate another particle while surveying the specimen, the particles effectively disappear because of the higher threshold current required with the rapid scan (i.e., shorter pixel dwell time).

The first improvement that can be made to this situation is to remove as much of the substrate as possible. The particles can be deposited onto a thin carbon film (~10 nm thick) supported on a metal grid. (Procedures for transferring particles from a filter to a carbon film on a grid are given in Chapter 11.) These film-on-grid preparations, available from several of the firms that provide microscopy supplies, are surprisingly strong. After the particles are deposited on the film/grid, a final carbon film (5–10 nm) is deposited to provide a discharge path and to improve the mechanical stability of the preparation. The grid, which is typically 3 mm in diameter, is then placed over a hole slightly smaller in diameter which has been drilled through a carbon block. (It is important to vent the hole from the side or bottom to avoid problems with entrapped air rushing out during the pump-down cycle and dislodging the grid.) The particles can then be imaged in the normal manner using any of the detectors, for example, E-T, through the lens (TTL), or dedicated BSE detectors.

A further improvement for the detection and visualization of particles in SEM images is to employ the technique of scanning transmission electron microscopy (STEM) in the SEM. By placing a suitable detector below the specimen on its thin-carbon-film support, as shown in Fig. 5.9, the transmitted electron signal can be collected and used to form an image. The specimen stage must be altered to allow access to the region below the specimen to place the detector. The STEM detector can be constructed from a scintillator–light pipe or solid state detector. Alternatively, a simple, inexpensive STEM detector, illustrated schematically in Fig. 5.10, can be made by placing a high-atomic-number scattering surface below the specimen and tilted so that the transmitted electrons are scattered toward the conventional E–T detector in the specimen chamber. A crude aperture below the specimen serves to create STEM contrast by cutting off a portion of the scattered signal. This "poor man's STEM detector" is especially useful with a FEG SEM equipped with a TTL detector for SE. The TTL detector can sample the signal above the grid, and the E-T detector can

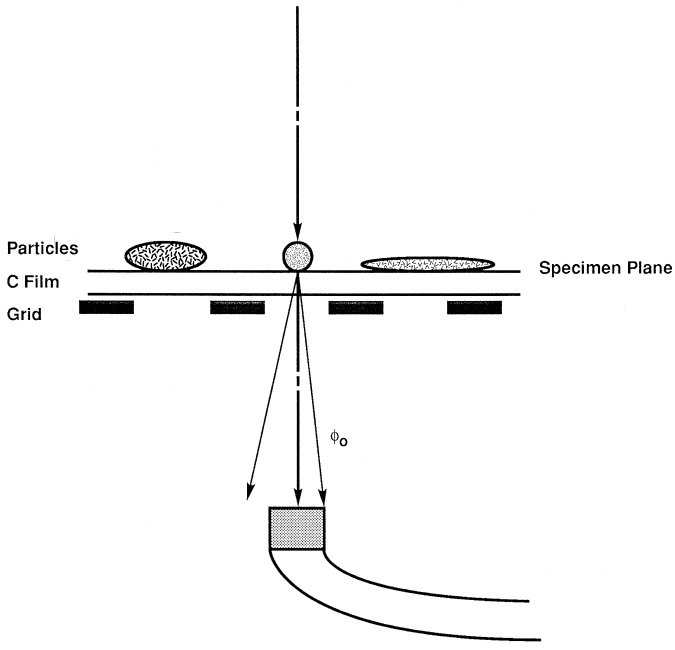


Figure 5.9. Schematic illustration of a detector for STEM in SEM.

be used to measure the transmitted electron signal, with both the conventional TTL and STEM-in-SEM images available simultaneously. The TTL image is nearly a pure SE_1 image in this case because the specimen thickness is reduced to the point where backscattering, and thus the SE_2 signal component, is eliminated.

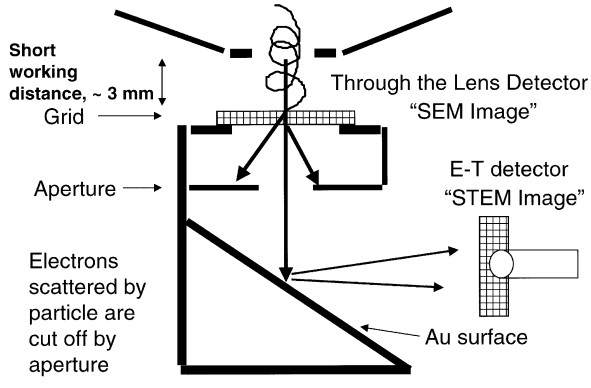


Figure 5.10. "Poor man's STEM-in-SEM" detector. An aperture placed below the specimen cuts off the transmitted, scattered electrons from reaching the detector. The unscattered transmitted beam passes through the aperture and strikes the gold-covered silicon surface, creating a high BSE and SE signal, which is collected by the conventional ET detector. The TTL detector can also operate to collect the above the specimen signal, which is mostly SE_1 electrons because backscattering and SE_2 production are limited by the specimen thickness.

An example of the use of the STEM-in-SEM mode is shown in Figs. 5.11a and 5.11b. A field of clay mineral particles is viewed simultaneously in the STEM-in-SEM mode and in the conventional E–T (positive bias) detector mode and recorded for the same time. The images show a similar array of particles and are complementary, in that bright areas in the conventional E–T image, which show areas of strong scattering, appear dark in the STEM image because these scattered electrons are lost to the transmitted signal. The striking thing about the images in Fig. 5.11 is that the conventional E–T image is so very much more noisy than the STEM-in-SEM mode image. Figure 5.11c shows an E–T detector image that is similar in quality to the STEM-in-SEM image, but this image required approximately a factor of 25 longer recording time to establish a similar level of signal to noise, and hence a similar threshold current. Small, low-mass objects have very low scattering power, so that the particle produces an SE_1 signal, which is desirable, but very little backscattering occurs. When the situation for the STEM detector is considered, even modest scattering from the specimen produces large contrast in the STEM signal, which measures the predominantly forward scattering component of elastic scattering. When operating in the STEM-in-SEM mode, it is best to use the maximum beam energy available in the SEM to improve both the electron penetration through the particles and the brightness of the source for optimum electron-optic performance.

For very fine particles, the "poor man's STEM-in-SEM" detector can produce some outstanding images when operated in conjunction with the TTL detector. Figure 5.12a shows the SEM TTL and Fig. 5.12b shows the STEM in SEM ("poor man's detector," Fig. 5.10) of fine particles attached to lacey carbon film with a beam energy of 25 keV. The internal structure as well as the surface structure can be simultaneously observed.

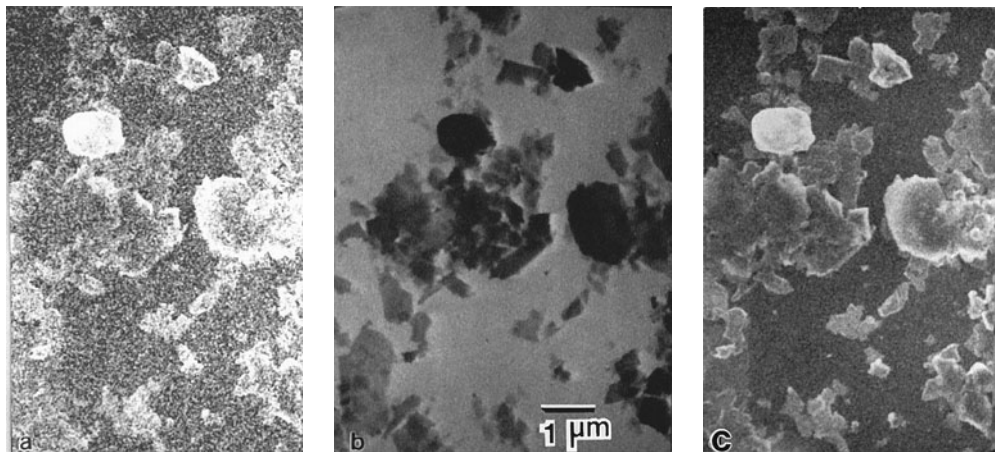


Figure 5.11. Clay mineral particles dispersed on a thin (20 nm) carbon film and viewed (a) by a conventional positively biased E–T detector located above the specimen, (b) a passive scintillator STEM detector placed 4 cm below the specimen. The images in (a) and (b) were recorded simultaneously for the same integration time. (c) Same as (a), but recorded for a factor of 25 more time. Beam energy = 25 keV.

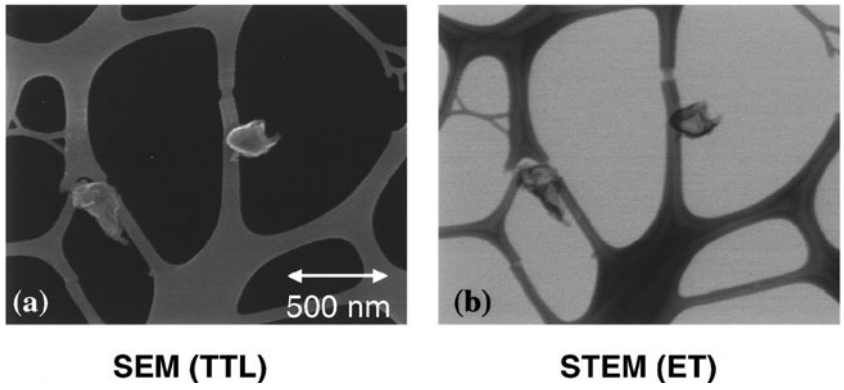


Figure 5.12. (a) SEM TTL and (b) STEM in SEM ("poor man's detector," Fig. 5.10) of fine particles attached to lacey carbon film; beam energy = 25 keV. (Courtesy of J. Small, NIST)

For very small particles, a simple theory can be developed to predict the contrast in the STEM image. This discussion is found in the Enrichments for Chapter 5, Sec 2, on the accompanying CD.

5.3. Surface Imaging at Low Voltage

Many processes involved in physical and biological phenomena as well as in technological systems are controlled by reactions and structures at the surface of a sample. The surface can be considered to be the region in which the atoms are in a different state of energy compared to those buried deep inside. Because of the lack of confining atoms above the surface, the first few layers of atoms of the surface region can vibrate in modes not possible when an atom is buried deep in the material and is surrounded by the regular arrangement of atoms in a crystal or by the three-dimensional atom distribution of an amorphous solid. Due to their position and elevated energy state, surface atoms are the first to react when exposed to other substances.

Considering the vacuum levels in the conventional SEM, the surface atoms of a specimen are usually not the same species as those in the interior, but rather consist of oxides of those atoms or even completely foreign material present as contaminants. Most "pure" elements, for example, are covered by a thin "native" oxide, about 1–10 nm thick, that forms quickly during exposure to laboratory atmosphere when fresh surfaces are prepared. Even if a fresh surface could be introduced into a conventional SEM (chamber pressure $\sim 10^{-5}$–10^{-3} Pa, or 10^{-7}–10^{-5} torr) or prepared *in situ,* a monolayer of gas is deposited each second at 10^{-5} Pa. The abundance of residual gas atoms in the SEM vacuum would therefore still lead to native oxide formation in a few seconds to minutes.

Despite these constraints, there is increasing interest in imaging the surface of specimens with the SEM. The strategy of lowering the beam energy to improve the lateral spatial resolution by restricting the range of the beam electrons is also useful for "surface" imaging. As E_0 decreases

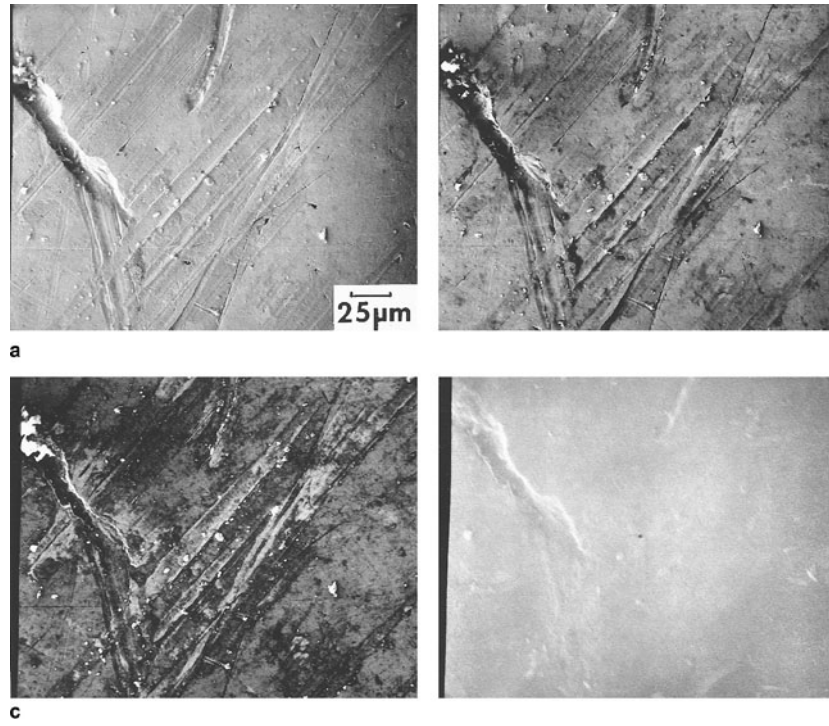


Figure 5.13. Beam energy series of the surface of a platinum-aperture, positively biased E–T detector. (a) 20, (b) 10, (c) 5, (d) 1 keV. Note the strong changes in contrast. At 20 keV the surface topography consisting of scratches is visible. At 10 and 5 keV, topography and a discontinuous surface layer are visible. At 1 keV, the imaging is entirely within the surface layer, and underlying topography is nearly lost.

and the BSE_2 and SE_2 ranges approach the BSE_1 and SE_1 ranges, the image sampling depth is also greatly restricted. Because we are often dealing with an unknown surface chemistry/structure both in depth and laterally, the best approach may be to consider a systematic imaging study of the material as a function of beam energy. By comparing a series of images to search for changes, surface features such as discontinuous surface layers can often be revealed.

Figure 5.13 shows such a beam energy image series taken with a positively biased E–T detector at the same location on the surface of platinum. Note the strong changes in contrast with beam energy. At 20 keV ($R_{KO} = 773$ nm) the surface topography consisting of scratches is visible. At 10 keV ($R_{KO} = 243$ nm) and 5 keV ($R_{KO} = 76$ nm), both this topography and a surface layer that appears to have variable thickness are visible. At 1 keV ($R_{KO} = 5$ nm), the imaging is constrained by the beam range to be entirely within the surface layer, and underlying topography is nearly lost.

When the same area is compared during a beam energy image series, care must be taken not to confuse contamination deposited during the earlier stages of the image series with the true specimen information. Minimum-dose microscopy principles should be followed, with specimen exposure

only occurring during image collection. If the contamination rate is too high, as observed in consecutive images taken at the same energy, then it may be necessary to choose a new area for each image and compare only general features. Since the low-beam energy images are the most sensitive to prior contamination, these images should be recorded first in any beam energy sequence, finishing with the high-beam-energy images.

5.4. Making Dimensional Measurements in the SEM

The microscopist often has the task of determining the size of a feature observed in an SEM image. With early SEMs, such measurements were usually made with a transparent ruler placed over a photograph of the SEM CRT display, or directly on the CRT itself. Computer-assisted microscopy provides numerous software tools for simple linear measurements of features (length, diameter, etc.) as well as more elaborate perimeter and area measurements. The measurement may require intervention by the microscopist, or it may be carried out automatically on individual objects or all objects identified in a field of view. Such digital measurement tools usually incorporate the magnification calibration directly, so that the results are reported in units appropriate to the magnification employed: millimeters, micrometers, to nanometers for linear dimensions as the magnification increases.

A rigorous measurement procedure requires provision of statements of accuracy and precision to accompany the reported results. Accuracy refers to the discrepancy between the true value and the value established by measurement. Precision refers to the agreement among repeated measurements of the same quantity.

To estimate the accuracy of a measurement, we require that the measurement must be established relative to an independently calibrated length standard. Such a standard can be provided through the use of a "stage micrometer," a series of independently calibrated measured distances (with a certified error range) between physical artifacts of some type. Stage micrometers include: mechanical scratches or other marks scribed with a diamond point onto a polished metal surface with a traveling engine; repetitive features produced by electroplating thin layers of different metals (e.g., nickel and gold) that produce strong atomic number contrast (Fig. 5.14a); arrays of features etched into a material such as silicon and reproduced in multiple copies through the use of lithography (Fig. 5.14b); and discrete objects such as spherical particles (e.g., monodispersed polystyrene latex spheres created with a very narrow distribution of diameters). Except for the polystyrene latex spheres, these stage micrometers are flat objects that are usually placed normal to the beam (i.e., 0° tilt) to calibrate the scanned image.

Calibrated dimensional objects are typically available as standards from national measurement laboratories, or in the case of secondary standards, objects available from commercial sources are derived from and

traceable to such national standards. Such national/international standards are typically supported by detailed instructions for use, including restrictions, as well as a description of how the object is traceable to the international fundamental standard of linear dimension, the wavelength of light. This information can be used to estimate the accuracy which can be achieved with the standard object as part of the measurement process. The same standard object can be used to estimate the precision for an optimum measurement situation (i.e., the standard object is usually contrived to produce high contrast and to have carefully defined edges) by repeating a specific measurement procedure and examining the distribution of the measurement results with an appropriate statistical analysis. When we wish to extend this established measurement protocol to actual objects of interest,

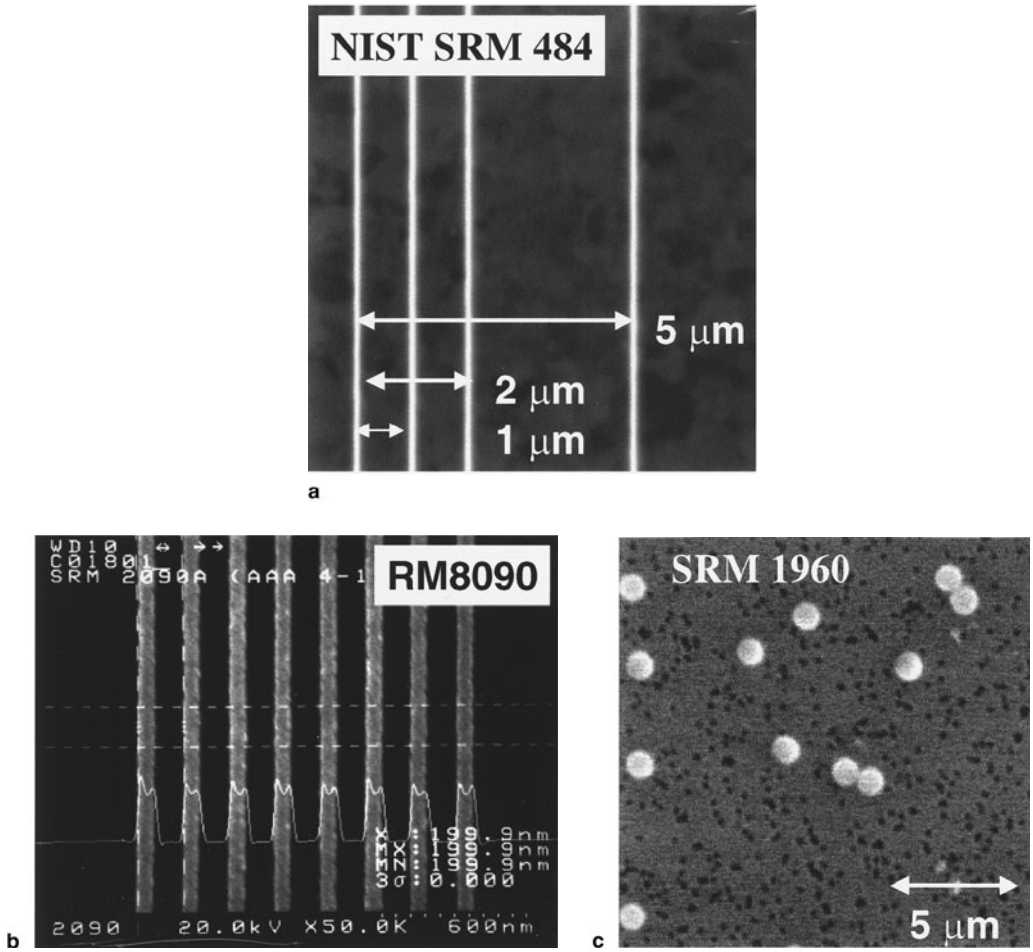


Figure 5.14. Examples of stage micrometers. (a) NIST SRM 484, Au–Ni electroplated layers; (b) RM 8090, lithographically produced meaurement pattern; (c) SRM 1960, polystyrene latex spheres (National Institute of Standards and Technology, Standard Reference Materials Program Office; www.nist.gov).

we often encounter significant limitations and larger measurement errors, which are imposed by the lower contrast of the real features and by details of the interaction of the scanned beam with the actual interface. One of the most significant sources of error derives from the difficulty of determining where a structure actually starts and ends. A physical feature, such as a phase boundary between two different compositions, may be atomically sharp, but the signal response as the beam is scanned across that boundary changes continuously over a distance that is a significant fraction of the electron range, which may be 1μm or more. The uncertainty that this signal spreading introduces may be insignificant for measurements with low magnifications less than 1000×, but as the magnification increases, the uncertainty in the position of the boundary will eventually dominate the errors. Advanced SEM image deconvolution methods based upon modeling of the interactions are necessary in support of high-magnification/high-resolution distance measurements (Lowney *et al.,* 1994; Vladar *et al.,* 2001).

A much larger source of error, and one that is often so out of control that a useful error estimate cannot be made, arises when we wish to measure three-dimensional objects with complex topography, which is to say almost any irregular object. As a critical user of such measurement tools, the microscopist must always be aware of the fundamental constraints that determine the basic validity of the measurements. In Chapter 4 the SEM imaging process was described as a point-by-point mapping procedure that transfers information between the locations of the scan field on the specimen to corresponding locations on the display CRT or to addresses in a computer memory. Figure 4.5 shows the ideal imaging situation where the shapes in the specimen field are transferred faithfully to the display/recording space. The "stretching" action of the SEM magnification process is equivalent to the projection process shown in Fig. 4.11. Note that only flat objects that lie in the plane perpendicular to the optic axis are transferred faithfully in the magnified image, whereas three-dimensional objects that extend along the optic axis are inevitably distorted by being projected into the two-dimensional planar image. Such projection distortion is called "foreshortening" because tilted surfaces appear in the image to have smaller linear dimensions than really exist. Foreshortening can introduce *huge* errors into linear, perimeter, and area measurements depending on the degree of inclination (local tilt). As an example of the magnitude of the distortion introduced by foreshortening, consider the object with various features shown in cross section in Fig. 4.13. In the projected image (SEM image), all of the features have the apparent same size, when in reality they are all of different size! [The construction in Fig. 4.13 has been made by parallel projection. At sufficiently high-magnifications, the cone angle of the scan in Fig. 4.13 becomes so small that we can consider the scans rays to be effectively parallel. For example, at 500× (scan field 200 μm along an edge) with a working distance of 10 mm, the semicone angle of the scan is only 0.01 rad or 0.57 deg.] With this condition of parallel projection, only face B of the object in Fig. 4.13 is projected without distortion in the SEM image because B is perpendicular to the optic axis. The other faces suffer the distortion of foreshortening, which can be quite severe. For feature D,

the distortion is a compression factor of three! That is, a linear measurement of a detail lying down the face of feature D would be *underestimated by a factor of three* in length compared to the same feature lying at any orientation within the face of feature B. To make valid measurements of features tilted at random to the beam, we must make use of the procedures of quantitative stereomicroscopy described in Section 5.5.

5.5. Recovering the Third Dimension: Stereomicroscopy

Scanning electron microscopy displays the images it generates of three-dimensional specimens on a two-dimensional viewing screen or sheet of photographic or printer paper. This is unfortunate because many objects of interest viewed in the SEM are likely to have a topographically rough surface and frequently it is the three-dimensional form rather than its two-dimensional projection that is of interest. Although most other forms of microscopy are also two-dimensional, the problem is particularly acute in the SEM because the high depth of field of the instrument further obscures depth information by removing any tell-tale clues such as areas which are out of focus compared to their neighbors. One solution to this dilemma is to devise techniques which can recover information about the missing dimension either directly in the form of a suitably processed micrograph that can be viewed by the operator, or indirectly as height or depth information that can be correlated with the beam position on the image. Such techniques are examples of stereoscopic imaging (Wells, 1960).

5.5.1. Qualitative Stereo Imaging and Presentation

The data required for our brain to derive depth information are produced by exploiting the phenomenon of parallax. Our two eyes are laterally separated by a spacing of 6–10 cm, so that the right eye and the left eye see slightly different views of the scene in front of them. The apparent position of an object is different in the two views by an amount, known as the *parallax,* which varies with the distance of that object from the viewer. If the object is close, then the parallax is large, whereas if the object is distant from the observer, then the parallax is small. Our brain then processes this parallax information and converts it into depth perception. Fortunately for everyone except perhaps the very youngest infants, the process is completely automatic and requires no conscious effort. In fact the brain is so adept at this operation that it can even perform it from memory. For example, if an object is viewed using TV rate scanning on the SEM while it is rotated or tilted, then a perception of the three-dimensional nature of the surface is often obtained because the brain is apparently able to compare the live image with a memory of a previous view and so obtain some parallax information. This perceived depth effect is immediately lost when the mechanical motion ceases and the stream of images becomes constant.

The classic method for producing stereoscopic images is called photogrammetry. Imagine, for example, an aircraft carrying a camera pointed straight down as it flies at a constant height over the surface of the earth. If at regular intervals the camera records an image and if the time between these snapshots is sufficiently small, then any sequential pair of such images will have partially overlapping fields of view so that some features will appear in both pictures, although seen from different viewpoints. If one of this pair of images is presented to the left eye and the other is shown to the right eye, then the parallax in the overlap region will result in the production of depth perception in the brain. This technique is in widespread use, both by aircraft and more recently by satellites, for mapping and measuring the earth's surface. Note, however, that because the two viewpoints producing the parallax are generally separated by a distance that is large compared to our interocular spacing, the stereo effect is exaggerated and is often referred to as "hyper-stereo." In principle this technique could also be applied in the SEM by recording a pair of images and shifting the sample between views. In practice this method is not widely used because it is not an efficient technique because to provide adequate parallax, no more than half of the field of view can appear in the overlap region. Moreover, this method must be used at very low magnifications in order to produce adequate parallax due to the motion.

Fortunately, viewing an object from two separated viewpoints which subtend an angle θ at that object is equivalent to taking two images from a single viewpoint, but rotating the object through the same angle θ between the images. It is this procedure that is usually employed in the SEM: Two micrographs are recorded from the same area of a specimen, but the sample is tilted between exposures. The effect of depth perception then depends on the choice of the angle θ. For the average pair of human eyes viewing something at a typical comfortable reading distance, θ is about 6° and this value produces good stereo effects. Larger values will produce hyper-stereo, which is dramatic, but tiring to view, whereas smaller angles may produce useful, but rather flattened results. The change in angle can be accomplished either by mechanically tilting the specimen in a goniometer stage or by holding the sample fixed and tilting the incident beam. Because beam tilting can be done electronically, real-time TV stereo imaging is possible in this way (Pawley, 1988).

Taking a stereo pair of images in the SEM is straightforward, and can be done in several different ways, all of which produce equivalent results. The important point is that it is necessary to proceed systematically in every case so that the depth perception that is created is unambiguous. Otherwise, if the right-hand and left-hand images in a stereo pair are swapped, then hills become valleys and valleys become hills. The simplest procedure is as follows:

1. Using TV scan rate, observe the general area of interest at low magnification and pick a feature that is easily recognizable. Swing the specimen stage tilt control through a range of 10 or 20 deg and note the direction in which the chosen feature is moving.

2. Turn on the scan rotation control and adjust the scan rotation angle so that when the stage is tilted the chosen feature moves exactly horizontally on the screen. The tilt axis is now vertical on the screen, placing it in the direction expected by our eyes (which are "tilted" about an axis running down parallel to our nose).

3. Decide on the stereo tilt value to be used. If 6 deg is chosen, then this can be obtained by taking the pair of images with the specimen stage tilt control adjusted to give the desired change in tilt angle, for example, set first at 0 and then at 6 deg, or at −3 and +3 deg, at 30 and 36 deg, or any other combination. Experience shows that a pair of images symmetric about normal incidence (e.g., −3 and +3 deg) is best, but negative tilts are not possible on all SEM stages.

4. Set the stage tilt to the lower of the two angles that will be used (e.g., −3 deg, 0 deg, or whatever). The *low*-tilt-angle image will be the *left*-hand component of the stereo pair (remember **L**eft = **L**ow) using the convention defined by Howell (1975).

5. Record the chosen field of view. Mark the micrograph, or identify the stored image, with a letter L to ensure that there is no confusion later (Fig. 5.15a).

6. Choose a prominent feature on the sample (something that you can easily recognize again). Mark the position of this feature on the viewing CRT screen with an easily erasable marker. Some SEMs provide controls to position a spot, or a cross, on the screen and this can also be used if available.

7. It is now necessary to change the tilt and record a second image. Whenever the specimen stage is tilted, it is very likely to translate the specimen laterally under the beam (unless a special eucentric stage is available),

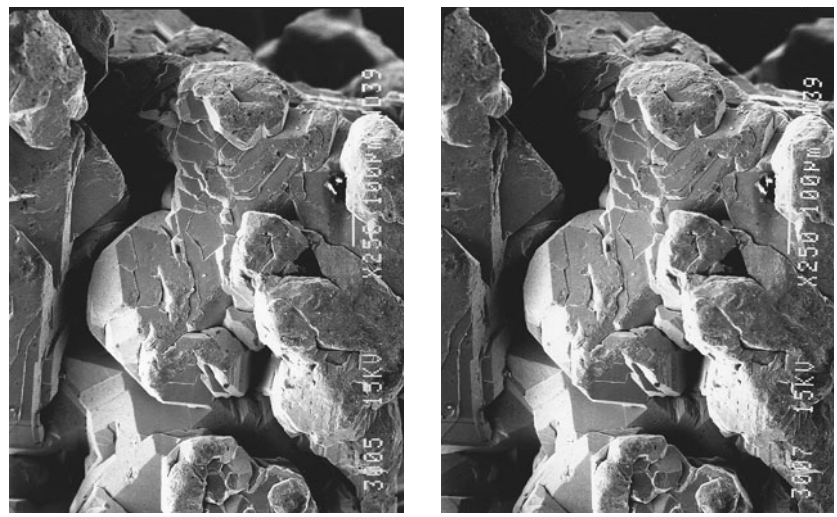


Figure 5.15. Stereo pair of crystals of native silver. A typical application of qualitative stereo imaging to view a complicated three-dimensional structure. Note that the apparent illumination comes from the left side of the image rather than at the top, a consequence of the necessity of orienting the tilt axis along the vertical.

so it is therefore a good idea *before* changing the tilt to lower the magnification by a factor of perhaps $\times 5$ and record a second image that clearly allows the target area to be identified. This additional precaution is especially important if the specimen has relatively homogeneous topographic features over large areas, making it easy to become lost.

8. While staying at the lower magnification, change the tilt to the second value required. Again locate the general area of interest and return to the original magnification used for the left-hand image to match up the areas as exactly as possible, placing your chosen feature under the temporary marked location on the CRT screen or electronic cursor spot.

9. Tilting the specimen will probably have changed its vertical position. Do not refocus the image by changing the lens excitation because this will change the magnification of the right-hand compared to the left-hand image as well as introduce an unwanted rotation. Use the z-control (vertical motion) on the specimen stage to bring the object back into sharp focus.

10. Record the second image and mark it R, for right hand (Fig. 5.15b). (Note: This procedure should be confirmed on a particular SEM by preparing a stereo pair of known topography, for example, the raised letters/date on a coin. This right/left selection that gives the proper topography establishes the proper orientation.)

Since the stage tilt axis on the SEM is usually designed to tilt the sample toward the ET detector, which is normally positioned at the top of the viewing screen, this stereo technique has the disadvantage that, after applying scan rotation to place the tilt axis along the vertical, the apparent direction of sample illumination will be from the side rather than from the top. This is evident in Fig. 5.15, which shows a stereo pair of native silver. Although the illumination is coming from the left, the utility of the image is not spoilt and when examined through a stereo viewer the complicated three-dimensional structure is strikingly visible.

The problem of the orientation of the illumination can be avoided by using a combination of mechanical tilting and rotation (Lane, 1970a). The rotation of the specimen about its own axis R, the tilt θ, and the desired parallax angle P are related by the formula

$$\sin P = \sin R \sin \theta. \tag{5.1}$$

As an example, for a tilt of $\theta = 30$ deg and a desired parallax angle of $P = 6$ deg, the rotation R that is required is about 12 deg. With a stage giving clockwise rotation for an increasing reading of the rotation control, the left member of the stereo pair is the image taken at the *higher* rotation reading. Also, the left-hand micrograph must be rotated clockwise relative to the right-hand micrograph in order to match their respective fields of view. This difficulty can be avoided if electrical image rotation is used to compensate for the mechanical stage rotation.

Although almost everybody with normal or correctable vision must possess the skill of depth perception (or else they would constantly be walking into the furniture), not everyone finds it equally easy to "see" the depth information in stereo images. Depending on the expertise of the viewer, therefore, a variety of tools may be necessary to permit satisfactory

imaging from stereo pairs. Skilled observers can often visualize the depth information in a stereo pair using nothing else than their own eyes, and the trick to do this can easily be acquired. There are two approaches, both of which work well, but not usually for the same people. The first is the "parallel" viewing method, which requires the lines of sight of the eyes to be parallel with each other, and the other is the "cross-eyed" technique. To try the parallel technique, hold the stereo pairs in front of you, but below eye level, and stare at an object across the room or out of the window for a few seconds. This relaxes the eyes to focus at infinity. Now raise the stereo pair into your field of view, but without actually looking at them and you will see the left- and right-hand images together with a central overlap image. These may be blurred, but do not try and focus on them. Instead, move the stereo pair forward and backward a little and allow the eyes to find their own focal position. The central image will then show the stereo information. To try the cross-eyed method, hold up a finger between your eyes and the stereo pair, and stare at its tip. This will cause your eyes to cross. Now allow your field of view to shift to the stereo pair, and three out-of-focus images will be seen, the left and right images and the phantom central image. The eyes will quickly adjust and bring the images into focus with the central image displaying the stereoscopy. The stereo effects achieved using parallel and cross-eyed viewing are the opposites of each other, so it will be necessary to invert the right and left images for consistent interpretation.

Most observers, however, find it more comfortable to use some optical aid to view a stereo pair. At its simplest such a device is just two lenses of short focal length, arranged like a pair of spectacles, but mounted on a stand which keeps them at the correct distance from the micrographs to achieve focus. The lenses help to restrict the field of view and so prevent the left eye from seeing the right-hand image and vice versa. The corresponding drawback is that only relatively small (5 cm square) fields of view can be employed. More elaborate viewers use mirrors as well as lenses. This makes it possible to view much larger images such as standard size 3.5 × 4.5-in. prints or full-size presentation prints. Since every pair of eyes is different from every other pair, it is always preferable to keep the left and right pictures of the stereo pair physically separate and movable (as opposed to rigidly attached to a backing) so that the observer can adjust one relative to the other while attempting to fuse the stereo image through the viewer.

One of the oldest and most popular ways of visualizing stereo is the anaglyph technique in which the left and right images are printed on top of one another in different colors and are then viewed through colored glasses to separate the images at the eyes (Judge, 1950). In the United States the left image is printed in red and the right image is printed in green, and the viewer uses red and green glasses; in Europe the left image is printed in red and the right image is printed in blue and so the viewer uses red- and blue-colored glasses. With the advent of personal computers this method has been rediscovered because it is easy to implement using image processing software and because it works effectively on a computer screen and can be readily transferred to hard copy with a color printer. To make an anaglyph image from a digitally stored left- and right-hand pair of images using, for

example, the NIH IMAGE (Macintosh) or SCION IMAGE (PC) software tools:

1. Load the left-hand image.
2. Load the right-hand image. The left- and right-hand images must be of the same pixel size, and the images must be oriented so that the tilt axis (axis across which the parallax occurs) is vertical.
3. Generate a blank (black) image of the same pixel size as the other two images using the "New Image" command.
4. From the Options menu select "Make Stack." This will group the three separate images into a single composite.
5. From the Options menu select "Make RGB." The first image in the three-layer stack, that is, the left stereo view, will be colored red (R), the second (right-hand) view will be colored green (G), and the blank third image will be colored blue (B).
6. The red–green anaglyph is now complete and can be viewed with the appropriate glasses.

To generate a European-style anaglyph, use the same procedure, but load the images in the sequence left, blank, right before making the stack. This forms a red–blue rather than a red–green image. The anaglyph method is highly effective for stereo microscopy and has the advantage that it can be printed, or projected, as a single image. Depending on the exact shade of the reds, greens, and blues, the color of the anaglyph when viewed through the glasses may be brownish and gray rather than black and white, but this is not usually a significant problem.

Several examples of anaglyph stereo pairs as finished color overlays as well as the individual component left and right images can be found in the Enrichments section, Chapter 5, Section 5, of the accompanying CD for the user to try out the image processing software. Colored glasses can be purchased from many vendors of SEM accessories and supplies.

5.5.2. Quantitative Stereo Microscopy

Quantitation of the topography of features in SEM micrographs can be carried out by measurements with stereo pairs (Boyde, 1973, 1974a, b; Wells, 1974). Such numerical measurements can be performed even if the operator is not personally able to perceive the stereo effect despite the aid of optical devices (as is the case for about 10–15% of the population). The only measurements that can be made from the micrograph are the x and y coordinates of the same feature of interest in both micrographs. These coordinates can be referred to any point that can be identified in both micrographs, usually chosen as some very obvious feature. This reference point will then be arbitrarily assigned the X, Y, Z coordinates (0, 0, 0) and all subsequent height measurements will be with respect to this point. In the following discussion, lower case coordinate pairs (x, y) will refer to measurements made in the individual images of the stereo pairs with a two-dimensional coordinate system in the image plane, and upper case coordinate triples

(X, Y, Z) will refer to calculated coordinates in a three-dimensional coordinate system in the real space in which the object actually resides.

The only measured parameters required are the (x, y) coordinates of a feature relative to orthogonal coordinate axes established in each of the images of the stereo pair and the tilt angle difference between the two halves of the stereo pair. The magnification M must also be known accurately if absolute values are needed. At magnifications above about 100× in an SEM, it can usually be assumed that the scan is effectively moving parallel to the optic axis; very simple formulas can then be used for quantification. With reference to a specific feature in both images, the three-dimensional coordinates (X, Y, Z) of the chosen feature are given by

$$MZ = \frac{P}{2\sin(\alpha/2)} \tag{5.2a}$$

$$MX = x_{\mathrm{L}} - (P/2) = x_{\mathrm{R}} + (P/2) \tag{5.2b}$$

$$MY = y_{\mathrm{L}} = y_{\mathrm{R}} \tag{5.2c}$$

$$P = x_{\mathrm{L}} - x_{\mathrm{R}}, \tag{5.2d}$$

where α is the difference in tilt angle, $\alpha = \theta_2 - \theta_1$, where θ_2 is the higher of the tilt angles (tilt $= 0^\circ$ means the tilt plane is perpendicular to the optic axis). The subscripts L and R refer to the measured coordinates in the left-hand and right-hand micrographs, respectively. With this convention points lying above the tilt axis will have positive parallax values P. Note that if the measured coordinates y_{L} and y_{R} are not the same, then this implies that the tilt axis is not accurately parallel to the y axis and the axes must then be rotated to correct this error. The steps of the measurement process are illustrated schematically in Figs. 5.16a and 5.16b:

1. Select a feature that can be recognized in both images to serve as a reference point and assign this point coordinates ($x = 0$, $y = 0$) in the SEM images and ($X = 0$, $Y = 0$, $Z = 0$) in real space.
2. Create orthogonal axes x, y through the reference point in both the left and right images. Image manipulation tools within digital imaging systems are very useful for creating these x_{L}, y_{L} and x_{R}, y_{R} axes and subsequently making the (x, y) coordinate measurements.
3. Select the feature of interest that can be recognized in both images whose (X, Y, Z) coordinates are to be determined.
4. Measure (x_{L}, y_{L}) in the left image and (x_{R}, y_{R}) in the right image for the point of interest. Note that as an internal consistency check, if all steps in preparing the stereo pair have been carried out correctly, $y_{\mathrm{L}} = y_{\mathrm{R}}$ because the tilt axis is required to be parallel to the y axis.

Let us now apply this procedure to a real SEM stereo pair. We will determine the difference in height between crystal faces A and B of the galena crystal shown in Fig. 5.16c, the left (low-tilt) image, and fig. 5.16d, the right (high-tilt, $+10^\circ$) image. The images were prepared according to the convention described above and oriented so that the parallax axis is vertical. It is a good idea to inspect the stereo pair with an optical viewer or with the anaglyph method to ensure that the stereo pair is properly

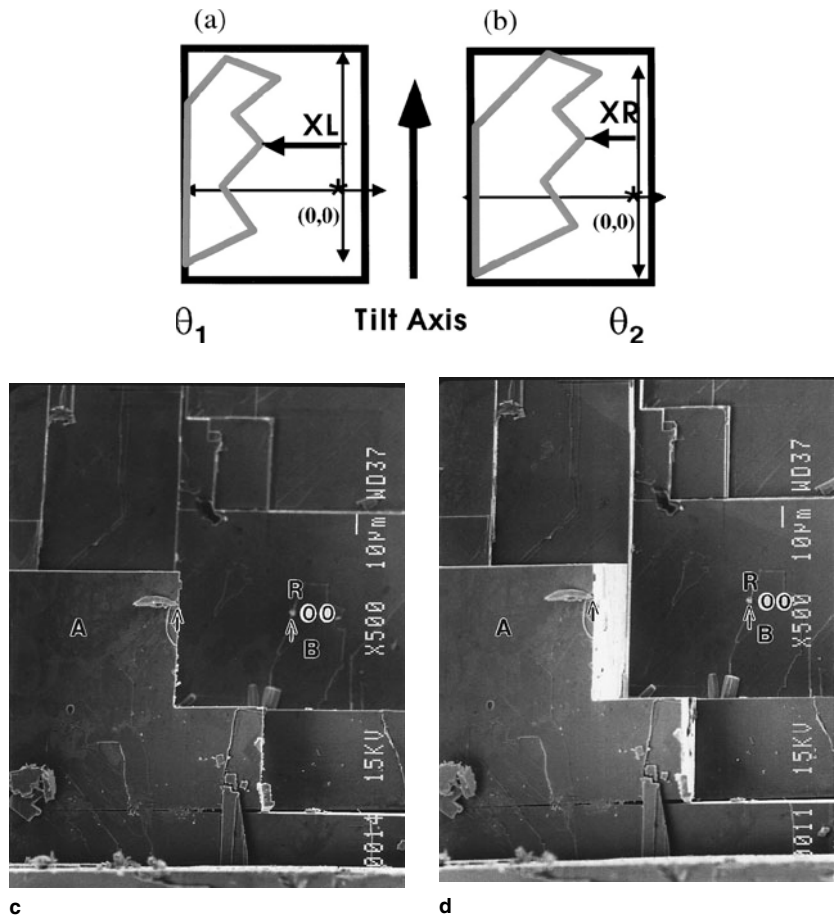


Figure 5.16. Quantitative stereo microscopy (a, b) Steps in the measurement process. (c, d) Stereo pair of a crystal of galena with surface steps. The height of the fragment that marks the edge of crystal face A relative to the point (0, 0) on crystal face B is calculated from measurements as shown. The tilt difference between the two images is 10 deg. (Note that because of the arrangement of the tilt axis and E–T detector in this instrument, the lighting appears to come from the right side after rotating the images to place the tilt axis vertical.)

arranged, and to qualitatively assess the nature of the topography, that is, determine how features are arranged up and down relative to each other. Establish a set of x (horizontal) and y (vertical) axes. With this arrangement, the parallax axis is parallel to the y axis, so all parallax measurements will be made along the x axis. To measure the height difference between faces A and B, as shown in Fig. 5.4, we first locate a distinctive feature on each face, the end of the flake on A and the particle on B noted on Figs. 5.4c and 5.4d. Define one feature (arbitrarily) as the reference point ($x = 0$, $y = 0$); the particle on B is chosen. In each photograph, a set of x–y axes is established. The x coordinates of the end of the flake are then found as

$$x_{\mathrm{L}} = -26\,\mathrm{mm}, \qquad x_{\mathrm{R}} = -35\,\mathrm{mm}.$$

The parallax is then $P = x_L - x_R = -26\ \text{mm} - (-35\ \text{mm}) = +9\,\text{mm}$.

From Eq. (5.2a), the Z coordinate of the end of the flake is

$$Z = \frac{P}{2M\ \sin(\alpha/2)} = \frac{9\,\text{mm}}{2 \times 500\ \sin(10/2)} = 0.103\,\text{mm} = 103\,\mu\text{m}. \tag{5.3}$$

Because the sign of the parallax is positive, face A is 103 μm *above* face B. This sense of the topography is confirmed by the qualitative stereo view.

As a more complicated example, if we wish to measure the length of a linear feature, then we must determine the full (X, Y, Z) coordinate set for each endpoint of the feature. The length L of the feature can then be calculated using a standard equation of analytical geometry:

$$L = \sqrt{(X_R - X_L)^2 + (Y_R - Y_L)^2 + (Z_R - Z_L)^2}. \tag{5.4}$$

If the tilt angle used is very large (i.e., more than 10 deg) or if the depth variations are large compared with the field of view of the micrograph, then the parallel-projection approximation assumed for Eq. (5.2a)–(5.2c) does not apply, and more general equations must be employed. For details of these more general equations the reader should consult, Wells (1960), Lane (1970a), and Boyde (1973, 1974a, b).

It does not take much extra effort to take a stereo pair as compared with producing a single micrograph, but the additional information gained is often invaluable, particularly on specimens where the topography is complex and where the high depth of field of the SEM may lead to a quite mistaken view of the actual relative height of various features or even the sense of the topography. When x-ray microanalysis is to be performed from an area with significant surface roughness it is always good practice to first take a stereo pair to ascertain that the take-off path to the detector is not obstructed by surrounding topography.

5.6. Variable-Pressure and Environmental SEM

One thing that all electron microscopes have always shared is the requirement that the electron-optic column and the specimen chamber be under high vacuum, about 10^{-3} Pa (10^{-5} torr) or lower. Although this condition ensures that the electron beam can travel from the source to the sample without being scattered by residual gas atoms, it means that many specimens of potential interest, such as biological tissue, samples that are damp or dirty, liquids, colloids, etc., cannot be observed in their natural state in the SEM. Either these samples are fundamentally incompatible with the vacuum environment, or their presence would contaminate the vacuum system. In this chapter a type of scanning electron microscope is described which can operate while its specimen chamber contains a gas or a vapor at a pressure in the range from approximately 10 to 2500 Pa (0.1–20 torr). This arrangement results in an instrument which has such practical convenience, versatile capabilities, and unique features that it has captured more than 50% of the present market for scanning electron microscopes.

Development of this type of instrument started even before conventional SEMs became commercially available. In 1960 Thornley described an arrangement which made it possible to image liquid droplets in the SEM. This was accomplished by sandwiching the liquid between two layers of carbon film, which helped prevent evaporation of the liquid and minimized contamination of the vacuum. This approach was further developed by J. A. Swift and his group at Unilever in England and produced some useful results. An alternative approach was used by Lane (1970b), who made an important advance by building a self-contained substage assembly for an SEM. This was a closed box, except for a small aperture to admit the incident electron beam and a controllable water vapor leak, and it contained both the sample to be imaged and a backscattered electron detector. This device made it possible to vary the pressure around the specimen without seriously degrading the vacuum in the normal SEM specimen chamber which surrounded it. Lane not only was able to image liquid surfaces, but also became the first person to observe the phenomenon of gas amplification, which is discussed later in this chapter. Soon afterward Robinson (1974) took the next step and converted the complete specimen chamber of a JEOL JSM-2 SEM into a substage, isolating it from the column vacuum system by means of a differential aperture. This enabled the pressure in the chamber to be a factor of up to 1000 times greater than that in the column without affecting the electron gun while maintaining normal access to the specimen. Commercial SEMs based on this concept were announced by ISI and JEOL (the GEOSEM) in 1978, and in 1987 Electroscan introduced the first differentially pumped SEM capable of relatively high pressure operation to 2700 Pa (20 torr).

5.6.1. Current Instruments

There are now two distinguishable varieties of SEM with the ability to operate with the sample in gas. The Environmental SEM is a trade-mark of Philips/FEI/Electroscan and refers to an instrument which can operate with a pressure in the specimen chamber of up to 2,700 Pa (20 torr) and which offers a proprietary secondary electron detector for use in the gas. Instruments from other manufacturers are usually referred to as variable-pressure SEMs (VPSEMs), nature SEMs, or some similar term. Typically, although not exclusively, such machines are limited to a maximum pressure in the specimen chamber of about 266 Pa (2 torr) and may not offer a secondary electron detector for use when there is gas in the chamber, relying only on BSE detection for imaging. However, the similarities among all of the instruments are more important than any differences and so their design and operation can be discussed without reference to any particular version.

Typically these instruments use a relatively modest electron-optical system and a basic thermionic gun because the performance of the machine is ultimately limited by the electron–gas interaction, discussed below, rather than by the lenses. Instruments employing field emission guns and more advanced lenses are now being built to offer higher levels of resolution. VPSEMs and ESEMs are also usually designed to operate optimally in the

Table 5.1. Basic Gas Pressure Definitions

1 atm = 760 mm Hg = 1 bar
1 torr = 1 mm Hg
1 torr = 133 pascal (Pa) = 1/760 atm
1 mbar = 1/1000 atm = 0.76 torr = 100 Pa

energy range 10–30 keV rather than at the lower energy of 1–5 keV favored by conventional high-vacuum machines. In all other respects, however, VPSEMs and ESEMs and conventional SEMs are identical.

These instruments offer significant advantages when compared to conventional SEMs. The relaxed vacuum environment means that many samples that would otherwise be unsuitable for observation, for example, materials such as bone that are porous and hard to pump down, specimens that are damp or wet, or biological specimens that cannot be maintained in their original state if they are allowed to dehydrate, can be imaged safely and conveniently. The gaseous environment allows poorly conducting and insulating samples to be imaged in a stable manner at any desired beam energy without the need to coat them with a conducting metal layer. The gas in the specimen chamber also creates a microenvironment which can be used for a wide variety of *in situ* experiments (e.g., in corrosion and oxidation). Moreover, the gas provides new modes of imaging, which can generate new types of information (Griffin, 2000). Finally, with minor modifications, these instruments can also operate as conventional high-vacuum SEMs and so offer great flexibility to the user who is confronted with a wide variety of problems and specimens.

5.6.2. Gas in the Specimen Chamber

5.6.2.1. Units of Gas Pressure

One of the problems of discussing VPSEMs and ESEMs is the rich variety of different units in which gas pressures are quoted. For convenience a brief listing is provided in Table 5.1. All of these units may be found in common use on microscopes, but the most widely used are the torr and the pascal. Because the pascal is the official SI unit, it will be used in this chapter, and torr units will follow in parentheses.

5.6.2.2. The Vacuum System

Figure 5.17 shows a schematic diagram of a VPSEM or an ESEM. In essence it is identical to that of a conventional SEM except for the additions to the vacuum system. The pressure-limiting aperture (PLA) or differential aperture is usually positioned at the bottom of the objective lens so that the distance that electrons have to travel through the gas to reach the sample (the gas path length, GPL) is no more than the working distance (WD). Some instruments, however, mount this aperture on an extension tube beneath the lens so that the gas path length can be much smaller than the working distance of the lens. In either case the aperture is typically a few hundred micrometers in diameter, a figure which is sufficiently small

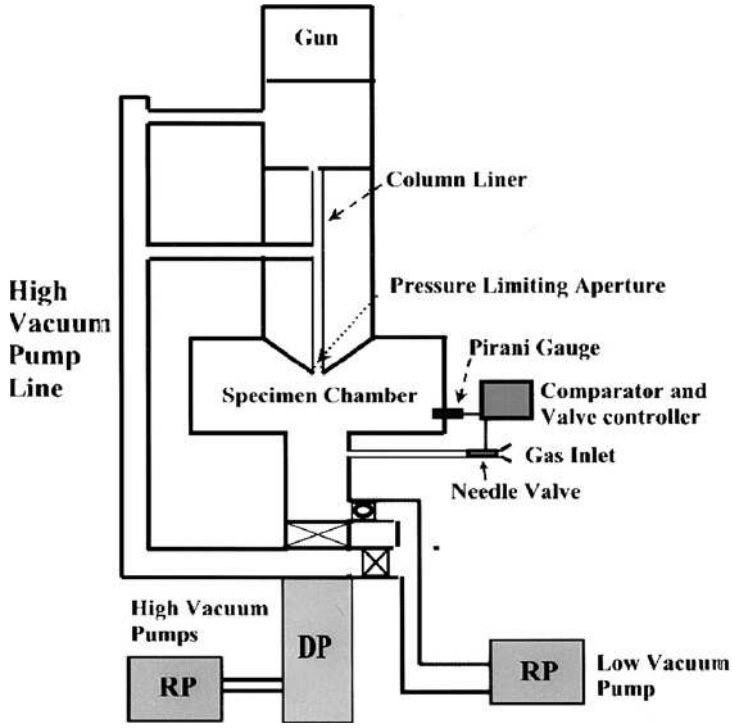


Figure 5.17. Schematic diagram of the arrangement of an environmental or a variable-pressure SEM showing differential pumping. The specimen chamber is isolated from the high vacuum of the column by a pressure-limiting aperture, and is pumped by a separate pumping line. The pressure in the specimen chamber is controlled by a motorized needle leak valve.

to allow a pressure differential of from 100 to 1000 times to be maintained between the specimen chamber and the column. In ESEMs, where pressures as high as 2700 Pa (20 torr) are employed, and in instruments where high-performance electron sources, either LaB_6 or field emitters, which have demanding vacuum requirements, are used, then two or more differential apertures may be arranged in series together with intermediate pumps to ensure that gas flow into the column is restricted to a safe value. Monte Carlo gas dynamics simulations of the gas flow around the PLA (Danilatos, 1993) show that the change in gas pressure occurs abruptly over a region no wider than the diameter of the aperture. As a consequence there are hypersonic gas velocities in the vicinity of the aperture and instabilities in both the gas pressure and temperature extending for a significant distance away from this region.

The isolation of the column from the specimen chamber can be improved by reducing the diameter of the PLA, but this leads to a problem when the instrument is used for low-magnification operation because the scanned rays then intersect the edge of the aperture causing vignetting of the image. Depending on the working distance, the lowest useful magnification is therefore typically 50× or more unless the instrument is specially modified (Wight and Taylor, 1995).

In a VPSEM instrument, the specimen chamber is pumped by the normal diffusion pump and roughing pump combination for conventional high-vacuum use, but when gas is present the sample region is typically pumped only by a rotary pump, leaving the diffusion pump to evacuate the column and gun. The chamber pressure is then controlled by means of a computer-operated leak valve, which can be either fully open or completely shut. The flow of gas into the chamber can be adjusted by varying the ratio of the time that the valve is open to the time that it is shut. The pressure is monitored and the leak valve is adjusted by a feedback loop until the target value (1–300 Pa) is achieved. Commonly the pressure is measured by a low-cost Pirani (i.e., a hot wire) vacuum gauge. Although this gauge reads accurately in the desired pressure regime, it responds only slowly to pressure variations and so it can take 30 s or more for the pressure to restabilize after an adjustment is made. It must also be noted that the pressure indicated by a Pirani gauge depends on the gas that it is measuring. A solution to both of these difficulties can be found by using a capacitance manometer gauge instead. This is a solid state, digital device which reads very rapidly and whose indicated pressure value is independent of the gas species.

In the ESEM a more sophisticated vacuum system is employed which offers greater control of the specimen environment as well as the ability to manage higher pressures up to 2700 Pa (~20 torr). This pressure range is achieved by having several successive levels of differential apertures and intermediate pumps to minimize gas flow back to the gun region. The upper pressure limit of 2700 Pa is chosen because, as shown in Fig. 5.18, this is the saturated vapor pressure of water at 22°C. With this pressure in the sample chamber, liquid water can be stabilized for observation. Fully hydrated biological tissue and samples that are moist or wet can be observed without any tendency for them to dry and shrink. In addition the ESEM has a facility to flood the sample chamber with water mist prior to experiments in which liquids or hydrated materials are being observed to ensure that full saturation is achieved so that no drying occurs during the pump down of

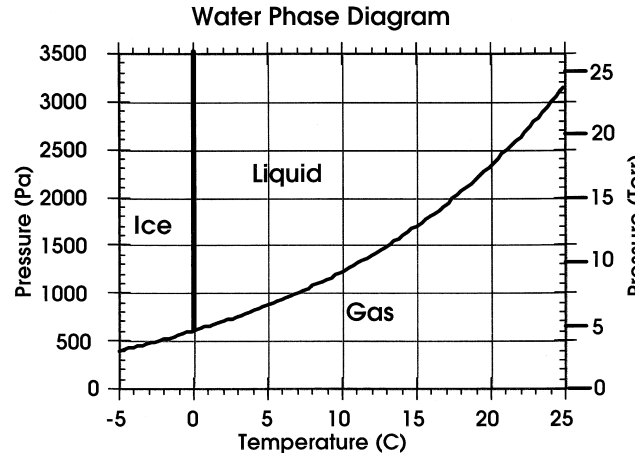


Figure 5.18. Equilibrium phase diagram for water/ice/water vapor.

the system. In VPSEM machines, where the maximum chamber pressure is limited to 300 Pa (~2.25 torr), liquid water cannot be maintained in equilibrium with water vapor, as can be seen from the equilibrium phase diagram for water in Fig. 5.18. Even if the specimen is cooled close to 0°C, the triple point where all three phases, gas, liquid, and solid, are in equilibrium occurs at 609.1 Pa (4.58 torr) and 0°C. Liquid water can only be observed in VPSEM machines in nonequilibrium situations where the liquid is actually evaporating to the gas phase and continuously drying out the specimen. Cooling the specimen to near 0°C before pumping in the VPSEM may slow the kinetics of this liquid–gas transformation enough to observe liquid water and to allow time to observe a wet specimen before it is significantly altered by drying. More-viscous liquids often exhibit pressure–temperature (P–T) relations that permit equilibrium with the gas phase under P–T conditions that can be achieved in the VPSEM.

5.6.3. Electron Interactions with Gases

Electrons interact with gases in the same sorts of ways as they interact with solids (Chapter 3), but the scale of distance over which the interaction occurs is very different because low-pressure gases have a density 1 million times less than that of a solid. Table 5.2 shows the mean free path (MFP) for 25-keV electrons traveling through an atmosphere of oxygen for a minimum scattering angle of 1 deg. [For electrons of some other energy E (keV) these values must be multiplied by a factor $(E/25)^{1/2}$.]

When the pressure is in the 10^{-4}-Pa (10^{-6}-torr) range, typical of the vacuum level found in a conventional SEM, then the electron mean free path is of the order of 10 km and so the probability of any given electron being scattered on its way along the roughly 1-m path from the gun to the specimen is negligible. However, when the pressure is raised to 100 Pa, the mean free path drops to about 1 cm, which is of the order of the working distance of a normal SEM. For still higher pressures the mean free path becomes 1 mm or less. The effect of this change in the mean free path can be understood by considering a simple model. The number of collisions m suffered by an electron traveling a distance s when the gas mean free path is λ is $m = s/\lambda$. Because the distance traveled by each electron in the VPSEM or ESEM is the gas path length (GPL), the average number of collisions per electron $m = \mathrm{GPL}/\lambda$. It can also be shown that the chance of an electron reaching the specimen without being scattered at all is $\exp(-m)$. Depending on the value of m, different imaging situations are encountered.

Table 5.2. Mean Free Path (MFP) for Electrons Traveling through Oxygen[a]

Pressure (torr)	Pressure (Pa)	MFP
10^{-6}	10^{-4}	10 km
1	133	1.4 cm
10	1330	1.4 mm
50	6650	0.28 mm

[a] 25 keV; minimum scattering angle is 1°.

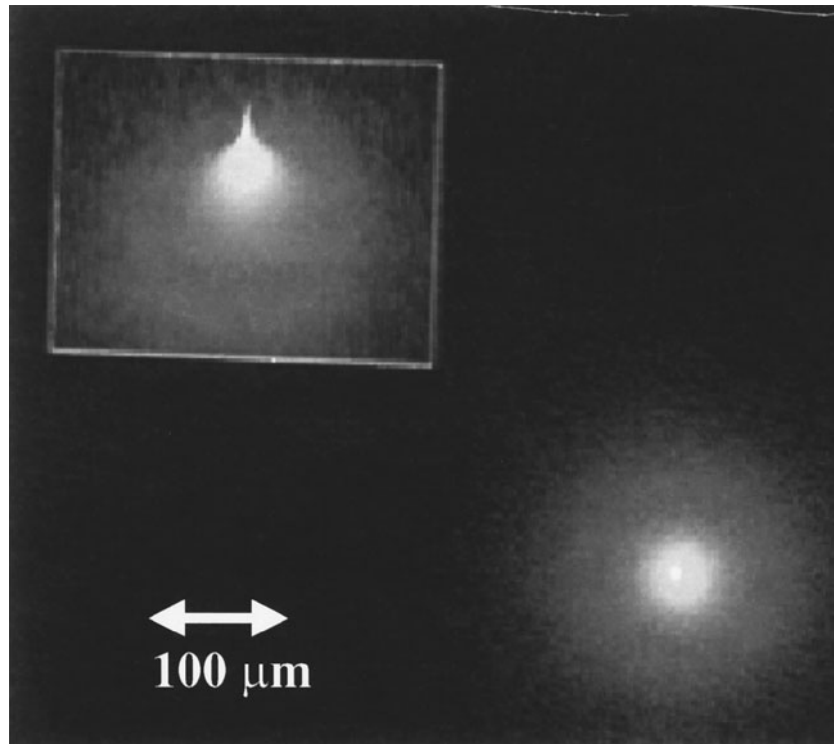


Figure 5.19. Image of beam profile in ESEM (Wight and Zeissler, 2000).

If the chamber is maintained at high vacuum ($<10^{-3}$ Pa), then for a normal working distance or gas path length of 15 mm, m will be 0.0001 or less. In this case essentially all of the beam current leaving the lens will reach the specimen, and the beam profile will remain undisturbed because scattering events will be rare. This is the situation encountered in a normal SEM. If the chamber pressure is high, however, for example, 400 Pa ($\sim$3 torr), then the mean free path shrinks 5 mm and so for a gas path length of 15 mm, m now has a value of 3. In this case only about $\exp(-3) = 0.05 = 5\%$ of the beam is unscattered, and the focused beam is almost completely destroyed, so scanning electron microscopy would be impossible. If, however, the path length is reduced to a few millimeters or the gas pressure is lowered to 100 Pa or so, then the value of m becomes about unity. In this case a significant portion of the beam [$\exp(-1) = 0.37 = 37\%$] remains unscattered, and a focused probe is still formed, although it is surrounded by a "skirt" of scattered electrons. The effect is analogous to that of viewing a car headlight through fog, where the main beam of light would still be visible, but would be surrounded by a halo or skirt of scattered light. Figure 5.19 shows a visualization of this kind of beam (Wight and Zeissler, 2000) obtained in an ESEM. The very sharp and intense central focused beam can be seen to be surrounded by a broad and less intense skirt region extending for a significant distance away.

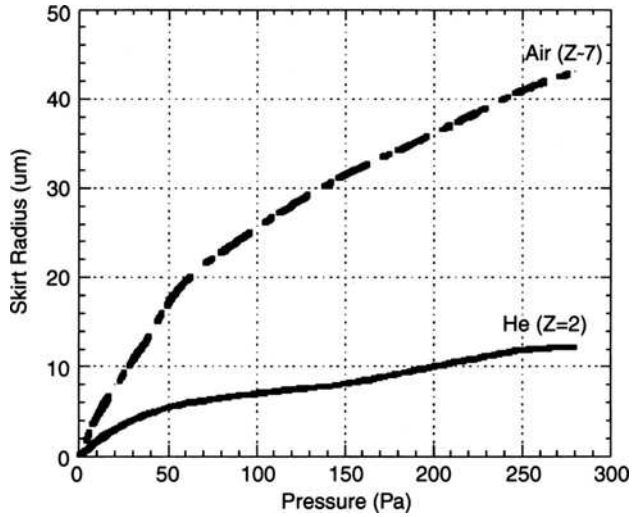


Figure 5.20. Plot of beam broadening (skirt radius) as a function of gas species and pressure. The gas path length is 5 mm and beam energy is 20 keV.

The scattering skirt is an extremely important effect in VPSEM/ESEMs. The width r_s of the skirt region can be estimated from a simple analytical model given by Danilatos (1988):

$$r_s = (364Z/E)(p/T)^{1/2}\,\mathrm{GPL}^{3/2}, \tag{5.5}$$

where r_s and the gas path length GPL are in meters, p is the pressure in pascals, T is the temperature in degrees Kelvin, Z is the atomic number of the gas, and E is the energy of the beam in electron volts. Figures 5.20 and 5.21

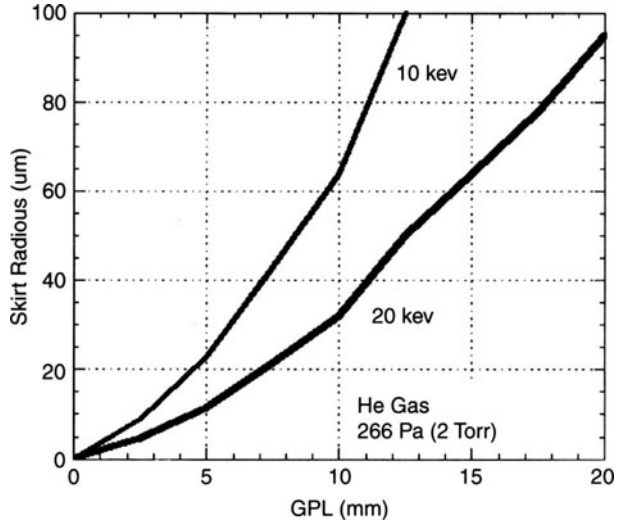


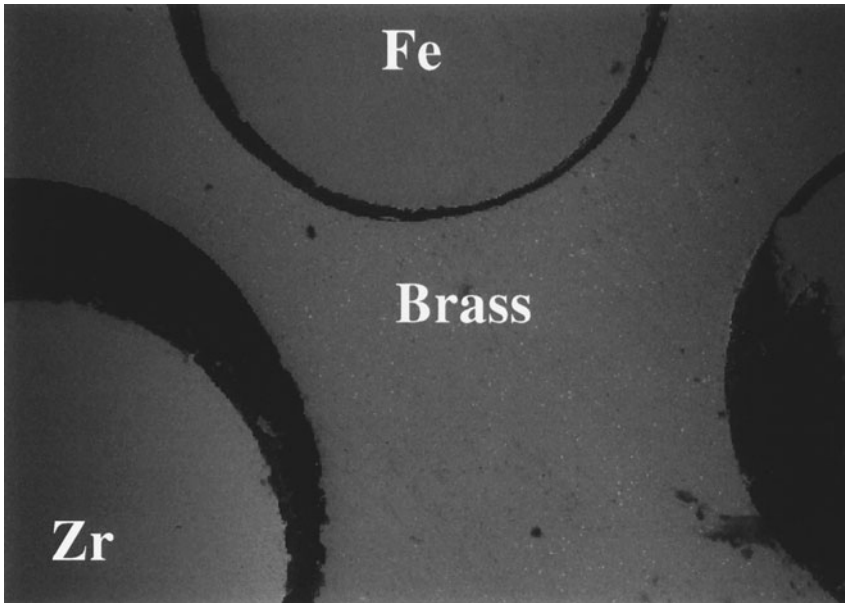
Figure 5.21. Plot of beam broadening (skirt radius) as a function of gas path length (GPL) and beam energy in helium.

show how the skirt radius varies as a function of the working distance, the beam energy, the pressure, and the type of gas. It can be seen that under all conditions the diameter of the skirt is typically tens of micrometers, and so is much greater than the size of the beam interaction volume in the solid. The spatial resolution of x-ray microanalysis will therefore always be defined by the spreading that occurs in the gas, and this resolution may be a factor of 10 or even 100 times worse than would be the case if the gas were absent. This is illustrated in Fig. 5.22, which shows spectra from microanalysis performed on a microprobe standards block. The electron beam is placed about 50 μm from the edge of the iron standard, which is surrounded by brass (copper and zinc). A further 500 μm away is a zirconium standard (Fig. 5.22a). The beam energy is 20 keV, and the gas path length and working distance are 12 mm. At high vacuum (Fig. 5.22b) only iron x-rays are detected, as would be expected because the beam interaction volume at 20 keV is only of the order of 3 μm. However, with an atmosphere of 20 Pa (0.15 torr) of air (Fig. 5.22c), copper and zinc x-ray lines begin to appear in the spectrum, indicating that the beam skirt already extends into the brass. When the pressure is increased to 50 Pa (0.38 torr) and then 110 Pa (0.83 torr), as shown in Fig. 5.22d, not only do the copper and zinc signals become much larger relative to iron, but a faint zirconium signal is observed, demonstrating that at least some electrons are being scattered by as much as 500 μm from the beam axis.

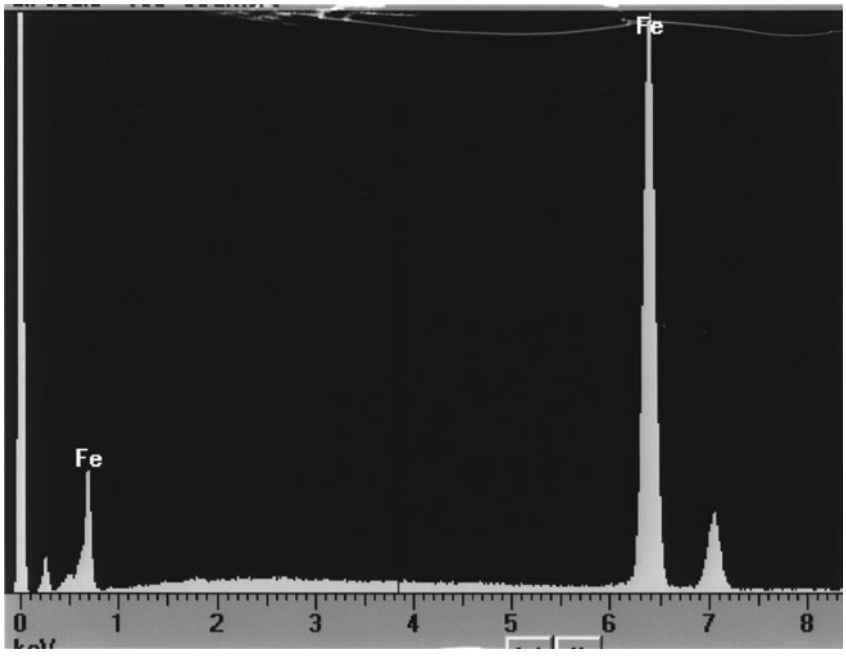
In order to reduce such scattering, as can be seen from Eq. (5.5), it is necessary to minimize the GPL and the gas pressure and to use the highest available beam energy. Since, for some modes, such as EDS microanalysis, the specimen must be at a fixed working distance to accommodate the EDS detector, the only way that this can be accomplished is to make the gas path length much smaller than the working distance. The high-vacuum region can be extended down toward the specimen through a small beam tube, as shown in Fig. 5.23, so that the pressure-limiting aperture is moved from the bottom of the lens to the bottom of the tube. In this way the GPL can be as small as 1 or 2 mm while the working distance remains at 12 or 15 mm, and so the spatial resolution is greatly enhanced.

The beam broadening does not reduce the resolution of the SEM image because a large fraction of the beam intensity remains unscattered and in the form of a focused probe. Under most imaging conditions the diameter of the skirt is comparable to the field of view of the total image. As the beam is rastered across the specimen, the contribution from the skirt stays essentially constant and the only contribution to signal information comes from the probe region. Thus, although the image contrast is reduced because of this background, and the signal-to-noise ratio is lowered because of the reduction in incident beam current caused by gas scattering, the resolution remains unchanged. Images recorded under the same beam conditions, but at gas pressures of less than 1 Pa (0.0075 torr) and about 200 Pa (1.5 torr) are compared in Fig. 5.24.

The simple model outlined above is satisfactory for computing the spatial resolution for microanalysis, but it provides no information about

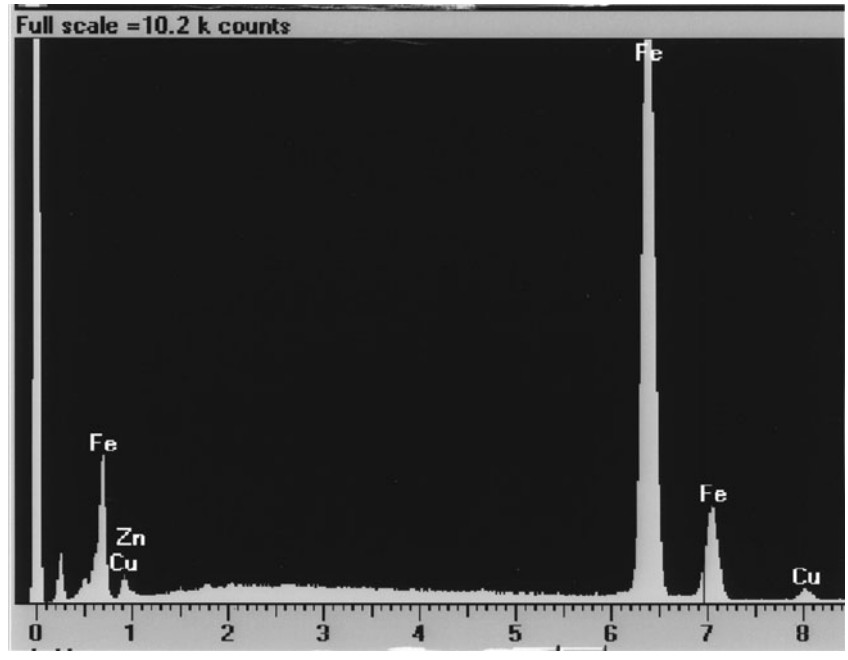


a

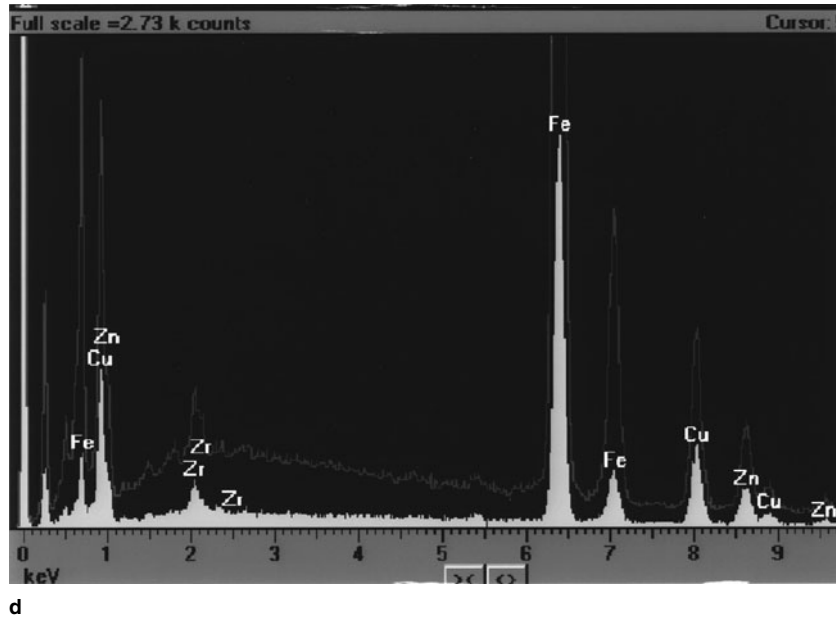


b

Figure 5.22. Beam broadening experiment using EDS to detect remote, gas-scatttered electrons. (a) Spatial layout of iron (Fe) and zirconium (Zr) standards in brass (Cu–Zn) holder. The beam is positioned near the edge of the iron standard. EDS spectra from Fe: (b) high vacuum, (c) 20 Pa of air, (d) 110 Pa of air. A change of scale of the EDS spectrum in (d) shows the Zr peaks and continuum background in more detail (gray trace). Working distance = 12 mm, beam energy = 20 keV.



c



d

Figure 5.22. (*Continued*)

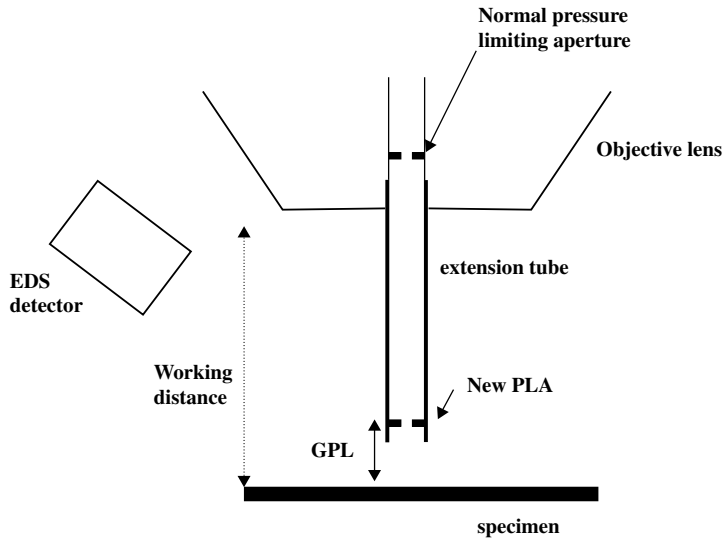
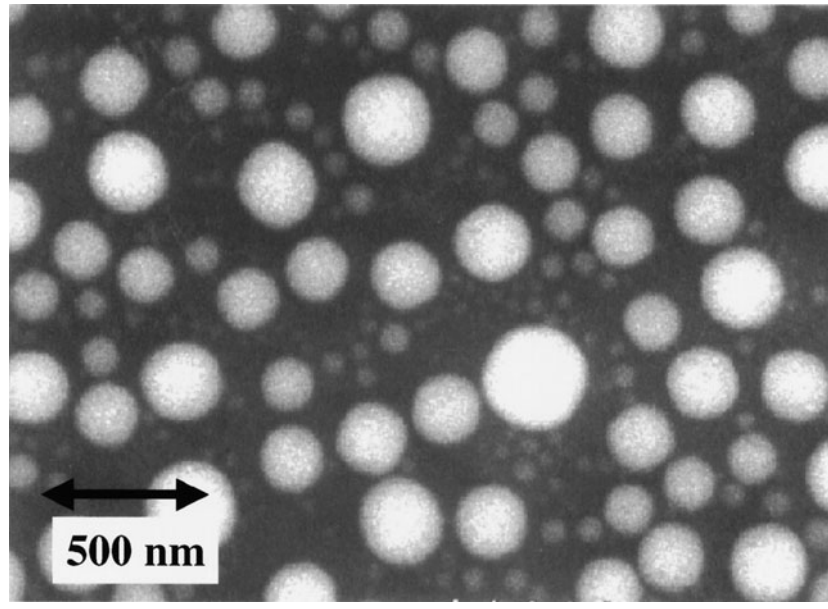


Figure 5.23. Schematic diagram of EDS tube used to shield the beam from chamber pressure and thus to shorten the gas path length.

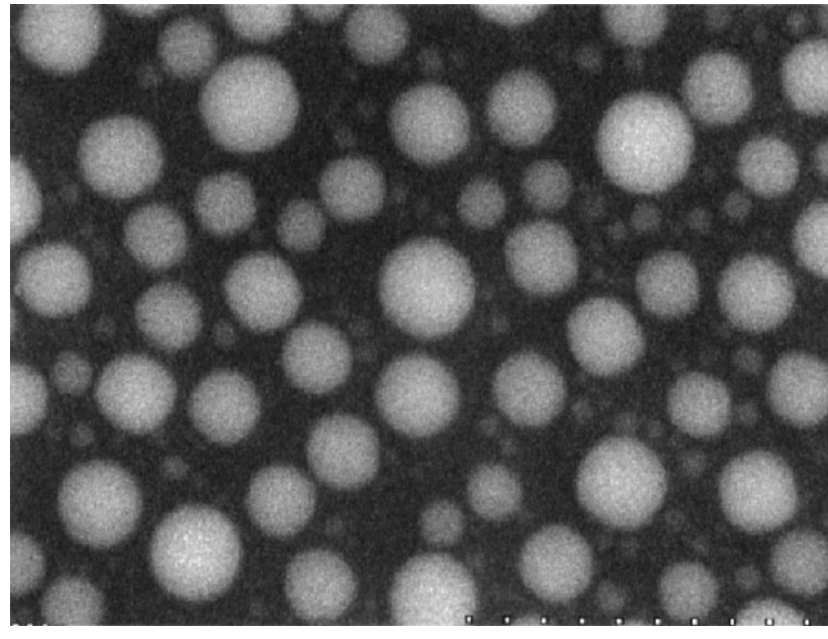
the intensity profile of the scattered beam and so is of little use in understanding the effects of the gas on imaging. More detailed models (Thiel *et al.,* 1996, Joy 1996) using Monte Carlo methods have been developed which can predict the shape of the beam profile as well as the magnitude of the beam broadening. As would be expected from the previous discussion, these profiles show an intense and unscattered central peak surrounded by a broad skirt of scattered electrons. As the pressure and the GPL increase, the intensity of the central peak decays and the width and the intensity of the skirt increase. Eventually all of the beam is scattered and no focused component remains. The agreement between skirt diameters computed by either the simple model or Monte Carlo methods and experimental measurements (Wight and Zeissler, 2000) is generally good, although the comparison of experimental and computed beam profiles is less encouraging. More work therefore must be done to refine such models before they can reliably be used to make predictions.

5.6.4. The Effect of the Gas on Charging

As discussed in Chapter 4, insulating or poorly conducting specimens often acquire an electric charge when they are imaged in the SEM. Whether or not this occurs depends on the "charge balance" at the sample. If the number of incident electrons impinging on the surface in unit time equals the number of electrons (secondary and backscattered) leaving the sample, then the specimen is in charge balance, and a sample that starts off electrically neutral will remain in that state. If the incident beam current exceeds the number of emitted electrons, however, then the sample will acquire a negative charge. Because the field from this charge affects the



a



b

Figure 5.24. Images of metal spheres recorded at 20 keV and a gas path length in air of 8 mm at gas pressures of below 1 Pa (0.0075 torr) (a) and about 200 Pa (1.5 torr) (b).

collection of secondary electrons, producing instabilities in the signal level and extremes of contrast, and may even deflect the incident beam, causing a distortion of the image, charging is always found to be a major problem for scanning electron microscopy.

The presence of gas in the specimen chamber of the ESEM or VPSEM can overcome the problem of charging, making it possible to observe all materials however insulating they may be. This is because the interactions between electrons and the gas result in the production of ions, which can change the charge balance at the specimen. Electrons can ionize the gas by stripping off an outer shell electron producing a positively charged nucleus and a electron (e.g., in helium $He \rightarrow He^+ + e^-$) or by breaking up a molecule to form two charged components (e.g., in water vapor $H_2O \rightarrow H^+ + OH^-$). In either case the interaction results in the production of particles with both positive and negative charges. The positively charged ions will drift to negatively charged areas of the specimen, and negative ions drift toward positively charged areas. If sufficient ions are produced, then the charging of the specimen can be neutralized by the flood of positive ions.

Although any electron–gas interaction could produce an ionization event, the probability of this happening depends strongly on the energy of the beam electron. The probability that an energetic electron will eject a bound outer shell electron from an atom increases from zero when the electron's energy just equals the ionization energy for that bound atomic electron to a maximum at about a factor of four times the ionization energy, beyond which the probability decreases with further increases in the primary electron energy. From the data in Table 5.2 we see at a gas pressure of 133 Pa (1 torr) a 25-keV incident electron would travel 10–15 mm before colliding with a gas molecule and causing an ionization. Depending on the beam gas path length, a beam electron might not produce an ionization at all. However, a backscattered electron in the low end of the BSE energy distribution with an energy of a few keV, traveling through the same gas density, would only go 3–4 mm before generating an ionization in a collision. A secondary electron in the high end of the SE energy distribution (0–50 eV), with an energy of just tens of eV, would experience a collision with a gas atom and produce an ionization within a few micrometers of leaving the surface. Consequently it is the lowest energy electrons that are most likely to generate an ionization. If in addition an electric field of a few hundred volts potential is applied, the SE will be repeatedly accelerated after each ionizing collision, producing a cascade of ejected electrons and positive ions. If the sample is charging negative, as will normally be the case, or if the positive bias field is applied to a nearby detector, the electric field will accelerate SE and negative ions toward the detector or lens, while positive ions are accelerated back toward the negatively charged areas on the specimen. As these electrons and ions accelerate and increase in kinetic energy, they can in turn produce other ionization events, and these electrons and ions are then in their turn accelerated producing still more ions. This "gas amplification" effect thus produces a much larger number of electrons and positive ions than the initial ionization events.

When imaging an insulating specimen it is only necessary for the operator to supply sufficient gas to produce a stable image. The positive ions

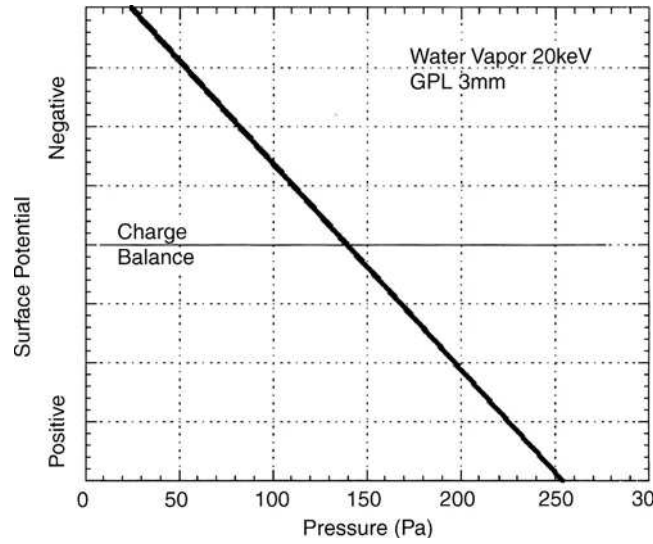


Figure 5.25. Schematic of variation in charging of an insulator with gas pressure. Gas path length = 3 mm.

will drift toward negatively charged regions until they are neutralized, so the process is self-controlling and does not require continuous intervention by the operator. If the conditions in the specimen chamber are varied over a wide range, then, as shown schematically in Fig. 5.25, it is found, for a given sample, gas, and beam energy, that the charge state of a sample varies from negative to positive as a function of the product of the gas pressure and the gas path length. Thus charge balance can be achieved by using either a higher pressure or a longer gas path length. This is reflected by the differences between the nominal operating conditions in the ESEM and the VPSEM. In the ESEM the GPL is typically short (2–3 mm), so a relatively high gas pressure (~150 Pa or more) is required for charge balance, whereas in a VPSEM, where the gas path length is usually 10–15 mm, the gas pressure can be reduced to 20–30 Pa. However, as shown from Eq. (5.5), because since the beam broadening varies as the product $p^{1/2} \cdot \mathrm{GPL}^{3/2}$; whereas the condition for charge balance depends on $p \cdot \mathrm{GPL}$, it is clear that the optimum condition is to use the shortest GPL and the highest pressure. Note also that if too high a pressure or too long a GPL is used, then the specimen can be driven into positive charging (Griffin and Nockolds, 1999), which may be as unwelcome as negative charging. This problem can be avoided by monitoring the specimen current while bleeding in gas, because charge balance has been achieved when the average specimen current is zero.

The choice of the gas in the specimen chamber is also an important variable. By default the residual gas in the microscope after pump down will be the laboratory mixture of air and water vapor, and so it is this mixture that is typically used for the majority of work. It ionizes easily and so is efficient for the control of charging. The effective atomic number of the mixture is reasonably low, so that beam broadening, as determined from

Eq. (5.5), is acceptably small. For more specialized applications, several other gases are often used. When biological samples, or any other type of material which must be kept hydrated, are being examined it is desirable to use saturated water vapor as the chamber gas. This is possible in the ESEM because, as shown in Fig. 5.18, the partial pressure of water at room temperature is about 2600 Pa (20 torr), and lower pressures can be used by modestly cooling the specimen to ~5°C. Because water vapor ionizes readily and produces acceptable beam broadening, this choice is generally satisfactory. However, if the partial pressure of water vapor in the system is deliberately set very high, for example, to ensure that a hydrated sample does not dry, then positive charging may occur and the drift of the negative ion OH^- species from the water can result in severe radiation damage to polymers, soft tissue, and similar materials (Fig. 5.26). Where the requirement is to minimize beam broadening as much as possible (e.g., for x-ray microanalysis) then the gas of choice is helium (Stowe and Robinson, 1998) because this has a low atomic number, and is also inert and so is not prone to cause ionization damage to fragile specimens. Unfortunately, pure helium does not ionize easily, which can make charge control and imaging

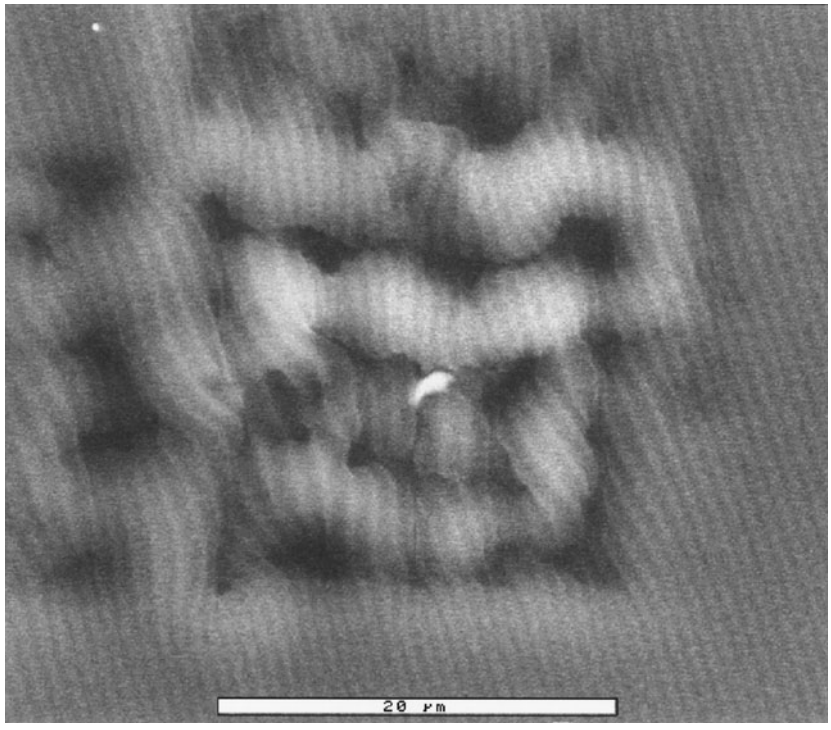


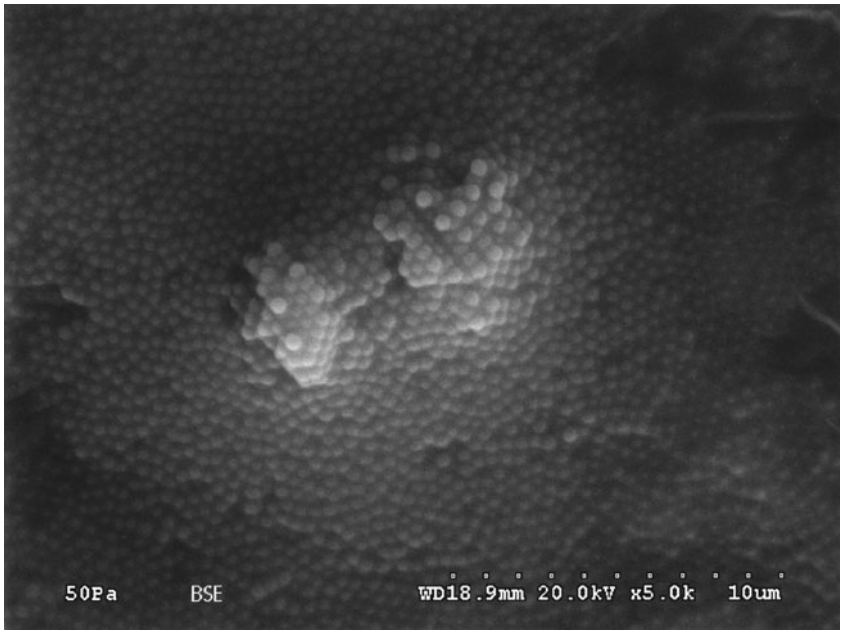
Figure 5.26. Water vapor abetted radiation damage of epoxy polymer grating in VPSEM/ESEM; $E_0 = 20$keV; pressure = 730 Pa (5.5 Torr). The undamaged grating can be seen at the edge of the field of view. Extreme distortion of the grating is observed at the center of the field which was previously scanned at higher magnifications (i.e., smaller scan fields). Bar = 20 micrometers.

difficult, although this can usually be overcome by adding a few percent of methane or a similar gas to the chamber.

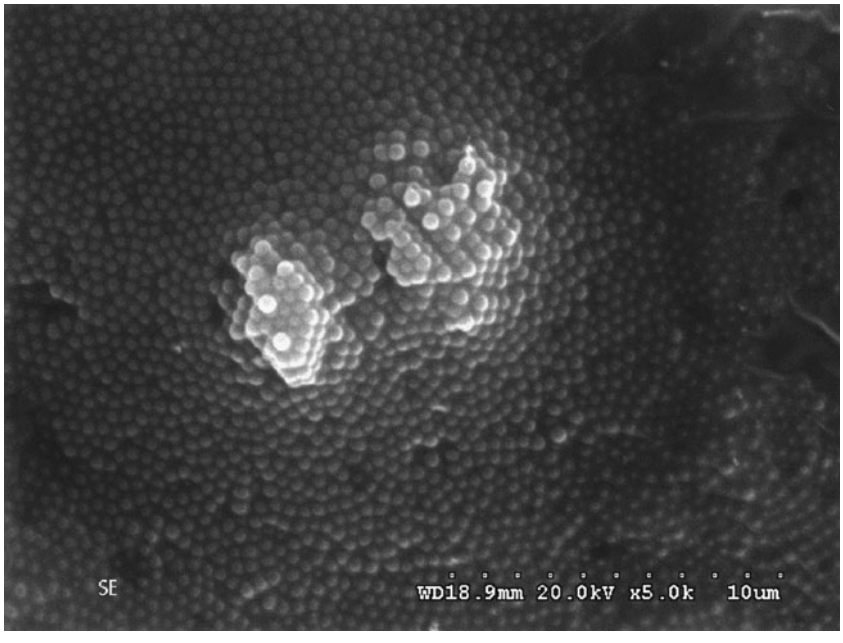
5.6.5. Imaging in the ESEM and the VPSEM

The incident beam of electrons hitting the surface of the specimen in the ESEM or the VPSEM produces the same result as in the high-vacuum SEM, so, in principle, the usual backscattered and secondary electron signals are available. However, although there is no problem in collecting the backscattered electron signal, the secondary electron signal is not available because, as noted above, the SE only travel a very short distance from the surface of the sample before they interact with and ionize a gas molecule and so it is difficult to collect them. Moreover, the conventional E-T detector used for SE collection relies on a high (+10 kV) bias voltage applied to the scintillator to accelerate the SE to sufficient kinetic energy to stimulate light emission from the scintillator. In the high-gas pressure environment of the specimen chamber, where a potential of more than just a few hundred volts would cause the gas to ionize and the detector to arc over to ground, the E-T is thus not a viable option. It is thus impossible to collect a SE signal in the conventional manner. Because images prepared with a positively biased E-T detector (consisting of SE and BSE contributions) account for more than 90% of all SEM images, this is clearly a problem.

VPSEMs switch off the E-T detector when there is gas in the chamber and rely on the passive backscattered electron detector. At first sight this may appear to be a poor exchange, but in fact for many purposes a BSE detector is a quite acceptable substitute route for imaging. First, VPSEMs are typically used at beam energies of 10 keV and higher (to minimize beam broadening) and for almost all materials the yield of BSE is higher in this energy range than the SE yield. Because modern BSE detectors are very efficient, the quality of the BSE image is in every way comparable with the corresponding SE image. Second, most VPSEM operation occurs at magnifications below 20,000×. In this operational range the pixel size of the image is of the order of a fraction of 1 μm or more, and consequently the information in the SE image comes predominantly from the SE_2 electrons (i.e., those secondaries generated by the backscattered electrons) because the generation volume of the SE_2 signal is comparable with the pixel size. The SE_2 and the BSE signals are therefore almost identical in information content and resolution because one is essentially a copy of the other (Fig. 5.27). This does not necessarily imply that the images will look identical because the BSE and SE detectors are usually placed in rather different locations relative to the specimen (remember that the apparent source of illumination always comes from the detector), but it shows that there is no fundamental loss of imaging capability in this condition. Finally, because BSE detectors are much less sensitive to charging effects than are SE detectors, the use of a BSE detector in the VPSEM simplifies the challenge of overcoming charging. The strategy of providing only a BSE detector in a VPSEM is therefore not a major limitation to what can be achieved in the instrument.



a



b

Figure 5.27. A comparison of (a) a backscattered image and (b) a secondary electron image from a specimen of opal. In the range of operating energies and magnifications normally used in the VPSEM and the ESEM the secondary electron signal is SE_2 in type, that is, produced by the backscattered electrons.

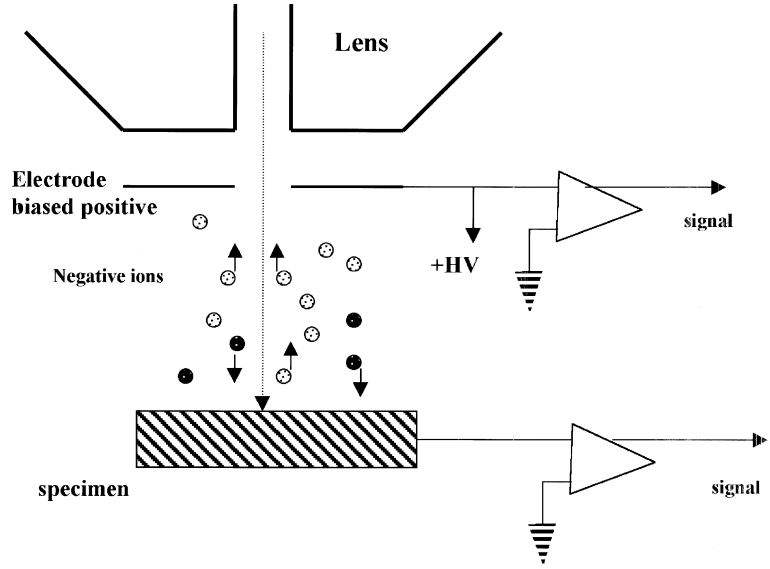


Figure 5.28. Schematic diagram for gaseous SE detectors (GSED) or environmental SE detectors (ESED).

Nevertheless there are instances where a true secondary image is desirable, for example, to image the actual surface of a specimen, and so in the ESEM and in some VPSEM instruments, special devices known as environmental SE detectors (ESED, Danilatos, 1990) or gaseous SE detectors (GSED; Farley and Shah 1991; Mohan *et al.,* 1998) are now provided to fill this need. Although the exact details vary from one system to another, the fundamental principle behind all of these devices is identical (Fig. 5.28). As seen above, it is the secondary electrons that are responsible for the majority of the gas ionization, so the amount of ionization varies with the SE yield. In the presence of an electric field, produced by positively biasing an electrode above the sample, the SE are accelerated and eventually gain enough energy (10–30 eV) to ionize another gas molecule. The two electrons that result from this process are again accelerated and each in turn may generate two more electrons and so on, producing an avalanche effect known as gas amplification. The magnitude of the amplification depends on the ease of ionization for the chosen gas and on the field available. With potentials of the order of a few hundred volts (low enough to avoid electrical breakdown of the gas) and electrode-to-sample distances of a few millimeters, the electron current collected on the biased electrode is hundreds or thousands of times larger than the original signal, which is a fraction (5–20 %) of the beam current. The positive ions produced in this cascade are themselves accelerated away from the positive electrode toward the specimen, forming an ion current equal in magnitude to the current of electrons moving in the opposite direction. For the same kinetic energy, ions move much more slowly than

electrons because of the great difference in mass (a factor of 1837 per proton equivalent, or about 33,000 for an H_2O^+ ion). The slowly moving ions determine the time-dependent response characteristics of this class of detector.

The electrode and the sample effectively form the two plates of a capacitor with the gas as a dielectric, and the drift motion of the ions between these plates induces a current in the external circuit, which can then be amplified and used to produce the required image. It is not necessary for the ions (or even any residual electrons) to be physically collected by the detector because the motion of any charged particles between the plates will result in a signal current (Ramo, 1939). Because the initial source of the ionization was the secondary electrons generated by the incident beam, the signal from this system is in every respect a secondary electron image, although it was carried by ions. A comparison of high-vacuum SE images (Fig. 5.27a) and the corresponding images generated by ESED or GSED systems in gas (Fig. 5.29) confirms that the images are identical except for changes that can be attributed to the different positions of these detectors relative to the specimen. There is, however, one practical difference between ESED/GSED images and those recorded using an E-T detector. Because ions travel much more slowly than electrons of similar energy, unless slower scan rates are used the image appears blurred because information from one pixel dwell time period is still being

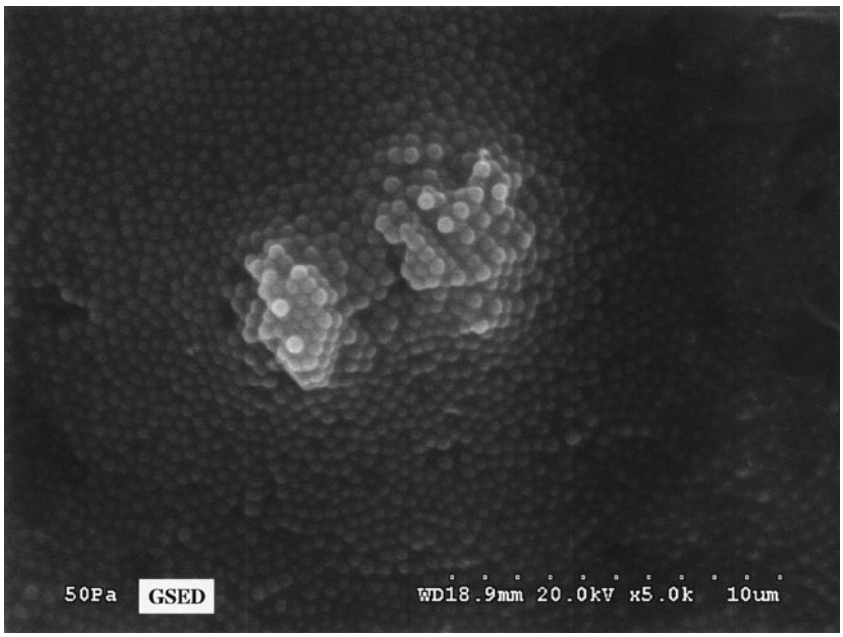


Figure 5.29. GSED image of an opal crystal, 20 keV beam energy, 50 Pa (0.38 torr) of air. A comparison of this image with the corresponding high-vacuum SE image of the same area (Fig. 5.27) shows that the detail and contrast information are essentially identical.

collected when the beam has already stepped onto the next pixel (Newbury, 1996).

The ESED and GSED detectors give ESEM and VSPEM instruments all of the imaging power and flexibility of the conventional E–T detector, and they also provide other benefits. Because surface charging can always be eliminated by the gas, the images that are obtained do not show the distortions, instabilities, and extremes of contrast due to charging that are encountered on conventional images. An excellent example of this (Fig. 5.30) are images from deep holes in a surface (Newbury, 1996). In a conventional SE or BSE image the bottom of the hole in an insulator is dark because charging on the walls of the hole traps electrons and prevents them from leaving. In the presence of gas, and when using an ESED or a GSED system, however, not only does the absence of charging allow electrons to leave freely, but the added gas path length experienced by the ions that are generated increases the amplification with the result that the bottom of the hole is actually brighter than the surface itself.

An additional benefit of the ESED/GSED detectors is that, unlike the conventional E–T detector, they are not sensitive to light. They can therefore be used to image samples that are at a high temperature (i.e., above about 750°C, where the specimen begins to visibly glow) without the annoyance of a strong and noisy background signal. This ability is exploited in the ESEM by the provision of stages capable of taking specimens to 1500°C or even higher.

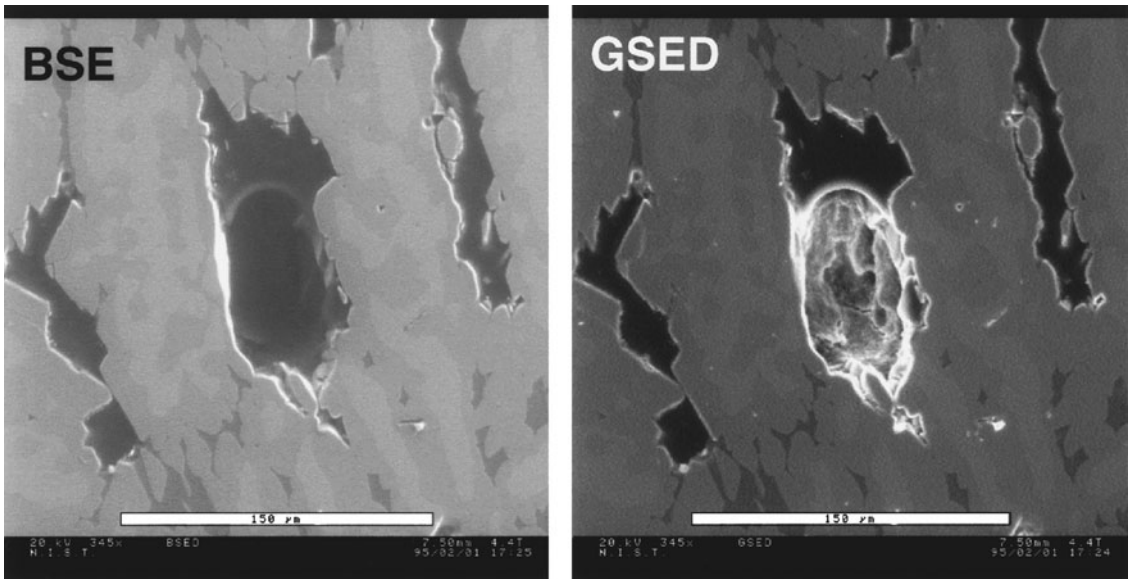


Figure 5.30. Deep holes imaged in Raney nickel alloy by (a) conventional passive BSE detector and (b) GSED detector (Newbury, 1996).

The beam broadening that occurs when the incident beam passes through a gas has a major effect on the x-ray analysis that can be performed because x-rays will be generated everywhere that electrons strike the surface. Because, at a typical working distance of 15 mm and a gas pressure of 30 Pa (0.22 torr), the beam broadening is tens or even hundreds of micrometers (see Fig. 5.20 and 5.21), the term "microanalysis" is not always appropriate. In addition electrons scattered out of the beam contribute to a fall in count rate and may lead to the production of spurious "systems" peaks from the stage or other pieces of hardware in the chamber. However, this drawback must be balanced against the advantage that an ESEM or a VPSEM need not be restricted to low beam energies in order to analyze nonconducting samples and so more usable lines and higher count rates are available.

A variety of techniques has been suggested to permit meaningful qualitative and even quantitative microanalysis to be carried out under these conditions. In one example (Doehne, 1997), a sequence of spectra is recorded

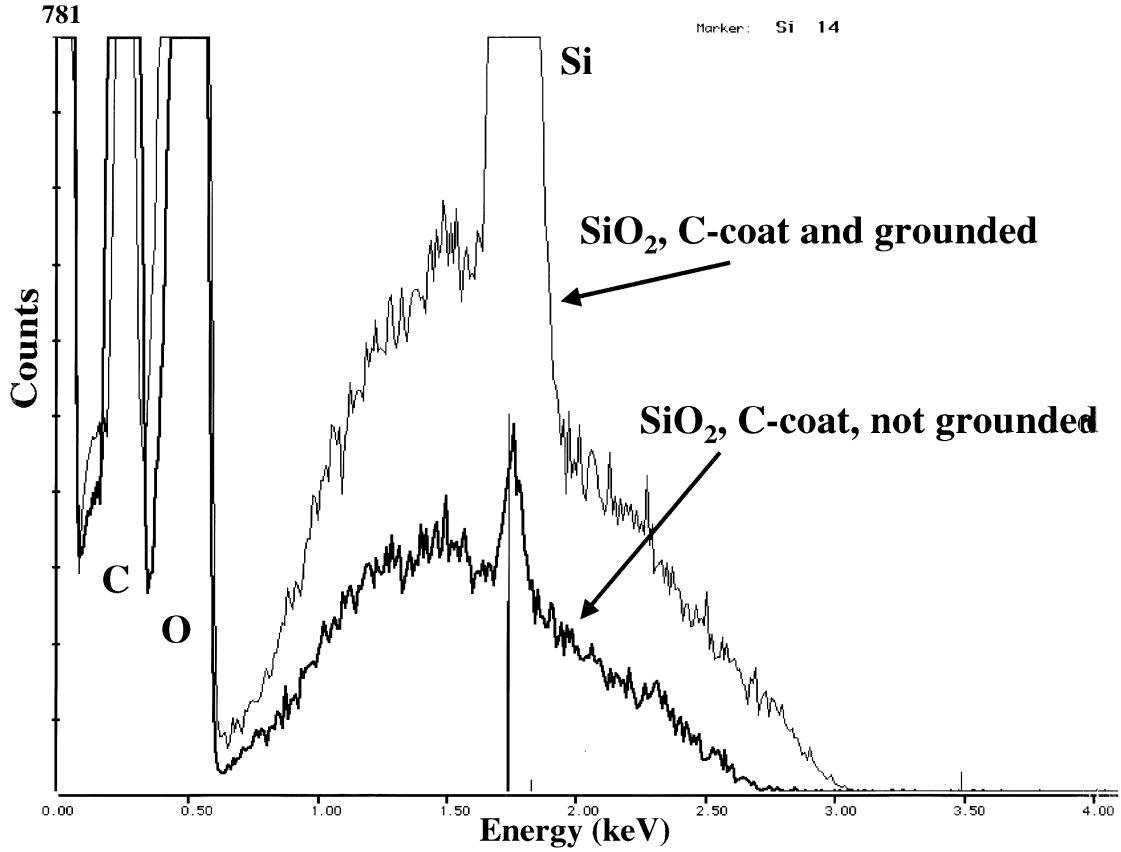


Figure 5.31. Use of Duane–Hunt cutoff for checking E_0, showing charging effects in SiO_2 films (Newbury, 2001).

at different gas pressures while keeping the beam positioned over the feature of interest. Peak intensities in the spectra are then extrapolated back to simulate a high-vacuum condition. However, careful tests of this and other published procedures (Mansfield, 1997) on well-characterized samples show that none of them is completely reliable and that great care must be taken in quoting and interpreting any results obtained under such conditions. Until sufficiently accurate and detailed models of the beam scattering profiles are available to be used for deconvolution, or some more reliable protocol for data correction is developed, the least hazardous approach is to treat all spectral data as being effectively global in extent. Useful, spatially resolved information about major constituents ($C > 0.1$ mass fraction) can still be obtained by x-ray mapping techniques because the incident electron dose (electrons per unit area) remains strongly peaked about the nominal beam position and so intensity maxima in the map will correlate with localized elemental distributions. Mapping sensitivity to trace ($C < 0.01$) and minor ($0.01 \leq C \leq 0.1$) constituents is usually lost due to gas scattering.

Because many of the samples that will be chosen for examination in an ESEM or a VPSEM will be insulators or poor conductors, added care must be taken in the interpretation of all x-ray data because, if the specimen is not brought to charge balance, then it might charge significantly. If the sample charges negatively, then the effective incident beam energy is reduced, which can result in some lines that would otherwise be present becoming either very weak or even invisible because the overvoltage becomes too low. Under conditions of excessive gas pressure (Griffin and Nockolds, 1999) specimens can also acquire a significant positive charge, which increases the energy of the electrons and could again lead to errors and ambiguities unless some account is made for the effect. The solution in either case is to monitor dynamically the Duane–Hunt high-energy cutoff of the continuum (Fig. 5.31) because this gives a precise indication of the true landing energy of the electrons during the analysis (Newbury, 2001).

Analyses performed at high pressure may also result in the production of emission lines from the gas (e.g., oxygen and nitrogen from air/water vapor) appearing in the spectrum. Because the magnitude of these peaks will depend directly on the gas pressure and path length, their origin can easily be determined. Finally, it should be noted that, despite some claims to the contrary, there is no significant absorption of emitted x-rays as a result of the gaseous environment. Because the density of gases at room temperature and a pressure of 100 Pa (0.75 torr) is only of the order of 10^{-5}, even soft x-rays will suffer minimal absorption for an exit path length of 10 mm to the detector.

5.7. Special Contrast Mechanisms

Topographic and compositional contrast provide the foundation of image information in scanning electron microscopy. These contrast mechanisms are encountered with virtually all specimens from the physical and

biological sciences as well as from engineering processes and materials and the environment. However, there exist several special contrast mechanisms that can only be observed with particular classes of solids or structures (Newbury *et al.,* 1986). These mechanisms carry information on specimen electric and magnetic fields and crystal structure and can be imaged on any SEM equipped with a conventional tungsten filament gun or an LaB_6 gun, which are capable of high-current operation. The key characteristics of these mechanisms and a guide to selecting the SEM operating parameters to observe them will be presented in this section along with references to more detailed information. As a general observation, all of the contrast mechanisms discussed below are relatively weak, which places special demands on specimen preparation. Specimen topography, which creates strong contrast, must be eliminated by appropriate preparation techniques to optimize detection of weak contrast mechanisms (see Chapter 11).

5.7.1. Electric Fields

Insulating specimens and engineering structures such as integrated circuits can develop electric fields as a result of externally applied potential(s) or in response to the charge injected by the primary electron beam. Variations in the local surface potential on such specimens can be detected as image contrast through the effects of that potential on the collection of secondary electrons, one of the earliest phenomena recognized with the SEM (Knoll, 1941). The Everhart–Thornley detector uses an applied positive potential of a few hundred volts to collect SE, but it is typically located 5 cm or more from the beam impact point on the specimen, so that the E-T collection field at the specimen is about 100 V/cm or less. For a noncharging specimen, this is an adequate field to collect SE with high efficiency. However, when even modest charging occurs, for example, to a few volts potential, in the beam impact region on the specimen, because of the small dimensions (typically micrometers) of the beam-affected region, the local electric field increases to orders of magnitude greater (10^3–10^5 V/cm) than the E-T electric field. Because of the inherently low kinetic energy of SE, typically less than 5 eV, these high local fields can strongly affect the SE trajectories. A positive surface potential acts to reduce the collection efficiency of the E-T. With a sufficiently high positive potential, the effect is to attract the SE back to the surface, thus recollecting them, which makes the image appear dark or even black in that area. Conversely, a negative surface potential acts to repel SE from the specimen and increase the apparent efficiency of the E-T, so that the region appears bright in the image. Such contrast effects in images due to differences in surface potentials from one location to another are designated "voltage contrast."

Engineered structures such as integrated circuits often involve features that are placed at different potentials during operation. These portions of the device can be directly observed by means of voltage contrast, provided that protective packaging layers can be removed. Complex circuits can be effectively studied with this approach, including stroboscopic methods to observe time-dependent sequencing (Gopinath *et al.,* 1978). An example

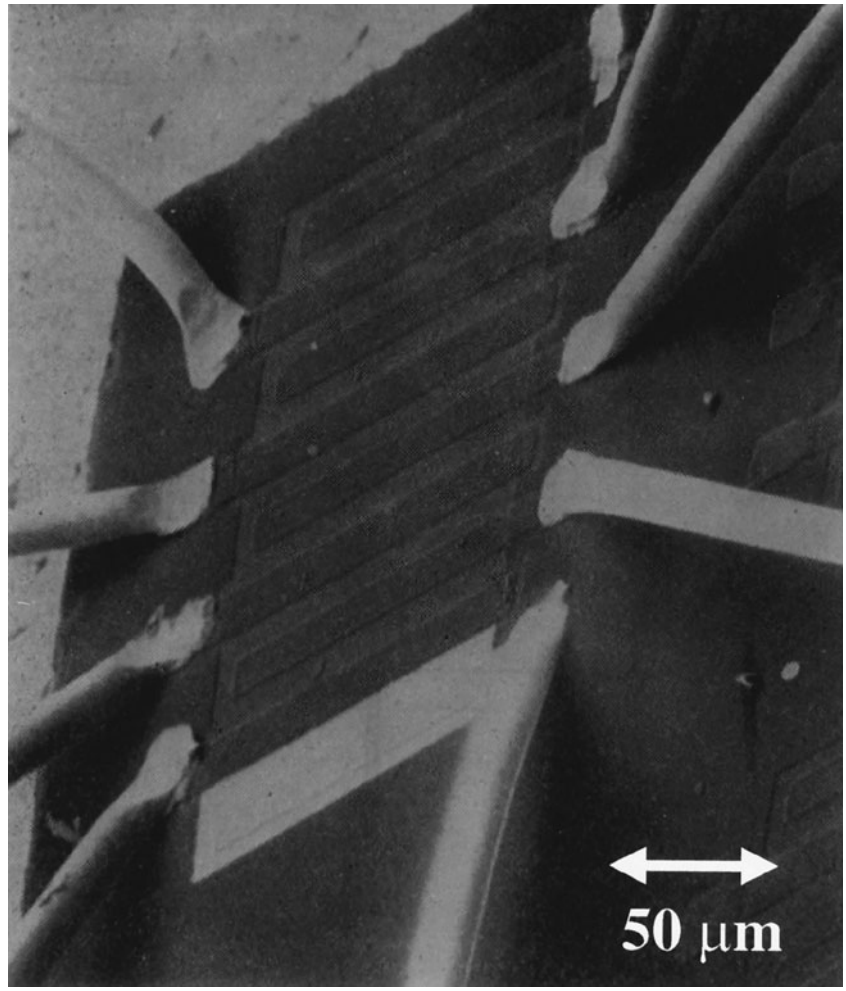


Figure 5.32. Voltage contrast of a circuit as viewed with the secondary electron signal measured with an Everhart–Thornley (+) detector.

of such an image is shown in Fig. 5.32, where the brighter a region is, the more negative is its potential.

The second type of voltage contrast we might encounter occurs due to localized specimen charging under the beam of insulating specimens, or conducting specimens with insulating inclusions. The pathological effects of specimen charging have been described previously in Chapter 4. It was noted that for an insulating specimen, the beam can inject sufficient numbers of electrons into the target to locally develop a surface potential equal to the beam potential, with the result that the specimen becomes an electron mirror, actually reflecting the primary beam and preventing it from reaching the surface. Charging due to the primary beam is a complex, dynamic phenomenon, highly dependent on the bulk and surface conductivities

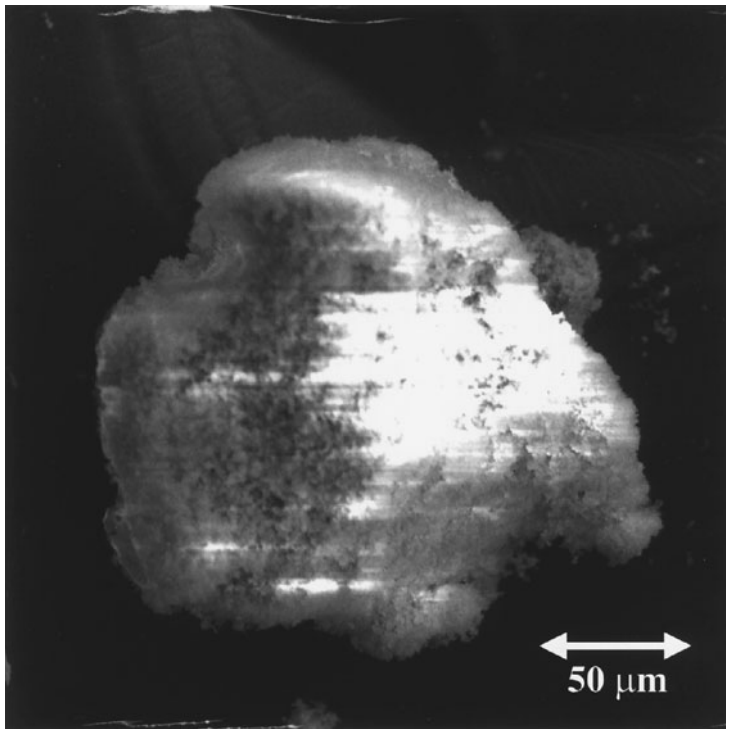
of the specimen, the immediate environment, and beam parameters including energy, current, scan speed, and angle relative to the specimen surface. Local charging that produces surface potentials of a few volts can often be observed as stable features that persist as long as the beam returns regularly during the scan. An example of such voltage contrast is shown in Fig. 5.33 for the commonly observed case of a dust particle. If the dose rate is changed by altering the scan speed or beam current, the charging pattern manifested as voltage contrast may change. Such changes are often dramatically evident when the scan speed is changed from rapid for viewing to slow for photographic recording. When this problem is encountered, the use of a digital framestore, which permits image recording by sampling at a continuous high scan rate, can often overcome the dynamic effects of charging. When we interpret such images, it is critical to recognize and distinguish features that arise from voltage contrast and those from topographic or compositional contrast.

5.7.2. Magnetic Fields

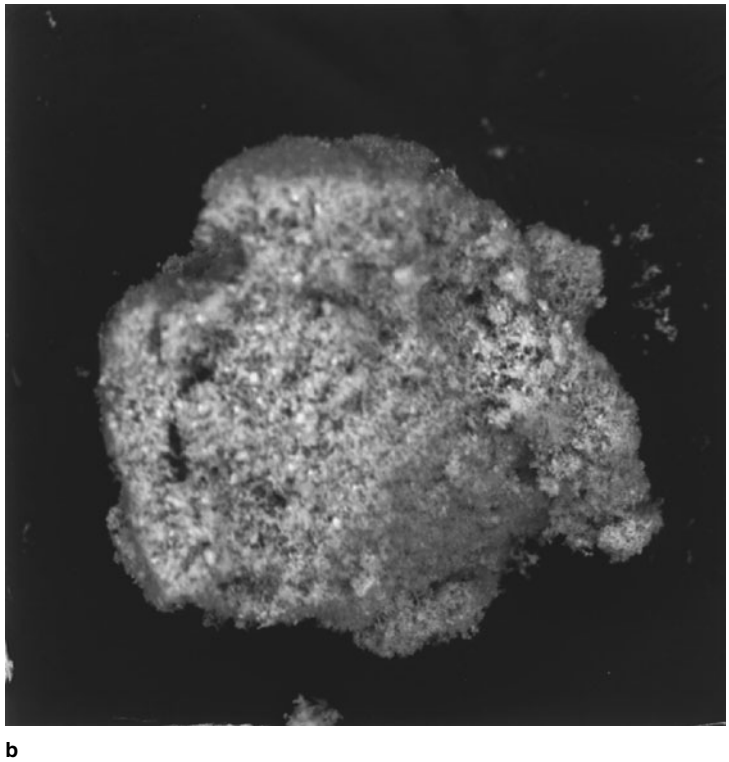
Natural and synthetic magnetic materials fall into two distinct classes based upon the behavior of the magnetic field at the surface of the material. Inside the magnetic material, the magnetic spins of certain atoms are aligned over large regions called domains in which the magnetic field is highly directional. For certain natural magnetic materials, such as cobalt, and nearly all synthetic engineered magnetic materials, such as magnetic recording tape or computer disk, as the domain intersects the surface, the magnetic field passes through the surface, passes through free space (creating a "leakage" field), and reenters the material at another domain of opposite magnetization. For materials of cubic crystal form such as iron, the multiplicity of possible crystal directions along which the magnetic field can align is so large that as the bulk domain intersects the surface, the crystal can form a surface "closure domain" in which the field runs parallel to and below the surface, preventing magnetic field lines from exiting the surface.

5.7.2.1. Type 1 Magnetic Contrast

Type 1 magnetic contrast arises from interactions of the secondary electrons after they have exited the specimen surface with the leakage magnetic field. This phenomenon was first observed from magnetic tape by Dorsey (1961) and from natural magnetic crystals by Banbury and Nixon (1967) and Joy and Jakubovics (1968). An electric charge moving through a magnetic field experiences a vector force $\mathbf{F} = -e\mathbf{v} \times \mathbf{B}$, where $\mathbf{v}$ is the velocity vector of the SE, $\mathbf{B}$ is the magnetic field, and e is the electron charge. The effect of $\mathbf{F}$ is to accelerate the SE, changing its velocity, especially the directional component of the velocity. If the acceleration imparted to the SE carries it away from the E-T detector, the signal collected is reduced, and that domain appears darker than the average. If this is the case, then above domains of opposite magnetization, the SE will be deflected toward the E-T



a



b

Figure 5.33. Voltage contrast of a dust particle as viewed (a) with the secondary electron signal measured with an Everhart–Thornley (+) detector and (b) simultaneously with a BSE detector.

detector and that domain will appear brighter than the average. From areas where the deflection is parallel to the face of the E-T detector, no increase or decrease of the SE signal occurs. Examples of Type 1 magnetic contrast from a cobalt crystal and from a square wave written on magnetic tape are shown in Fig. 5.34.

Type 1 magnetic contrast can be observed across the full range of primary beam energies, although lower beam energies create larger yields of SE. Because the contrast depends on the magnetic deflection relative to the ET detector, rotation of the specimen relative to the ET detector may be necessary to maximize the contrast.

5.7.2.2. Type 2 Magnetic Contrast

Type 2 magnetic contrast arises from the effect of the internal magnetic field upon the beam electrons as they initially scatter within the specimen (Fathers et al., 1973 a,b). Beam electrons experience a force $\mathbf{F} = -e\mathbf{v} \times \mathbf{B}$. For certain orientations of the specimen magnetic field relative to the beam velocity, the effect of the magnetic force within domains of one polarity is to deflect the beam electrons toward the surface, thus increasing the backscatter coefficent, whereas in the domains of opposite polarity the deflection tends to drive the beam electrons deeper into the specimen, thus reducing the backscatter coefficient and creating contrast. Type 2 magnetic contrast is maximized when the specimen is tilted to approximately 55° from the horizontal, and when the magnetic field is oriented along the tilt axis, so that it may be necessary to rotate the tilted specimen around its surface normal. Type 2 contrast increases with the beam energy, so the highest value available should be chosen, and with the strength of the magnetic field, for example, for natural ferromagnetic materials, the order of increasing strength is Co < Ni < Fe. Even under optimum conditions, the maximum contrast for iron, the strongest natural ferromagnet, is only 0.3% at 30 keV. Such weak contrast demands an extremely high threshold current, typically 500 pA or more, to establish visibility under minimal search mode conditions (1 frame/s). Extensive black-level processing (analog or digital) is required to render the contrast visible to an observer. Finally, a BSE detector must be used, but because the specimen must be highly tilted, the normal position of the BSE detector on the bottom of the objective lens is not appropriate for capturing type 2 contrast. By tilting the specimen toward the positively biased E-T detector, BSE can be detected with high efficiency and weak type 2 contrast detected. An example of type 2 magnetic contrast in iron is shown in Fig. 5.35, where the domains that are visible are those with the magnetic field parallel to the tilt axis.

5.7.3. Crystallographic Contrast

The regular arrangement of atoms in crystalline solids can influence the backscattering of electrons compared to those same atoms placed in

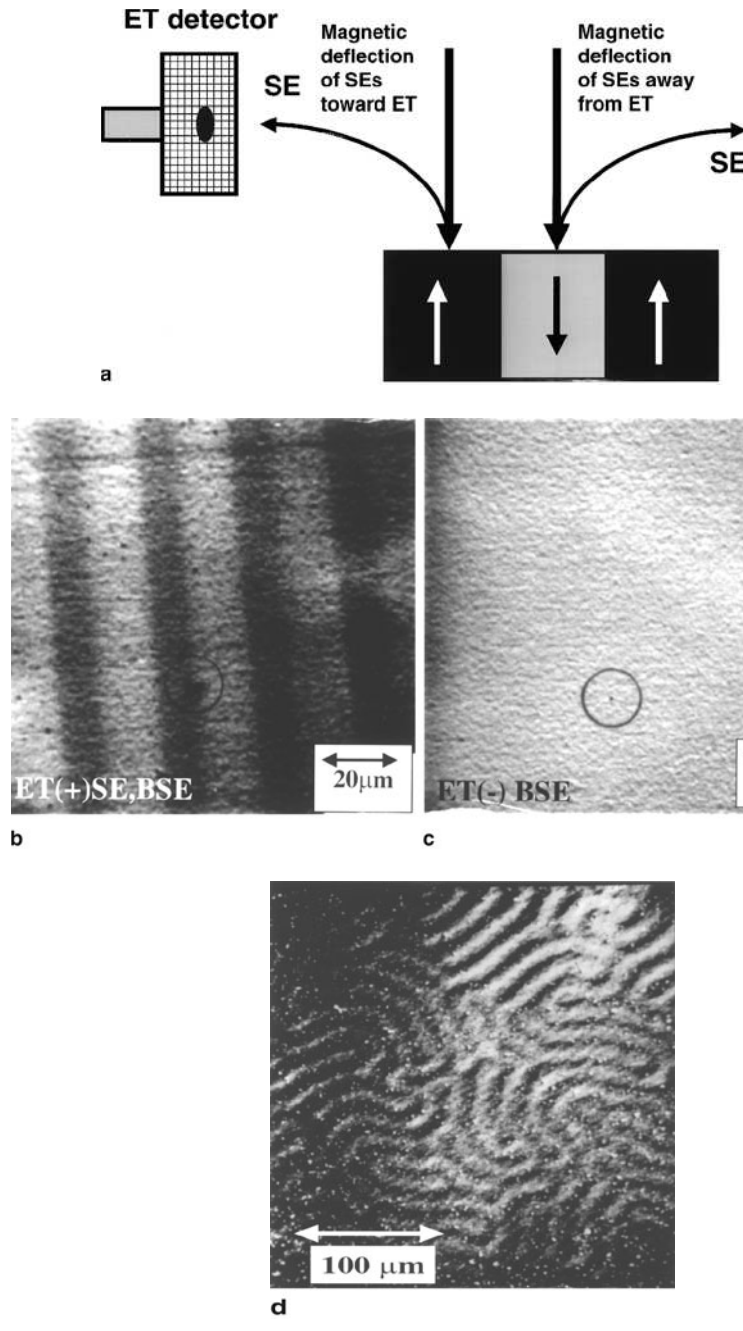
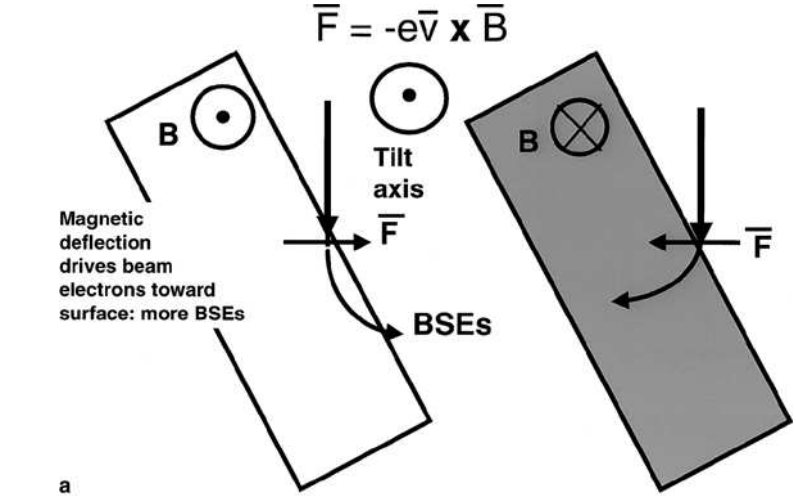


Figure 5.34. Type 1 magnetic contrast. (a) Schematic diagram showing origin. (b, c) Type 1 magnetic contrast appearance as bands in magnetic tape recording of square waves in SE signal collected with an Everhart–Thornley (+) detector and elimination of contrast in BSE signal collected with an Everhart–Thornley (−) detector. (d) Type 1 magnetic contrast in single-crystal cobalt.



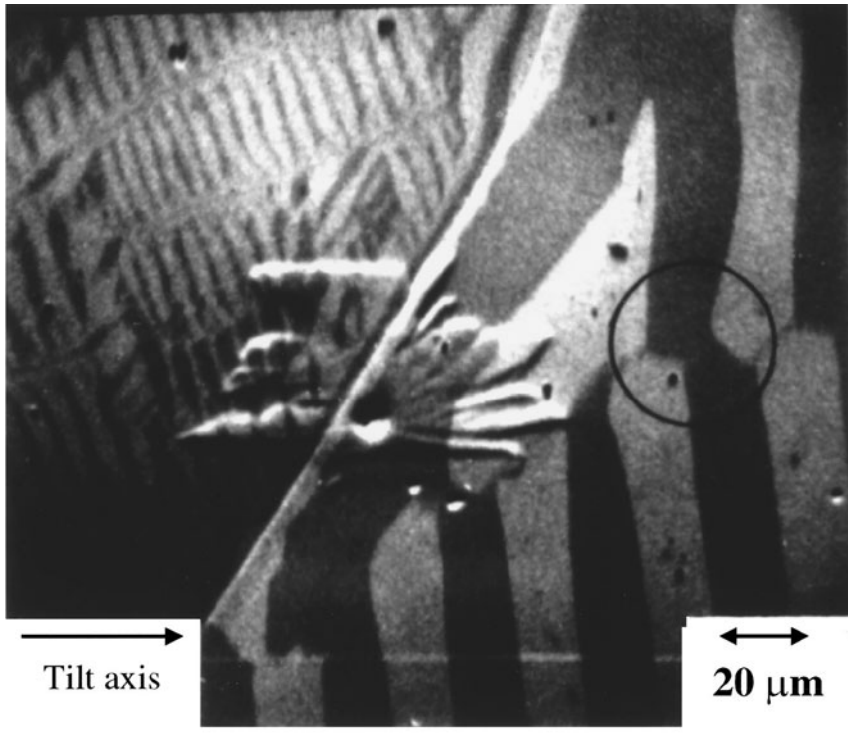
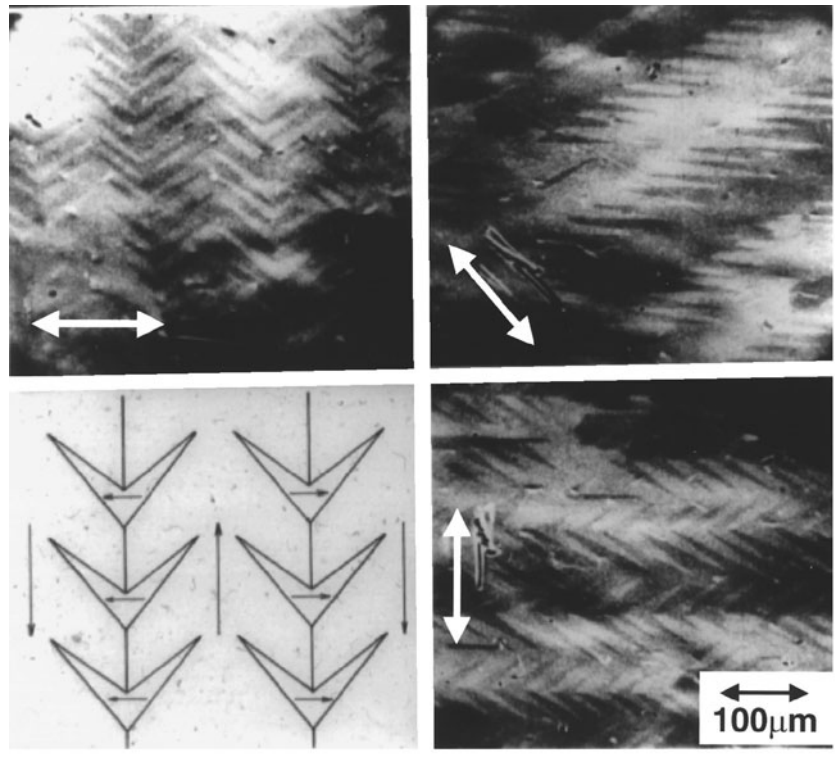


Figure 5.35. Type 2 magnetic contrast. (a) Schematic diagram showing origin. (b) Type 2 magnetic contrast appearance as fine domains in Fe–3% Si transformer steel. (c) Type 2 magnetic contrast in iron crystal, showing change in appearance as the crystal is rotated (rotation axis orientation is indicated) and magnetization parallel to the tilt axis is detected.



c

Figure 5.35. (*Continued*)

the near-random three-dimensional distribution of an amorphous solid. If a well-collimated (i.e., highly parallel) electron beam is directed at a crystal array of atoms along a series of different directions, the density of atoms that the beam encounters will vary with the crystal orientation. Along certain directions, paths of very low atomic density are found, the so-called "channels," which permit at least a fraction of the beam electrons to penetrate more deeply into the crystal before beginning to scatter. When beam electrons start scattering deeper in the crystal, the probability that they will return to the surface as backscattered electrons is reduced compared to the amorphous target case, and so the measured backscatter coefficient is lowered compared to the average value from the amorphous target. For other crystal orientations, a denser atom packing is found, and the beam electrons begin to scatter immediately at the surface, increasing the backscatter coefficient relative to the amorphous target case. The modulation of the backscatter coefficient between the maximum and the minimum channeling cases is small, only about a 5% difference. Nevertheless, this crystallographic or electron channeling contrast can be used to derive microstructural information about the important class of crystalline materials.

To determine the likelihood of electron channeling, the properties of the electron beam (i.e., its energy E or, equivalently, wavelength λ) are

related to the critical crystal property, namely the spacing of the atomic planes d, through the Bragg diffraction relation:

$$n\lambda = 2d \sin \theta_B, \tag{5.6}$$

where n is the integer order of diffraction ($n = 1, 2, 3$, etc.). Equation (5.6) defines special angles (referred to as "Bragg angles" θ_B) at which the channeling condition changes sharply from weak to strong for a given set of crystal planes. Because a real crystal contains many different possible sets of atom planes, the degree of channeling encountered for an arbitrary orientation of the beam to the crystal is actually the sum of contributions of many planes. Generally, these contributions tend to cancel, giving the amorphous average backscattering, except when the Bragg condition for a certain set of planes dominates the mix, giving rise to a sharp change in the channeling condition.

Detecting weak channeling contrast effects requires that the threshold current be exceeded, which is about 10 nA for a frame rate useful for searching. Channeling contrast is carried by the high-energy fraction of the backscattered electrons, and to maximize that signal, a large-solid-angle BSE detector (solid state or passive scintillator) is the best choice with the specimen set normal to the beam. The positively biased E-T detector, with its high efficiency for BSE through capture of SE_3 signals, is also satisfactory. Because of the weak contrast, the image must be enhanced by "black-level" processing by either analog or digital means to spread the channeling contrast over the full black-to-white range of the final display to render it visible.

We can see the influence of crystallographic channeling for the special case of a large single crystal viewed at a low magnification. As shown schematically in Fig. 5.36, not only does the act of scanning move the beam laterally in x and y, but the angle of the beam relative to the surface normal changes. For a 2-cm working distance from the scan rocking point and a

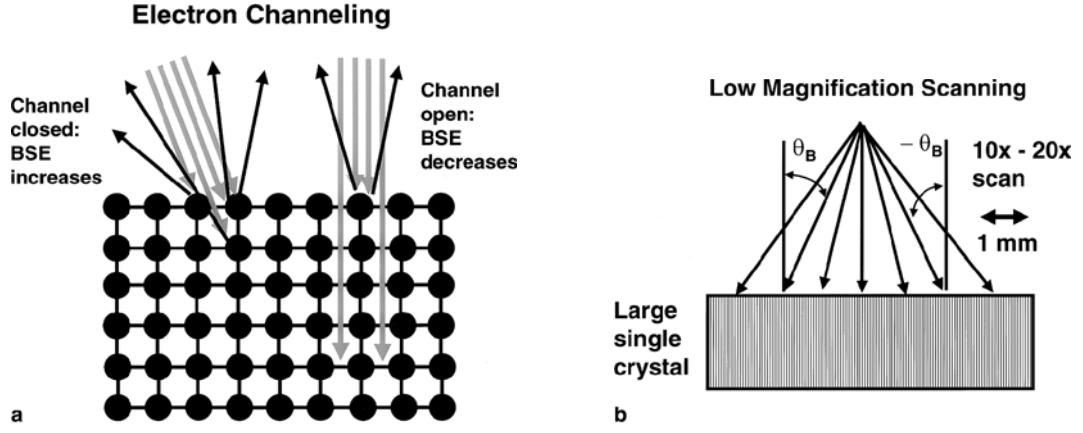


Figure 5.36. Channeling contrast. (a) Schematic diagram showing the origin of the channeling effect as penetration of the beam electrons along the crystal planes; (b) scanning at low magnification both moves the beam on the surface and changes the angle substantially across the field.

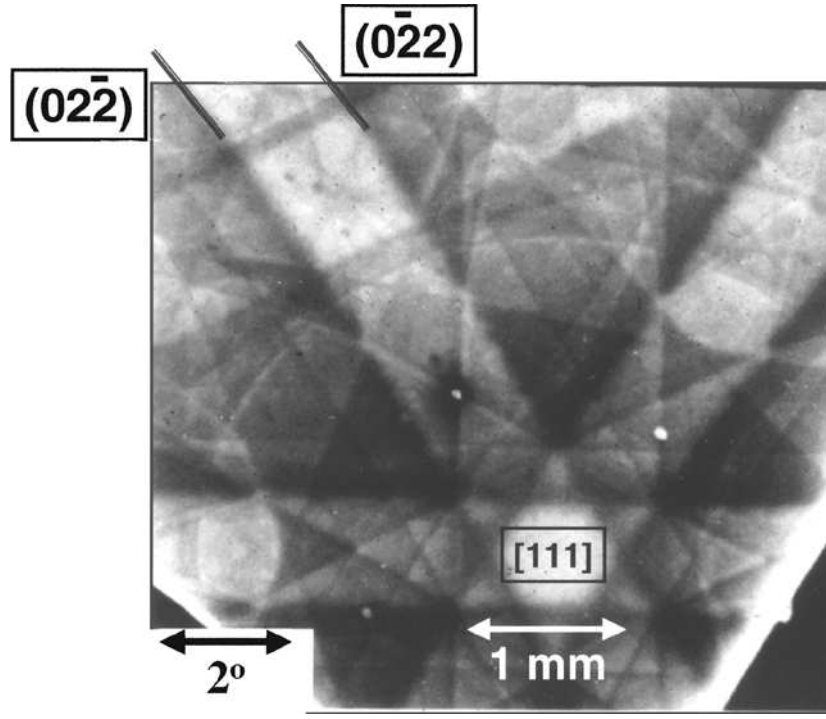


Figure 5.37. Electron channeling pattern (ECP) observed on a silicon single crystal with the [111] pole nearly parallel to the surface normal.

10× scan giving a 1-cm scan excursion on the edge, the beam angle changes by ±14 deg across the field. The Bragg angles for a 20-keV electron beam interacting with crystals such as silicon have values of a few degrees, so that a scan angle of ±14 deg will certainly cause the beam to pass through the Bragg condition for some planes. This condition is shown in Fig. 5.37, which shows a flat, topographically featureless silicon crystal with a pattern of lines and bands, a so-called channeling pattern. The channeling effect appears as a series of lines because the intersection of the crystal planes with the surface defines a line.

Most crystalline materials are not single crystals. During solidification from a molten state, numerous crystals nucleate randomly in the liquid and grow in size until they encounter each other, producing a three-dimensional microstructure of randomly oriented crystalline units called "grains." Materials fabrication processes such as rolling, extrusion, forging, etc., and subsequent thermal treatment further modify the crystalline microstructure, resulting in grain sizes ranging from centimeters to nanometers, depending on composition and thermomechanical history. What happens to the channeling contrast image as the grain size is reduced from the large single-crystal case? Figure 5.38 shows a coarse polycrystal with grain dimensions in the millimeter to centimeter range, again viewed at low magnification. The scan angle is nearly the same as for Fig. 5.37, but the presence of boundaries between the crystals ("grain boundaries")

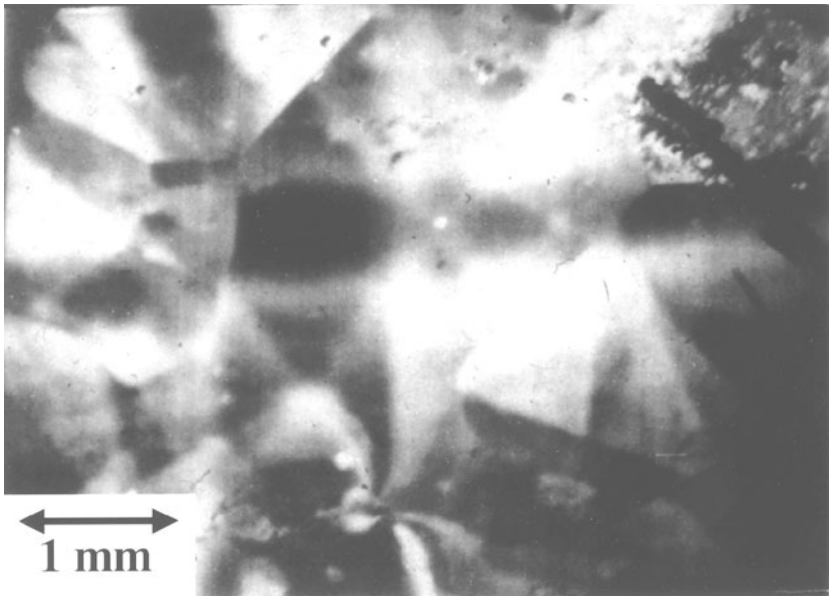


Figure 5.38. Electron channeling contrast from a coarse polycrystal (Fe–3% Si transformer steel). Compare with Fig. 5.37. The individual crystals are too small to show more than a few degrees of the ECP.

interrupts the channeling pattern so that we can no longer observe enough of the pattern to determine the orientation. We can, however, detect the grains themselves because of the pattern discontinuities. In Fig. 5.39, the grain size has been further reduced, so that the magnification must be increased so that several grains fill the image. At higher magnification the scan angle is reduced, so that very little of the channeling pattern in seen across the grain, only vaguely mottled contrast. At still smaller grain size and higher magnification, Fig. 5.40, the grains can be seen as having uniform shades of gray because each grain orientation provides an easy, intermediate, or hard orientation for channeling. How far down in scale can crystalline features be observed? Fig. 5.41 shows submicrometer spurs on special crystalline boundaries called annealing twins. With special high-resolution SEMs and energy-selecting BSE detectors, individual crystal dislocations have been seen.

Finally, the effects of plastic deformation, which introduces defects and residual stress patterns into ductile metals, can be directly observed in channeling contrast images. Figure 5.42 shows a contrast pattern around a diamond hardness indent in polycrystalline nickel, revealing the extent of plastic deformation around the indent. Figure 5.43 shows a residual strain pattern remaining in some grains of deformed polycrystalline Fe–3% Si which has been partially heat-treated to remove residual strain.

Crystallographic contrast by electron channeling provides images of the crystallographic microstructure of materials. For many applications in materials science, it is also important to measure the actual orientation of the microstructure on a local basis with high spatial resolution of

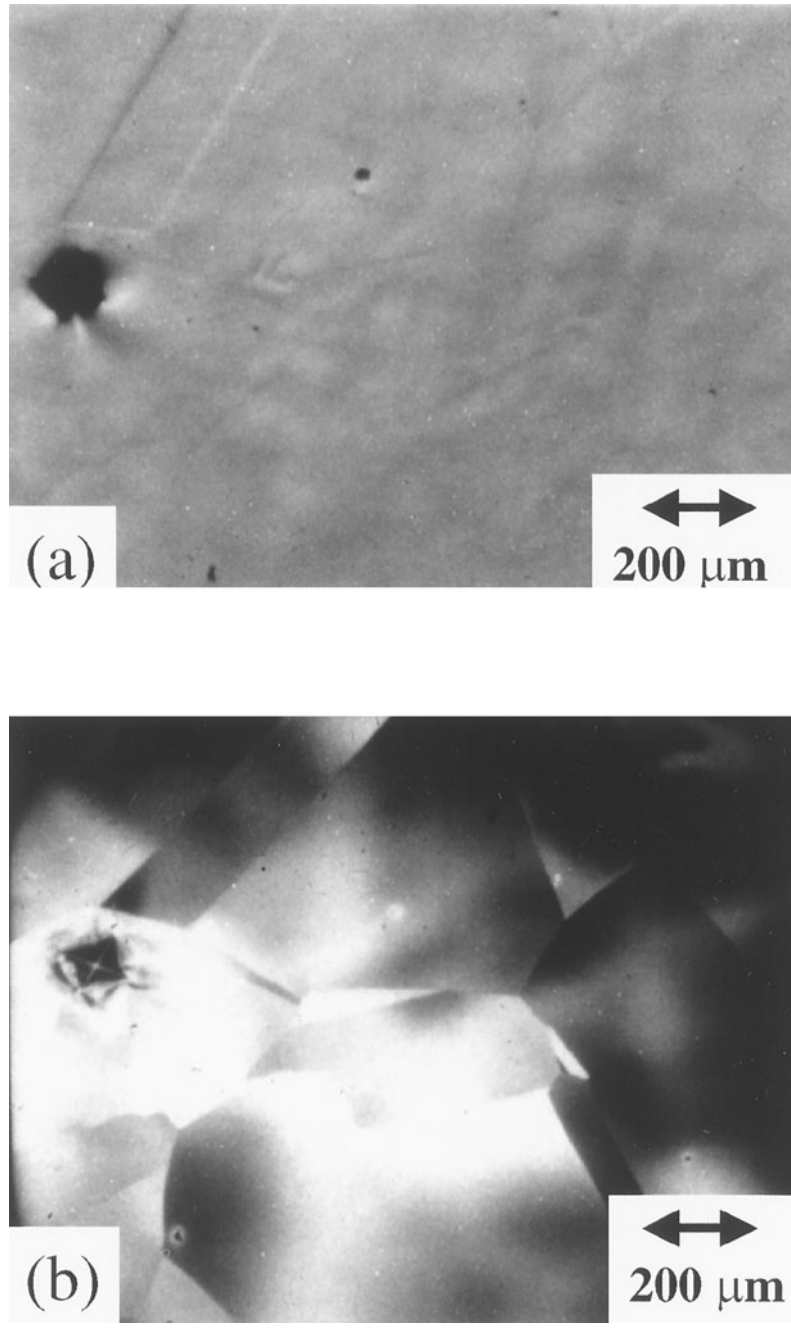


Figure 5.39. (a) Optical micrograph of a polished nickel polycrystalline specimen with a diamond hardness indent; (b) SEM electron channeling contrast of the same area. The grain size is now too small to show any distinct ECP, but the grain boundaries are defined.

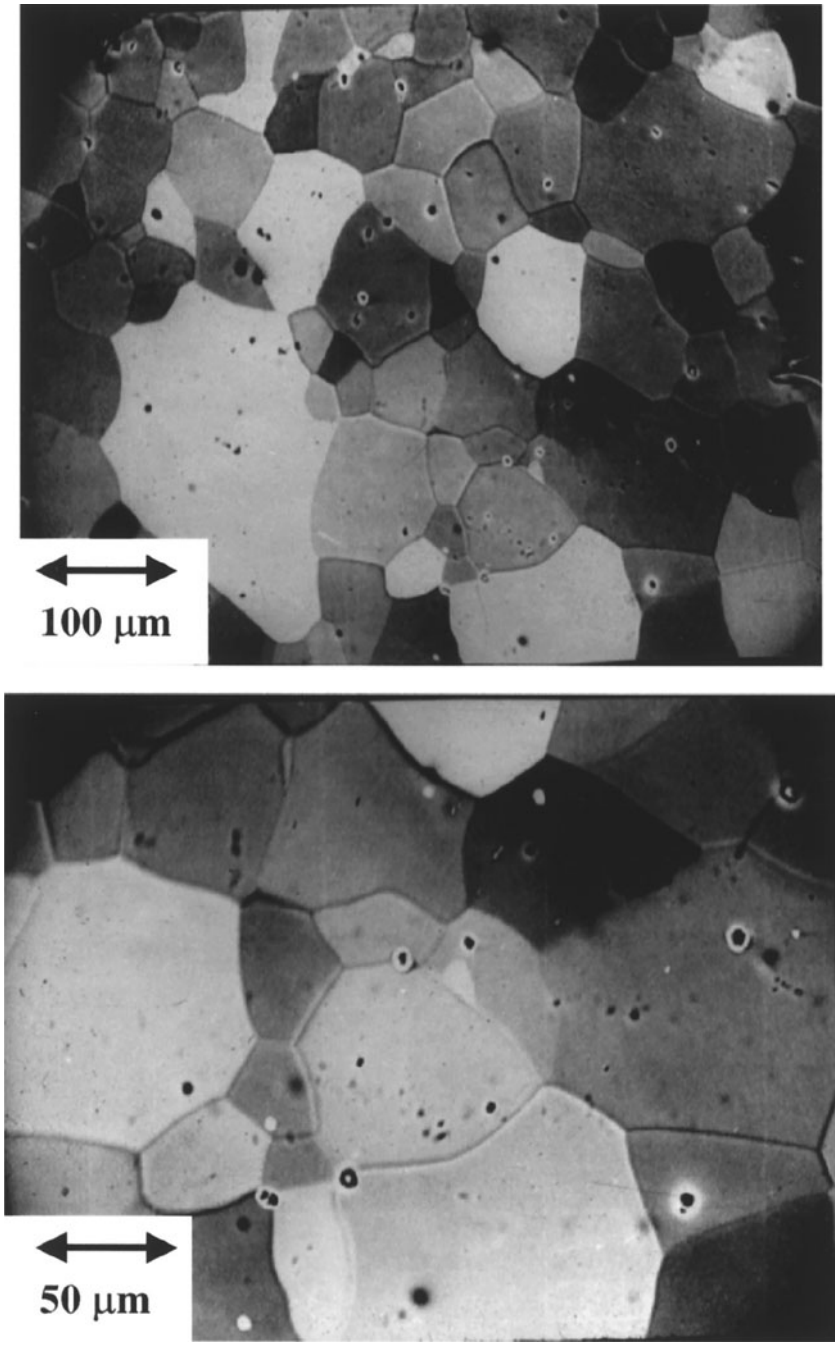
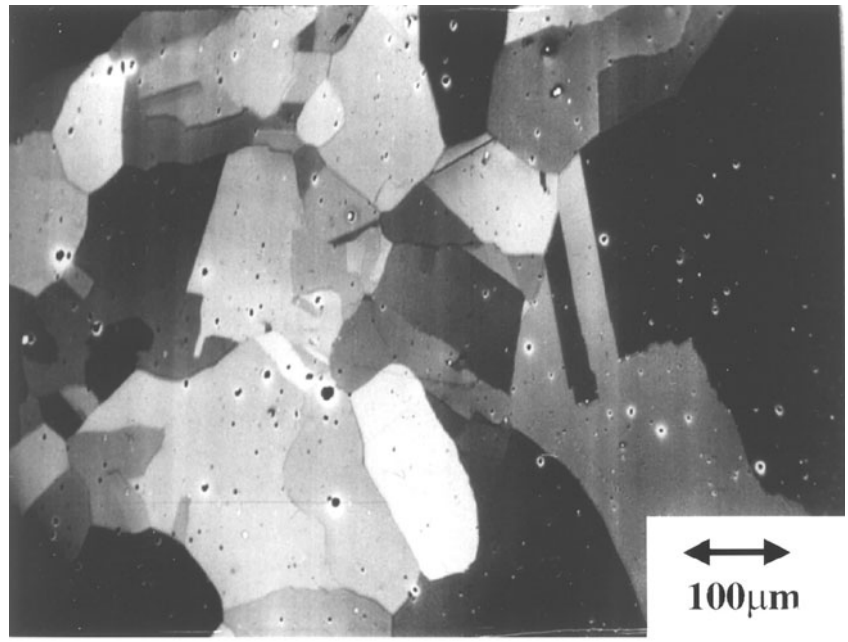


Figure 5.40. Electron channeling contrast from polycrystalline tin at two magnifications. The grains are now so small that each grain effectively shows only a tiny part of the ECP.



a

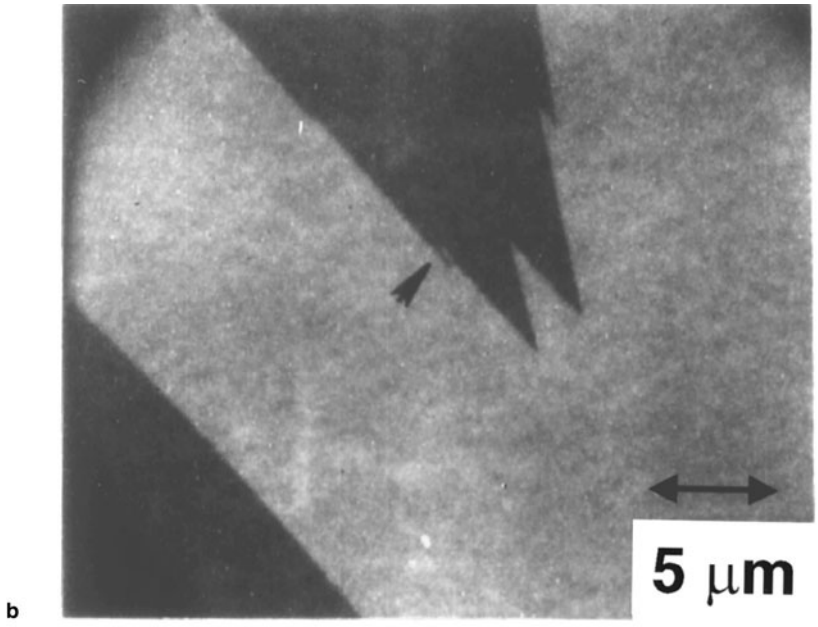
Figure 5.41. (a) Electron channeling contrast from polycrystalline Ni showing annealing twins with crystallographic fine structure. (b, c) Higher magnification images showing spikelike structure with a resolution of 100–200 nm.

1 μm laterally or even finer. Observation of electron channeling patterns obtained from small, selected areas through special scanning techniques was an early method used to provide such orientation information (Joy *et al.,* 1982). Due to limitations imposed by electron-optical considerations, the selected-area electron channeling pattern method has been nearly entirely replaced by the technique of electron backscatter patterns, described in the next section. Such orientation patterns provide the ideal complement to channeling contrast microscopy.

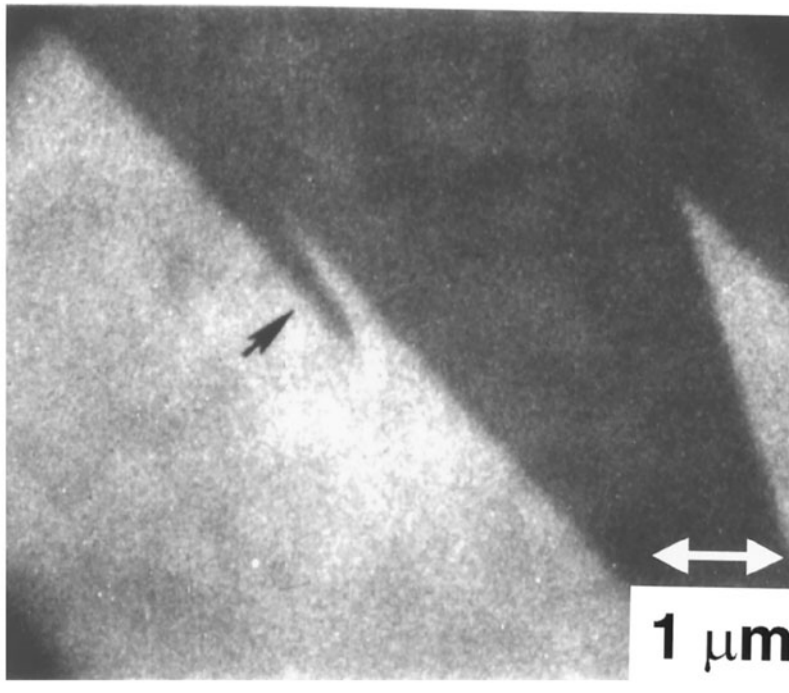
5.8. Electron Backscatter Patterns

An understanding of the crystallography of a material is important to a full understanding of structure/property relationships. By linking microstructure and the crystallography of the sample, a more complete picture of the sample is developed, which permits a truer understanding of materials performance. The development of electron backscattering diffraction (EBSD) into a commercial tool for phase identification and orientation determination has provided a new insight into the links between microstructure, crystallography, and materials physical properties that was not previously attainable.

EBSD in the SEM is used for two main purposes. The oldest application of EBSD is the measurement of the individual orientations (collectively



b



c

Figure 5.41. (*Continued*)

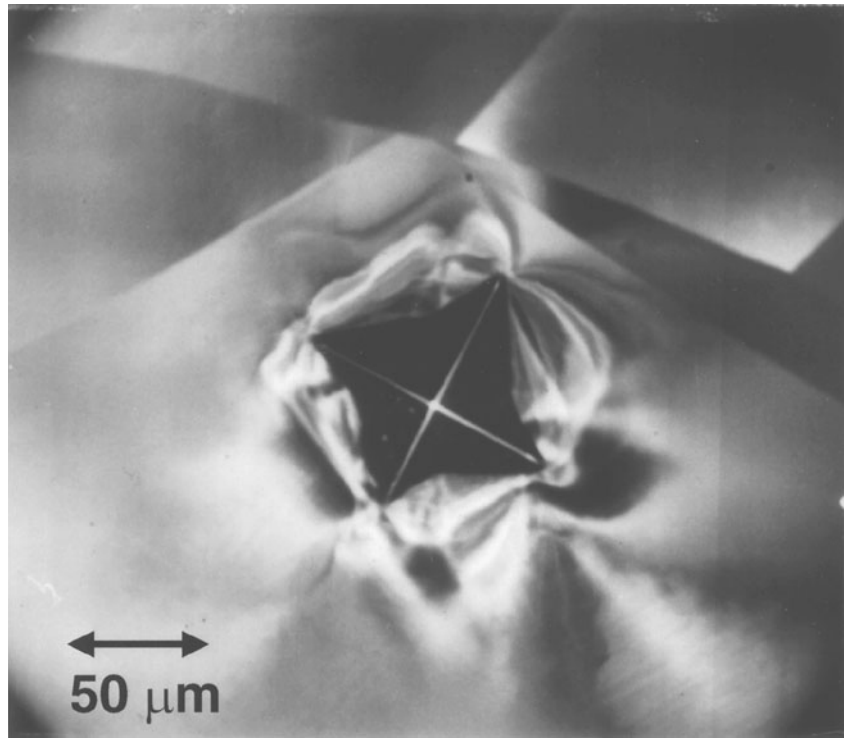


Figure 5.42. Deformation of polycrystalline Ni induced by a diamond hardness indent.

referred to as a "texture") of a group of contiguous microscopic crystals ("grains") that make up the basic crystallographic units of a specimen. This measurement performed on a grain-by-grain basis gives the local microtexture (Randle and Engler, 2000). The other use of EBSD is for the identification of micrometer or submicrometer crystalline phases through determination of the characteristic crystallographic parameters such as crystal plane spacing, angles between planes, and crystal symmetry elements (Michael, 2000).

Crystallographic contrast, discussed in Section 5.7.3 and shown in Figs. 5.38–5.40, permits grains to be imaged with different intensities based on the crystallographic orientation of the grain. However, the different intensities observed in an image that shows crystallographic contrast or channeling contrast tell us nothing quantitative about the actual crystallographic orientation of the grains in the image. EBSD provides this important quantitative information about the orientation of each grain in an image. Microtexture is defined as a population of individual orientations that are usually related to some feature of the sample microstructure. A simple example of this is the relationship of grain size to grain orientation. The concept of microtexture may also be extended to include the misorientation between grains, often termed the mesotexture. The misorientation of

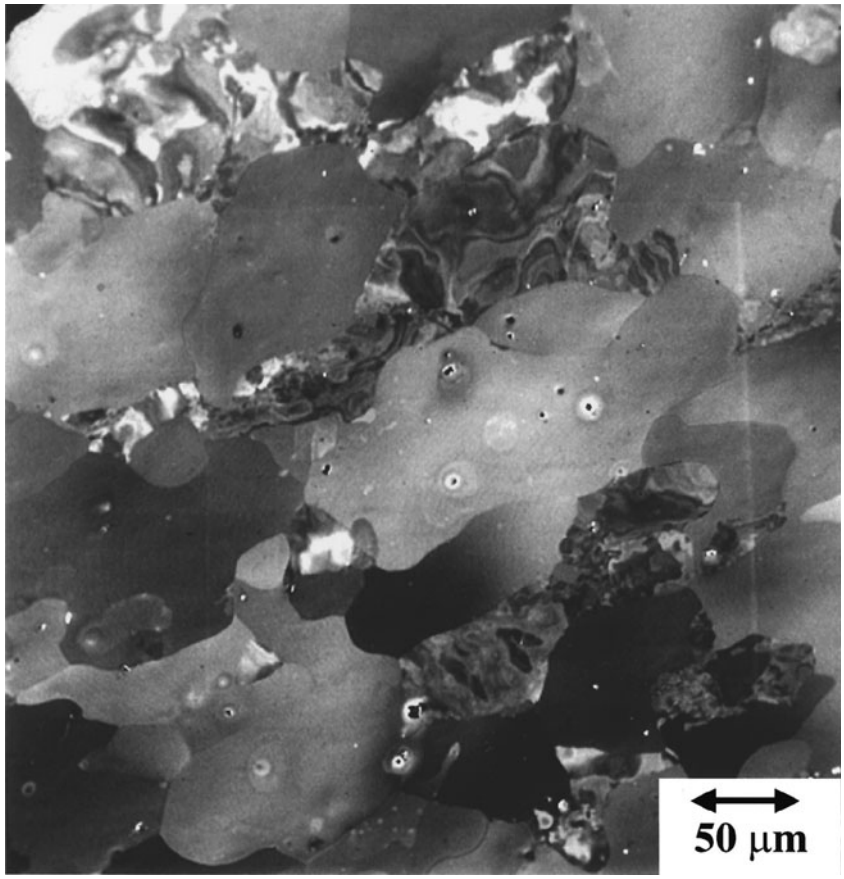


Figure 5.43. Remaining deformed grains (contrast contours) in polycrystalline Fe–3% Si transformer steel after partial recrystallization of highly deformed material.

all the grain boundaries in a given area may be determined. It is now possible using EBSD to collect and analyze thousands of orientations per hour, thus allowing excellent statistics in various distributions to be obtained. The ability to link texture and microstructure has enabled significant progress to be made in the understanding of recrystallization, grain boundary structure and properties, grain growth, and many other important physical phenomena that are orientation-dependent.

The identification of phases in the SEM has usually been accomplished with x-ray microanalysis by measuring the elemental composition. This measurement is subject to the inherent inaccuracies in the quantitative analysis technique (see Chapter 9), but it is certainly not definitive when one is attempting to identify a specific phase when two or more crystal structures are possible for exactly the same composition. For example, TiO_2 has three different crystal structures. Another example is the identification of austenite or ferrite, which often have

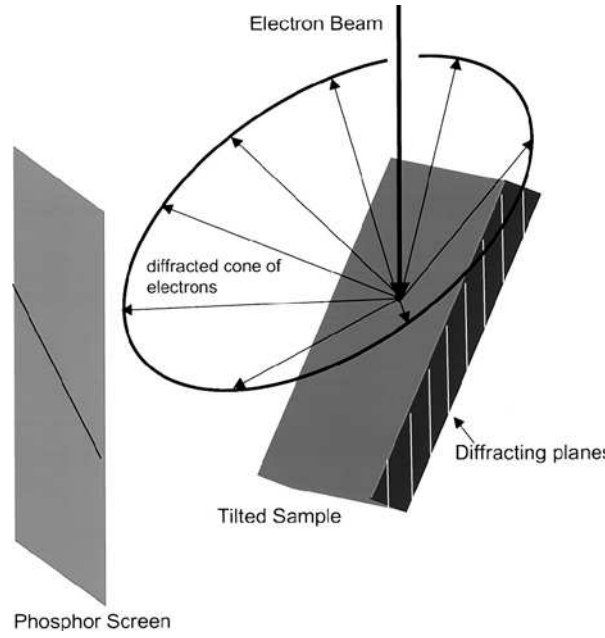


Figure 5.44. Schematic of the experimental arrangement of the sample and the phosphor screen placement within the SEM for obtaining electron backscattering diffraction (EBSD) patterns. Also shown is a single diffracted cone and its intersection with the phosphor screen.

similar chemistries in engineering steels. Austenite has a face-centered-cubic crystal structure and ferrite is body-centered-cubic. The distribution of these two phases can significantly alter the properties of a steel alloy.

5.8.1. Origin of EBSD Patterns

EBSD patterns are obtained in the SEM by illuminating a highly tilted specimen with a stationary electron beam. A typical arrangement of the specimen with respect to the electron beam and a pattern detector is shown in Fig. 5.44. Both diffraction and channeling of electrons have been used to describe how these patterns are formed. Fortunately, we do not need to understand either mechanism in detail to make use of this technique (Wells, 1999; Reimer, 1998). In either mechanism, the diffraction or channeling of the electrons causes the electrons to form flat cones of intensity above the sample as shown in Fig. 5.44. These cones are very flat (i.e., a large apex angle) with an apex semiangle of $(90^\circ - \theta_B)$, where θ_B is the angle for Bragg reflection to occur [see Eq. (5.6)]. Because the wavelength of the electron is small, the Bragg angle is typically small, less than 2°. Due to diffraction from both the front and the back of the atomic planes, we observe two cones per atom plane. These pairs of cones intercept the imaging plane and are imaged as two nearly straight lines separated by an angle of $2\theta_B$.

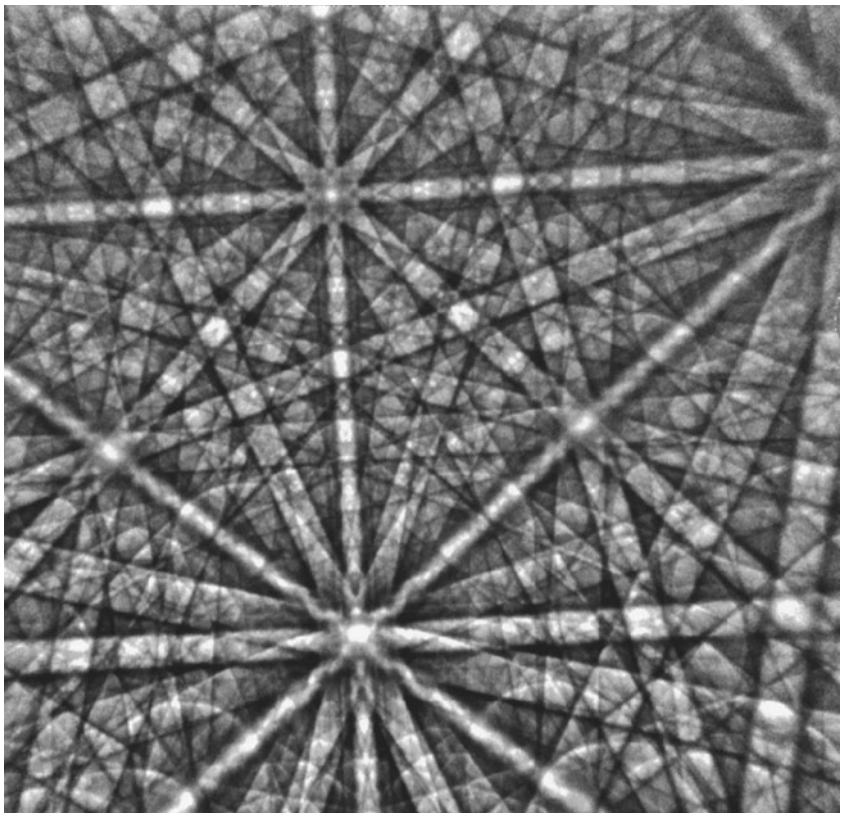


Figure 5.45. Typical EBSD pattern obtained from the mineral wulfenite ($PbMoO_4$) obtained at 20 kV with a CCD-based camera.

A typical electron backscatter diffraction pattern is shown in Fig. 5.45. The pattern consists of a large number of parallel lines ("Kikuchi lines"). These lines may appear as discrete lines or may appear as a band of increased intensity in the patterns. Frequently, as shown in Fig. 5.45, one of the pair of a set of lines will appear dark relative to the other. This is a result of the unequal angles between the scattered beam and the primary beam, which lead to an unequal intensity transfer to one of the diffracted or channeled cones. There are many places where a number of the line pairs intersect. The intersections are zone axes and are related to specific crystallographic directions in the crystal. The line pairs represent planes within the crystal structure and the zone axes represent crystallographic directions within the crystal. EBSD patterns are a map of the angular relationships that exist within the crystal. Distances on the pattern represent angles within the crystal.

The width of a pair of lines is inversely related to the atomic spacing of those planes within the crystal. The EBSD pattern is an angular map of the crystal, so we relate the distance between any pair of Kikuchi lines to an angle through the Bragg angle given by Eq. (5.6). From this equation, it

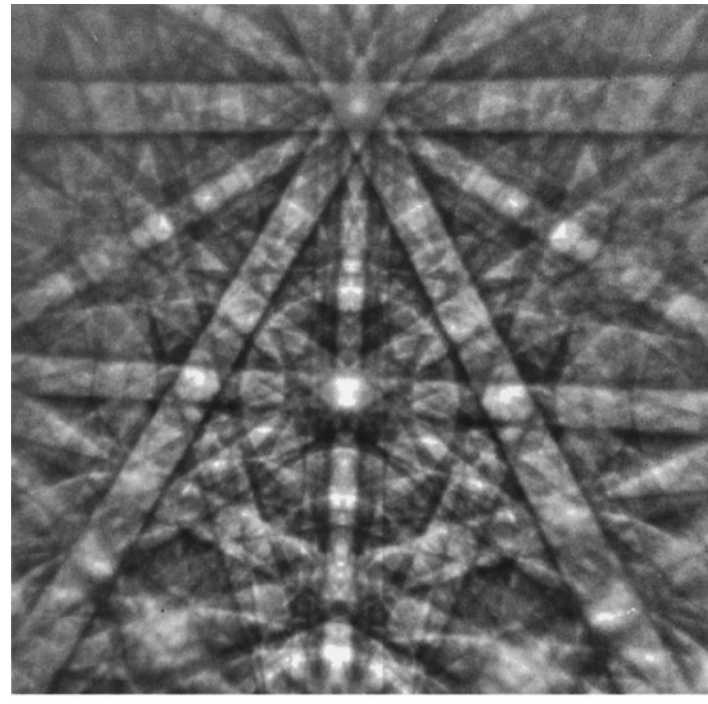
can be clearly seen that as the interatomic spacing d increases, the Bragg angle must decrease. Thus, a smaller atomic spacing will result in wider line pairs.

Equation (5.6) can also be used to determine the effect of changes in the microscope accelerating voltage because the wavelength of electrons increases as the accelerating voltage is decreased. The Bragg angle will also increase as the accelerating voltage is decreased for a given set of diffracting planes. Thus, as the accelerating voltage of the microscope is decreased, the width of the bands will increase. This effect is shown in Fig. 5.46 for patterns obtained from the mineral hematite (Fe_2O_3) at 10 and 30 keV. Also note that as the accelerating voltage is varied, the positions of the zone axes do not change because the positions of the zone axes are representative of the crystal orientation and the orientation does not change with accelerating voltage.

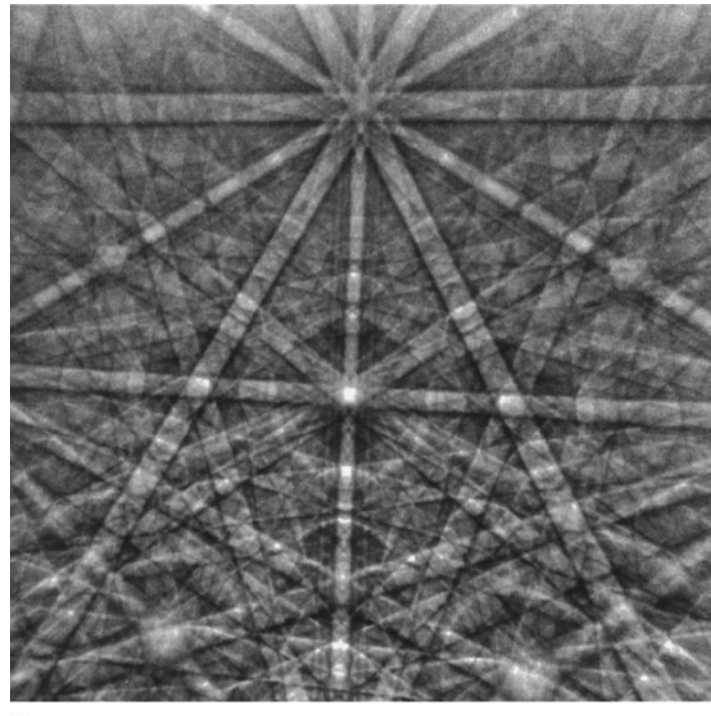
5.8.2. Hardware for EBSD

The hardware required for performing EBSD in the SEM includes a suitable camera or pattern detector system and a computer for the control of the electron beam and the analysis of the EBSD patterns. There have been three types of detectors used for imaging of EBSD patterns, including photographic emulsions, video rate cameras, and slow-scan charge-coupled-device (CCD)-based cameras. Photographic recording yields very good clear patterns, but suffers all the difficulties and disadvantages of the use of film in an electron microscope. The most common current approach to recording EBSD patterns is to convert the electrons into light using a phosphor screen. The phosphor screen is imaged with a video rate camera or a CCD-based camera. Coupling of the camera to the phosphor screen is accomplished using an optical camera lens or fiber optic bundles or a combination of these. The amount of light produced by the phosphor may be quite low (only just visible to the unaided eye in a darkened room) and therefore low-light-level cameras or cameras with image intensifiers have been used. The advantage of video-rate recording of the EBSD patterns is speed. Video images can be collected at many frames per second, making the video camera the imaging device of choice for orientation mapping of polycrystalline samples where many patterns are needed to cover the area of interest on the sample. CCD-based cameras for EBSD patterns provide images that are comparable to film quality, but without the problems of film in the vacuum of the SEM. However, CCD cameras have two disadvantages: First, the rate at which patterns can be collected is quite slow, and second, the CCD-based detectors are quite expensive compared to video-rate systems. Thus, slow-scan CCD-based detector systems are mainly utilized in phase identification studies where high-quality patterns are absolutely required (Eades, 2000).

At the present time, phase identification is accomplished in an analogous fashion to the collection of EDS information from a single point. Future increases in computer processor speed and better algorithms will



a



b

Figure 5.46. Comparison of EBSD for hematite (Fe_2O_3) at (a) 10 and (b) 30 keV, showing change in bandwidths as a function of energy. Note that the positions of the band intersections do not change.

allow phase mapping to be accomplished. Commercially available orientation mapping systems consist of a suitable video camera interfaced to a computer equipped with a frame grabber and an external scan generator. The scan generator controls the position of the electron beam during the collection of the EBSD patterns. Automated orientation mapping is accomplished in the following manner. The computer is programmed to step the electron beam pixel-by-pixel over the area to be mapped. At each pixel the beam is stopped and an EBSD pattern is obtained and automatically indexed and the orientation calculated. The beam is moved to the next position in the scan. In this manner an orientation map is produced. The orientations maps are usually displayed in color with various colors used to represent the corresponding crystal orientations.

5.8.3. Resolution of EBSD

5.8.3.1. Lateral Spatial Resolution

As with any analytical technique, it is important to understand where the signal is generated within the sample. The spatial resolution of EBSD is strongly influenced by sample tilt. Large sample tilt results in an asymmetric spatial resolution along the beam direction. This asymmetric spatial resolution also influences the depth from which the EBSD pattern is generated (see Chapter 3, Section 3.3). Note that due to the sample tilt (as shown in Fig. 5.44) the longitudinal resolution is much worse than the lateral resolution. The resolution perpendicular to the tilt axis is roughly three times the resolution parallel to the tilt axis for a tilt angle of 70° and increases to 5.75 times for a tilt angle of 80°. Examples of backscattered electron distributions for normal and tilted incident beams are shown in Figs. 3.14a and 3.14b.

Monte Carlo electron trajectory simulations have been used to study the energy distribution of the electrons that exit the sample and an example of this is shown in Fig. 5.47, where the energy distribution of the backscattered electrons from a Ni specimen is plotted for an untilted sample (specimen plane normal to the beam) and one tilted 70° from the horizontal position for an initial electron beam energy of 20 keV. The number of initial trajectories was the same for each sample orientation. There are two important points to note. First, tilting the sample to a high angle with respect to the normal results in a much higher backscatter yield, as discussed in Chapter 3, Section 3.4, and Eq. (3.11). This is good because the increased signal improves the ability to collect patterns. The other important point is the shape of the backscattered electron energy distribution. The sample that is normal to the electron beam has a broad spectrum of energies associated with the backscattered electrons. It is apparent from Fig. 5.47 that many of the backscattered electrons from the tilted sample have nearly the initial electron beam energy. The electrons in the sharp energy peak contribute to the crystallographic information in the EBSD pattern, and the electrons in the remainder of the distribution contribute to the overall background intensity of the EBSD pattern.

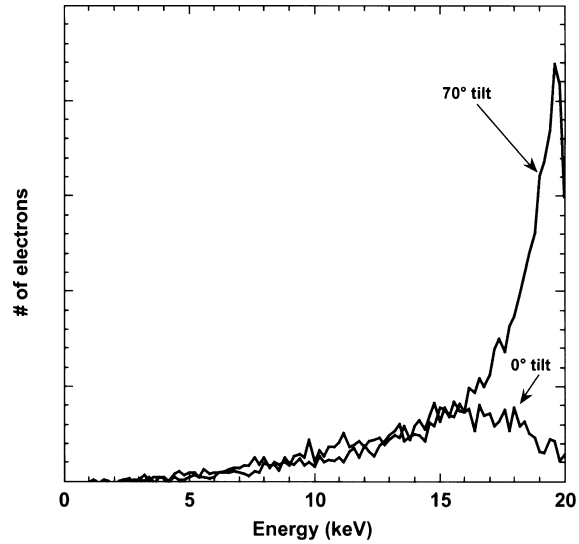


Figure 5.47. Comparison of energy distributions from a sample perpendicular to the electron beam and tilted 70° with respect to the electron beam. Note the increase in intensity and the increase in the number of electrons that have lost a small fraction of the initial beam energy.

The spatial resolution of EBSD is much higher than what can be achieved in normal backscattered electron imaging. This may seem surprising, but it is a consequence of the way the diffraction pattern is formed. In Chapter 3, Section 3.4, it was shown that the lateral spatial resolution of all BSE is large compared to the initial beam size. In the case of EBSD however, it can be inferred from the sharpness of the features in the EBSD pattern that only those electrons that have lost a small fraction of their energy contribute to the diffraction pattern. Monte Carlo electron trajectory simulations have showed that the low-loss backscattered electrons emerge from the specimen surface very near the initial beam impact point. Thus, the resolution of EBSD is related directly to the electron probe size. This point is discussed further in Chapter 3, Section 3.4. From this discussion it is apparent that higher resolution can be obtained from SEM instruments with field emission electron guns due to the smaller beam size, provided sufficient beam current is available (EBSD patterns have been collected with as little at 0.1 nA, well within the capabilities of most FEG sources). The spatial resolution is also a function of the accelerating voltage and the atomic number of the sample. Higher accelerating voltages and lower atomic number samples will degrade the spatial resolution of EBSD.

There have been a few studies to determine the resolution of EBSD. Many of these studies have collected patterns as a function of the distance of the beam from a grain boundary. When two overlapping patterns are observed, the distance to the grain boundary is assumed to be the spatial resolution. One study using an FEG electron source operated at 20 kV found spatial resolutions of 30 nm in Al and 10 nm in a brass

alloy (Humphries *et al.,* 1999). Other studies using slightly different resolution criteria measured lateral resolutions in Ni of 100 and 50 nm at 20 and 10 kV, respectively (Isabell and Dravid, 1997). A thorough study utilizing experimental measurements and Monte Carlo electron trajectory simulations produced lateral resolutions of less than less than 60 nm for Al and less than 20 nm for Pd at 10 kV (Ren *et al.,* 1998). Figure 5.48 shows an example of EBSD patterns obtained from 100-nm Au islands at 20 kV.

5.8.3.2. Depth Resolution

The sample tilt also influences the depth resolution of EBSD patterns. In Chapter 3, Section 3.3, it was shown that as the angle of tilt of a specimen surface increases, the interaction volume becomes smaller and asymmetric. An example of this phenomenon is shown in Fig. 3.6. At high sample tilts the electrons do not penetrate as deeply into the sample, reducing the depth of the interaction volume. For EBSD we are only interested in those electrons that have lost very little energy and these are the electrons that escape from a region very close to the sample surface. Few experimental studies have been conducted to measure the depth resolution of EBSD, but it is thought to be quite small and on the order of 10–100 nm.

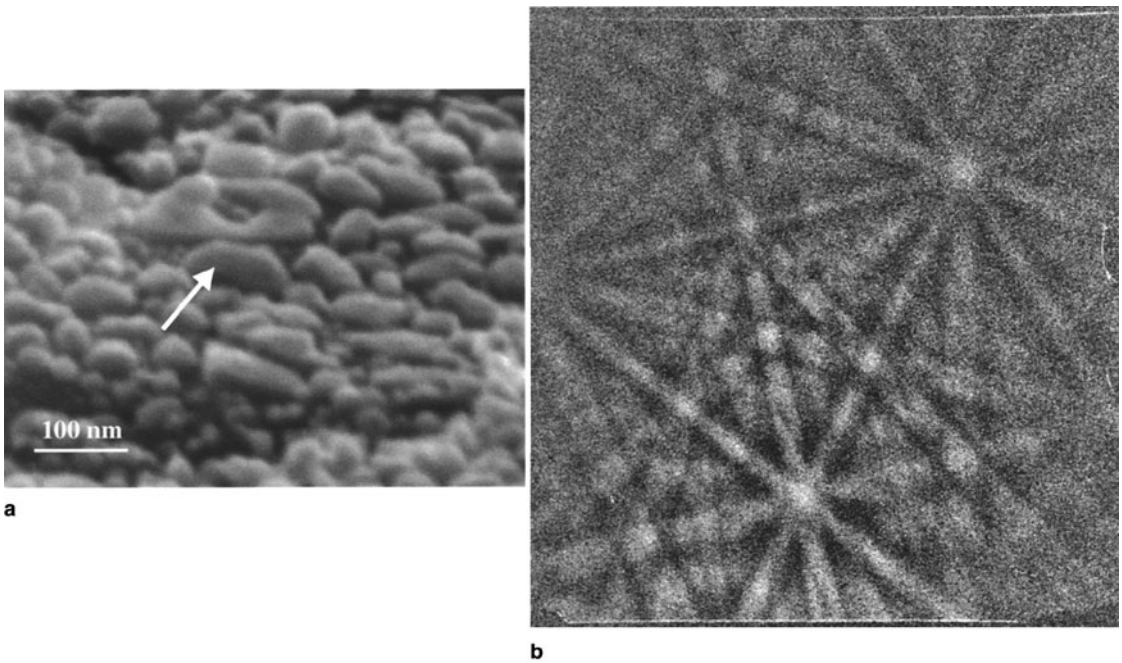


Figure 5.48. EBSD patterns may be obtained from single Au islands approximately 0.1 μm in diameter. (a) Au island indicated by arrow, (b) EBSD pattern collected from the Au island at 20 kV.

5.8.4. Applications

5.8.4.1. Orientation Mapping

Electrodeposition of metals into photolithographically defined cavities in photoresist is a promising way to produce very small parts for micromechanical systems. It is important to understand the mechanical properties of these parts after exposure to elevated annealing temperatures. One important aspect of the microstructural evolution is the crystallographic texture of the parts after deposition and following various annealing treatments. Figure 5.49 (See color figure on the following page.) demonstrates the use of orientation mapping on electrodeposited Ni. The images are cross sections of the deposits and are color-coded using the stereographic triangle shown in Fig. 5.49d and are relative to the growth direction of the deposit. It is clear from Fig. 5.49a that the as-deposited material has a clear ⟨100⟩ growth direction. The anneal at 400°C (Fig. 5.49b) is of interest as it shows the formation of recrystallized grains at the surface of the deposit (blue grains near the top of the image). These grains are of particular interest as they do not have a random texture, but have formed with a ⟨111⟩ direction parallel to the growth direction. Figure 5.49c shows the results of annealing at 500°C and shows that more ⟨111⟩-orientated grains have formed. Thus, EBSD can be used to relate the microstructural features (e.g., the recrystallized grains near the surface) with the crystallographic orientation of those features.

5.8.4.2. Phase Identification

Stainless steels that are intended for long-term exposure to high temperatures can fail in a brittle manner due to the precipitation of brittle intermetallic compounds. Figure 5.50a is a polished cross section through a weld that had failed in a brittle manner after long-term elevated temperature exposure. The microstructure is seen to consist of a matrix phase and a precipitate phase. The precipitate phase was shown to contain Fe and Cr only, whereas the matrix contained Fe, Cr, and Ni. Figure 5.50b is an EBSD pattern obtained from the precipitate phase. The phase identification procedure calculated the unit cell volume to be 39.7Å^3. The ICDD Powder Diffraction File crystallographic database was searched for compounds that contained only Fe and Cr and had a unit cell volume of some multiple of 39.7Å^3 with an error of 10%. The search returned one match, which was tetragonal FeCr. The pattern was indexed using these data and then simulated to demonstrate the line pair positions expected from the selected match. Figure 5.50c shows the simulation overlaid upon the experimentally determined pattern. There is excellent agreement between the simulation and the experimental pattern. The intermetallic compound that formed was a well-known brittle phase called sigma. The failure of the weld resulted from precipitation of sigma phase during long-term elevated temperature exposure.

EBSD in the SEM is one of the newest and most exciting developments in the field since the first x-ray detector was interfaced to a SEM.

EBSD now provides the SEM with a complete set of characterization tools. A sample may now be imaged in any one of the many imaging modes available, the chemistry of small areas of the sample may be determined by x-ray microanalysis, and then the crystallography (either for orientation or phase identification) of the sample may easily be determined using EBSD.

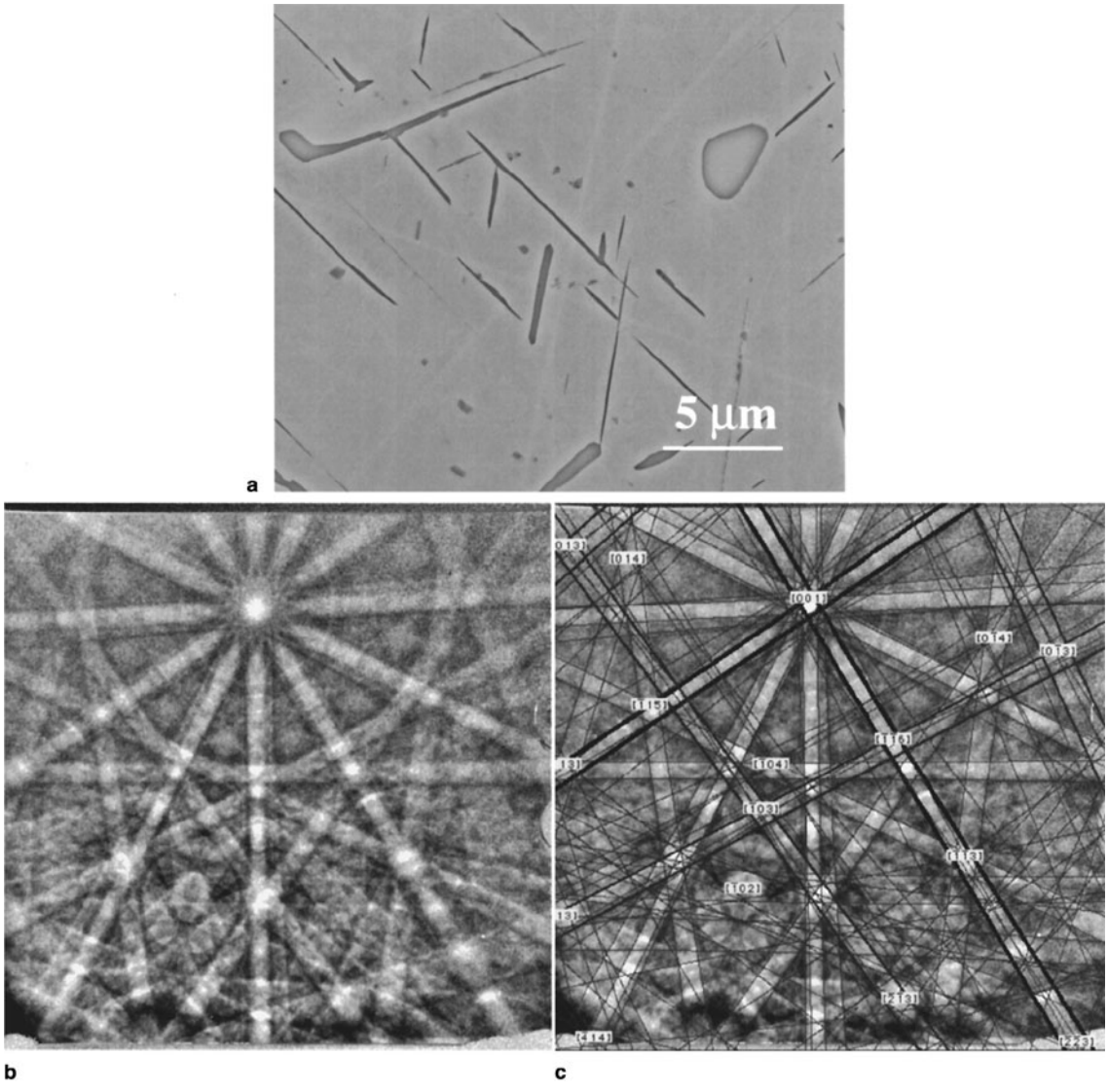


Figure 5.50. Identification of sigma phase in a weld in a stainless steel alloy. (a) SE image with precipitates clearly visible in the microstructure. (b) EBSD pattern collected from precipitate at 20 kV using a CCD-based camera. (c) Phase identified as FeCr (sigma phase) with simulated pattern overlaid on the EBSD pattern.

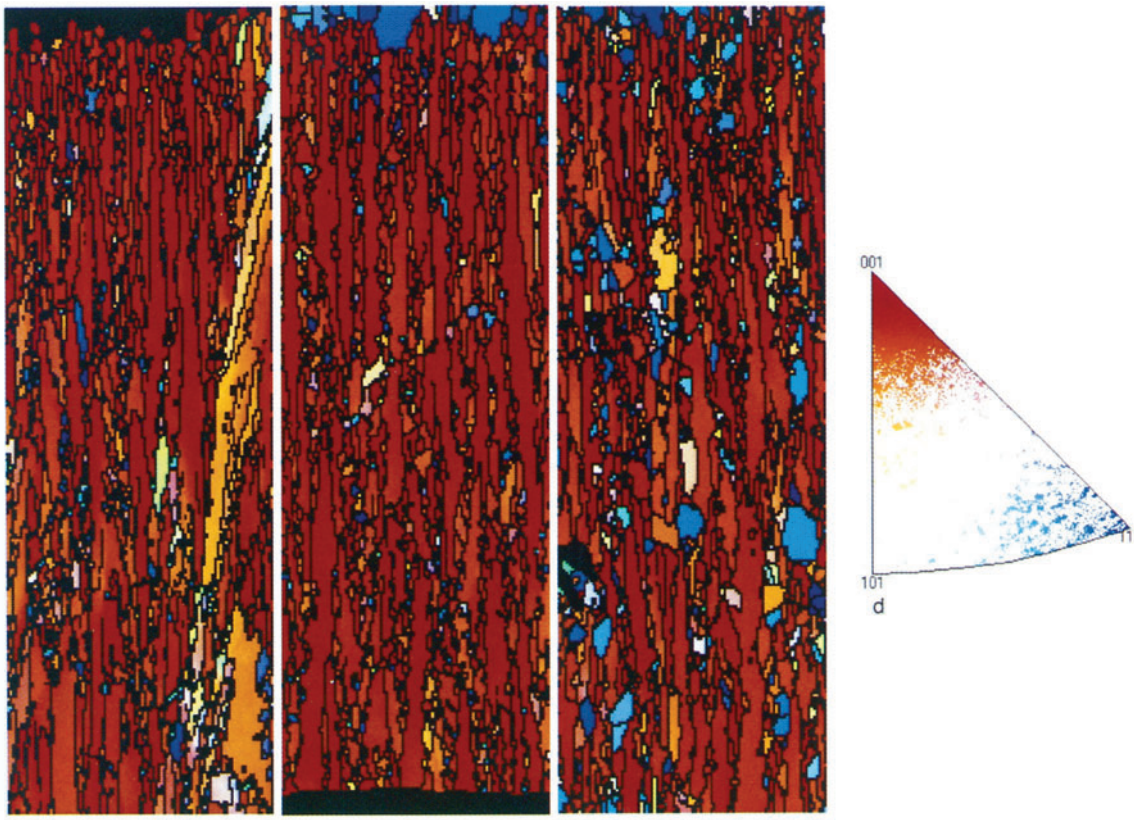


Figure 5.49. Microtexture maps of electrodeposited Ni. Note the change in texture as a function of annealing temperature. The width of the images is 70 μm. (a) As-deposited Ni, (b) 400°C, (c) 500°C, (d) color key for microtexture maps.

Banbury, J. R., and W. C. Nixon (1967). *J. Sci. Instr.* **44,** 889.
Boyde, A. (1973). *J. Microsc.* **98,** 452.
Boyde, A. (1974a). In *SEM/1974,* IIT Research Institute, Chicago, p. 101.
Boyde, A. (1974b). In *SEM/1974,* IIT Research Institute, Chicago, p. 93.
Danilatos, G. D. (1988). In *Advances in Electronics and Electron Physics,* Vol. 71, Academic Press, New York, p. 109.
Danilatos, G. D. (1990). In *Advances in Electronics and Electron Physics,* Vol. 78, Academic Press, New York, p. 1.
Danilatos, G. D. (1993). *Scanning Microsc.* **7,** 57.
Doehne, E. (1997). *Scanning* **19,** 75.
Dorsey, J. R. (1961). *Adv. Electronics Electron Phys. Suppl.* **6,** 291.
Eades, J. A. (2000). In *Electron Backscatter Diffraction in Materials Science* (A. J. Schwartz, M. Kumar, and B. L. Adams, eds.), Kluwer/Academic Press, New York, p. 123.
Farley, A. N., and J. S. Shah (1991). *J. Microsc.* **164,** 107.
Fathers, D. J., J. P. Jakubovics, D. C. Joy, D. E. Newbury, and H. Yakowitz (1973a). *Phys. Stat. Sol. A* **20,** 535.
Fathers, D. J., J. P. Jakubovics, D. C. Joy, D. E. Newbury, and H. Yakowitz (1973b). *Phys. Stat. Sol. A* **22,** 609.
Hein, L. R., F. A. Silva, A. M. M. Nazar, and J. J. Ammann (1999). *Scanning* **21,** 253.
Howell, P. G. T. (1975). In *SEM/1975,* IIT Research Institute, Chicago, p. 697.
Humphries, F. J., Y. Huang, I. Brough, and C. Harris (1999). *J. Microsc.* **195,** 212.
Isabell, T. C., and V. P. Dravid (1997). *Ultramicroscopy* **67,** 59.
Gopinath, A., K. G. Gopinathan, and P. R. Thomas (1978). In *SEM/1978/I,* SEM, Inc., AMF O'Hare, Illinois, p. 375.
Griffin, B. J. (2000). *Scanning* **22,** 234.
Griffin, B. J., and C. Nockolds (1999). In *Proceedings 14th International Conference on Electron Microscop* (Cancun), Vol. 1, p. 359.
Joy, D. C. (1984). *J. Microsc.* **136,** 241.
Joy, D. C. (1996). In *Proceedings of the Annual Meeting of the Microscopy Society of America* (G. W. Bailey, ed.), San Francisco Press, San Francisco, p. 836.
Joy, D. C., and J. P. Jakubovics (1968). *Philos. Mag.* **17,** 61.
Joy, D. C., D. E. Newbury, and D. L. Davidson (1982). *J. Appl. Phys.* **53,** R81.
Knoll, M. (1941). *Naturwissenschaften* **29,** 335.
Judge, A. W. (1950). *Stereographic Photography,* 3rd ed., Chapman and Hall, London.
Lane, W. C. (1970a). In *Proceedings 3rd Annual Stereoscan Colloguium,* Kent Cambridge Scientific, Morton Grove, Illinois, p. 83.
Lane, W. C. (1970b). In *Proceedings SEM Symposium* (O. Johari, ed.), IITRI, Chicago, p. 43.
Lowney, J. R., A. E. Vladar, W. J. Keery, E. Marx, and R. D. Larrabee (1994). *Scanning* **16**(Suppl. IV), 56.
Mansfield, J. (1997). *Microsc. Microanal.* **3**(Suppl. 2), 1207.
Michael, J. R. (2000). In *Electron Backscatter Diffraction in Materials Science* (A. J. Schwartz, M. Kumar, and B. L. Adams, eds.), Kluwer/Academic Press, New York, p. 75.
Mohan, A., N. Khanna, J. Hwu, and D. C. Joy (1998). *Scanning* **20,** 436.
Newbury, D. E. (1996). *Scanning* **18,** 474.
Newbury, D. E. (2001). *Microsc. Microanal*. **7**(Suppl. 2: Proceedings), 702.
Newbury, D. E., D. C. Joy, P. Echlin, C. E. Fiori, and J. I. Goldstein (1986). *Advanced Scanning Electron Microscopy and X-ray Microanalysis,* Plenum Press, New York, p. 45.
Pawley, J. B. (1984). *J. Microsc.* **136,** 45.
Pawley, J. B. (1988). *Inst. Phys. Conf. Ser. 93* **1,** 233.
Peters, K.-R. (1982). In *SEM/1982/IV,* SEM, Inc., AMF O'Hare, Illinois, p. 1359.
Peters, K.-R. (1985). In *SEM/1985/IV,* SEM, Inc., AMF O'Hare, Illinois, p. 1519.

Ramo, S. (1939). *Proc. IRE* **27,** 584.
Randle, V., and O. Engler (2000). *Introduction to Texture Analysis: Macrotexture, Microtexture and Orientation Mapping,* Gordon and Breach, New York.
Reimer, L. (1998). *Scanning Electron Microscopy: Physics of Image Formation and Microanalysis,* Springer, New York.
Ren, S. X., E. A. Kenik, K. B. Alexander, and A. Goyal (1998). *Microsc. Microanal,* **4,** 15.
Robinson, V. N. E. (1974). In *Proceedings of the 8th International Congress on Electron Microscopy* (J .V. Sanders, ed.), Australian Academy of Sciences, Canberra, Australia, p. 50.
Stowe, S. J., and V. N. E. Robinson (1998). *Scanning* **20,** 57.
Thiel, B. L., I. C. Bache, A. L. Fletcher, P. Meridith, and A. M. Donald (1996). In *Proceedings of the Annual Meeting of the Microscopy Society of America* (G. W. Bailey, ed.), San Francisco Press, San Francisco, p. 834.
Thornley, R. F. M. (1960). Ph. D. Thesis, University of Cambridge, Cambridge, England.
Vladar, A. E., M. T. Postek, N. Zhang, R. Larrabee, and S. Jones (2001). Reference Material 8091: New Scanning Electron Microscope Sharpness Standard, *Proc. SPIE* 4344-104.
Wells, O. C. (1960). *Br. J. Appl. Phys.* **11,** 199.
Wells, O. C., (1974), "Scanning Electron Microscopy", McGraw-Hill, N.Y.
Wells, O. C. (1999). *Scanning* **21,** 368.
Wight, S. A., and M. A. Taylor (1995). *Microbeam Anal.* **2,** 391.
Wight, S. A., and C. J. Zeissler (2000). *Microsc. Microanal.* **6,** 798.

6

Generation of X-Rays in the SEM Specimen

The electron beam generates x-ray photons in the beam–specimen interaction volume beneath the specimen surface. X-ray photons emerging from the specimen have energies specific to the elements in the specimen; these are the characteristic x-rays that provide the SEM's analytical capabilities (see Fig. 6.1). Other photons have no relationship to specimen elements and constitute the continuum background of the spectrum. The x-rays we analyze in the SEM usually have energies between 0.1and 20 keV. Our task in this chapter is to understand the physical basis for the features in an x-ray spectrum like that shown in Fig. 6.1.

6.1. Continuum X-Ray Production (Bremsstrahlung)

Beam electrons can undergo deceleration in the Coulombic field of the specimen atoms, which is the positive field of the nucleus modified by the negative field of the bound electrons, as shown in Fig. 6.2. The loss in electron energy ΔE that occurs in such a deceleration event is emitted as a photon. The energy of this photon is $\Delta E = h\nu$, where h is Planck's constant and ν is the frequency of the electromagnetic radiation. This radiation is referred to as bremsstrahlung, or "braking radiation." Because the interactions are random, the electron may lose any amount of energy in a single deceleration event. Therefore, the bremsstrahlung can take on any energy value from zero up to the original energy of the incident electron E_0, forming a continuous electromagnetic spectrum.

A calculated x-ray spectrum, without absorption and peak broadening effects, is shown in Fig. 6.3. The x-ray spectrum consists of both the continuous and characteristic components that would be generated by an electron beam within the interaction volume of a solid copper specimen. Note that the bremsstrahlung continuum is an electron-generated radiation. (X-ray spectra from a radioactive source or from an x-ray fluorescent source will not have an associated continuum.)

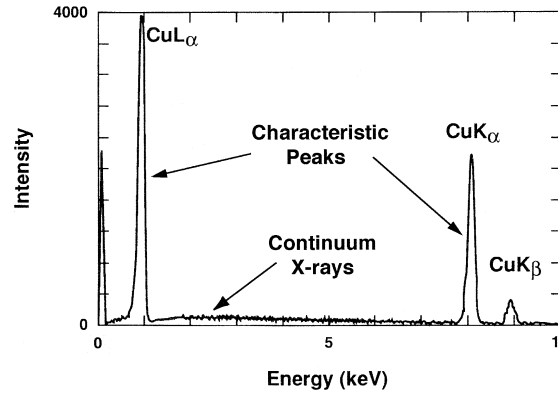


Figure 6.1. X-ray spectrum of copper showing *K*-series and *L*-series x-ray peaks and the continuous x-ray spectrum (bremsstrahlung or continuum) obtained with a Si(Li) EDS detector with an ultrathin (diamond) x-ray window. Natural widths of peaks are much narrower than measured here. A noise peak is measured at very low energies.

In describing x-rays, we make use of both the energy of the x-ray E (keV) and its associated wavelength λ (nm), which are related by the following expression:

$$\lambda = \frac{1.2398}{E} \quad \text{nm}, \tag{6.1}$$

where E is x-ray photon energy in keV.

The maximum x-ray energy observed in the spectrum corresponds to beam electrons that have lost all of their incident energy in a single event (Fig. 6.2). Since x-ray wavelength is inversely proportional to energy, the most energetic x-rays will have have a minimum wavelength λ_{min} or λ_{SWL}, called the short-wavelength limit or Duane–Hunt limit (Duane and Hunt, 1915), which can be related to the incident energy E_0 through Eq. (6.1). Measurement of the Duane–Hunt limit can be accomplished with any calibrated x-ray spectrometer and provides an unambiguous measure of the true electron beam energy as the electrons actually reach the specimen. Figure 6.4 shows an EDS x-ray spectrum of a carbon specimen at electron

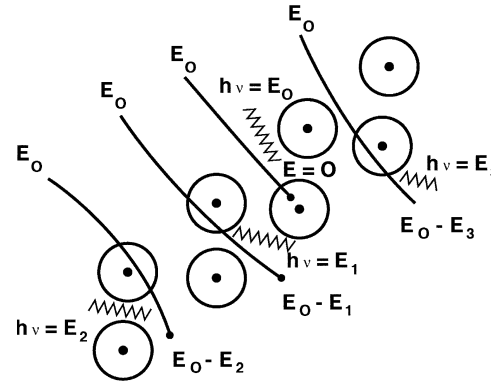


Figure 6.2. Schematic illustration of the origin of the x-ray continuum which results from the deceleration of beam electrons in the Coulombic field of the atoms.

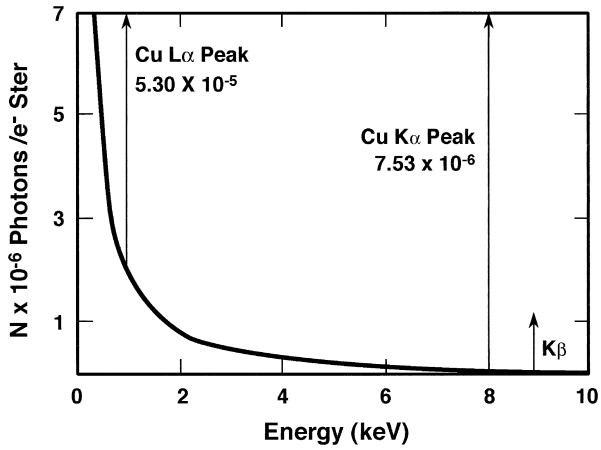


Figure 6.3. Calculated x-ray spectrum "as generated" in a copper target by a 20-keV electron beam. The continuum background (heavy line) and the Cu $K\alpha$, Cu $K\beta$, and Cu $L\alpha$ characteristic x-ray lines are shown. Compare the background intensity with that of Fig. 6.1 and note that most of the low-energy coninuum is absorbed in the specimen or the x-ray window when the spectrum is measured.

beam energies of 10, 15, and 20 keV. The Duane–Hunt limit is the energy where the extrapolated x-ray background goes to zero.

The intensity of the x-ray continuum I_{cm} at any energy or wavelength has been quantified by Kramers (1923) as

$$I_{\text{cm}} \approx i_{\text{p}} \overline{Z} \, \frac{E_0 - E_\nu}{E_\nu}, \tag{6.2}$$

where i_{p} is the electron probe current, $\overline{Z}$ is the average atomic number based upon mass (weight) fractions of the elemental constituents of the specimen, E_0 is the incident beam energy, and E_ν is the continuum photon energy at some point in the spectrum. At low photon energies, the intensity increases rapidly because of the greater probability for slight deviations in

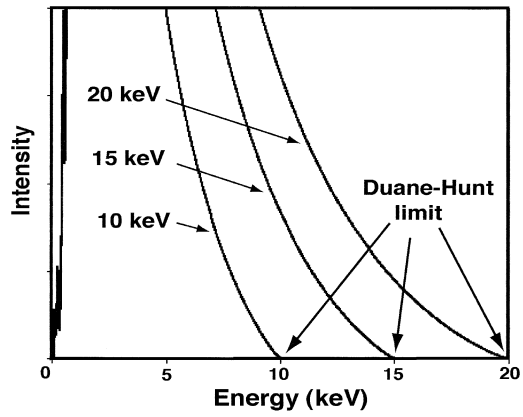


Figure 6.4. Duane–Hunt limit for simulated spectra from a carbon specimen at beam energies of 10, 15, and 20 keV. The Duane–Hunt limit is the energy where the continuum background intensity goes to zero.

trajectory caused by the Coulombic field of the atoms. The intensity of the continuum increases with increasing beam current, atomic number of the specimen, and beam energy. The intensity of the continuum increases with the atomic number because of the increased Coulombic field of the nucleus (more charge).

The intensity of the continuum radiation is important in analytical x-ray spectrometry because it forms a background under the characteristic peaks. Once a photon is created with a specific energy, it is impossible to determine whether it is a continuum or a characteristic x-ray. Thus, the background intensity due to the continuum process, occurring at the same energy as a characteristic x-ray, sets a limit to the minimum amount of an element that can be detected. The continuum is therefore usually regarded as a nuisance to the analyst. However, it should be noted from Eq. (6.2) that the continuum carries information about the average atomic number in the specimen and hence the overall composition. Thus regions of different average atomic number in a specimen will emit different amounts of continuum intensity at *all* x-ray energies. This fact can prove useful in measuring background with wavelength-dispersive x-ray spectrometers (WDS) and forms the basis for an important correction scheme (Marshall–Hall method, peak-to-background method) for quantitative analysis of particles, rough surfaces, and biological specimens. The atomic number dependence of the bremsstrahlung also causes an important artifact in x-ray mapping of minor constituents, which can lead to serious misinterpretation if not recognized and corrected.

6.2. Characteristic X-Ray Production

6.2.1. Origin

A beam electron can interact with the tightly bound inner shell electrons of a specimen atom, ejecting an electron from a shell. The atom is left as an ion in an excited, energetic state, as shown in Fig. 6.5. The incident beam electron leaves the atom having lost at least E_K, where E_K is the binding energy of the electron to the K shell. The ejected orbital electron leaves the atom with a kinetic energy of a few eV to several keV, depending on the interaction. The atom itself is left in an excited state with a missing inner shell electron. The atom relaxes to its ground state (lowest energy) within approximately 1 ps through a limited set of allowed transitions of outer shell electron(s) filling the inner-shell vacancy. The energies of electrons in the shells (atomic energy levels) are sharply defined with values characteristic of a specific element. The energy difference between electron shells is a specific or characteristic value for each element. The excess energy can be released from the atom during relaxation in one of two ways (two branches in Fig. 6.5). In the Auger process, the difference in shell energies can be transmitted to another outer shell electron, ejecting it from the atom as an electron with a specific kinetic energy. In the characteristic x-ray process, the difference in energy is expressed as a photon of electromagnetic radiation which has a sharply defined energy. For the case of the neon atom shown schematically in Fig. 6.5, creation of a

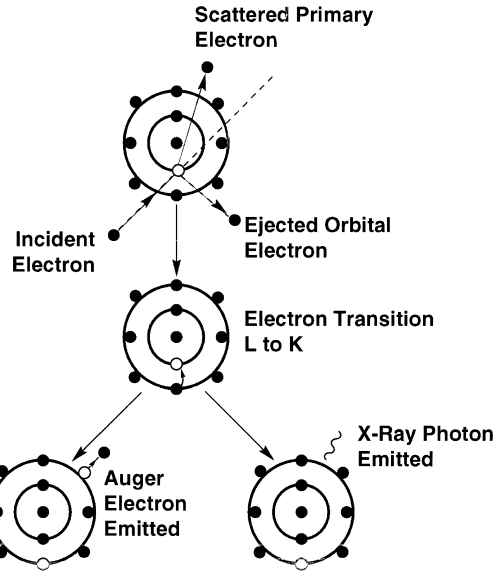


Figure 6.5. Inner shell electron ionization in an atom and subsequent de-excitation by electron transitions. The incident electron is elastically scattered. The unscattered direction of the incident electron is shown by the dotted line. The difference in energy from an electron transition is expressed either as the ejection of an energetic electron with characteristic energy (Auger process) or by the emission of a characteristic x-ray photon.

K-series x-ray involves filling the vacant state in the innermost electron shell (K shell) with an electon transition from the next shell out (L shell). Thus, the energy of the $K\alpha$ x-ray produced is equal to the difference in energy between the K shell and the L shell. The actual situation is more complicated than this because the L shell and the M shell are split into subshells, as we will see shortly. A comprehensive treatment of the properties of characteristic x-rays is beyond the scope of this book, and the interested reader is referred to the literature (e.g., Bertin, 1975). However, certain basic concepts that are fundamental to x-ray microanalysis will be discussed in this chapter.

6.2.2. Fluorescence Yield

The partitioning of the de-excitation process between the x-ray and Auger branches (Fig. 6.5) is described by the fluorescence yield ω. For the production of K radiation the fluorescence yield is

$$\omega_K = \frac{\#K \text{ photons produced}}{\#K\text{-shell ionizations}}. \tag{6.3}$$

The x-ray process is not favored for low atomic numbers; for example, $\omega_K \sim 0.005$ for the carbon K shell (Fig. 6.6). The characteristic x-ray process dominates for high atomic numbers; for example, $\omega_K \simeq 0.5$ for germanium, increasing to near unity for the heaviest elements. The fluorescence yields of the L and M shells are also shown in Fig. 6.6.

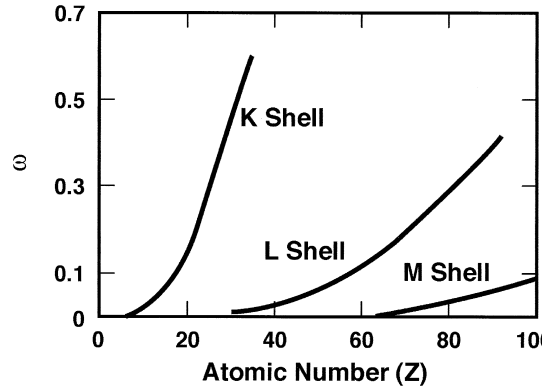


Figure 6.6. Fluorescence yield ω as a function of atomic number for electron ionization within the K, L, and M electron shells.

6.2.3. Electron Shells

Electrons of an atom occupy electron shells around the atom that have specific energies. In order of increasing distance from the atomic nucleus, these shells are designated the K shell, the L shell, the M shell, etc. (Fig. 6.5). These shells are directly related to the quantum numbers of atomic physics. For shells beyond the K shell, the shells are divided into subshells. For example, the L shell is composed of three subshells that are closely spaced in energy, and the M shell has five subshells. Electrons populate these subshells in a definite scheme as shown in Table 6.1.

Table 6.1. Electron Shells and Subshells of Atoms

X-ray notation	Modern notation	Electrons that may fill vacant K states	Electrons that may fill vacant L states	Electrons that may fill vacant M states	Maximum electron subshell population
K	$1s$				2
L_{I}	$2s$				2
L_{II}	$2p^{1/2}$	x			2
L_{III}	$2p^{3/2}$	x			4
M_{I}	$3s$				2
M_{II}	$3p^{1/2}$				2
M_{III}	$3p^{3/2}$	x	x		4
M_{V}	$3d^{3/2}$	x	x		4
M_{IV}	$3d^{5/2}$	x	x		6
N_{I}	$4s$				2
N_{II}	$4p^{1/2}$				2
N_{III}	$4p^{3/2}$	x	x	x	4
N_{IV}	$4d^{3/2}$	x	x	x	4
N_{V}	$4d^{5/2}$	x	x	x	6
N_{VI}	$4f^{5/2}$	x	x	x	6
N_{VII}	$4f^{7/2}$	x	x	x	8

More detailed information can be found in the Enhancements CD, Chapter 6.

6.2.4. Energy-Level Diagram

Figure 6.5 schematically shows electrons in orbits around the nucleus; an alternative way of representing the energies of specific electrons is with an energy-level diagram such as Fig. 6.7. Such diagrams originate from the more accurate quantum mechanical picture of an atom that results from the solution of the Schrödinger wave equation. Here the energy of an atom is represented by the energies of the various vacant states (electrons removed) that can be created by the electron beam. For example, ionization of an electron in the K shell increases the energy of the atom to that of the "K level." If an L-shell electron moves into the K shell to fill the vacancy, the energy of the atom decreases to the "L level," but there is still now a vacant state in the L shell, which would be filled by an electron of some lower energy, and so on.

6.2.5. Electron Transitions

Characteristic x-ray lines result from transitions between subshells; however, atomic theory tells us that only transitions between certain subshells are allowed (see Table 6.1). This means that not all transitions between

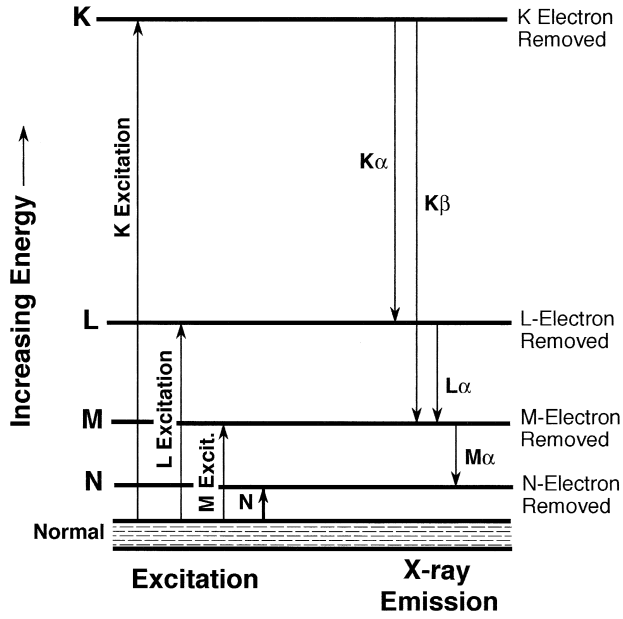


Figure 6.7. Energy level diagram for an atom. The energy of the atom increases upon ionization of the K, L, M, or N shell (excitation). As the atom's energy returns to normal, $K\alpha$, $K\beta$, $L\alpha$, and $M\alpha$ x-rays are emitted from the atom. Each horizontal line represents the energy of an electron state. Zero energy represents an atom at rest with no electrons missing (normal).

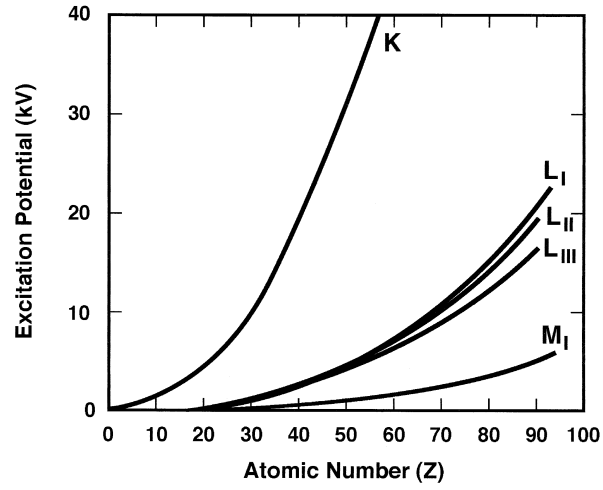


Figure 6.8. Critical excitation energies (absorption edge energies) for the K, L, and M series as a function of atomic number.

shells result in characteristic x-rays. Returning to the example of the copper K series, electron transitions (note Table 6.1) to fill a vacant K state from the L shell in the Cu atom are as follows:

Electron moves from $L_{\mathrm{III}} \rightarrow K$	$K\alpha_1$ x-ray of $E_{K\alpha_1} = E_K - E_{L_{\mathrm{III}}}$
Electron moves from $L_{\mathrm{II}} \rightarrow K$	$K\alpha_2$ x-ray of $E_{K\alpha_2} = E_K - E_{L_{\mathrm{II}}}$
Electron moves from $L_{\mathrm{I}} \rightarrow K$	Not allowed; no emission

6.2.6. Critical Ionization Energy

Because the energy of each shell and subshell is sharply defined, the minimum energy necessary to remove an electron from a specific shell has a sharply defined value as well. This energy is called the critical ionization or excitation energy E_{c}, also known as the excitation potential or x-ray absorption edge energy for example, E_K and $\mathrm{E}_{L_{III}}$. Each shell and subshell of an atom requires a different critical ionization energy for electron removal. Figure 6.8 shows the critical excitation energies for the K, L, and M series as a function of atomic number. As an example, consider the wide range in critical ionization energies for the K, L, and M shells and subshells of platinum ($Z = 78$) listed in Table 6.2. A 20-keV electron beam can ionize the L and M shells of Pt, but not the K shell. As the atomic number decreases, as shown in Table 6.2 for Nb ($Z = 41$) and Si ($Z = 14$), the ionization energies decrease. The critical ionization energy is an important parameter in calculating characteristic x-ray intensities. As discussed in following chapters, for x-ray microanalysis we typically operate the SEM at energies two to three times the critical ionization energies of the elements of interest. Tabulations of the critical ionization energy and characteristic

Table 6.2. Critical Ionization Energies for Pt, Nb, and Si

Shell	E_c for Pt (keV)	E_c for Nb (keV)	E_c for Si (keV)
K	78.39	18.99	1.839
L_I	13.88	2.697	0.149
L_{II}	13.27	2.464	0.099
L_{III}	11.56	2.370	0.099
M_I	3.296	0.468	
M_{II}	3.026	0.378	
M_{III}	2.645	0.363	
M_{IV}	2.202	0.207	
M_V	2.122	0.204	

x-ray energies for the K, L, and M shells are available in the tables (Bearden, 1967a,b) given in the accompanying CD.

6.2.7. Moseley's Law

X-rays emitted during an electron transition are called characteristic x-rays because their specific energies and wavelengths are characteristic of the particular element which is excited. The energies of the electron shells vary in a discrete fashion with atomic number, so that the x-rays emitted in the process have energies characteristic of that atomic number. The difference in energy between shells changes in a regular step when the atomic number changes by one unit. This fact was discovered by Moseley (1913, 1914) and can be expressed in the form of an equation:

$$E = A(Z - C)^2, \tag{6.4}$$

where E is the energy of the x-ray line, and A and C are constants which differ for each x-ray series ($C = 1.13$ for the K series and approximately 7 for the L series). Thus, Eq. (6.4) may be used to find the energy of any element K or L line. Moseley's law forms the basis for qualitative x-ray analysis, that is, the identification of elemental constituents.

6.2.8. Families of Characteristic Lines

For elements with atomic numbers ≥ 11 (sodium), the shell structure is sufficiently complex that when an ionization occurs in the K shell, the transition to fill that vacant state can occur from more than one outer shell. As shown in Fig. 6.7, following ionization of a K-shell electron, a transition to fill the vacant state can occur from either the L shell or the M shell. X-rays resulting from transitions of electrons from the M shell to the K shell are designated $K\beta$ x-rays. Because the energy difference between the K and M shells is larger than between the K and L shells, the $K\beta$ x-ray energy is larger than that of the $K\alpha$. For example, for copper, the

$K\alpha$ x-ray has an energy of 8.04 keV and the $K\beta$ x-ray has an energy of 8.90 keV.

Primary ionization involves removing an electron from a bound state in a shell to an effective infinity outside the atom. However, characteristic x-rays are formed by transitions between bound states in shells. Therefore, the energy of the characteristic x-ray is *always less* than the critical ionization energy for the shell from which the original electron was removed. Thus, $E_{K\alpha} = E_K - E_L$ and $E_{K\beta} = E_K - E_M$; therefore, $E_{K\alpha} \leq E_{K\beta} \leq E_K$.

As noted in Table 6.1 and Fig. 6.9, the L, M, and N shells may be further divided into subshells of slightly different energy, giving rise to

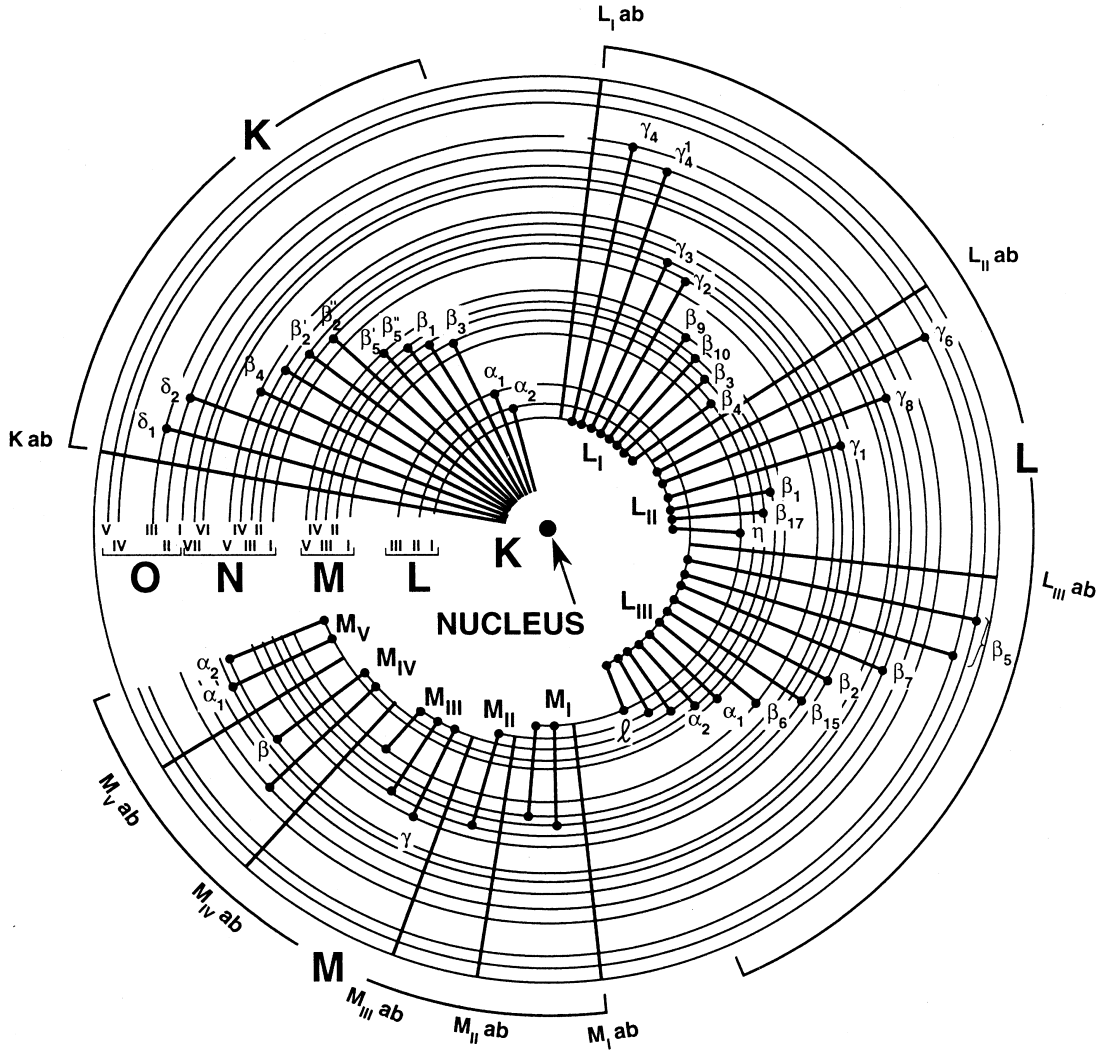


Figure 6.9. Electron transitions in the atom that give rise to M-series, L-series, and K-series x-rays. By reference to a table of atomic energy levels, the x-ray energy of each photon shown can be calculated as the difference between the associated energy levels (Woldseth, 1973).

additional x-ray lines. For copper, various transitions into the K shell produce $K\alpha_1$ (8.048 keV, a transition from the L_{III} subshell) and $K\alpha_2$ (8.028 keV, a transition from the L_{II} subshell) and several $K\beta$ lines from the M and N shells: $K\beta_1$ (8.905 keV, from M_{III}), $K\beta_2$ (8.976 keV, from $N_{II,III}$), and $K\beta_5$ (8.977 keV, from $M_{IV,V}$) (see Fig. 6.9). Not all observable x-ray lines have been assigned Greek letter designations. A number of x-ray lines with low intensity relative to the principal lines are designated in an alternative style which is used in atomic physics. In this scheme, the first letter denotes the shell in which the ionization initially takes place, with the subscript indicating the specific subshell, and the second letter and number indicates the shell and subshell from which the inner shell vacant state is filled. Thus, the weak $M_{II}N_{IV}$ line arises from an ionization in the M_{II} subshell followed by an electron transition from the N_{IV} subshell.

Within the SEM beam energy range, each element can emit x-rays of the K series only (light elements), the K series plus the L series (intermediate elements), or the L series plus the M series (heavy elements). Figure 6.9 shows the electron transitions responsible for essentially all the possible x-ray lines. Because of the relatively low resolution of the energy-dispersive spectrometer (EDS), only some of the possible x-ray lines are observed (see chapter 8, Fig. 8.1). Many more of the x-ray lines shown in Fig. 6.9 are observed with the wavelength-dispersive spectrometer (WDS). As the atomic number of the atom increases above 10, the K-shell x-rays split into $K\alpha$ and $K\beta$ pairs. Above atomic number 21, an additional family, the L shell, begins to become measurable at 0.2 keV. When atomic number 50 is reached, the M-shell family begins to appear above 0.2 keV. The complexity of the shells is extended further by the splitting of lines due to the energy divisions of subshells. Figure 6.10 shows elemental x-ray spectra at the energy resolution typical of an EDS system for several elements of increasing atomic number. These spectra show that K-, L-, and M-series x-rays increase in energy with increasing atomic number from Zn ($Z = 30$) to Nb ($Z = 41$) to La ($Z = 57$) to Pt ($Z = 78$). Note that many x-ray lines for the lighter elements are not resolved (e.g., $L\alpha$, $L\beta$, $L\gamma$) because of the close proximity in energy of the peaks and the relatively poor resolution of the energy-dispersive x-ray spectrometer.

When a beam electron has sufficient energy to ionize a particular electron shell in an atom to produce characteristic x-ray lines, all other characteristic x-rays of lower energy for the atom will also be excited, provided there are electrons available for such transitions. The production of lower energy x-rays occurs because of (1) direct ionization of those lower energy shells by beam electrons and (2) propagation of the vacant state outward (K to L to M shells) as the atom returns to the ground state.

6.2.9. Natural Width of Characteristic X-Ray Lines

Energy levels are so sharply defined that the energy width of a characteristic x-ray is a small fraction of the line energy. For example, elements

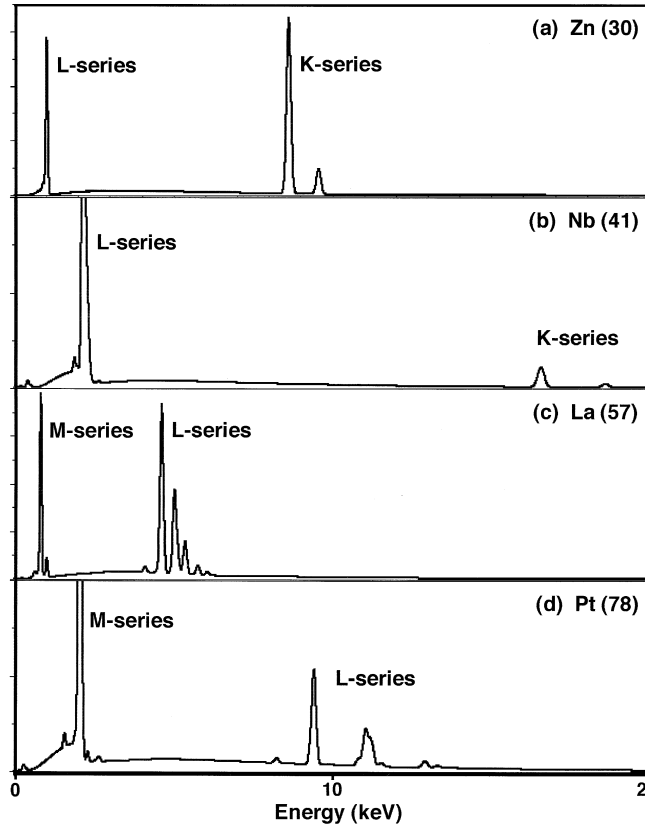


Figure 6.10. EDS spectra for families of x-ray lines (a) Zinc ($Z = 30$), (b) niobium ($Z = 41$), (c) lanthanum ($Z = 57$), and (d) platinum ($Z = 78$). Note that for zinc the L series is an unresolved single peak, but as the atomic number increases the individual L-series peaks spread out into separate lines.

such as calcium ($Z = 20$) have an x-ray linewidth at half-height intensity of only 2 eV (Salem and Lee, 1976). This narrow, linelike character of peaks superimposed on the broad, slowly changing x-ray background has led to the frequent use of the term x-ray "lines" in the literature when referring to characteristic x-ray peaks. Note that the width of measured peaks in Fig. 6.1 is governed by the resolution of the EDS x-ray spectrometer used to acquire the spectrum. Typically the peak is about 70 times wider than the natural linewidth.

6.2.10. Weights of Lines

Although many possible electron transitions can occur in high-atomic-number elements to fill a vacancy in an inner shell, which in turn gives rise to the families of x-rays (Fig. 6.9), the probability for each type of transition varies considerably. The relative transition probabilities for the lines arising

Table 6.3. Weights of Lines

Family	Approximate line weights for K, L, or M shells
K	$K\alpha = 1$, $K\beta = 0.1$
L	$L\alpha = 1$, $L\beta_1 = 0.7$, $L\beta_2 = 0.2$, $L\gamma_1 = 0.08$, $L\gamma_3 = 0.03$, $L_l = 0.04$
M	$M\alpha = 1$, $M\beta = 0.6$, $M\zeta = 0.06$, $M\gamma = 0.05$, $M_{\mathrm{II}}N_{\mathrm{IV}} = 0.01$

from an ionization of a specific shell are termed the "weights of lines." The weights of lines are dependent upon atomic number and vary in a complex fashion, but in general, the greater the energy difference in the electron transition, the less probable and less intense is the x-ray line. Thus, the $K\beta$ lines are less intense than the $K\alpha$ lines. The intrafamily weights of K lines are the best known; for atomic numbers above $Z = 18$ the $K\alpha$ to $K\beta$ ratio is about 10:1. General values are presented in Table 6.3 for the lines of significant intensity that can be readily observed in energy-dispersive x-ray spectra. Although these values may not be exact for a specific element, these weights are a useful guide in interpreting spectra and assigning peak identifications in energy-dispersive x-ray spectrometry. Note that the values in Table 6.3 are weights of lines for one element. It is difficult to compare x-ray line intensities of different elements even for the same spectral series because the intensity depends on a variety of factors including the fluorescent yield, absorption (see Section 6.4), and the excitation energy.

6.2.11. Cross Section for Inner Shell Ionization

Numerous cross sections expressing the probability for inner shell ionization can be found in the literature; these have been reviewed by Powell (1976a, b; 1990). The basic form of the cross section is that derived by Bethe (1930):

$$Q = 6.51 \times 10^{-20} \frac{n_{\mathrm{S}} b_{\mathrm{S}}}{U E_{\mathrm{c}}^2} \log_e(c_{\mathrm{S}} U), \tag{6.5}$$

where n_{s} is the number of electrons in a shell or subshell (e.g., $n_{\mathrm{s}} = 2$ for a K shell), b_{s} and c_{s} are constants for a particular shell, E_{c} is the critical ionization energy (keV) of the shell, and U is the overvoltage:

$$U = \frac{E}{E_{\mathrm{c}}}, \tag{6.6}$$

where E is the instantaneous beam energy. The cross section for K-shell ionization for silicon is plotted as a function of overvoltage in Fig. 6.11. The cross section increases rapidly from an overvoltage of 1 to a maximum at about $U \simeq 3$. As beam electrons lose their energy because of inelastic scattering, they can interact with atomic inner shells down to overvoltages as low as $U = 1$.

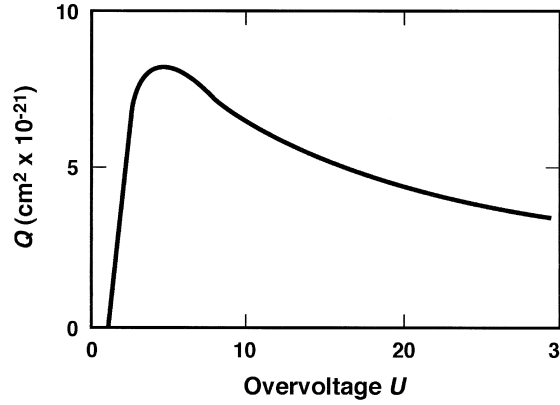


Figure 6.11. Plot of the cross section for inner shell ionization of the silicon K shell as a function of overvoltage $U = E/E_c$.

6.2.12. X-Ray Production in Thin Foils

X-ray production from a thin foil can be estimated using the cross section for inner shell ionization. A "thin" foil has a thickness which is small compared to the elastic mean free path of a beam electron. In a "thin" foil the average beam electron passes through the foil along its incident trajectory without significant deflection. The energy loss due to inelastic scattering is also negligible when the foil is thin. To convert the ionization cross section given in Eq. (6.5), which has dimensions of ionizations/e^-/(atom/cm^2), to x-ray production per electron n_x, in units of photons/e^-, the following dimensional argument is used:

$$\frac{\text{photons}}{e^-} = \frac{\text{ionizations}}{e^-(\text{atom/cm}^2)} \cdot \frac{\text{photons}}{\text{ionization}} \cdot \frac{\text{atoms}}{\text{mole}} \cdot \frac{\text{moles}}{\text{g}} \cdot \frac{\text{g}}{\text{cm}^3} \cdot \text{cm}$$

$$n_x = Q\omega N_0 \frac{1}{A}\rho t, \tag{6.7}$$

where ω is the fluorescence yield, N_0 is Avogadro's number, A is the atomic weight, ρ is the density, and t is the thickness.

6.2.13. X-Ray Production in Thick Targets

A thick specimen target is one for which elastic scattering, inelastic scattering, and energy loss are significant. A thick specimen has a thickness several times the elastic mean free path, or nominally about 100 nm or more. Specimen targets greater than 10 μm in thickness are effectively infinite in thickness to beam electrons at typical energies for the SEM ($\leq$30 keV). To accurately calculate x-ray production for thick targets, one must consider that the x-ray cross section varies over the entire range of electron energy from E_0 down to E_c. The calculation of the x-ray intensity under these

conditions will be presented in Chapter 9 as part of the physical basis for quantitative x-ray analysis.

A number of workers have reported experimental measurements of the generated characteristic x-ray intensity I_c. This is the intensity produced in the specimen prior to its absorption as the radiation propagates through the specimen (Green, 1963; Lifshin *et al.,* 1980). The experimental expressions have the general form

$$I_c = i_p\, a \left(\frac{E_0 - E_c}{E_c} \right)^n = i_p\, a\, (U-1)^n\,, \tag{6.8}$$

where i_p is the electron probe current and a and n are constants for a particular element and shell. The value of n is normally in the range of 1.5–2.0. Thus, for a particular characteristic x-ray line, the line intensity increases as the beam energy increases.

6.2.14. X-Ray Peak-to-Background Ratio

The most important factor in determining the limits of detection in x-ray spectrometric analysis is the presence of the continuum background, that is, noncharacteristic radiation at the same energy as the characteristic radiation of interest. By dividing I_c by I_{cm} [Eq. (6.8) by Eq. (6.2)], we can calculate the peak-to-background ratio and observe the variables that influence the limits of detection:

$$\frac{P}{B} = \frac{I_c}{I_{cm}} = \frac{1}{Z} \left(\frac{E_0 - E_c}{E_c} \right)^{n-1}. \tag{6.9}$$

From Eq. (6.9), the peak-to-background ratio increases as the difference $E_0 - E_c$ increases. An increase in P/B allows the analyst to measure a smaller elemental mass fraction in a specimen. It would therefore seem to be advantageous to make E_0 as large as possible for a specific excitation energy E_c of an element in the specimen. However, Eq. (6.9) only accounts for generated x-rays. As noted in the discussion of interaction volume in Chapter 3, beam electrons penetrate deeper into the specimen as the beam energy increases. Correspondingly, x-rays are produced deeper into the specimen as the beam energy is increased. Besides degrading the spatial resolution of analysis, another consequence of increasing the beam energy is an increased absorption of x-rays (see Section 6.4) within the specimen before they are measured in the x-ray detector. This absorption reduces the measured x-ray intensity, degrades the limit of detection, and increases the uncertainty of the correction for absorption that must be applied in the quantitative analysis procedure (Chapter 9). Thus, for a thick, bulk specimen there is an optimum beam energy beyond which further increases in beam energy actually degrade analytical performance. This limit depends on the energy of the characteristic x-rays and the composition of the specimen. An overvoltage U of 2–3 is usually optimum for a given element.

6.3. Depth of X-Ray Production (X-Ray Range)

Depending on the critical excitation energy E_c, characteristic x-rays may be generated over a substantial fraction of the electron interaction volume as shown in Fig. 3.5. To predict the depth of x-ray production (x-ray range) and the lateral x-ray source size (x-ray spatial resolution), the starting point is the electron range, such as given by the Kanaya–Okayama range expression Eq. (3.4). Electron range expressions have the following general form:

$$\rho R = K E_0^n, \tag{6.10}$$

where E_0 is the incident electron beam energy, ρ is the density, K depends on material parameters, and n is a constant between 1.2 and 1.7. This formulation of the range considers electrons that lose all their energy while scattering in the specimen. Characteristic x-rays can only be produced within that portion of the electron trajectories for which the energy exceeds E_c for a particular x-ray line, whereas bremsstrahlung x-rays continue to be produced until the electron energy equals zero. The range of direct primary x-ray generation is therefore always smaller than the electron range. To account for the energy limit of x-ray production, the range equation is modified to the form

$$\rho R = K\left(E_0^n - E_c^n\right). \tag{6.11}$$

6.3.1. Anderson–Hasler X-Ray Range

By fitting Eq. (6.11) to experimental data, Anderson and Hasler (1966) evaluated the constants K and n, obtaining an analytical expression for the x-ray range useful for most elements:

$$R_x = \frac{0.064}{\rho}\left(E_0^{1.68} - E_c^{1.68}\right), \tag{6.12}$$

where R_x has units of μm when E is in keV and ρ is in g/cm^3. Figure 6.12 shows Al $K\alpha$ and Cu $K\alpha$ ranges in an Al specimen and Cu $L\alpha$ and Cu $K\alpha$ in a Cu specimen as a function of beam energy. Note that the x-ray ranges in Al are all less than the electron range.

6.3.2. X-Ray Spatial Resolution

For an electron beam normal to a surface, the maximum width of the electron interaction volume or x-ray generation volume projected up to the specimen surface is approximately equal to the spatial resolution. Figure 6.13 shows how the x-ray range and the x-ray spatial resolution L_x are defined in cross section. As the atomic number and density of the target increase, the depth of production for the principal line decreases. The depth of production is also a function of the critical ionization energy of the line. These effects are also shown in Fig. 6.13. For the same beam energy the figure shows how the x-ray range and the x-ray spatial resolution vary for Al $K\alpha$ and Cu $K\alpha$ in a matrix of density $\sim$3 g/cm^3 (e.g., aluminum) and in a matrix of density $\sim$10 g/cm^3 (e.g., copper). In the low-density sample,

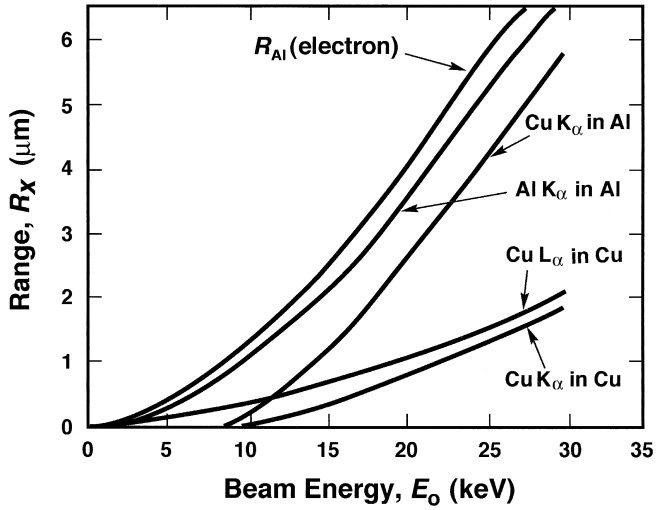


Figure 6.12. X-ray generation range (Anderson–Hasler) for Al $K\alpha$ and Cu $K\alpha$ generated in an Al specimen and Cu $K\alpha$ and Cu $L\alpha$ in a Cu specimen as a function of beam energy. Note that the electron range for aluminum R_{Al} is larger than the x-ray range of the x-ray lines generated by the electron beam.

a trace amount of Cu is present in an Al sample, whereas in the high-density sample, a trace element amount of Al is present in a Cu sample. Both Al $K\alpha$ and Cu $K\alpha$ are produced at greater depths in the low-density matrix than in the high-density matrix. The shapes of the interaction and x-ray generation volumes differ considerably in the two targets, with the low-density matrix having a pear shape and the high-density matrix giving a less distinct neck.

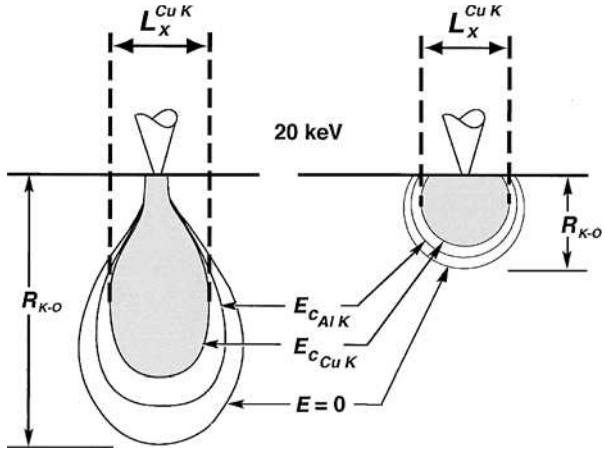


Figure 6.13. Comparison of x-ray production regions from specimens with densities of 3 g/cm^3 (left) and 10 g/cm^3 (right) at a beam energy of 20 keV. The x-ray spatial resolution L_x is found by projecting the maximum diameter of the x-ray distribution to the surface of the specimen.

6.3.3. Sampling Volume and Specimen Homogeneity

The range of primary x-ray generation is the critical parameter in estimating the "sampling volume" for x-ray microanalysis. From Figs. 6.12 and 6.13, it is clear that the sampling volume for x-ray microanalysis depends on the beam energy and the energy of the x-ray line measured. When two different analytical lines are available for a given element with widely differing energies (Cu $K\alpha$ and Cu $L\alpha$ in Cu, Fig. 6.12), the depth of the sampling volume may differ by a factor of two or more. If the specimen is heterogeneous in composition over a depth equivalent to the electron range, the effect of differing sampling volumes will often lead to incorrect results. Therefore, a fundamental requirement for accurate results by conventional quantitative x-ray microanalysis is that the specimen must be homogeneous over the electron range.

6.3.4. Depth Distribution of X-Ray Production, $\phi(\rho z)$

Although the single numerical value given by the x-ray range defines the effective spatial limit of x-ray production, even a cursory examination of the Monte Carlo electron trajectory/x-ray plots in Fig. 3.5 suggests that the depth distribution of x-ray production is not uniform within the x-ray range. In fact, the distribution of x-ray generation beneath the surface is nonuniform both laterally and in depth. The lateral distribution, as projected up to the surface plane, is important in defining the spatial resolution of x-ray microanalysis. The x-ray distribution in depth is important because the x-rays produced below the surface must propagate through the specimen to escape, and some fraction of these x-rays is lost because of photoelectric absorption (described below). Figure 6.14 shows a Monte Carlo simulation

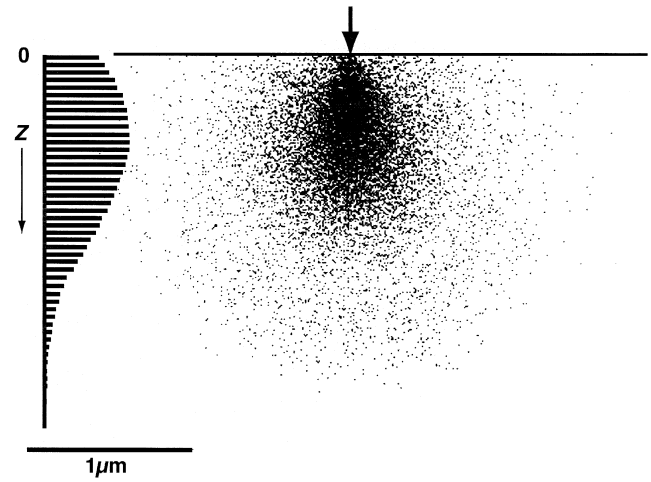


Figure 6.14. (Right) Monte Carlo simulation of the generation of Al $K\alpha$ x-rays in pure Al at 30 keV. The electron beam enters the specimen at the surface, 0, and progresses into the specimen as a function of distance Z. (Left) A histogram of the x-ray generation function with depth, $\phi(\rho z)$. X-ray generation increases to a maximum just below the surface and then decreases as fewer x-rays are generated deeper within the specimen.

of the generation of Al $K\alpha$ x-rays in pure Al at 30 keV. Each dot represents a generated x-ray. The histogram on the left of the figure shows the number of x-rays generated at various depths z into the specimen. The x-ray depth distribution function, called $\phi(\rho z)$, is an important tool in developing quantitative x-ray microanalysis correction procedures (Castaing, 1951) and is discussed in Chapter 9.

6.4. X-Ray Absorption

X-rays, as photons of electromagnetic radiation, can undergo the phenomenon of photoelectric absorption upon interacting with an atom. That is, the photon is absorbed and the energy is completely transferred to an orbital electron, which is ejected with a kinetic energy equal to the photon energy minus the binding energy (critical ionization energy) by which the electron is held to the atom. For x-rays of an incident intensity I_0 in photons per second propagating through a slab of thickness t and density ρ, the intensity on the exit surface is attenuated according to the expression

$$I = I_0 \exp\left[-\left(\frac{\mu}{\rho}\right)_{\text{absorber}}^{\text{x-ray}} (\rho t)\right], \tag{6.13}$$

where (μ/ρ) is the mass absorption coefficient (cm^2/g) of the sample atoms (absorber) for the specific x-ray energy of interest. Fortunately, there is no change in the energy of the x-ray photon that passes through the absorber. For the micrometer dimensions of interest in electron beam x-ray microanalysis, inelastic scattering of x-rays is insignificant. Typical values of mass absorption coefficients for Cu $K\alpha$ radiation traveling through various pure elements are listed in Table 6.4. Note that for Cu $K\alpha$ radiation the mass absorption coefficient is high for Co as an absorber. The energy of Cu $K\alpha$ is slightly higher than E_c for Co, but is much lower than E_c for Cu as an absorber. Because the energy of an element's characteristic radiation is always less than E_c, the absorption coefficient of an element for its own radiation is low. Therefore, an element passes its own characteristic line with little absorption. Extensive compilations of mass absorption coefficients are available in the literature (Heinrich 1966, 1986). A selection of mass absorption coefficients for $K\alpha$ and $L\alpha$ characteristic lines is tabulated in the accompanying CD.

Table 6.4. Mass Absorption Coefficient (μ/ρ) of Several Elements for Cu $K\alpha$

Element (atomic number)	X-ray energy (keV)			(μ/ρ) of Cu $K\alpha$ in given element (cm^2/g)
	$K\alpha$	$K\beta$	$E_c = E_{K\text{edge}}$	
Mn (25)	5.895	6.492	6.537	272
Fe (26)	6.400	7.059	7.111	306
Co (27)	6.925	7.649	7.709	329
Ni (28)	7.472	8.265	8.331	49
Cu (29)	8.041	8.907	8.980	52

As an example, one can use Eq. (6.13) to calculate the amount of x-ray absorption in the Be window of an EDS x-ray detector. Suppose fluorine $K\alpha$ radiation is passing through an 8-μm thick Be window. The mass absorption coefficient for F $K\alpha$ radiation in Be is 2039 cm^2/g:

$$I/I_0 = \exp[-(2039\ \text{cm}^2/\text{g})(1.848\ \text{g/cm}^3)(8 \times 10^{-4}\ \text{cm})]$$
$$I/I_0 = 0.05.$$

Only 5% of the F $K\alpha$ radiation incident on the detector window passes through. Therefore fluorine is not considered an element that can be practically measured with an EDS detector employing a Be window.

6.4.1. Mass Absorption Coefficient for an Element

Photoelectric absorption by electrons in a specific shell requires that the photon energy exceed the electron binding energy for that shell. When the photon energy is slightly greater than the binding energy of the electron in the absorber (E_c), the probability for absorption is highest. For a specific absorber, mass absorption coefficients generally decrease in a smooth fashion with increasing x-ray energy given approximately by

$$\left(\frac{\mu}{\rho}\right) = KZ^4\left(\frac{1}{E}\right)^3. \tag{6.14}$$

However, there is a sharp jump in absorption coefficient in the energy region just greater than the critical excitation energy for each shell of the absorber. Figure 6.15 shows a plot of (μ/ρ) for the element lanthanum ($Z = 57$) as an absorber of various x-ray energies. Sharp jumps in (μ/ρ) occur at the energy of the K edge at 38.9 keV, the L edges at ~5.9 keV, and the M edges at ~1.1 keV. These jumps are referred to as "x-ray absorption edges." X-rays with an energy slightly greater than the x-ray absorption edge (critical ionization energy) can eject a bound electron and therefore are strongly absorbed themselves.

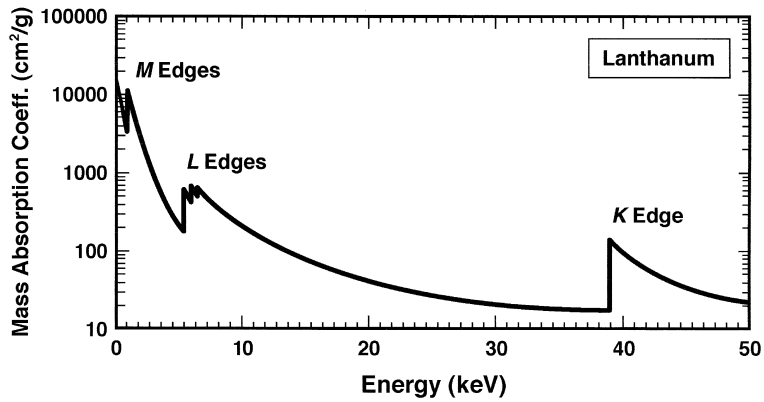


Figure 6.15. Mass absorption coefficient as a function of x-ray energy in an absorber of lanthanum (atomic number 57).

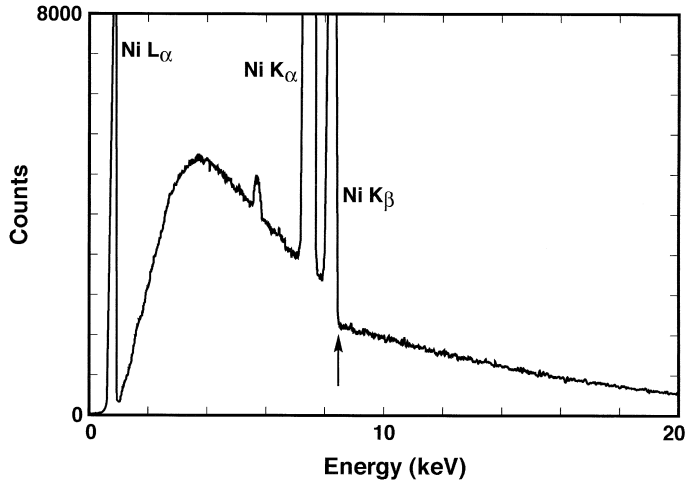


Figure 6.16. Energy-dispersive x-ray spectrum of nickel generated with a primary electron beam energy of 40 keV. The spectrum shows a sharp step in the x-ray continuum background due to increased absorption just above the Ni K absorption edge (arrow).

6.4.2. Effect of Absorption Edge on Spectrum

Absorption edges can be directly observed when the x-ray spectrum energy range spans the critical excitation energy for the absorber element in the target. The electron-excited x-ray bremsstrahlung provides a continuous distribution in energy. At the absorption edge, the bremsstrahlung intensity abruptly decreases for x-ray energies slightly above the edge because the mass absorption coefficient increases abruptly at the absorption edge. An example for nickel excited with a 40-keV electron beam is shown in Fig. 6.16. Just above the Ni $K\alpha$ and Ni $K\beta$ lines, at the position of the critical excitation energy for the Ni K shell ($E_c = E_K = 8.33$ keV), the x-ray bremsstrahlung is much lower than what would be expected if the bremsstrahlung below the peaks were extrapolated to higher energy. This sudden decrease in the continuum corresponds to a sudden increase in the mass absorption coefficient for Ni as shown in Fig. 6.17. The presence of absorption edges in spectra may be difficult to observe if they are hidden under an x-ray peak that has been broadened by the detection process as is often the case in EDS analysis.

6.4.3. Absorption Coefficient for Mixed-Element Absorbers

The mass absorption coefficient for a sample containing a mixture of elements is found by taking the summation of the mass absorption coefficient for each element multiplied by its weight (mass) fraction:

$$\left(\frac{\mu}{\rho}\right)^{i}_{\text{spec}} = \sum_{j} \left(\frac{\mu}{\rho}\right)^{i}_{j} C_j, \tag{6.15}$$

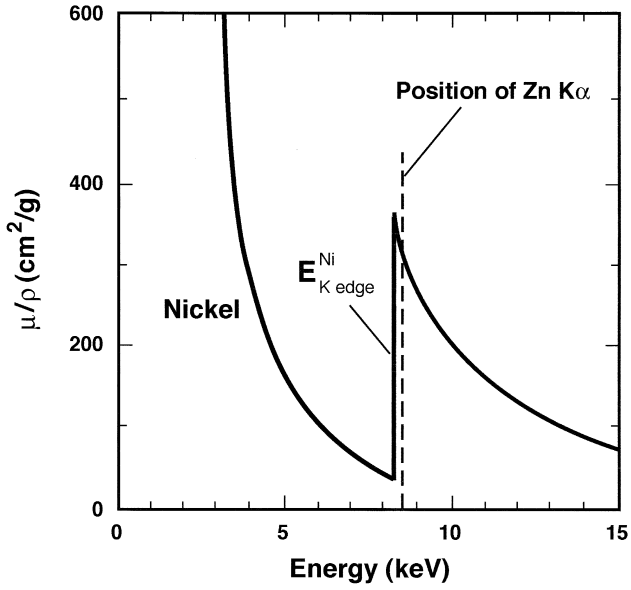


Figure 6.17. Mass absorption coefficient of a Ni absorber as a function of x-ray energy. A step increase in μ/ρ occurs at an energy above the Ni K edge. The position of the Zn $K\alpha$ x-ray line energy is noted on the drawing. The Zn $K\alpha$ line is just slightly more energetic than the Ni K edge (the E_c for the Ni K series) and thus causes fluorescence of Ni K radiation. At the same time the Zn $K\alpha$ line will be heavily absorbed in Ni.

where $(\mu/\rho)^i_j$ is the mass absorption coefficient for radiation from element i passing through element j and C_j is the weight (mass) fraction of element j. Thus, for Cu $K\alpha$ radiation propagating through a sheet of SiO_2 the mass absorption coefficient may be calculated as

$$\left(\frac{\mu}{\rho}\right)^{\mathrm{Cu}K\alpha}_{\mathrm{SiO_2}} = (\text{wt. fraction Si})\left(\frac{\mu}{\rho}\right)^{\mathrm{Cu}K\alpha}_{\mathrm{Si}} + (\text{wt. fraction O})\left(\frac{\mu}{\rho}\right)^{\mathrm{Cu}K\alpha}_{\mathrm{O}}$$

$$\left(\frac{\mu}{\rho}\right)^{\mathrm{Cu}K\alpha}_{\mathrm{SiO_2}} = 0.468\left(63.7\frac{\mathrm{cm}^2}{\mathrm{g}}\right) + 0.533\left(11.2\frac{\mathrm{cm}^2}{\mathrm{g}}\right) = 35.8\frac{\mathrm{cm}^2}{\mathrm{g}}. \tag{6.16}$$

It is important to note that x-ray photoelectric absorption is an "all-or-nothing" process. Either the photon is completely absorbed in a single absorption event or else it continues to propagate without modification of its energy. In the absorption process, characteristic x-rays that escape the specimen retain their specific energy.

6.5. X-Ray Fluorescence

For some element pairs in the specimen the primary x-ray of element A created by the electron beam can generate a secondary x-ray of element B

by fluorescence. This process is an inevitable consequence of photoelectric absorption of primary A x-rays by a specimen B atom with the ejection of a bound inner shell electron from the absorbing B atom. The primary A photon is absorbed by the specimen B atom and its energy is transferred to the kinetic energy of the ejected electron, the photoelectron. The B atom is left in the same excited state as that produced by inner shell ionization directly by the electron beam. Subsequent deexcitation as the atom returns to the ground state by electron transitions is the same for both cases: The excited atom will follow the same routes to deexcitation, producing either characteristic x-rays or characteristic electrons (Auger electrons). X-ray-induced emission of x-rays is referred to as "x-ray fluorescence." To distinguish the effects, "primary radiation" will refer to the x-rays produced by electron-beam ionization of atoms and "secondary radiation" will refer to x-rays produced through ionization of atoms by higher energy x-rays. Secondary radiation can be created from both characteristic x-rays and continuum (bremsstrahlung) x-rays. Because the x-ray photon causing fluorescence must have at least as much energy as the critical excitation energy of the atom in question, the energy of the secondary radiation will always be less than the energy of the photon that is responsible for ionizing the inner shell.

6.5.1. Characteristic Fluorescence

If a mixed sample consists of atom species A and B and the energy of the characteristic radiation from element A exceeds the critical excitation energy for element B, then characteristic fluorescence of B by the radiation of A will occur. The fluorescence effect depends on how close the photon energy of A is above the critical excitation energy for B, with the maximum effect occurring when E_A just exceeds E_c for B. The absorption process during characteristic fluorescence is illustrated in Fig. 6.17, where element A is Zn and element B is Ni. Nickel $K\alpha$ fluorescent radiation is produced, as Zn $K\alpha$ is strongly absorbed by the sample.

To examine this situation for a complex alloy, consider a sample containing a sequence of transition elements, for example, manganese, iron, cobalt, and nickel (Table 6.5). The critical excitation energy for manganese is lower than the $K\alpha$ energies for cobalt and nickel, and therefore characteristic manganese fluorescence will occur from Co and Ni x-ray radiation. The $K\beta$ energies of iron, cobalt, and nickel exceed the critical excitation energy for manganese, and so these radiations also contribute to the

Table 6.5. Secondary Fluorescence in a Sample Containing Mn, Fe, Co, and Ni

Fluoresced element (atomic number)	Radiation causing fluorescence
Mn (25)	Fe $K\beta$, Co $K\alpha$, Co $K\beta$, Ni $K\alpha$, Ni $K\beta$
Fe (26)	Co $K\beta$, Ni $K\alpha$, Ni $K\beta$
Co (27)	Ni $K\beta$
Ni (28)	None

secondary fluorescence of manganese. These arguments can be repeated for each element in the complex alloy. As shown in Table 6.5, the situation for the generation of secondary fluorescence is different for each of the constituents.

For characteristic fluorescence to be a significant effect, the primary x-ray radiation must be strongly absorbed, that is, the absorber must have a high mass absorption coefficient for the primary radiation. Examination of Tables 6.4 reveals how the absorption/fluorescence phenomena depend on proximity to the critical excitation energy. Cu $K\alpha$ radiation (8.04 keV) is strongly absorbed in cobalt ($Z = 27$, $E_c = 7.709$ keV, $\mu/\rho = 326$ cm^2/g) with consequent emission of fluorescent Co $K\alpha$ and $K\beta$; in nickel, which is only one atomic number higher, the absorption is reduced by almost a factor of seven ($Z = 28$, $E_c = 8.331$ keV, $\mu/\rho = 49$ cm^2/g). The Cu $K\alpha$ line (8.04 keV) is not energetic enough to ionize the K shell of Ni. The magnitude of the fluorescence effect from different x-ray energies on a given element can be estimated by comparing the mass absorption coefficients.

6.5.2. Continuum Fluorescence

Characteristic-induced fluorescence can only occur if the primary radiation in question is more energetic than the critical excitation energy of the element of interest. In practice, the fluorescence effect is only significant if the characteristic energy is within approximately 3 keV of the critical excitation energy. On the other hand, the continuum (bremsstrahlung) radiation provides a source of x-rays at all energies up to the incident beam energy. Continuum-induced fluorescence will therefore always be a component of the measured characteristic radiation from a specimen. The calculation (Henoc, 1968) of the intensity of continuum fluorescence involves integrating the contributions of all photon energies above the critical excitation energy for the edge of interest. In practice, the extra intensity ranges from 1 to 7% for $Z = 20$ to 30 at $E_0 = 20$ keV.

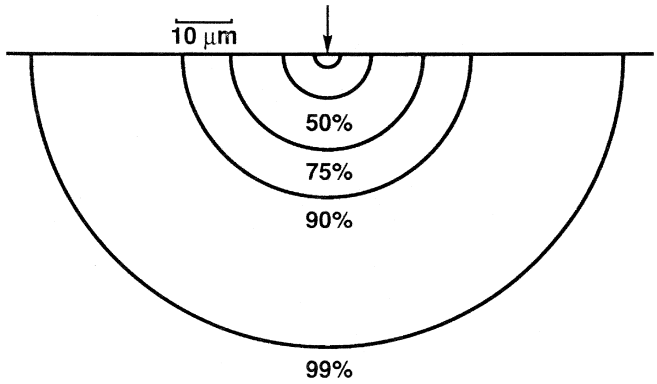


Figure 6.18. Range of the secondary fluorescence of Fe $K\alpha$ by Ni $K\alpha$ in an alloy of composition Ni–10% Fe, $E_0 = 20$ keV. The innermost semicircle shows the extent of direct electron-excited production of Ni $K\alpha$ and Fe $K\alpha$ by beam electrons. The 99% semicircle denotes locations deep in the specimen where some fluorescence of Fe by Ni $K\alpha$ is still possible.

Direct, electron-induced characteristic x-ray production is constrained to lie within the interaction volume of the electrons in the target. A good estimate of range of production is given by the x-ray range, Eq. (6.12). However, x-rays penetrate into matter much farther than electrons, so the range of x-ray-induced fluorescence is correspondingly much greater. Consider the case of a Ni–10% Fe alloy, where the Ni $K\alpha$ radiation can induce fluorescence of iron K radiation. The range of Ni $K\alpha$ in the alloy can be calculated with Eqs. (6.13) and (6.15) using the data in Table 6.4. Consider that the Ni $K\alpha$ radiation propagates uniformly in all directions from the source, which is the hemisphere containing the direct, electron-induced Ni $K\alpha$ production. Based on these calculations, the electron and x-ray ranges are compared in Fig. 6.18. The radius of the hemisphere that contains 90% of the Fe $K\alpha$ induced by Ni $K\alpha$ is approximately 10 times greater, and the volume approximately 1000 times greater, than the hemisphere that contains all of the direct electron-induced Fe $K\alpha$ production. The dimensions of the 50%, 75%, and 99% hemispheres are also shown in Fig. 6.18. Clearly, the x-ray-induced fluorescence from both characteristic and continuum contributions originates in a much larger volume than the electron-induced characteristic radiation. This has the effect of degrading the spatial resolution of x-ray microanalysis.

References

Anderson, C. A., and M. F. Hasler (1966). In *Proceedings of the 4th International Conference on X-ray Optics and Microanalysis* (R. Castaing, P. Deschamps, and J. Philibert, eds.), Hermann, Paris, p. 310.

Bearden, J. A. (1967a). *Rev. Mod. Phys*. **39,** 78.

Bearden, J. A. (1967b). X-ray Wavelengths and X-ray Atomic Energy Levels, NSRDS-NBS 14, National Bureau of Standards, Washington, D.C.

Bertin, E. P. (1975). *Principles and Practice of X-ray Spectrometric Analysis,* 2nd ed. Plenum Press, New York.

Bethe, H. (1930). *Ann. Phys.* (Leipzig) **5,** 325.

Castaing, R. (1951). Ph.D. Thesis, University of Paris, Paris.

Duane, W., and F. L. Hunt (1915). *Phys. Rev.* **6,** 166.

Green, M. (1963). In *Proceedings of the 3rd International Symposium on X-ray Optics and X-ray Microanalysis* (H. H. Pattee, V. E. Cosslett, and A. Engstrom, eds.), Academic Press, New York, p. 361.

Heinrich, K. F. J. (1966a). In *The Electron Microprobe* (T. D. McKinley, K. F. J. Heinrich, and D. B. Wittry, eds.), Wiley, New York, p. 296.

Heinrich, K. F. J. (1986). In *Proceedings of the 11th International Conference on X-ray Optics and Microanalysis* (J. D. Brown and R. H. Packwood, eds.), University of Western Ontario, London, Ontario, Canada, p. 67.

Henoc, J. (1968). In *Quantitative Electron Probe Microanalysis* (K. F. J. Heinrich, ed.), National Bureau of Standards Special Publication 298, U.S. Government Printing Office, Washington, D.C., p. 197.

Kramers, H. A. (1923). *Phil. Mag*. **46,** 836.

Lifshin, E., Ciccarelli, M. F., and Bolon, R. B. (1980). In *Proceedings of the 8th International Conference on X-ray Optics and Microanalysis* (D. R. Beaman, R. E. Ogilvie, and D. B. Wittry, eds.), Pendell, Midland, Michigan, p. 141.

Moseley, H. G. J. (1913). *Phil. Mag*. **26,** 1024.
Moseley, H. G. J. (1914). *Phil. Mag*. **27,** 703.
Powell, C. J. (1976a). *Rev. Mod. Phys.* **48,** 33.
Powell, C. J. (1976b). In *Use of Monte Carlo Calculations in Electron Probe Microanalysis and Scanning Electron Microscopy* (K. F. J. Heinrich, D. E. Newbury, and H. Yakowitz, eds.), National Bureau of Standards Special Publication 460, U.S. Government Printing Office, Washington, D.C., p. 97.
Powell, C. J. (1990). In *Microbeam Analysis—1990* (J. R. Michael and P. Ingram, eds.), San Francisco Press, San Francisco, p. 13.
Salem, S. I., and P. L. Lee (1976). *Atomic Data Nucl. Data Tables* **18,** 233.
Woldseth, R. (1973). *X-ray Energy Spectrometry*, Kevex Corp., Foster City,California.

7

X-Ray Spectral Measurement: EDS and WDS

7.1. Introduction

Chemical analysis in the scanning electron microscope and electron microprobe is performed by measuring the energy and intensity distribution of the x-ray signal generated by a focused electron beam. The subject of x-ray production has already been introduced in Chapter 6, which describes the mechanisms for both characteristic and continuum x-ray production. This chapter is concerned with the methods for detecting and measuring these x-rays as well as converting them into a useful form for qualitative and quantitative analysis.

7.2. Energy-Dispersive X-Ray Spectrometer

7.2.1. Operating Principles

Fitzgerald *et al.* (1968) first described the use of a lithium-drifted silicon Si(Li) solid state x-ray detector mounted on an electron probe microanalyzer. Although their system was barely capable of resolving adjacent elements, it did demonstrate the feasibility of interfacing the x-ray detector to an electron beam instrument. Over the years there has been a significant improvement in detector resolution from 500 eV to less than 130 eV, thereby making the technique much more suitable for microanalysis requirements. Today solid state detectors, both Si(Li) energy-dispersive spectrometers (EDS) and intrinsic or high-purity Ge (HPGe) EDS, are the principal methods for x-ray measurement for mid-energy x-ray spectroscopy (0.2–12 keV) for scanning electron microscopes. EDS is also often used in combination with wavelength-dispersive spectrometers (WDS) with electron microprobe analyzers, whereas HPGe systems are used extensively with transmission electron microscopes because of their better collection efficiency for high-energy x-rays. This chapter will describe the principles

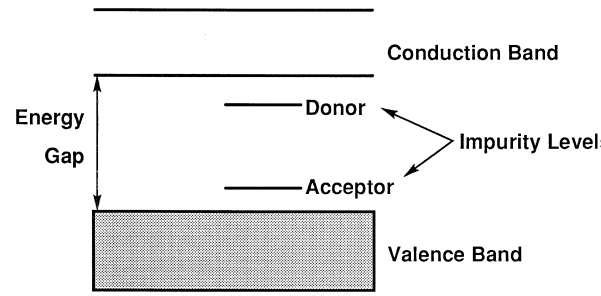


Figure 7.1. Schematic representation of energy levels in silicon.

of EDS operation primarily using examples with Si(Li) detectors, but the basic principles also apply to HPGe detectors.

In order to understand the operation of the Si(Li) detector it is first useful to consider a few basic points about semiconductors. Figure 7.1 is a simplified energy diagram of silicon. It illustrates the fact that in undoped silicon (which has four outer shell electrons) all of the valence band is full, whereas the next highest available energy levels in the conduction band are empty. The filled valence band reflects the fact that each silicon atom shares its outer electrons, one each, with its four nearest neighbors. Consequently, there are no excess electrons available for charge transport in the conduction band nor are there any holes in the valence band. Such a material is called an intrinsic semiconductor. It will not conduct current in an applied electric field unless it absorbs energy causing electrons to be promoted into the conduction band and leaving holes behind. For this reason intrinsic semiconductors make good radiation detectors.

Most electronic devices consist of junctions of doped *p*- and *n*-type silicon. Doping silicon with boron makes it a *p*-type semiconductor (holes are the majority carriers). Doping silicon with phosphorus makes it *n*-type (electrons are the majority carriers). Dopants create new energy levels in the band gap. Because a boron atom has three, rather than four, electrons in its outer shell, when boron is substituted for a silicon atom it will remove an electron from the valence band leaving a hole behind. On the other hand, phosphorus, with five electrons in its outer shell, can provide an excess electron into the conduction band when substituted for a silicon atom. In each case the net effect of having mobile charge carriers (excess holes or electrons) is to allow current to flow across such a device in the presence of an applied field. Therefore doped silicon is unsuitable for a detector because any change in conductivity with the absorption of radiation would be superimposed on an existing background level caused by the presence of mobile charge carriers.

In actual practice, sufficiently pure silicon to achieve the "intrinsic" condition is hard to obtain. Instead, detector crystals are made to behave like intrinsic silicon by a process developed by Pell (1960). Lithium (an *n*-type dopant) is applied to the surface of *p*-type silicon and caused to diffuse into the crystal forming a small *p–n* junction. Although the *p–n*

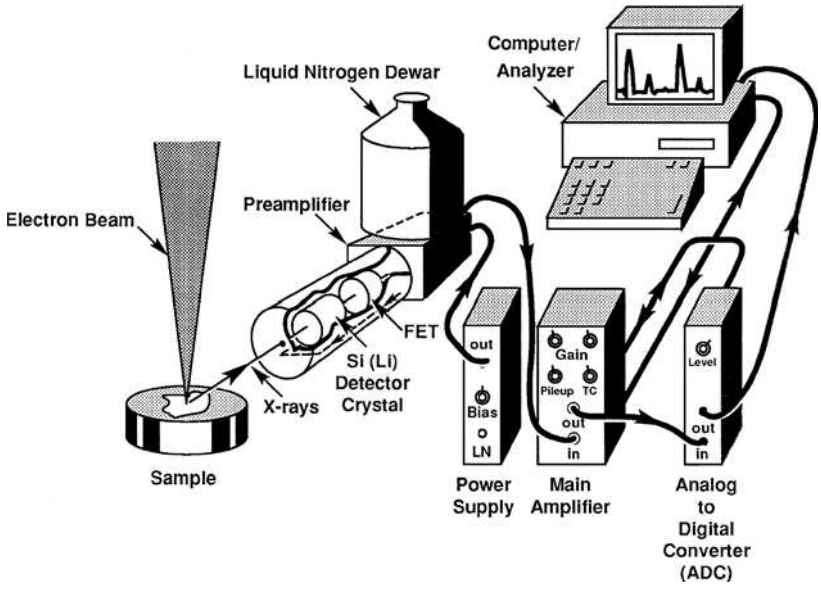


Figure 7.2. Schematic representation of an energy-dispersive spectrometer and associated electronics.

junction boundary is an intrinsic semiconductor, it is only a few micrometers thick. However, when an appropriate electric field (reverse bias) is applied to the *p*–*n* junction at elevated temperature, the intrinsic region can be enlarged to a few millimeters making it suitable for a detector material after most of the *p*-type material is removed. It is important to note that both steps are performed above room temperature, but even at room temperature the Li is mobile in the presence of an applied field. For this reason detectors can never be allowed to operate under bias except near liquid nitrogen temperatures.

The operating principles of a solid state detector system are illustrated in Fig. 7.2. X-ray photons from the sample pass through a thin window isolating the specimen chamber environment from the detector into a cooled, reverse-bias *p*–i–*n* (*p*-type, intrinsic, *n* type) Si(Li) crystal. Absorption of each individual x-ray photon leads to the ejection of a photoelectron, which gives up most of its energy to the formation of electron–hole pairs. They in turn are swept away by the applied bias to form a charge pulse, which is then converted into a voltage pulse by a charge-to-voltage converter (preamplifier). The signal is further amplified and shaped by a linear amplifier and finally passed to a computer x-ray analyzer (CXA), where the data are displayed as a histogram of intensity by voltage. The contents of the CXA memory in all recent instruments either reside directly in a computer or can be transmitted to a computer for further processing such as peak identification or quantification. The key to understanding how an EDS works is to recognize that each voltage pulse is proportional to the energy of the incoming x-ray photon. The role of the CXA is to establish this relationship and to present it in a form understandable to the operator.

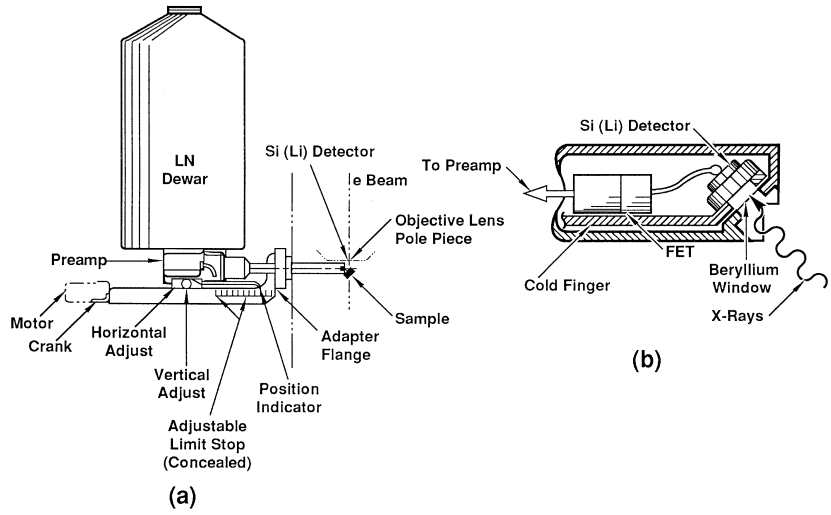


Figure 7.3. (a) Physical appearance of a retractable detector and associated preamplification electronics. (b) Detail of Si(Li) mounting assembly.

The typical physical appearance of a detector can be seen in more detail in Fig. 7.3. The lithium-drifted silicon crystal is mounted on a cold finger connected to a liquid nitrogen reservoir stored in the dewar. Because the detecting crystal is light-sensitive, it is essential to block visible radiation, preferably through the use of an opaque window. Windowless and ultrathin-window EDS can be used if the specimen chamber is light-tight and the specimen is not cathodoluminescent. Thin windows, especially atmospheric-pressure supporting thin windows that incorporate a grid, that can transmit low energy x-rays down to 200 eV now dominate the field. The detector chamber is also sealed under vacuum both to prevent contamination from the specimen region (especially when the specimen chamber is brought to air pressure) and to more easily maintain the low temperature essential for reducing noise. As mentioned previously, low temperature is also needed to limit the mobility of the lithium ions initially introduced in the silicon crystal to neutralize recombination centers. Again, under no conditions should the bias be applied to a non-cooled detector. Many systems, in fact, incorporate safety features to turn the bias off in the event of either the crystal warming up or a vacuum failure. Dewars presently available are capable of maintaining cryogenic conditions for several days without refilling. Detectors using Peltier cooling and other types of refrigeration rather than liquid nitrogen are also available. Note that the crystal and supporting cold finger are well separated from the housing assembly to prevent condensation on the latter and to provide electrical isolation. The feature of being able to mechanically move the detector housing in and out relative to the specimen without breaking the microscope vacuum is quite useful. As will be discussed later, the situation may arise where, for a fixed beam current, the x-ray signal has to be increased to obtain better counting statistics or decreased for better energy resolution. In many cases the desired count rate can be obtained by simply varying the crystal-to sample

distance, which changes the solid angle of detector acceptance, with the solid angle decreasing as the square of the specimen-to-detector distance. Some new instruments incorporate an adjustable diaphragm to limit the detector area at a fixed distance for this purpose.

The development of the EDS has made "easy" x-ray spectrometry available to virtually all types of electron beam instruments. However, one must recognize that, because of the nature of the technique, distortions are introduced into the ideal x-ray spectrum ("spectral artifacts") during the measurement process which must be dealt with in practical analytical spectrometry. In the discussion that follows, we will consider these artifacts at each stage of the detection and amplification process.

7.2.2. The Detection Process

The basic detection process by which this proportional conversion of photon energy into an electrical signal is accomplished is illustrated in Fig. 7.4. The active portion of the detector consists of intrinsic silicon, with a thin layer of *p*-type material on the front surface, called the "dead layer," coated with a thin gold electrical contact. When an energetic photon is captured, electrons are promoted into the conduction band, leaving holes in the valence band. Under an applied bias, these electrons and holes are swept apart and collected on the electrodes on the faces of the crystal. The process of x-ray capture is photoelectric absorption, and the x-ray photon is annihilated in the process (see Chapter 6). The incident x-ray photon with an energy $h\gamma$ is first absorbed by a silicon atom and an inner shell electron is ejected with an energy $h\gamma - E_c$, where E_c for silicon is 1.84 keV. This photoelectron then creates electron–hole pairs as it travels in the detector silicon and scatters inelastically. The silicon atom is left in an excited state because of the vacancy caused by the ejection of the photoelectron. As electrons from less tightly bound states fill the vacancy,

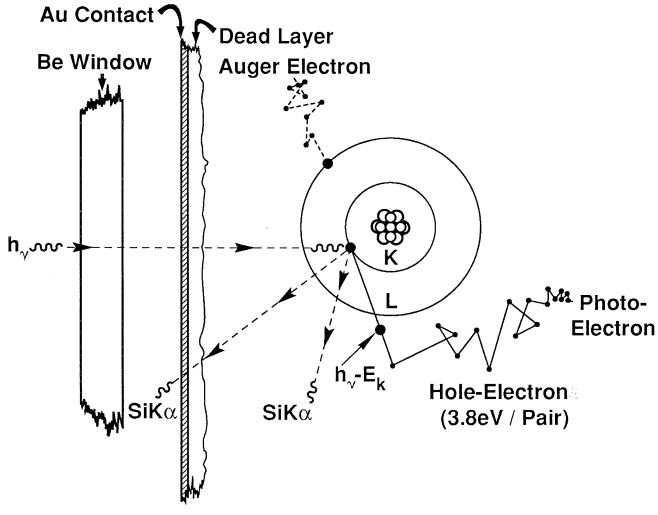


Figure 7.4. The x-ray detection process in the Si(Li) detector.

energy is subsequently released in the form of either an Auger electron or a silicon x-ray. The Auger electron scatters inelastically and also creates electron–hole pairs. The silicon x-ray can be reabsorbed, which initiates the process again, or it can be scattered inelastically. Thus, a sequence of events takes place leading to the deposition of all of the energy of the original photon in the detector, unless radiation generated during the sequence, such as a silicon $K\alpha$ photon, escapes the detector giving rise to the artifact known as the "escape peak," which will be discussed later.

The ideal number of charges n created in the detector per incident photon with energy E (eV) is given by

$$n = E/\varepsilon, \tag{7.1}$$

where $\varepsilon = 3.86$ eV for Si at 77 K ($\varepsilon = 2.96$ eV for Ge). For example, if the detector captures one photon having an energy of 5 keV, then from Eq. (7.1) the total number of electrons swept from the detector is approximately 1300, which represents a charge of 2×10^{-16} C. This is an extraordinarily small charge. To measure charge accurately, noise minimization is essential, hence the need for keeping the detector crystal close to liquid nitrogen temperature to reduce thermal noise contributions.

7.2.3. Charge-to-Voltage Conversion

Figure 7.5a is a representation of the detector, charge-to-voltage converter (often referred to as a preamplifier), and pulse-shaping linear amplifier from an electronic perspective. Once created, the charge from the

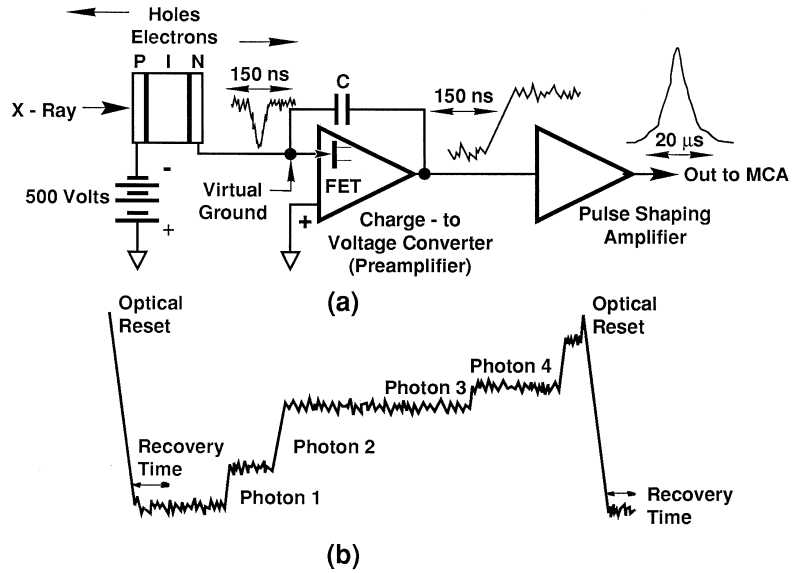


Figure 7.5. Charge-to-voltage conversion process. (a) Representation of the detector charge-to-voltage converter and pulse-shaping amplifier from an electronics perspective. (b) Output of the charge-to-voltage converter after the detection of a series of x-ray photons.

detector is rapidly swept into a virtual ground at the input of an operational amplifier, where it is deposited onto the capacitor, C. The first stage of the preamplifier is a carefully selected low-noise field-effect transistor (FET), which is also cooled and placed immediately behind the detector (as shown in Fig. 7.3b). The preamplifier output for each event is a voltage step proportional to the incident photon energy. The height of each step is on the order of a few millivolts and occurs over a time period of about 150 ns. It is important to recognize that the effect of electronic noise is the introduction of some fluctuation in the output voltage level on both sides of each step. This will be discussed in further detail in the section on factors determining energy resolution. Figure 7.5b shows the output of the charge-to-voltage converter after the detection of a series of x-ray photons. The difference in the step sizes indicates that they have different energies. An increase in the overall voltage level is also observed, and a point could eventually be reached when the amplifier would be saturated. To prevent this from happening it is necessary to discharge the capacitor C to return the operational amplifier output voltage back to zero. Note that only four photons are shown in the figure; however, in an actual system several hundred may be processed before reseting. One method to accomplish this is to employ a light-emitting diode which when turned on will cause a leakage current to flow in the FET. This technique is called pulsed optical feedback (POF). Because considerable noise is generated when the POF is turned on, it is necessary to gate the main amplifier for a period of about 1 ms. POF preamplifiers are now in common use. In other systems, special five-electrode FETs obviate the use of POF. It should be noted, however, that the present state of the art of detector manufacture is very near the theoretical resolution limit for silicon (about 120 eV at 5.6 keV).

7.2.4. Pulse-Shaping Linear Amplifier and Pileup Rejection Circuitry

It is the principal function of the pulse-shaping linear amplifier to transform each individual voltage step from the charge-to-voltage converter into a voltage pulse suitable for presentation to the CXA. As shown in Fig. 7.5a the linear amplifier output pulse starts at zero and rises to a maximum value of typically several volts and then returns to zero. The pulse height retains proportionality to the energy of the originating x-ray photon. The overall measurement process is particularly challenging because, as mentioned previously, the steps are on the order of millivolts and the voltage level on which they occur is noisy. This noise depends not only on the quality of the detector crystal, its environment, and associated preamplifier electronics, but also on the operating characteristics of the linear amplifier. In a Si(Li) detector system noise reduction is particularly critical because, as will be shown later, unlike the WDS, all of the spectral dispersion is done electronically.

In order for the linear amplifier to convert a step to a pulse requires, conceptually, averaging the noise on either side of the step. The noise intensity as a function of frequency is not uniform, but has a distinct minimum

at a particular frequency. The pulse-shaping circuitry is designed in such a way as to take maximum advantage of the noise minimum. [For more information on this topic see Goulding (1972) or Reed (1975)]. Up to a point, increasing the time that the noise level is analyzed improves the accuracy of the pulse height as a measure of the energy of the originating x-ray photon. In order to accomplish this, however, voltage step measuring times have to be extended to several or even tens of microseconds. This can cause a problem, called "pulse pileup," if the number of photons entering the detector exceeds a few thousand per second. Pulse pileup occurs if a photon arrives at the detector before the linear amplifier is finished processing the preceding photon. It appears as an increased output pulse height for the second photon because it is riding on the tail of the first, as shown in Fig. 7.6 for the 6- and 10-μs cases. Pulse pileup can also appear as a single large pulse representing the combined voltages of two pulses if the second photon arrives before the pulse from the first has reached its maximum value. In the most extreme case, two photons arrive at the detector almost simultaneously, and the output is a single combined pulse corresponding to the sum of the two photon energies. This phenomenon gives rise to what are called "sum peaks" (or "coincidence peaks") such as the $2K\alpha$ and the $K\alpha + K\beta$ peaks.

Pulse pileup can be reduced by decreasing the processing time for each photon because shorter linear amplifier output pulses are less likely to interfere with each other. This is accomplished electronically through user selection of the linear amplifier time constant (T.C.), which determines the output pulse width (sometimes designated as fast, medium, or slow). It should be recognized, however, that the T.C. is only an indication of pulse width because the acceptable time between pulses may be many times the value of the T.C. This point is illustrated in Fig. 7.6. It can be seen that pulse II arriving 20 μs after pulse I will correctly be evaluated as 4 V for the 1-μs T.C., but as 4.5 V for the 6 μs T.C. and 6.5 V for the l0-μs T.C.

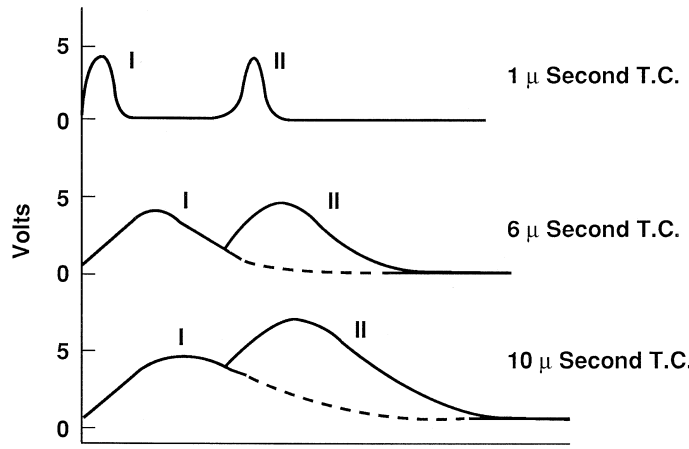


Figure 7.6. EDS main amplifier pulse shapes for different time constants (T.C.).

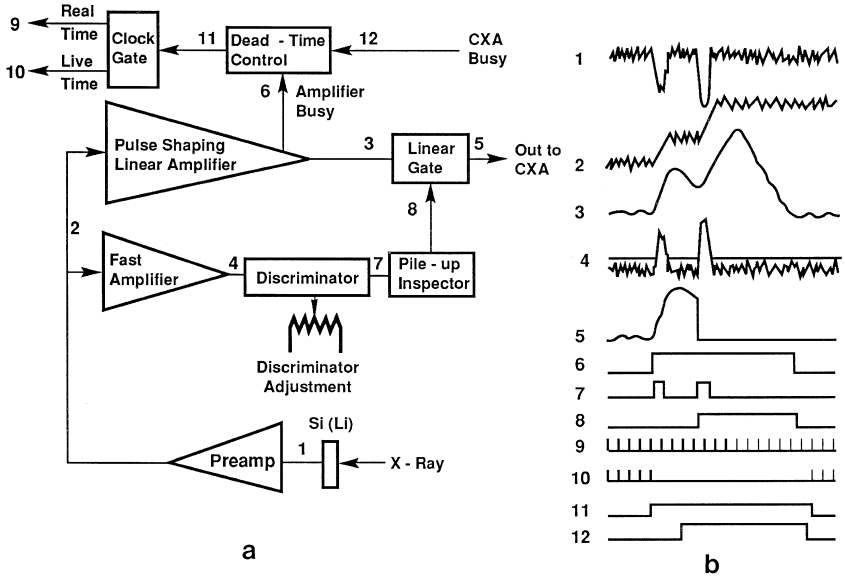


Figure 7.7. (a) Schematic representation of the principal components used with pulse-pileup rejection. (b) Waveforms at various key locations.

Pulse pileup situations will lead to the incorrect photon energy assignment because of inaccurate pulse height measurement.

The electronic elimination of such events is accomplished by means of a pulse-pileup rejection circuit of the type illustrated in Fig. 7.7a, which shows both the principal electronic components and signals at key points (Fig. 7.7b). A second amplifier, called a "fast amplifier," is placed in parallel with the slower pulse-shaping linear amplifier. It has a much shorter T.C., and thus can give rise to narrower output pulses at the expense of pulse height precision. This is acceptable because the fast amplifier is only used to sense the presence of each pulse arriving at the pulse-shaping linear amplifier (point 2) even if this amplifier is busy processing a previous pulse. If the fast amplifier pulse exceeds a predetermined discriminator setting (point 4), then a signal is sent to the pulse-pileup inspector which can prevent the output of the main amplifier from going to the CXA. The decision can be made to block both pulses if the second one arrives before the first one reaches its peak value or to only block the second one if the first has peaked and been processed by the CXA, but the signal has not returned to the baseline.

Figure 7.8 shows a comparison between two iron spectra taken with and without pulse-pileup rejection. Pulse pileup causes distortion in the background at energies above the Fe $K\alpha$ peak. Note the presence of the Fe $2K\alpha$ and Fe $K\alpha$ + Fe $K\beta$ peaks in both spectra. These peaks are due to the simultaneous arrival of two Fe $K\alpha$ photons or an Fe $K\alpha$ and an Fe $K\beta$ photon in the detector crystal creating electron–hole pairs in numbers corresponding to a single photon with the sum energy. They cannot be

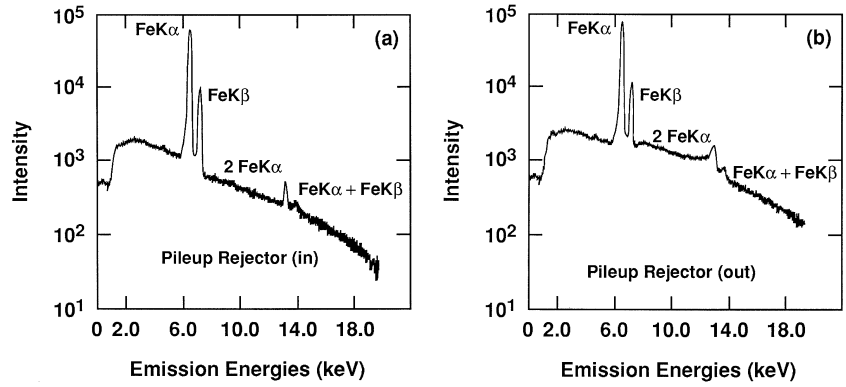


Figure 7.8. The use of pileup rejection to improve an iron spectrum. (a) Pileup rejector in. (b) Pileup rejector out; note more pronounced pileup peaks. (From Lifshin and Ciccarelli, 1973.)

eliminated by the pulse-pileup rejector, but the fraction of such peaks relative to the main characteristic peak can be reduced to a very low level by keeping the total count rate below a few thousand counts per second when using a longer T.C. When a quantitative analysis is performed the number of $K\alpha + K\beta$ and twice the number of $2K\alpha$ pulses should be added to the $K\alpha$ peak intensity to correct for their incorrect energy assignment.

Because of the shorter T.C. of the fast amplifier, its output pulses ride on a higher noise level than those from the pulse-shaping linear amplifier (see Fig. 7.7). Proper setting of the discriminator is crucial because, if it is too low, noise will be accepted as pulses, causing unnecessary pulse rejection; however, if it is too high, low-energy pulses might be missed. Pileup rejection is therefore more complicated for low-energy x-rays, which may be hard to separate from the background noise. Even in an optimally adjusted discriminator, there is an energy below which the desired pulses are so close to the noise that the discrimination becomes inadequate. For an example based on magnesium and silicon pileup see Enhancement 7.1 in the accompanying CD.

Operation of the pulse-pileup rejector results in a period of time during a given measurement when pulses are not measured, called "dead time." Dead time is often expressed as a percentage of "real" or "clock time." Conversely, the time during which pulses are measured is referred to as "live time." It is often necessary to compare spectral information taken under identical operating conditions, including live times. In a quantitative analysis, for example, live time must be used when ratios of x-ray intensities between samples and standards, taken under identical operating conditions, are required as inputs to the quantitative correction model. In most systems available today the operator sets the desired live time, and the counting interval is automatically extended to compensate for the dead time. Figure 7.7a shows how this is done. The evenly spaced clock pulses at point 9 (Fig. 7.7b) corresponding to actual elapsed time (real time) are passed through a clock gate and then are output (point 10) to a counter (not shown).

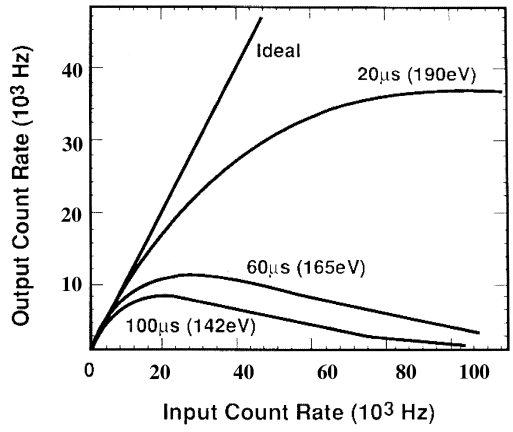


Figure 7.9. Relationship of the output and input count rates for three different values of pulse width and resolution.

Data collection is terminated when a preset number of clock pulses is collected. The decision to temporarily block the flow of clock pulses is determined by the dead-time control circuit inhibit pulse (point 11), activated by a combination of the amplifier busy (point 6) and CXA busy signals (as well as POF busy, also not shown). In the example shown the clock gate has only allowed 8 of 21 clock pulses to be counted, indicating significant dead time (compare the pulse trains at points 9 and 10).

The effect of this loss of pulses is illustrated in Fig. 7.9, which shows an example of the relationship between the linear amplifier output count rate (CXA input count rate) and the linear amplifier input count rate for different T.C.'s for a conventional analog electronic system. It can be seen that only at low count rates (less than a few thousand counts per second) are the two count rates equal. For reasons already described, as the amplifier input count rate increases, pulse-pileup effects become more serious, particularly with the large amplifier time constants. In a quantitative analysis it may therefore be necessary to count for a longer period of time than anticipated on the basis of real time to achieve a desired level of precision based on counting statistics. As shown in Fig. 7.9, increasing the amplifier input count rate by changing the probe current or moving the detector closer to the sample will initially result in a linear increase in CXA input count rate. However, this will be followed by a nonlinear region in which the rate of increase in CXA input is less than the rate of increase of linear amplifier input pulses. Eventually a point of diminishing returns is reached when increasing the main amplifier input count rate actually results in a decrease in the count rate seen by the CXA. Further increases beyond this point lead to essentially 100% dead time and consequently a total locking-up of the system. Figure 7.9 also shows that the onset of the various regions described is determined by a choice of operating curves based on acceptable resolution criteria. It is important to note that Fig. 7.9 is not a universal curve and similar curves must be established for each system.

The maximum input count rate is not simply that of the element of interest, but rather that for the total of all pulses measured across the full range of the spectrum. Therefore, system count rate performance is determined by the major elements and not by the specific minor constituents which may be under investigation. Clearly, achieving maximum count rate is critical to optimizing precision. Because pulse pileup is the principal limiting factor, research continues on how to increase throughput without paying a major premium in resolution. One approach is to truncate the linear amplifier pulses immediately after reaching their maximum value so that the likelihood of one pulse riding on the tail of the next is reduced. This approach, developed by Kandiah (1971), has been implemented in many systems.

7.2.5. The Computer X-Ray Analyzer

The previous section described how the output of the pulse-shaping amplifier must be a series of well-separated pulses whose magnitudes remain proportional to the energy of each corresponding photon. The role of the CXA is to present this information in a form understandable to the operator. In earlier times, the CXA was referred to as a multichannel analyzer or MCA, a separate piece of hardware. The evolution of computer-based systems has now integrated most functions into software commands and computer displays. In recognition of this change, the term, "computer x-ray analyzer," or CXA, is used in this book. The first step is to accurately determine the peak height of each pulse and to convert it into a digital value. This is done by an analog-to-digital converter (ADC) (ADC operation is described in Enhancement 7.2 in the accompanying CD).

In addition to an ADC, a computer x-ray analyzer system includes an acquisition processor, a memory, operator controls including key pads, and analog input devices. It may also include peripherals such as data and program storage devices, printers, and network connectivity. Since many of these capabilities are available in most personal computers and workstations, the latter are often modified to meet the needs of energy-dispersive spectrometry. This evolution has been taking place over several decades and represents a significant departure from the stand-alone hardware MCAs first used.

Figure 7.10a shows a simplified picture of the sorting process. For illustration purposes, 10 locations in a computer memory have been assigned to store the numbers of pulses corresponding to specific voltage sizes, for example, address 1001 contains the count of pulses with energies between 0 and 1 V, address 1002 contains the count of pulses with energies between 1 and 2 V, etc. Assume at the beginning of data collection that all of the memory contents are 0. The first pulse from the main amplifier is 5.2 V high, and the ADC sends a digital signal to the acquisition processor that contains this value. The acquisition processor then causes 1 to be added to the contents of address 1006 because that is where pulses between 5 and 6 V are counted. The display is a CRT, which provides a convenient way

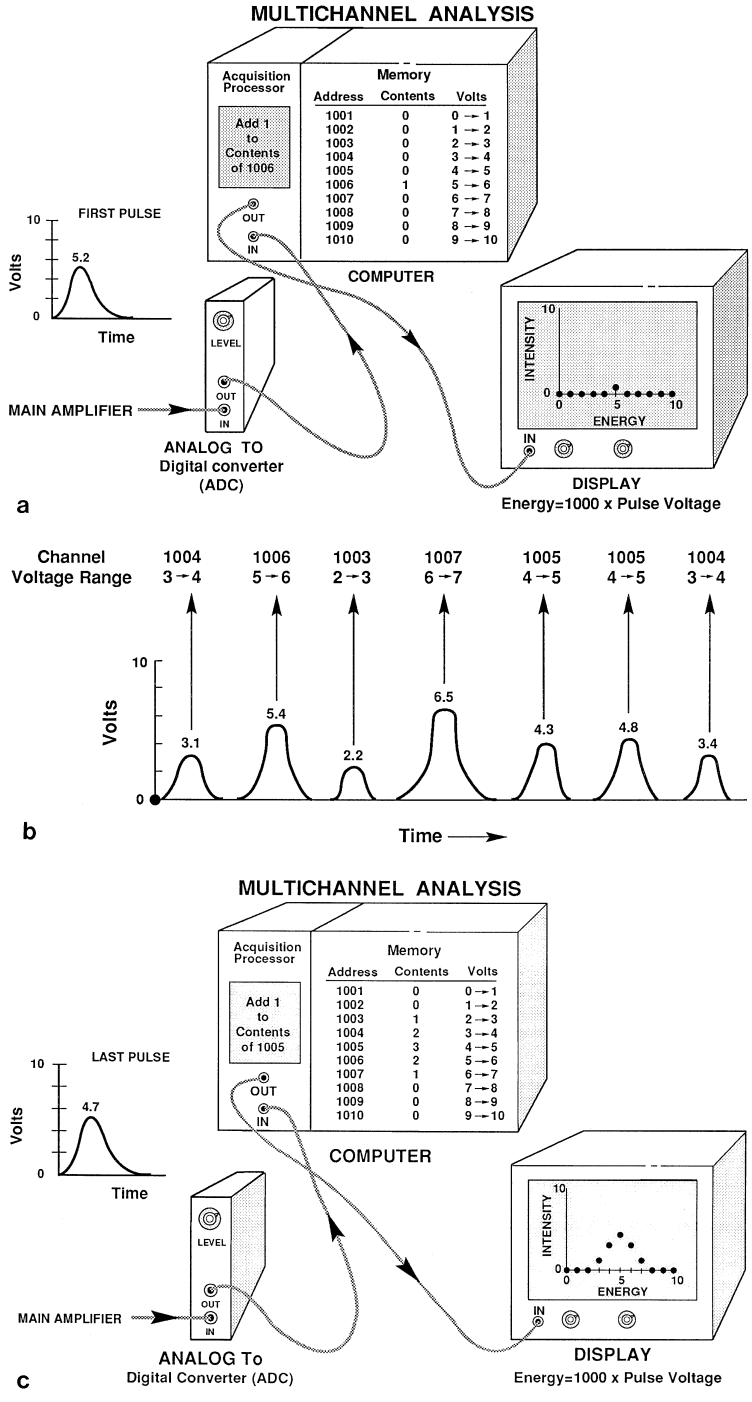


Figure 7.10. Schematic representation of a computer-based x-ray analyzer. (a) Components of the pulse processor. (b) Pulse sequences and channel destinations of the next seven pulses. (c) Final display of processed pulses.

for the operator to follow what is happening. The horizontal sweep occurs rapidly from 0 to 10 V full scale. It would appear as just a bright line, but is electronically chopped into a series of dots corresponding to integer voltage values. The vertical deflection of each dot is then modulated to reflect the number of counts in each memory address, which initially was zero, but now shows one count corresponding to 5.2 V.

Figure 7.10b shows a sequence of seven more pulses of different sizes and the ADC and acquisition processor's assignment of these pulses to the computer memory. Figure 7.10c shows a ninth pulse (the last before the data acquisition time runs out) and also a listing of the contents of the computer memory and what the contents of that memory looks like on the display. In this simple example something looking like a peak is beginning to emerge. In actual systems the total number of memory locations involved may be 1024, 2048, or higher and the number of counts stored in each location may be in the tens of thousands or even considerably higher.

As described thus far, pulses have been sorted and displayed by voltage, which, if everything has been done right, should be proportional to the energy of each corresponding x-ray photon in accordance with Eq. (7.1). Calibration is required to make this happen on the display of the CXA. Whereas the exact procedure may vary between systems, the operator first determines the energy range of interest, for example, 0–10 keV. An arbitrary decision is made that 5-V pulses correspond to 5 keV x-rays (as shown in Fig. 7.10c). At this point it would be totally fortuitous to have 5 keV x-rays produce 5 V pulses. However, if a source is available emitting x-rays of known energies, (e.g., Cu $K\alpha$ at 8.04 keV) then the linear amplifier gain can be adjusted until Cu $K\alpha$ will produce pulses that are 8.04 V high. The process could stop here, except that there is the possibility that Cu pulses are not referenced to 0 V. To make this correction requires adjusting the amplifier zero level and now there are two variables, gain and zero, which requires a second peak at low energy (such as Al $K\alpha$) to uniquely determine their values. This procedure requires an interactive process adjusting the gain and zero level until both peaks appear at the proper positions. Some systems can now do this process in a totally automated mode.

Currently available EDS systems are extremely flexible. As shown in Fig. 7.11, peak energies can easily be determined by a dial or mouse-controllable feature known as a "cursor." At the same time its position, in energy, as well as the corresponding number of counts can be read directly from a numerical display. Scale information and labels are also displayed directly on the CRT. Other features include regions of interest (ROI) displays, line markers, and automatic peak identification routines. The ROI approach allows the user, often with the help of the cursor, to define a series of energy bands in which a running count of the integrated intensities is recorded. In this manner integrated peak intensities can be monitored either as a prerequisite for quantitative analysis or pulses corresponding to specific elements can be coupled to SEM displays for line scans or x-ray maps. Because all CXAs are personal computer-based, a wide variety of high quality printers can be used to display output. Most CXA software

packages also use a Windows type of environment that makes it easy to cut and paste images and graphics into numerous types of qualitative and quantitative reports.

7.2.6. Digital Pulse Processing

One of the most significant changes in EDS in recent years has been the introduction of digital pulse processing made possible as a result of currently available high-speed circuitry. The approach involves taking the output stream from the charge-to-voltage converter and directly digitizing it rather than using analog pulse processing followed by an analog-to-digital conversion (Mott and Friel, 1995). This concept is illustrated in Fig. 7.12. Once the signal is in a digital form a variety of pulse processing techniques can be employed which both reduce noise and increase count rate performance. One method, for example, is adaptive pulse processing, in which the time between pulses is continuously monitored such that the time constant can vary for each pulse pair rather than use a single time constant. Other approaches use a fixed time constant, but take advantage of the

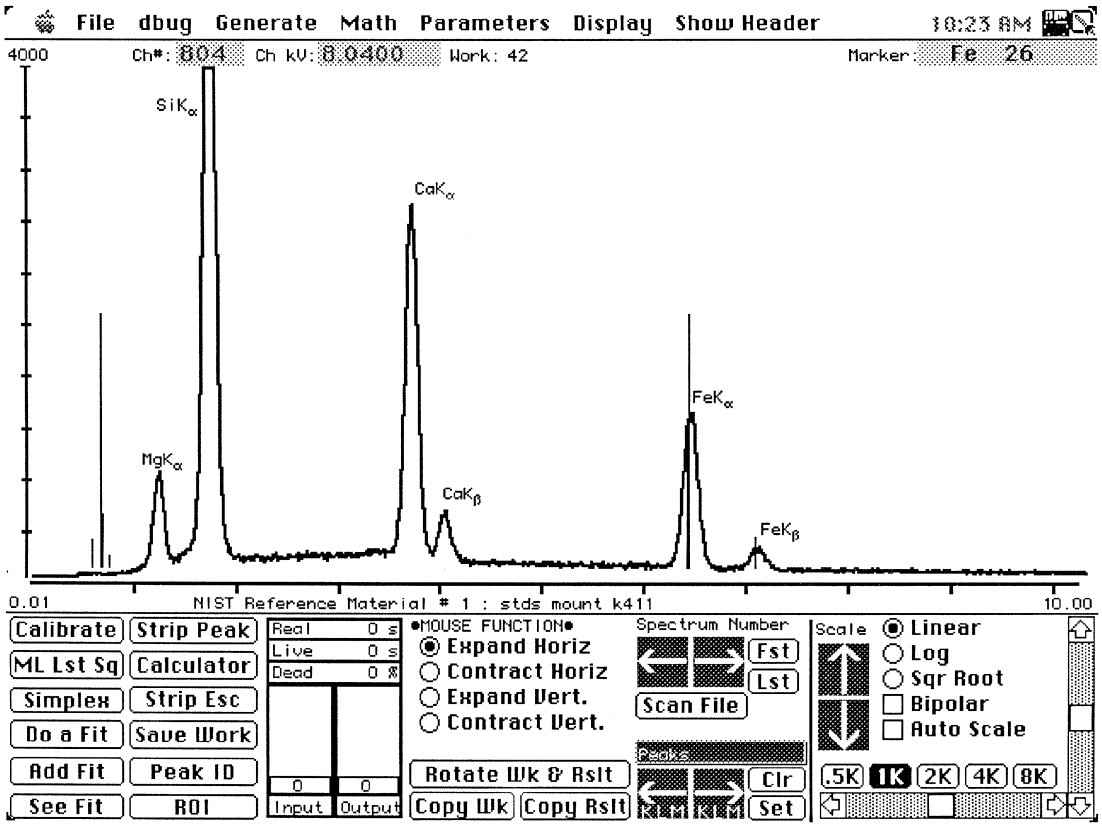


Figure 7.11. Features of an EDS CRT display illustrating the use of line markers and other capabilities. (Courtesy of NIST.)

high speed and flexibility of digital computation, which emulate or improve upon the analog pulse processing schemes that have been traditionally used to optimize signal-to-noise ratios and obtain high-count-rate performance. Figure 7.13 is an example of the improved count rate performance that is possible with digital pulse processing. It should also be pointed out that because the output pulse of a signal processor is already digital, a numerical value of the pulse height can be fed directly to the CXA because the analog-to-digital conversion has already been done at an earlier stage of the process.

7.2.7. Spectral Modification Resulting from the Detection Process

Deviations from the ideal detection process result in the appearance of spectral modifications, principally peak broadening, peak distortion, silicon x-ray escape peaks, sum peaks, silicon and gold absorption edges, and the silicon internal fluorescence peak.

7.2.7.1. Peak Broadening

The natural width of an x-ray peak is energy-dependent and is on the order of 2–10 eV measured at half the maximum of the peak intensity, full width at half maximum intensity, or "FWHM." The use of the FWHM

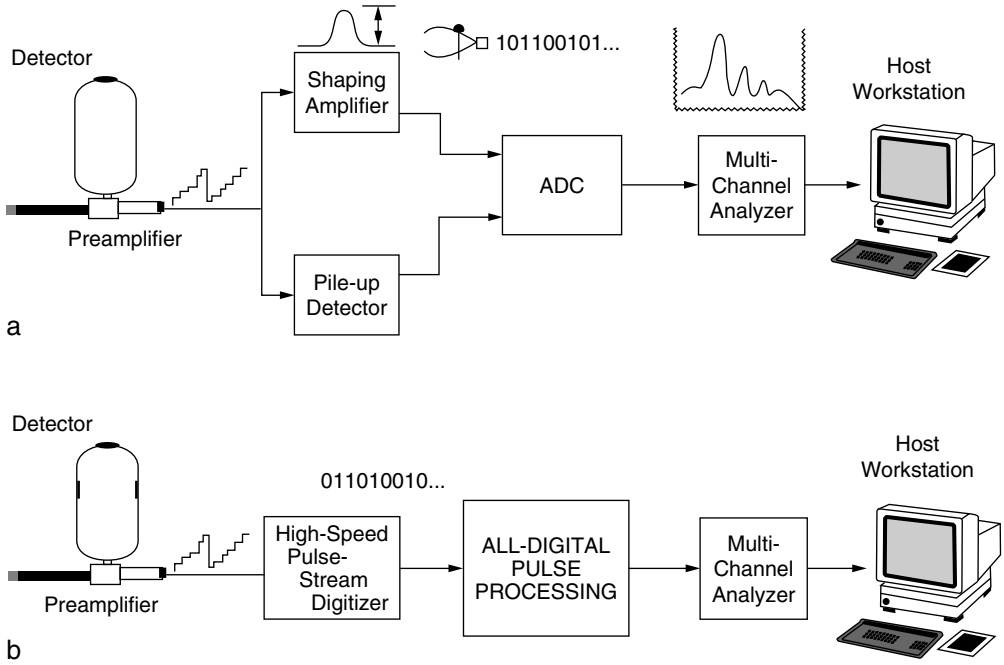


Figure 7.12. Comparison of digital and analog pulse processing. (From Mott and Friel, 1995; courtesy of Princeton Gamma Tech.)

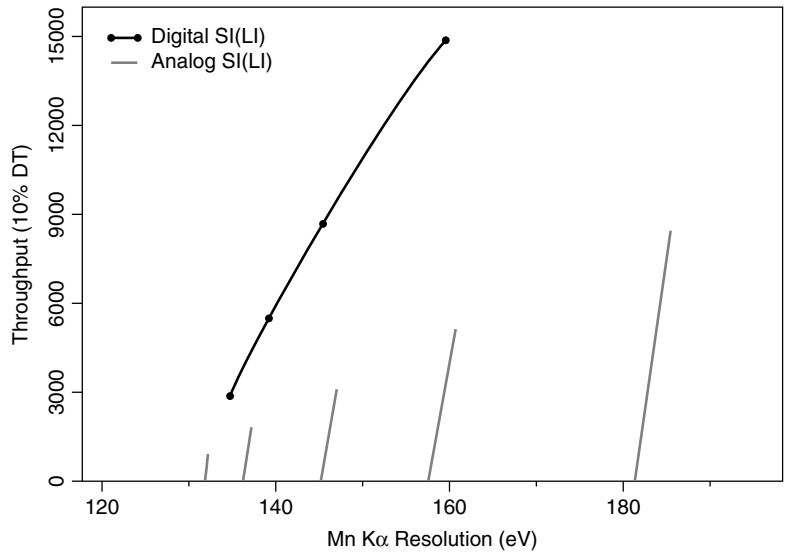


Figure 7.13. Improved count rate performance with digital pulse processing.

as a measure of peak width is traditional in spectroscopy because it is easily measured. The concept of a natural linewidth arises from the fact that there is a statistical distribution of x-ray photon energies associated with nominally identical transitions from an ensemble of atoms of the same type. This variation is a direct consequence of the Heisenberg uncertainty principle. For example, for manganese $K\alpha_1$ radiation (5.898 keV), the natural FWHM is approximately 2.3 eV, which makes the natural width about 0.039% of the peak energy. The measured peak width from the Si(Li) spectrometer is degraded by the convolution of the detector system response function with the natural linewidth to values of about 130 eV for Mn $K\alpha_{1,2}$, or 2.5% of the peak energy. The significant broadening of observed EDS peaks relative to their natural linewidths is the result of two factors. First, when a sequence of monoenergetic photons enters a detector, each photon does not necessarily create exactly the same number of electron–hole pairs as given in Eq. (7.1). Instead, there is a statistical distribution in the final number of charge carriers created and the value "3.8 eV per electron hole pair" only represents the average. Second, an uncertainty is introduced by the thermal noise of the amplification process, primarily from the FET. The distribution of numbers of charge carriers for a single photon energy Y is reasonably well described by a Gaussian distribution for x-ray peaks above 1 keV as shown schematically in Fig. 7.14 and in the following expression:

$$Y = A_{\mathrm{A}} \exp\left[-\frac{1}{2}\left(\frac{E_{\mathrm{A}} - E}{\sigma}\right)^2\right], \tag{7.2}$$

where A_{A} is the maximum x-ray intensity, E_{A} is the average peak energy, and E is the x-ray energy. The relationship between the standard deviation

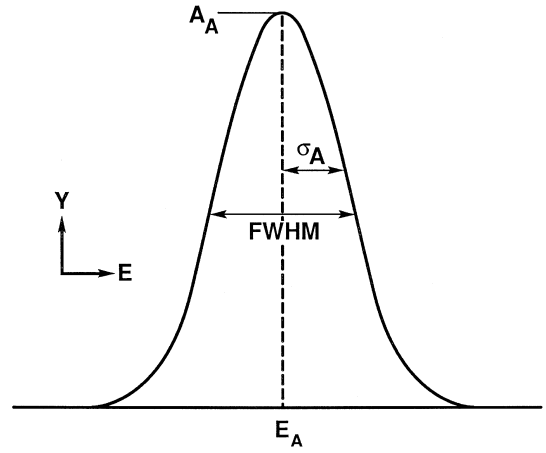


Figure 7.14. Gaussian peak shape representation for x-ray peaks above 1 keV.

σ used in statistics as a measure of "broadening" and "FWHM" used in spectroscopy is FWHM $= 2.355\sigma$.

The FWHM of this distribution can be calculated from the two sources of noise by quadrature addition (see Chapter 2 for an explanation of quadrature) according to the equation

$$\text{FWHM} \propto (C^2E + N^2)^{1/2}, \tag{7.3}$$

where C is the uncertainty in the formation of charge carriers, E is the x-ray energy, and N is the FWHM of the electronic noise of the amplification

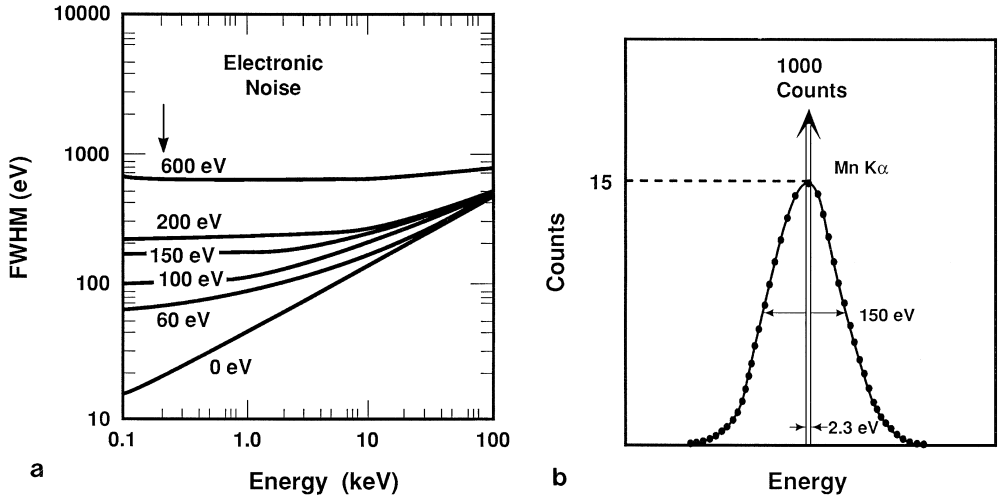


Figure 7.15. (a) Silicon energy resolution including intrinsic and electronic noise effects (adapted from Woldseth, 1973). (b) Redistribution of peak counts for the Mn $K\alpha$ peak with 150-eV resolution.

process. C is given by the expression

$$C = 2.35(F\varepsilon)^{1/2}, \tag{7.4}$$

where F is a constant known as the Fano factor (about 0.1 for Si) and $\varepsilon = 3.8$ eV for Si. Figure 7.15a shows values calculated by Woldseth (1973) of the observed FWHM as a function of energy for different contributions of electronic noise. It can readily be seen that even if the noise contribution were totally eliminated, the theoretical energy resolution limit would still be greater than 100 eV for Fe $K\alpha$ at 6.4 keV.

Fiori and Newbury (1978) derived an equation which takes advantage of the fact that the noise term is a constant for a given detector operated under a fixed set of conditions. Consequently, Eq. (7.3) can be restated in a form which is useful for comparing the width of a peak at an energy of interest to the width of a peak located at any other place in a spectrum:

$$\text{FWHM} = \left[2.5\,(E - E_{\text{ref}}) + \text{FWHM}_{\text{ref}}^2\right]^{1/2}. \tag{7.5}$$

All units in this expression are in electron volts.

The immediate consequence of the peak broadening associated with the detection process is a reduction in the height of the peak (counts per energy interval) as compared to the natural peak and an accompanying decrease in the peak-to-background ratio as measured at a given energy. Figure 7.15b shows the effect of detector broadening on the natural peak width of Mn $K\alpha$. In this example, a natural peak of width 2.3 eV and 1000 counts high is broadened into a peak of 150 eV width and 15 counts high.

A related effect is shown in Fig. 7.16, which is a computer simulation of a series of peaks containing equal numbers of counts measured at different energies. In this case the variation of FWHM described by Eq. (7.5) results in a variation of peak heights. This is an additional factor, like overvoltage effects described by Eq. (6.5), which points out the potential danger of estimating relative elemental concentrations by comparing peak heights between elements.

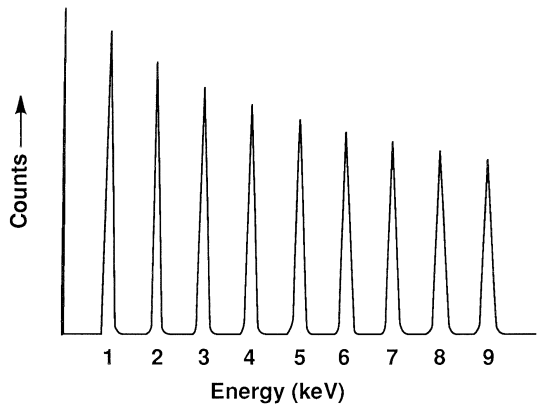


Figure 7.16. Resolution and peak intensity as a function of energy; Mn $K\alpha$ has a resolution of 150 eV.)

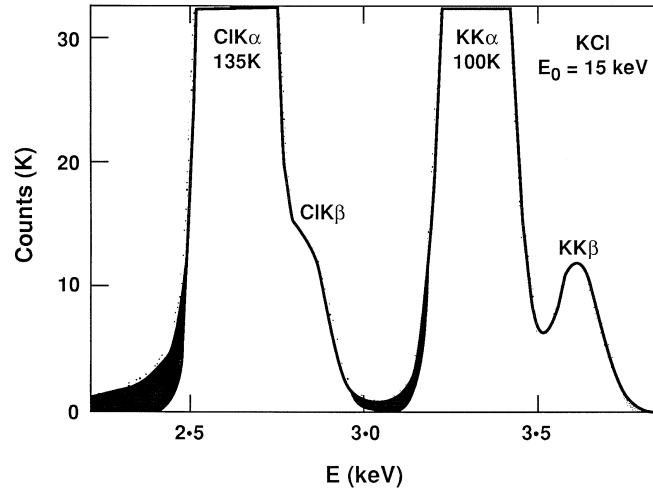


Figure 7.17. EDS spectrum of KCl illustrating peak overlap; K $K\alpha$ and $K\beta$ are nearly resolved. The solid line is a Gaussian fit to data points. The shaded area represents the deviation caused by incomplete charge collection.

The value of FWHM is useful in estimating the extent of the overlap of peaks which are close in energy. An estimate of the extent of overlap is vital when considering peak interferences in qualitative analysis, where the identification of a low-intensity peak near a high-intensity peak may be difficult (Chapter 8), and in quantitative analysis, where the removal of the interference is necessary for accurate determination of the composition (Chapter 9).

Peak overlap can also be a problem even for peaks in the same family for a given element. An example is given in the case of KCl in Fig. 7.17. With a detector of 170 eV resolution (Mn $K\alpha$), the potassium $K\alpha$ and $K\beta$ are nearly resolved, whereas the chlorine $K\alpha$ and $K\beta$ peaks are not.

7.2.7.2. Peak Distortion

Two different artifacts cause distortion; that is, deviation from a Gaussian shape on the low-energy side of a peak:

1. The collection of charge carriers created in certain regions of the detector near the faces and sides is imperfect due to trapping and recombination of the electron–hole pairs, leading to a reduction in the value of n predicted by Eq. (7.1) for the incident photon. The resulting distortion of the low-energy side of the peak is known as "incomplete charge collection" (Freund *et al.,* 1972; Elad *et al.,* 1973) and its effect is illustrated in Fig. 7.17 for the chlorine $K\alpha$ and potassium $K\alpha$ peaks. The deviation from a Gaussian distribution (shown as a solid line) is a function of energy. For example, the magnitude of the effect is significantly different for chlorine and potassium, which are separated by an atomic number difference of only 2.

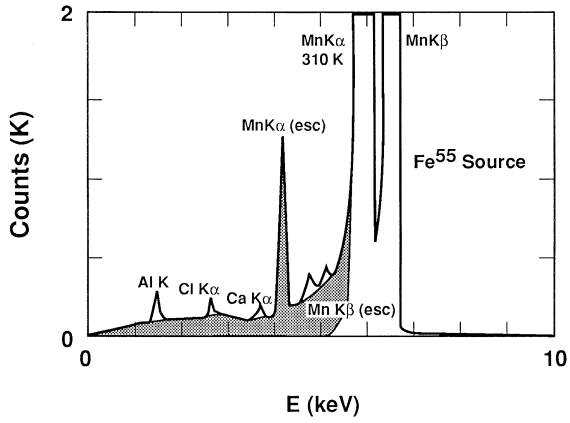


Figure 7.18. EDS spectrum derived from ^{55}Fe radioactive source (emits Mn K radiation). The background shelf is easily observable at energies below Mn $K\alpha$ down to the threshold at 300 eV. The Mn $K\alpha$ and $K\beta$ silicon escape peaks are noted. Extraneous characteristic peaks from the source holder are observed.

2. The background shelf, Fig. 7.18, is a phenomenon in which the presence of a peak increases the background at all energies below the peak value. This effect is more easily seen by observing the spectrum obtained from the radioactive source, ^{55}Fe, which emits Mn K lines. This source is chosen because it is readily available and is relatively background-free. Additional counts above the expected background result both from the incomplete charge collection phenomenon extending to low energy and from the escape from the detector of some of the continuum x-rays generated by the photoelectron as it scatters inelastically in the silicon. Any radiation lost from the detector reduces the number of charge carriers created. Both the incomplete charge collection and the loss of continuum x-radiation lead to a transfer of counts from the peak to the entire energy region down to zero. Typically, the background shelf at one-half the energy of a peak has a relative intensity of about 0.1% of the parent peak. The total number of counts lost to the full peak due to this effect is approximately l%. No Si(Li) detector, therefore, is 100% efficient. Although efficiency values exceeding 99% are common, poorly functioning detectors can have a 20% loss for peaks below a few keV, which can result in problems for both qualitative and quantitative analysis.

7.2.7.3. Silicon X-Ray Escape Peaks

Each x-ray photon detected produces a photoelectron and also leaves a silicon atom in the detector in an ionized state. The energy of the photoelectron is the difference between the energy of the x-ray photon and the bonding energy of the photoelectron. In the case of K-shell ionization, relaxation of an L-shell electron results in either the emission of a Si K-series x-ray photon or an Auger electron as shown in Fig. 7.4. The range of the Auger electron is only a fraction of 1 μm, and hence it is highly

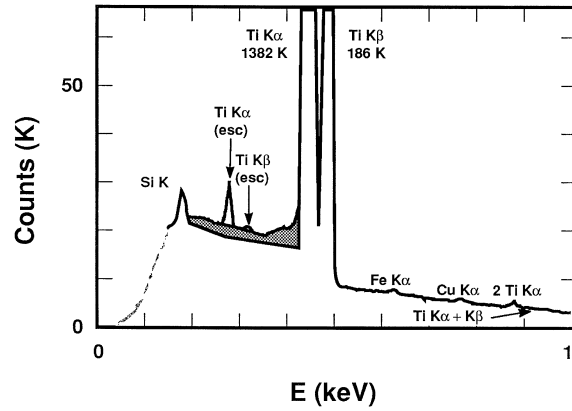


Figure 7.19. Electron-excited EDS spectrum of titanium. The Ti $K\alpha$ and $K\beta$ silicon x-ray escape peaks and the $2K\alpha$ and $(K\alpha + K\beta)$ sum peaks are noted. Extraneous peaks from the specimen chamber are also observed. Hatched area is the contribution due to incomplete charge collection.

probable that this electron will be reabsorbed in the detector. In this case the total energy deposited leads to creation of the number of electron–hole pairs given by Eq. (7.1). If a Si K x-ray is produced, one of two events will occur. In the first case, the Si K photon will be absorbed by the L shell of another Si atom which will be ionized and the ejected photoelectron will lead to electron–hole pair production with the total number again given by Eq. (7.1). In the second case, the Si K x-ray photon escapes from the detector because an x-ray can travel much further than an Auger electron of the same energy. This is more likely to happen if the Si K x-ray photon is produced near the surface. In this case, the total number of electron–hole pairs will be given by Eq. (7.1), provided the energy term E is replaced by $(E - E_{\mathrm{Si}K\alpha}) = (E - 1.74\,\mathrm{KeV})$. The reduction in the number of electron–hole pairs produced when an escape event takes place leads to the creation of an artifact peak called an "escape peak." It appears at an energy equal to the energy of the parent line minus that of the silicon x-ray, 1.74 keV. In principle, both Si $K\alpha$ and Si $K\beta$ escape peaks are formed, but the probability for $K\beta$ formation is about 2% that of the $K\alpha$, hence only one escape peak is usually observed per parent peak. Escape peaks are illustrated in Figs. 7.18 and 7.19. In Fig. 7.18 the Mn $K\alpha$ and Mn $K\beta$ escape peaks are observed. In Fig. 7.19 the parent peaks are Ti $K\alpha$ (4.51 keV) and Ti $K\beta$ (4.93 keV) and escape peaks are found at 2.77 keV (Ti $K\alpha$–Si $K\alpha$) and 3.19 keV (Ti $K\beta$–Si $K\alpha$). The magnitude of the escape peak relative to the parent peak varies from about 1.8% for phosphorus to 0.01% for zinc K x-rays (Fig. 7.20). Silicon x-ray escape peaks cannot occur for radiation below the excitation energy of the silicon K shell (1.838 keV).

7.2.7.4. Absorption Edges

The radiation emitted from the specimen penetrates several layers of "window" material before it arrives in the "active" part of the detector.

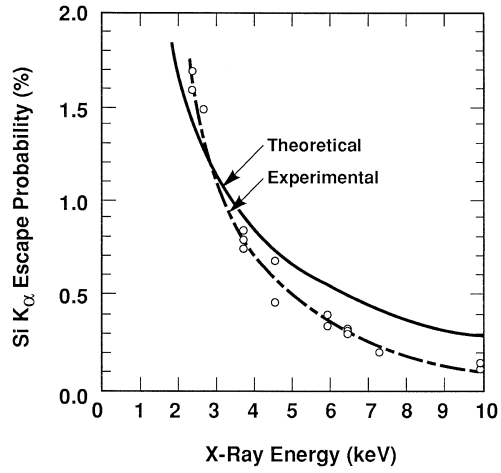


Figure 7.20. The escape probability of Si $K\alpha$ as a function of incident energy. (Adapted from Woldseth, 1973.)

These materials are either functional or artifacts related to detector fabrication, and their effects must be understood. The nominal purpose of the first window material is to protect the cooled detector chip from the relatively poor vacuum in the specimen chamber, which can be high in water vapor and organic components because of frequent exchanges of the specimen. It is very important that these components do not condense on the cooled detector. Even with a protective window, ice will build up over a period of time and will absorb lower energy x-rays. Some modern detectors have special heaters to remove this ice buildup, some of which actually originates from the detector assembly and cryostat at the time of manufacture. This problem can be overcome by the use of high-vacuum construction practices. In those cases where scanning and transmission electron microscopes are themselves ultra-high-vacuum devices the purpose of the window may actually be to protect the microscope vacuum from that of the detector.

Protective windows are made from a variety of different materials, each having advantages and disadvantages. Historically, beryllium, about 7.6 μm thick (0.3 mil), was used. This thickness has the advantage that it can withstand a full atmosphere pressure differential associated with specimen changes. Furthermore, the detector can be removed from the microscope without the need for isolation valves. It is also opaque to optical photons, which can easily saturate the detector during operation. The disadvantage is that it is relatively opaque to soft x-rays and thus makes detection of elements with low atomic numbers (sodium and below) difficult or impossible. Most current detectors use protective window materials with considerably less mass-thickness, but still able to withstand a full atmosphere of pressure differential. They are constructed of thin films on a support grid having approximately 85% open space. Low-energy x-rays, such as those from oxygen or carbon, will only pass through the unbacked portions of the film, whereas more energetic x-rays (>8 keV) can pass both

through the film and the support grid. The grid is usually pure silicon or a metal such as nickel and the window materials are boron nitride, silicon nitride, diamond or polymeric. Examples of transmission curves for various window materials currently in use are given in Fig. 7.21. Because thin films of these materials are transparent and the detector is also sensitive to visible light, it is necessary to either eliminate possible sources of light or apply an optically opaque layer to the protective window. Light sources include microscope illumination systems, viewing ports, and light leaks. Specimen cathodoluminescence is rarely intense enough to cause a serious problem, although it is encountered with some minerals and semiconductors.

The second window layer is an electrical contact (about 10–20 nm and usually gold). Gold has been observed to form "islands" during the evaporation process and hence is not uniform in thickness. The third window layer consists of inactive *p*-type silicon extending 100–200 nm or less into the detector. This layer is also not uniform and is often called the "dead layer" or "silicon dead zone." The combined effect of these Au and Si layers is often overlooked. In reality, their effect is as great as or greater than that of any of the new "thin" window materials. It is clear from Fig. 7.21a that there really is no such thing as a "windowless" detector.

During passage of x-rays through all of these window layers, absorption occurs. It is important to recall that photoelectric absorption refers to a process in which x-rays are diminished in number, but they do not lose energy, and thus the energies of the observed spectral lines are not altered while passing through the windows. In the case of 0.3-mil (7.6-μm) beryllium protective windows, nearly all x-rays below about 600 eV are eliminated due to absorption effects. Above 2 keV, virtually all x-rays are transmitted. Between these limits the absorption increases with decreasing energy such that at 1.5 keV about 70% of the x-rays are transmitted, whereas for an energy of 1 keV the transmission is 45%. Absorption by the gold and silicon layers is much less significant because of the small mass thickness of these layers. However, a noticeable change in the x-ray continuum is observed at the absorption edge of silicon and to a lesser degree at gold (Fig. 7.25). Just above the energy of the absorption edge, the mass absorption coefficient increases abruptly, resulting in a decrease in the measured continuum x-radiation. The height of the resulting step is an indication of the thickness of the layer. Note that the action of the broadening effect of the detection process causes the absorption edge, which in reality is a sharp change of absorption over a range of about 1 eV, to be smeared over a much broader range, typically 100 eV for the silicon absorption edge.

7.2.7.5. Silicon Internal Fluorescence Peak

The photoelectric absorption of x-rays by the silicon dead layer results in the emission of Si *K* x-rays from this layer into the active volume of the detector. These silicon x-rays, which do not originate in the sample, appear in the spectrum as a small silicon peak, the so-called silicon internal fluorescence peak. An example of this effect is shown in the spectrum of pure carbon, Fig. 7.22, which also contains a significant silicon absorption

edge. For many quantitative analysis situations, this fluorescence peak corresponds to an apparent concentration of approximately 0.2 wt% or less silicon in the specimen.

7.2.8. Artifacts from the Detector Environment

The user of an EDS system is usually responsible for the installation of the spectrometer and its associated electronics on the electron beam

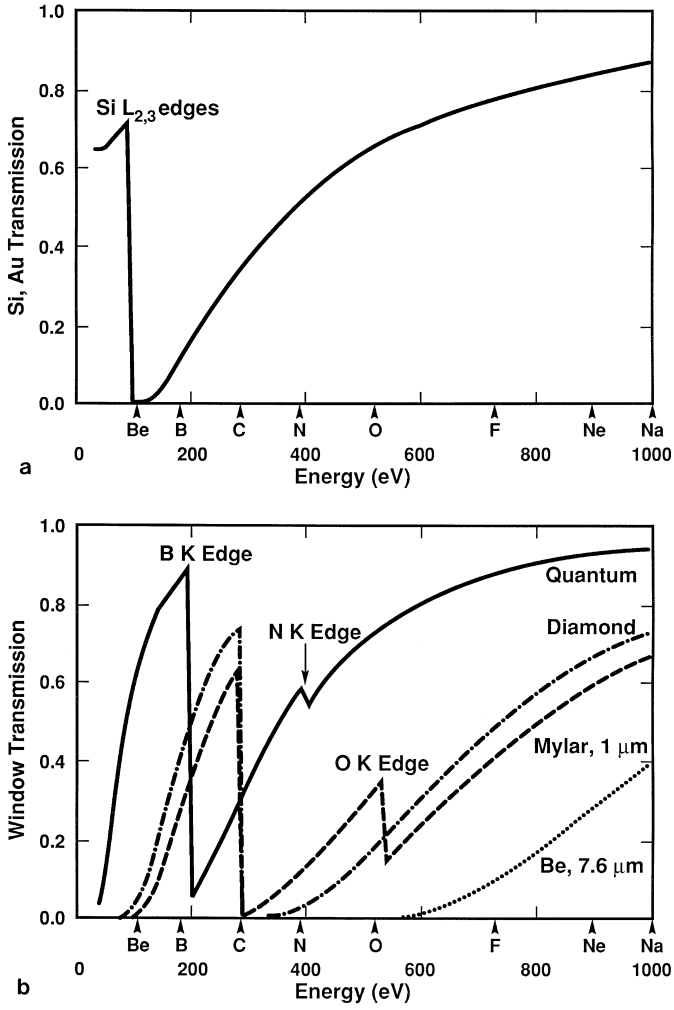


Figure 7.21. (a) Transmission curve for a "windowless" detector with a 0.01-μm Au contact and a 0.1-μm silicon dead layer. The energies of the light element K lines are shown along the energy axis. (b) Transmission curve for several commercially available window materials. The energies of the light-element K lines are shown along the energy axis. The actual transmission of x-rays into the active part of the detector is a combination of the transmission characteristics shown in (a) and (b). Quantum window = 0.25 μm; diamond window = 0.40 μm.

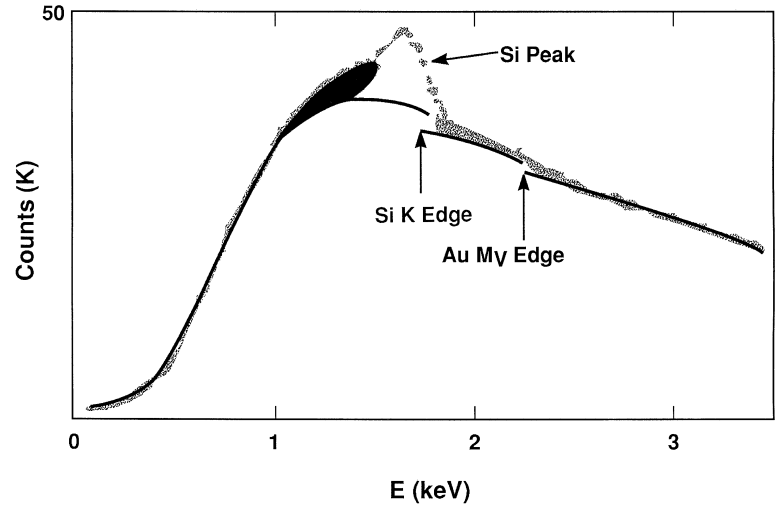


Figure 7.22. Electron-excited EDS spectrum from carbon. The silicon and gold absorption edges are illustrated. The solid line represents a theoretical fit to the continuum. $E_0 = 10$ keV. The silicon peak arises from internal fluorescence.

instrument. In order to obtain an optimum spectrum, the user may have to overcome a number of artifacts which result from interactions between the EDS system and its environment. These include microphonics, ground loops, accumulation of contaminants, and the entry of spurious radiation, including electrons, into the detector. See Enhancement 7.3 on the CD for a discussion of these effects and Enhancement 7.4 on the CD for a discussion of how to set up a detector to avoid some of these effects.

7.2.9. Summary of EDS Operation and Artifacts

The process by which the characteristic lines and continuum distribution are modified in going from the generated spectrum in a Mn sample to the observed spectrum in the EDS are summarized in Fig. 7.23. Point *a* and the associated spectrum corresponds to what the distribution of characteristic lines and continuum might look like if the measurement could be made by an ideal detector located within the sample. Point *b* shows the spectral distribution as it would exist in the emitted x-rays. Note that the principal change is due to absorption, which sharply truncates the low-energy portion of the continuum, introduces a discontinuity associated with continuum absorption at the Mn *K* absorption edge, and also slightly diminishes the characteristic peak intensities. At point *c* the signal has passed through the Be window, Au surface contact, and silicon dead layers causing a further reduction in intensity at low energies and also introducing some fine structure caused by absorption and fluorescence effects. Finally at point *d* the signal has been processed by the Si(Li) diode and associated electronics. The processing results in a smearing effect, which distributes most of the information corresponding to each sharply defined energy increment in the

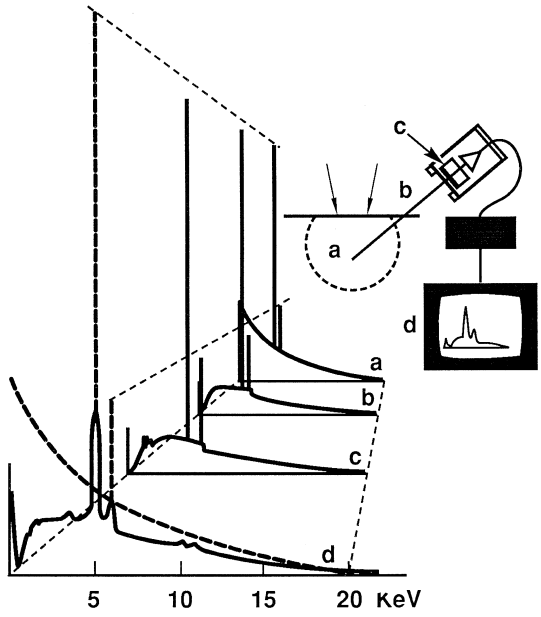


Figure 7.23. Schematic drawing of the relationship between the measured and generated spectral distributions of manganese. The letters *a*, *b*, *c*, and *d* indicate different points of observation in theory, although only *d* represents what the analysts actually sees.

CXA encompassing the FWHM. The effect is to noticeably broaden peaks and absorption edges well beyond their natural width, which leads to reduced peak-to-background ratios and the high probability of peak overlap. In addition, it is at this point that the electronic artifacts of escape peaks and peak-on-peak pileup make their appearance. Although the EDS does introduce a large number of modifications to the generated spectrum, understanding and accounting for these artifacts leads to spectral information critical to accurately assessing the identity and amount of the elemental constituents of a sample.

7.3. Wavelength-Dispersive Spectrometer

7.3.1. Introduction

Until 1968, when energy dispersive spectrometers (EDS) were first interfaced to SEMs, the wavelength-dispersive spectrometer (WDS) was used almost exclusively for x-ray spectral characterization. It is very important to recognize that although it is much less commonly used on SEMs for chemical analysis, it is still the principal mode of x-ray measurements for the analytical SEM configuration known as the electron probe microanalyzer (EPMA). The main reasons for the application of the WDS are superior energy resolution (often better by a factor of 10) as well as high count rate capability without compromising energy resolution (often better than

50,000 counts per second on the peak of interest). The WDS is generally more expensive, time-consuming, and difficult to use.

7.3.2. Basic Description

The basic components of the WDS are illustrated in Fig. 7.24. A small portion of the x-ray signal generated from the specimen impinges on an analyzing crystal, which consists of a regular array of atoms. The crystal is oriented such that a selected crystallographic plane of these atoms is parallel to its surface. For a particular spacing d and wavelength λ there exists an angle θ at which the x-rays are strongly scattered. The following relationship is known as Bragg's law:

$$n\lambda = 2d \sin\theta, \tag{7.6}$$

where n is an integer, 1,2,3,...; λ is the x-ray wavelength; d is the interplanar spacing of the crystal; and θ is the angle of incidence of the x-ray beam on the crystal. The x-rays will be diffracted (strongly scattered to a particular angle) and detected by a gas-filled proportional counter. The signal from the detector is amplified, converted to a standard pulse size by a single-channel analyzer (SCA), and then either counted with a scaler or displayed as a rate meter output on a recorder. As described in Chapter 6, each element in the sample can emit one or more series of unique spectral lines. A typical qualitative analysis therefore involves obtaining a recording of x-ray intensity as a function of crystal angle. The peaks observed in the spectra correspond to each detectable line, but may have their shape or position altered as a result of the attributes of the spectrometer itself, such as alignment and crystal quality. In the case of longer wavelength x-rays, the shape and position may also be altered by changes in the local chemical environment. Taking these factors into consideration, it is nevertheless

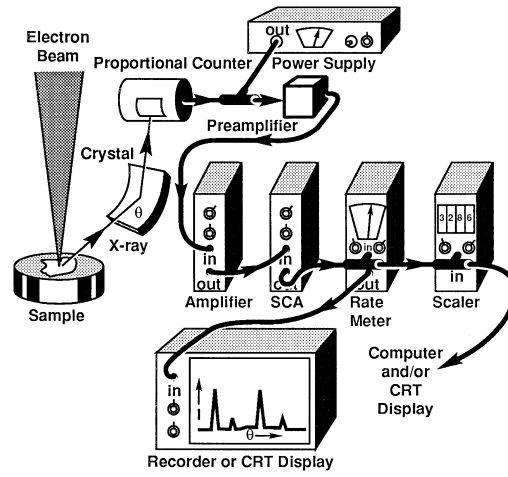


Figure 7.24. Schematic representation of a wavelength-dispersive spectrometer and associated electronics.

possible to convert peak positions to wavelengths through Bragg's law and then use the Moseley relationship [Eq. (6.4)] to identify the elemental constituents. In practice, crystal spectrometer readings are either proportional to wavelength or calibrated directly in wavelength. The data can often also be displayed as a graph of intensity versus energy on a CRT in a similar manner to the presentation of EDS data, therefore making peak identification and peak labeling a similar process to that discussed in connection with Fig. 7.11.

7.3.3. Diffraction Conditions

Bragg's law can be derived with the aid of Fig. 7.25. In this simplified drawing, the x-ray beam approaching the crystal is viewed as a wavefront rather than as an individual photon. The wavefront is then specularly reflected from parallel crystal planes spaced d units apart. Of the two x-ray paths shown in Fig. 7.25, the lower one is longer by an amount $ABC = 2d \sin\theta$. If this distance equals an integral number of wavelengths n, then the reflected beams will combine in phase and an intensity maximum will be detected by the proportional counter. If the intensity distribution is measured as a function of θ with a high-quality analyzing crystal, the effect of combined reflections from a large number of planes will result in relatively narrow peaks. For example, the measured full-width at half-maximum (FWHM) for Mn $K\alpha$ is about 10 eV, compared to the natural value of 2 eV. X-rays having a wavelength that will not satisfy the Bragg equation are absorbed by the crystal or pass through it into the crystal holder.

The x-ray signal in focused electron beam instruments is fairly weak and can be thought of as originating from a point source. Therefore, the signal is maximized at the detector by the use of fully focusing x-ray spectrometers. These spectrometers use curved crystals in preference to flat crystals of the type normally associated with tube-excited x-ray emission analysis. In a fully focusing spectrometer of the Johansson type, illustrated

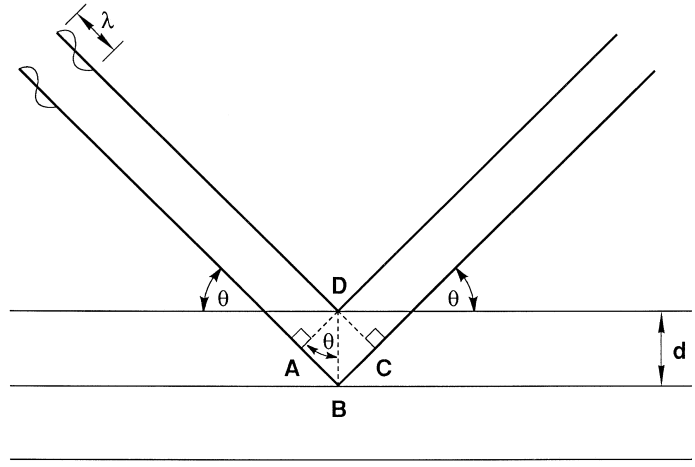


Figure 7.25. Diffraction according to Bragg's law. Strong scattering of x-rays of wavelength $n\lambda$ occurs only at angle θ. At all other angles scattering is weak.

in Fig. 7.26a, the x-ray point source, the specimen, the analyzing crystal, and the detector are all constrained to move on the same circle with radius R, called the focusing circle. Furthermore, the crystal planes are bent to a radius of curvature of $2R$. The surface of the crystal is also ground to a radius of curvature R. As a result of this geometry, all x-rays originating from the point source will have the same incident angle θ on the crystal and will be brought to a focus at the same point on the detector. This maximizes the overall collection efficiency of the spectrometer without sacrificing good wavelength resolution. Clearly, if a flat crystal were used, the angle of incidence of the x-ray beam would vary across the length of the crystal, causing a considerable loss in count rate. This is because only a narrow band of the crystal would satisfy the Bragg condition for the desired wavelength.

Figure 7.26 also illustrates the additional geometric requirement of a constant x-ray take-off angle ψ imposed by having a small region available above the sample and beneath the final lens. The fully focusing requirement is maintained by moving the analyzing crystal along a straight line away from the sample while rotating the crystal, as illustrated in Fig. 7.26b, and moving the detector through a complex path so as to make the focusing circle rotate about the point source. An interesting consequence of this arrangement is that the crystal-to-source distance L is directly proportional to the wavelength. This can be shown with the aid of Fig. 7.26: First we write

$$L/2 = R\sin\theta \quad \text{or} \quad L/2R = \sin\theta. \tag{7.7a}$$

Combining this with Bragg's law, we find that Eq. (7.6) gives

$$n\lambda = 2d\sin\theta = 2dL/2R, \tag{7.7b}$$

or for first-order reflections

$$\lambda = (d/R)L, \tag{7.7c}$$

with higher order reflections occurring simply at multiples of the L value for the first-order reflections. Most spectrometers read out directly in L

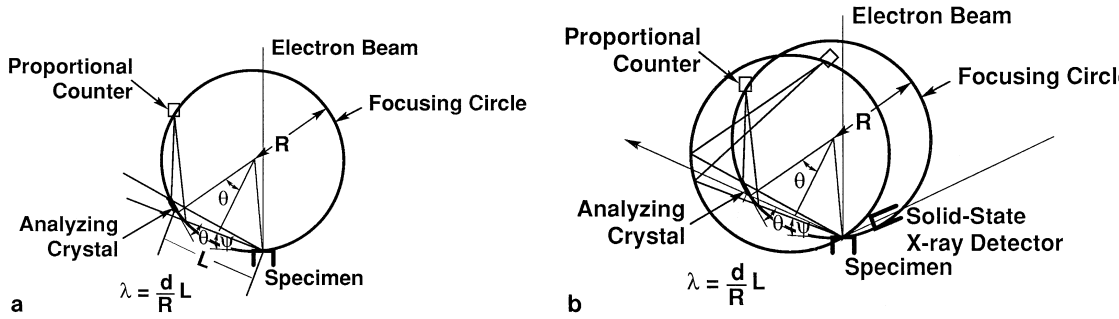


Figure 7.26. (a) Fully focusing wavelength spectrometer. R is the radius of the focusing circle, ψ is the take-off angle, and θ is the diffraction angle. (b) Movement of the focusing circle to change the diffraction angle. Note that the position of the EDS shown in the figure for comparison remains stationary because no scanning is required.

(mm or cm). Actually, the majority of fully focusing spectrometers use crystals that are only bent to a radius of curvature of $2R$, but not ground to be completely tangent to the focusing circle because grinding a crystal will tend to degrade its resolution by causing defects and a broader mosaic spread. This compromise, known as Johann optics, results in some defocusing of the image at the detector, but does not seriously impair resolution.

In instruments with fully focusing crystal spectrometers, there is a finite volume of the sample that can be considered to be "in focus." The shape and dimensions of this region depend on a number of factors, such as the design of the spectrometer, the choice and size of crystal, the wavelength being measured, and the criterion for focus (i.e., 99% or 95% of the maximum intensity). If the plane of the spectrometer is parallel to the electron beam ("vertical spectrometer") then the z-axis setting of the sample must be within a few micrometers for focus. In EPMAs this is normally accomplished by the use of a coaxial reflecting optics light microscope that has a very limited depth of focus. The optics are designed to allow the passage of the electron beam down the center, yet such that the sample can be viewed at the same time in either an SEM mode or through the optical microscope. The overall instruments are designed such that placing the sample in light-optical focus by z-axis adjustment ensures that it is on the focusing circle. Most SEMs equipped with WDS do not, however, have an optical microscope and the excellent depth of focus of the SEM is a liability because it is difficult to observe changes in working distance of a few micrometers, which is critical to the x-ray measurement. In this case the spectrometer is normally tuned in the setting corresponding to a known element in the sample and the z adjusted for maximum intensity to place the sample on the focusing circle. Although this approach is satisfactory for flat samples mounted normal to the electron beam, it may introduce errors for samples that are either tilted or have irregular surfaces as areas away from the initial point of focus are examined.

Even when examining flat samples at normal incidence, defocusing can occur from tens to hundreds of micrometers from the initial focusing point. Therefore if the electron beam is scanned to form an x-ray map, a loss of intensity can occur in the outer portions of the image, particularly at low magnifications. It is for this reason that the most accurate quantitative line scans or maps are done by scanning the stage under a static electron beam. The difficulties associated with the vertical placement of samples have led to the use of "inclined" spectrometers whose plane is no longer parallel to the electron beam. Although this can help the alleviate effects due to roughness or tilt mentioned above, this approach makes mapping even more difficult due to defocusing in one of the directions scanned.

7.3.4. Diffracting Crystals

Most electron microprobes and scanning electron microscopes can be equipped with more than one crystal spectrometer. Multiple spectrometers,

Table 7.1. Crystals Used in Diffraction

Name[a]	2*d* (nm)	Lowest atomic number diffracted	Resolution	Reflectivity
α-Quartz (1011)	0.6687	$K\alpha_1$ ^{15}P $L\alpha_1$ ^{40}Zr	High	High
KAP (1010)	2.6632	$K\alpha_1$ ^{8}O $L\alpha_1$ ^{23}V	Medium	Medium
LiF (200)	0.4028	$K\alpha_1$ ^{19}K $L\alpha_1$ ^{49}In	High	High
PbSt	10.04	$K\alpha_1$ ^{5}B	Medium	Low
PET	0.8742	$K\alpha_1$ ^{13}Al $L\alpha_1$ ^{36}Kr	Low	High
TAP (1010)	2.59	$K\alpha$ ^{8}O $L\alpha_1$ ^{23}V	Medium	Medium
Layered synthetic materials				
LDE1	6.0	$K\alpha$ ^{6}C	Low	High
LDEC	9.97	$K\alpha$ ^{5}B	Low	High
LDEB	14.6	$K\alpha$ ^{4}Be	Low	High

[a] KAP, . . . ; PbSt, . . . ; PET, . . . ; TAP, . . . ; LDE, ...

each containing several crystals, are necessary not only for analyzing more than one element at a time, but also for including the variety of crystals required for optimizing performance in different wavelength ranges. Table 7.1 lists some of the most commonly used analyzing crystals, showing their comparative 2*d* spacings and range of elements covered. Because $\sin\theta$ cannot exceed unity, Bragg's law establishes an upper limit of 2*d* for the maximum wavelength diffracted by any given crystal. More practical limits are imposed by the spectrometer design itself because it is obvious from Fig. 7.26 that for $\sin\theta = 1$, that is, $\theta = 90^\circ$, the detector would have to be at the x-ray source point inside the electron-optical column. A lower wavelength limit is imposed by Eq. (7.6) because it becomes impractical to physically move the analyzing crystal too close to the specimen.

It should be noted that most naturally occuring crystals like LiF are limited to shorter wavelengths because their maximum interplanar spacings are small. Synthetic crystals, like TAP, provide a very useful extension of the measurable wavelength range, but still do not have sufficiently large *d* spacings to cover the longest wavelengths encountered, for example, Be, B, C, and N $K\alpha$. For many years, the measurement of these wavelengths required the use of "pseudocrystals" grown by the Langmuir–Blodgett technique (Blodgett and Langmuir, 1937), illustrated in Fig. 7.27a. In this method a suitable substrate such as mica or glass is dipped into a trough containing a monolayer film of lead stearate floating on the surface of water. The lead stearate, or similar compound, is characterized by long-chain molecules with a heavy atom on one end. The initial dipping results in the attachment of the heavy metal with the organic tail perpendicular to the substrate. Removal of the substrate from the trough results in a second layer

deposited in which the hydrocarbon tails are attached and a row of heavy metal atoms forms the external surface. The "pseudocrystal" is then built up by repeated dippings. The organic tails thus serve as light-element spacers between the reflecting planes of heavy atoms. The interplanar spacing can therefore be established by using organic molecules with different chain lengths.

Currently the measurement of long-wavelength x-rays is most often done with the use of layered synthetic microstructure (LSM). Figure 7.27b is a transmission electron micrograph of a cross section of an LSM. These diffraction structures are built up by the physical vapor deposition of alternating layers of heavy and light elements (in this case carbon and tungsten). This fabrication technique permits deposition of layers with any value of the spacing required and with material choices optimized to scatter x-rays efficiently. LSM diffractors are typically an order of magnitude more efficient than pseudocrystal diffractors. As shown in Fig. 7.28, the LSMs generally offer considerably improved count rates, under the same instrument operating conditions, relative to the equivalent "pseudocrystal" with an acceptable loss of resolution. This loss of resolution is often an advantage because the effects of chemical shifts in wavelength are minimized.

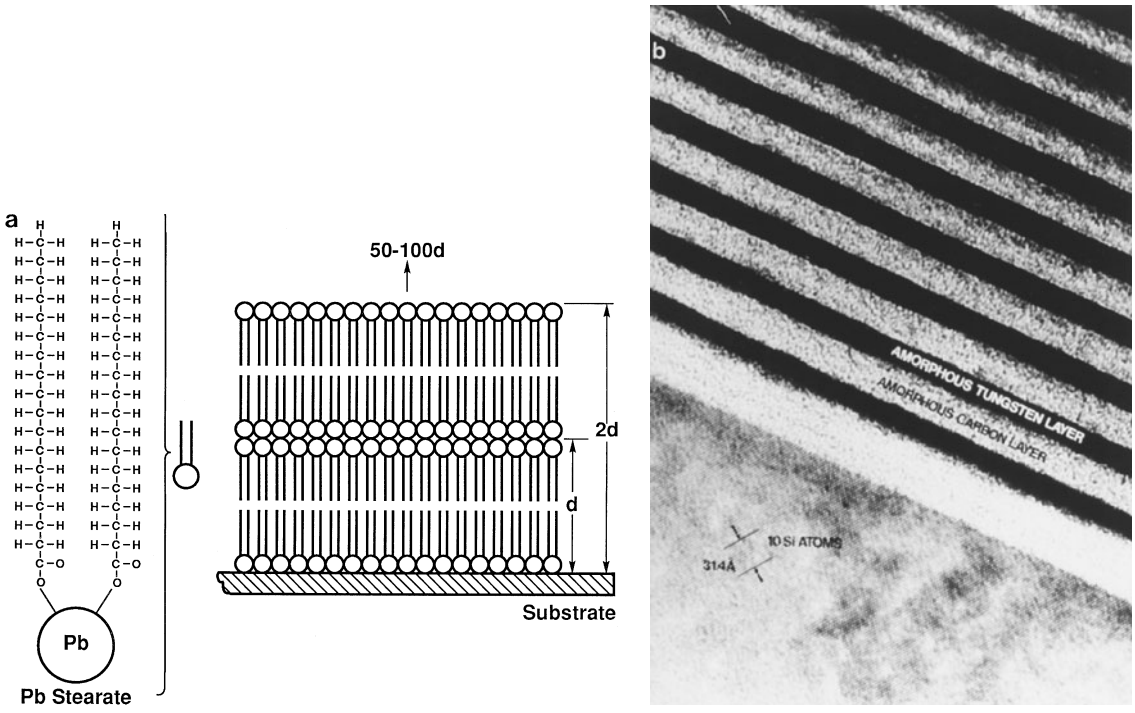


Figure 7.27. (a) Pseudocrystal grown by the Langmuir–Blodgett technique. (b) transmission electron micrograph of layered synthetic microstructure. (Courtesy of Ovonics.)

7.3.5. The X-Ray Proportional Counter

The most commonly used detector with microanalyzer crystal spectrometer systems is the gas proportional counter shown in Fig. 7.29. This detector has an excellent dynamic range (0–50,000 counts per second or more), covers a wide range of energies, and has a generally high collection efficiency. It consists of a gas-filled tube with a thin wire, usually tungsten, held at a l- to 3-kV potential, running down the center. When an x-ray photon enters the tube through a thin window on the side and is absorbed by an atom of the gas, it causes a photoelectron to be ejected, which then loses its energy by ionizing other gas atoms. The electrons thus released are then attracted to the central wire, giving rise to a charge pulse. If the gas fill used is P10 (90% argon–10% methane), approximately 28 eV is absorbed per

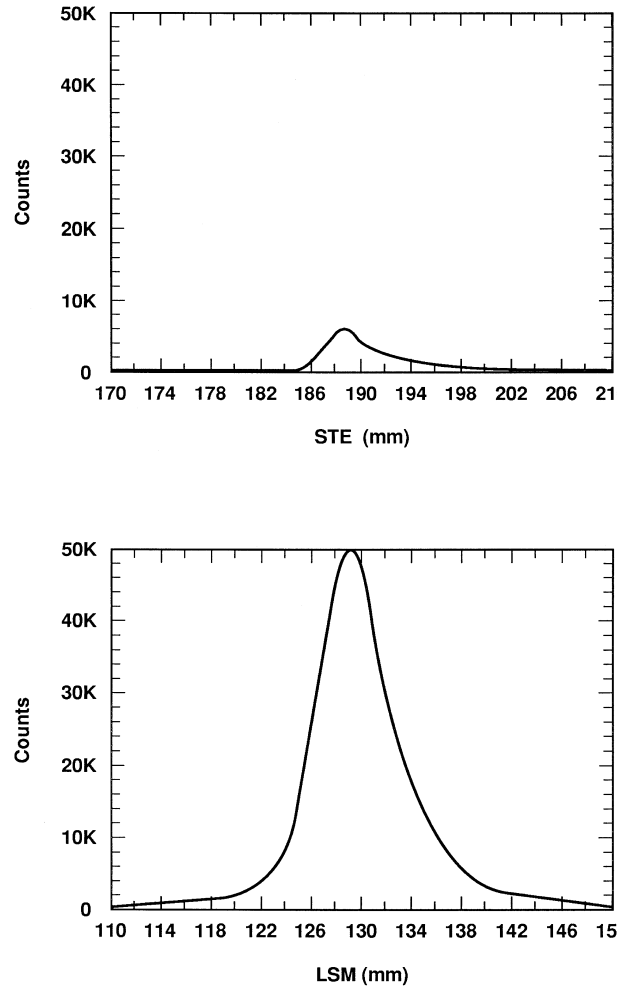


Figure 7.28. Intensity comparison of the boron spectrum between stearate (STE) and layered synthetic microstructure (LSM). (Courtesy of C. Nielsen, JEOL.)

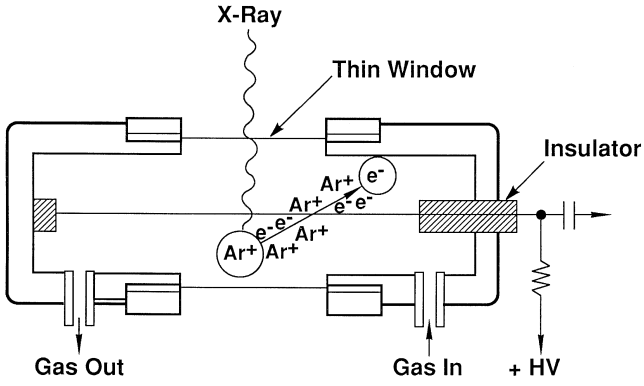


Figure 7.29. Schematic drawing of a gas flow proportional counter.

electron ion pair created. For Mn $K\alpha$, which has an energy of 5.895 keV, about 210 electrons will be directly created by the absorption of a single photon. This would be an extremely small amount of charge to detect without a special cryogenic preamplifier system (low temperature, low noise). However, if the positive potential of the anode wire is high enough, the electrons ejected in the primary events are accelerated sufficiently to ionize gas atoms, and secondary ionizations occur which can increase the total charge collected by several orders of magnitude. Figure 7.30 schematically shows the effect on the gas amplification factor of increasing the bias voltage applied to a tube. The initial increase corresponds to increasing primary charge collection until it is all collected (a gas amplification factor of 1) and then the curve levels off in the "ionization" region. Increasing the potential beyond this point initiates secondary ionization, the total charge collected increases drastically, and the counter tube enters what is termed the

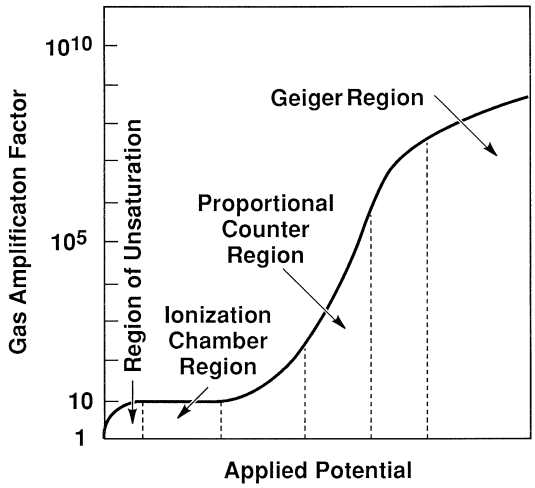


Figure 7.30. The effect of applied counter-tube bias on the gas amplification factor.

"proportional" region because the collected charge remains proportional to the energy of the incident photon. Further increasing the voltage causes the tube to enter the Geiger region, where each photon causes a discharge giving rise to a pulse of a fixed size independent of its initial energy, thereby losing the information necessary for pulse height analysis. A further disadvantage is that the counter "dead time," which is the time required for the tube to recover sufficiently to accept the next pulse, increases from a few to several hundred microseconds. Any increase in applied voltage beyond the Geiger region will result in permanent damage to the tube. In practice, operation in the lower part of the proportional region is preferred to minimize the effect of gain shifts with counting rate and minimize dead time.

The proportional counter shown in Fig. 7.29 is of the gas flow type, normally used for detecting soft x-rays ($\lambda > 3$ Å). A flowing gas, usually P10, is chosen because it is difficult to permanently seal the thin entrance windows necessary to reduce absorption losses. Because crystal spectrometers are kept under vacuum to eliminate absorption of the x-ray beam in the air, it is necessary to support ultrathin windows like Formvar and cellulose nitrate on fine wire screens in order to withstand a pressure differential of l atm; however, this causes an additional decrease in detector collection efficiency. Recently, unsupported stretched polypropylene films have come into use with considerable success. Sealed counters, with Be windows or other materials, containing krypton or xenon are used for shorter x-ray wavelengths because, as shown in Fig. 7.31, they have a higher quantum counter efficiency (the percentage of input pulses detected) than argon-filled detectors at l atm. The efficiency of argon-filled detectors for shorter x-ray wavelengths can, however, be increased by increasing the gas pressure to 2 or 3 atm.

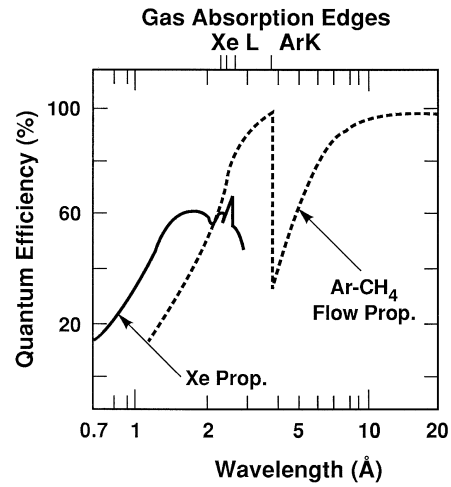


Figure 7.31. Representative collection efficiencies for proportional counters filled with Xe and Ar–CH_4 gases. The Ar–CH_4 counter has a 25-nm Formvar window.

The role of the detector electronics is to integrate the total charge produced by each x-ray photon and convert it into a single voltage pulse to be further processed for counting or display purposes. As is shown in Enhancement 7.5 on the accompanying CD, the size of the voltage pulse is directly proportional to the energy of x-ray photon detected by the flow proportional counter. Measurement of pulse shapes at various points in the detector electronics chain is described in Enhancement 7.6.

Information about the actual voltage distribution of pulses processed for a preselected time period can be readily obtained by means of a single-channel analyzer (SCA). Basically the SCA serves two functions. As a discriminator, it is used to select and transmit pulses within a predetermined voltage range for further processing. As an output driver the selected pulses trigger rectangular pulses of a fixed voltage and time duration compatible with scalar and rate meter input requirements. A typical SCA output pulse shape is given in Fig. 7.32c. In this example the output pulse is 6 V high and lasts for 0.2 μs. Note that Figs. 7.32a and 7.32b represent pulse shapes at the preamplifier and main amplifier outputs (see Fig. 7.24). Unlike 7.32c, the sizes of the pulses are proportional to the x-ray photon energy.

The process of pulse height analysis with an SCA is illustrated schematically in Fig. 7.33. The operator sets either a combination of a baseline voltage E_L and a window voltage ΔE by means of potentiometer controls in the SCA, or a baseline voltage E_L and an upper window voltage E_U. In the example shown, only pulses from 5 to 7 V (pulse II, 6 V) are accepted. Pulses larger (pulse III, 8 V) or smaller (pulse I, 4 V) are rejected. In practice pulse voltage distribution curves with an SCA can be obtained by tuning the WDS to a specific characteristic line, selecting an SCA window voltage ΔE of a few tenths of a volt, and recording the pulse intensity

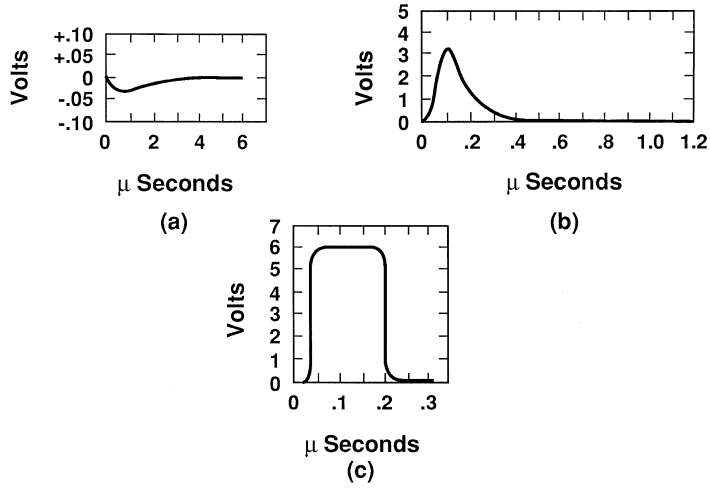


Figure 7.32. Typical WDS x-ray detection pulse shapes: (a) preamplifier, (b) main amplifier, (c) single-channel analyzer output.

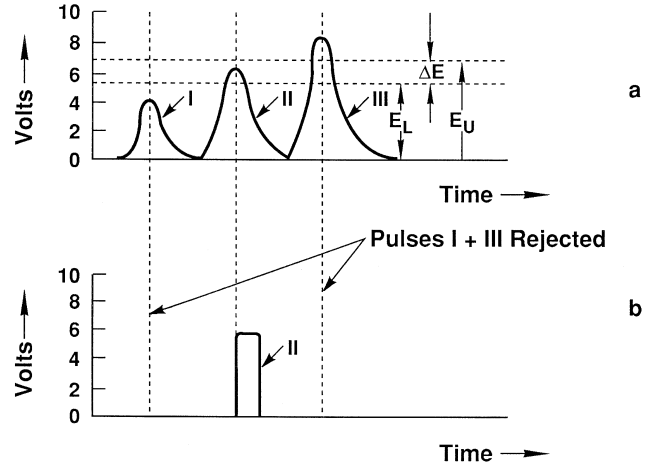


Figure 7.33. Schematic representation of pulse-height analyzer behavior. (a) Main amplifier output; (b) single-channel analyzer output. $E_{\mathrm{L}} = 5$ V; $\Delta E = 2$ V; $E_U = 7$ V.

as a function of the baseline, E_{L}, setting (see Fig. 7.34). This can be done by either scanning the baseline at a fixed rate from 0 to 10 V and recording the SCA output with a rate meter or manually stepping through a range of baseline values, counting the pulses with a scaler, and plotting the results.

The principal features to note in Fig. 7.34 are two relatively broad peaks. The larger of the two corresponds to the main Ni $K\alpha$ pulse distribution. The smaller peak is known as an "escape peak." Its center occurs at a pulse voltage corresponding to an x-ray energy equal to the Ni $K\alpha$ energy minus the characteristic line energy of the element used for counter tube gas, which in this case is xenon (xenon $L\alpha = 4.11$ keV) for the example shown in Fig. 7.34. The escape peak corresponds to a special type of proportional detector response during which either the incoming x-ray photon

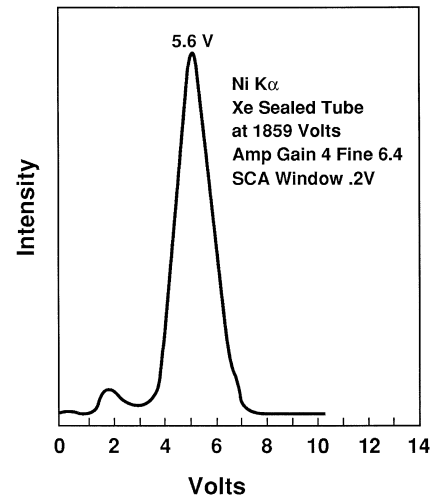


Figure 7.34. Pulse distribution of Ni $K\alpha$ determined by a single-channel analyzer.

or the primary photoelectron ionizes a core rather than outer shell electrons in the counter tube gas. This process is then followed by the emission of a characteristic x-ray photon, which will have a high probability of escaping from the counter tube entirely. If it does, the remaining energy for electron–ion pair production will be correspondingly diminished. For these events the amplifier output equation must be modified to the form

$$V_{\mathrm{A}}\ (\text{escape peak}) = K(E - E_{\mathrm{CT}}) \tag{7.8}$$

where E_{CT} is the characteristic line energy of the counter tube gas. If P10 gas is used, then $E_{\mathrm{CT}} = 2.96\,\text{keV}$, the energy of argon $K\alpha$. Note that the escape peak is analogous to that found in EDS analysis, but it does not appear in the spectrum, only in the pulse height distribution analysis.

The natural spread of both peaks arises from the fact that every monoenergetic photon entering the detector does not give rise to the same number of electron–ion pairs. Several competing processes occur by which the initial photoelectron may dissipate its energy. The percentage resolution of a detector is defined as 100 times the width of the pulse distribution curve at half-maximum divided by the mean peak voltage. The resolution of a properly functioning counter tube is about 15–20%. The pulse distribution should be approximately Gaussian and free of large asymmetric tails. It is desirable to periodically check this distribution because failure of the electronics or degradation of the counter tube can lead to a shift in the peak position, width, and/or symmetry, making any prior SCA settings improper.

There has always been some confusion about the precise role of the energy discrimination capability of the SCA. First, it cannot improve the energy selection of the spectrometer system for wavelengths close to that of the characteristic line being measured, for this has already been done by the diffraction process itself, which clearly has much higher energy resolution than the flow counter. The SCA can, however, eliminate both low- and high-energy noise as well as higher order reflections ($n > 1$ in Bragg's law) arising from the diffraction of higher energy characteristic or continuum x-rays with the same $n\lambda$ value because such events will occur at the same λ setting as the line of actual interest. SCA window settings should be approached with some caution because setting too narrow a window may actually lead to the rejection of useful pulses. This point is illustrated with the aid of Fig. 7.35, which shows a series of pulse distributions for crystal spectrometer settings corresponding to several elements readily accessible to a single LiF crystal during a typical scan. It can be seen in scanning from Ti to Cu that because pulse height is proportional to energy, the main pulse distribution shifts from a mean value of about 4.5 to 8.0 V. It is evident that if the SCA window is set to only include pulses from 4 to 6 V, then, although Ti would be measured, Cu pulses would be rejected. Therefore, the use of such a window is totally incompatible with qualitative line scanning and should only be used for fixed-spectrometer-setting quantitative analysis.

It should further be pointed out that even when narrow-band SCA settings are used for fixed spectrometer positions, as in the case of a quantitative analysis, the analyst is not free from problems. Under certain conditions the entire pulse distribution may shift to lower voltages with increasing

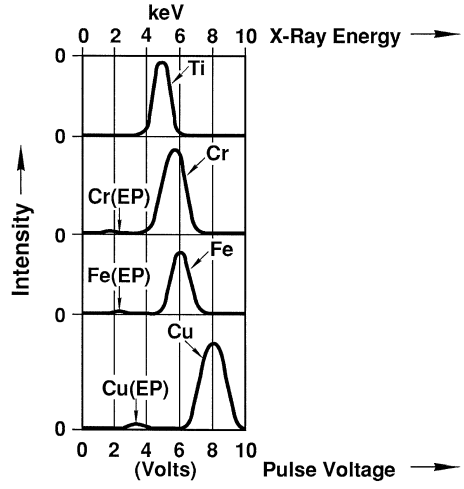


Figure 7.35. Four elemental pulse distributions determined from a proportional counter.

count rate. The nature of such a shift is strongly dependent on a number of parameters including the counter tube voltage, the main amplifier gain, and the condition of the counter tube including the gas fill (pressure and type). It is recommended that before quantitative data are collected the analyst perform a pulse distribution measurement at both high and low count rates (corresponding to standards and low-concentration samples) to see if the window will accept all pulses of the desired energy at either count rate.

A consequence of the finite time for an x-ray pulse to pass through the detector electronics is the possibility that during this time, another x-ray photon may enter the proportional counter. If so, this second photon is lost because the detector electronics has not yet returned to the baseline for measuring the next pulse. As the rate of photons reaching the detector increases, the probability for these coincident events also increases. The length of time that the detector is unavailable because of pulse processing is termed the "dead time," as discussed for EDS detectors. The loss of events due to pulse coincidence is termed "coincidence losses." For accurate measurement of x-ray intensities as the input to quantitative analysis routines, the dead time must be corrected, or errors of several percent will be introduced at high spectrometer count rates.

In the wavelength-dispersive spectrometer, the dead-time correction is applied after the intensity is measured. The dead-time relationship is

$$N = N'/(1 - \tau N'), \tag{7.9}$$

where N' is the measured count rate, N is the true count rate, which we wish to calculate, and τ is the dead time in seconds. To apply Eq. (7.9), the dead time τ must be known. One method for determining τ consists in plotting N' versus the measured beam current, which is directly proportional to N. The deviation of this plot from linearity can be fitted to Eq. (7.9) to determine τ.

Any intensity measured with the WDS must be corrected for dead time by the use of Eq. (7.9) before that intensity is used for quantitative

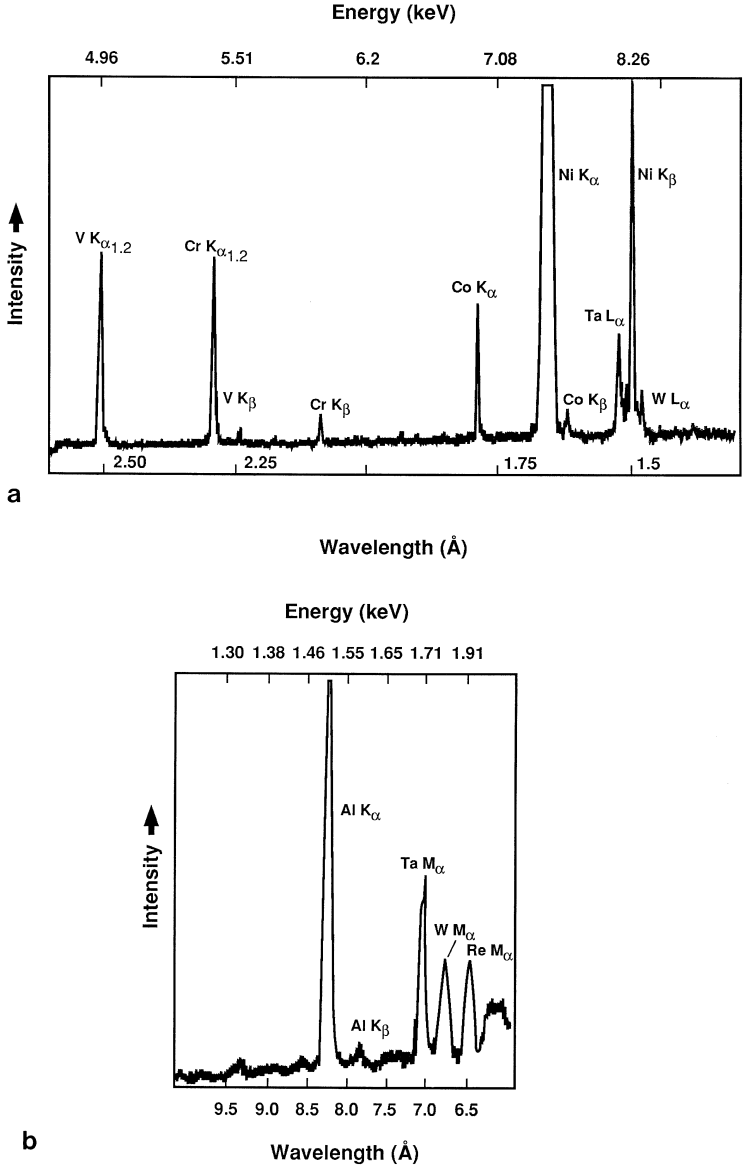


Figure 7.36. Wavelength-dispersive spectrometer scans of a nickel-based superalloy. (a) Scan using LiF crystal; (b) scan using a TAP crystal.

analysis. In a typical WDS, τ is approximately 2 μs, so that the correction for dead time is no more than 2% for count rates below 1×10^4 counts/s. It is possible to electronically "fix" the dead time to a value slightly larger than the longest dead time in the electronic chain. This method allows direct application of Eq. (7.9).

Typical crystal spectrometer scans of rate meter output versus wavelength for a nickel-base superalloy are illustrated in Fig. 7.36a for LiF and

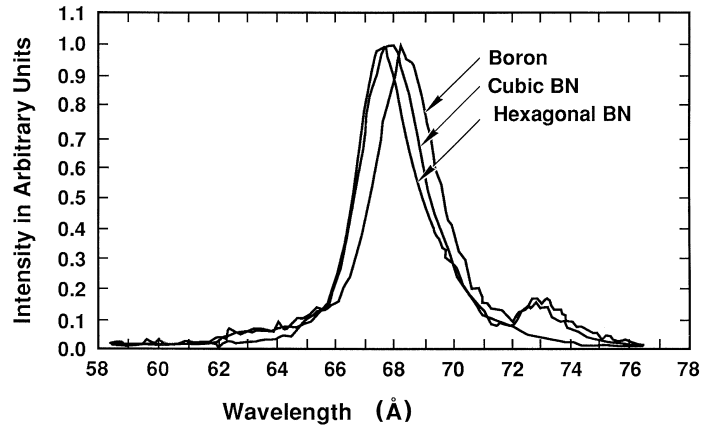


Figure 7.37. Boron $K\alpha$ scans obtained from pure boron, cubic boron nitride, and hexagonal boron nitride.

Fig. 7.36b for TAP. The Cr $K\alpha_{1,2}$ and V $K\beta$ peaks, which are separated by only 16 eV, can be separated with a crystal spectrometer. The shape of the diffracted peak is a Lorentzian, rather than the Gaussian found with the EDS which means that the tails of the distribution are wider. Two other capabilities, namely, light element detection and peak shift measurements, are illustrated in Fig. 7.37. This figure shows superimposed computer-controlled scans for boron $K\alpha$ in pure boron, cubic boron nitride, and hexagonal boron nitride. The peak shifts and satellite lines are due to shifts in the outer electron energy states associated with differences in the chemical bonding. Measurements of this type can also be used to fingerprint various cation oxidation states in metal oxides (Holliday, 1963).

In the case of instruments that utilize both EDS and WDS systems, WDS spectra are sometimes displayed as plots of intensity versus energy and can then be displayed and analyzed using the same software developed for EDS spectra. Figure 7.38 shows spectral data for the same energy range by both EDS and WDS software. The intensity scales have been adjusted so that both spectra can displayed at the same time. Many of the concepts for data processing and display that have been extensively developed for EDS are also now fairly standard for WDS as well. Intensity and wavelength (or energy) data can be directly fed into computer memory to allow a greater variety of display options such as peak identification, intensity measurement, background subtraction, correction of spectral overlap, multiple spectrometer display, etc.

In addition to performing chemical analysis at a fixed point it is often desirable to analyze variations in the x-ray intensity of one or more selected elements across a line on a specimen or even over a two-dimensional field of view. In the line scan mode illustrated in Fig. 7.39 the rate meter output corresponding to a fixed spectrometer setting is used to modulate the vertical deflection of the CRT as the electron beam is scanned across the sample. Multiple exposures of the CRT are used to superimpose the x-ray

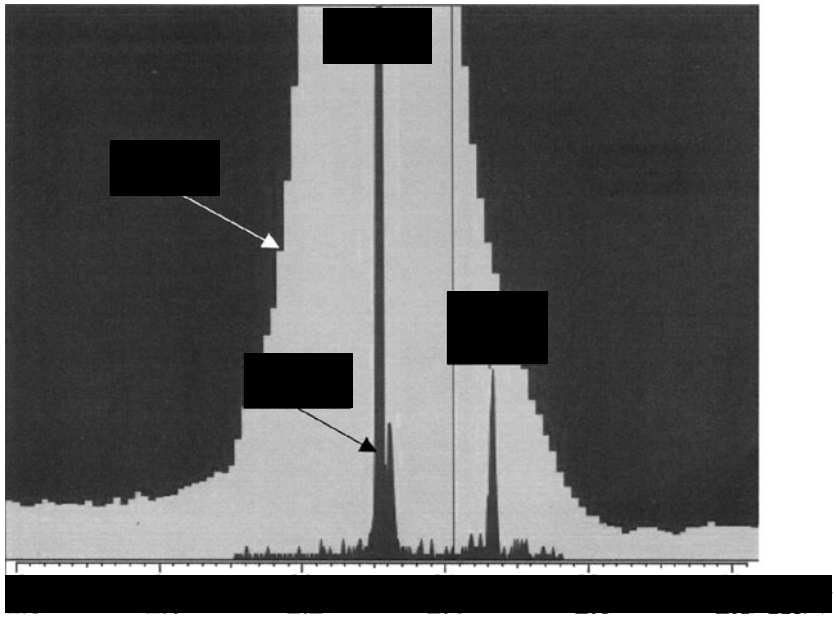


Figure 7.38. Display showing comparison of WDS and EDS spectra of sulfur plotted on the same energy scale with intensities normalized to fit on the same scale. (Courtesy of Oxford Instruments.)

and beam position information on an electron image so that the analyst can readily interpret the results. An example of a line scan is given in Fig. 7.40. Data presented in this manner give qualitative information because a complete quantitative evaluation requires conversion of the intensity data to chemical composition by one of the mathematical methods described in Chapter 9. Furthermore, because deflection of the beam can lead to defocusing of the x-ray spectrometers, quantitative line scans can be performed by stepping the sample under a static beam. Two-dimensional scanning, known as x-ray mapping, described in detail in Chapter 10, involves taking the output of the SCA and using it to modulate the brightness of the CRT during normal secondary-electron raster scanning. Each x-ray photon

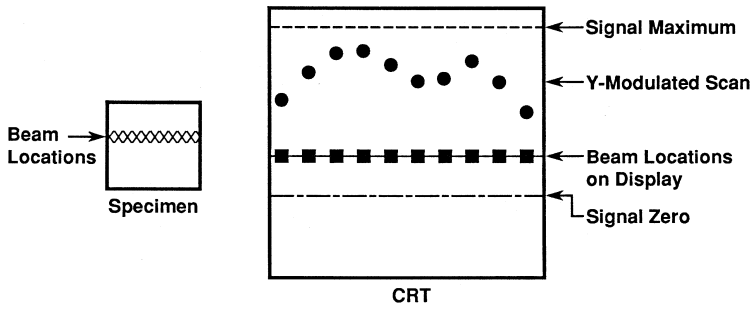


Figure 7.39. Schematic diagram showing how line scans are made.

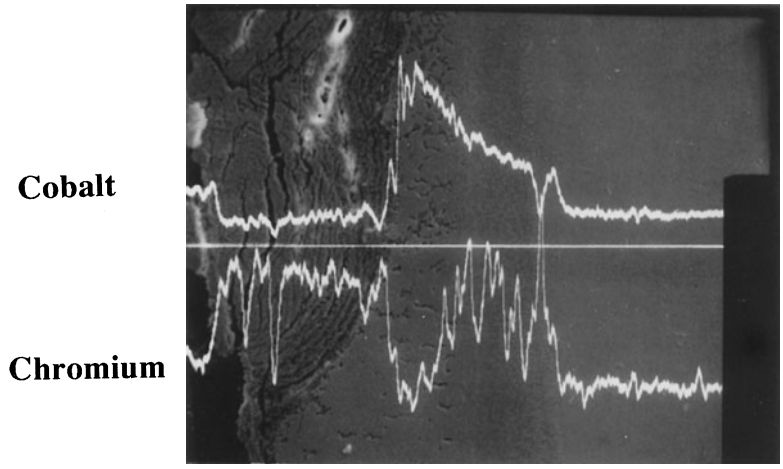


Figure 7.40. Co and Cr line scans across an oxidized high-temperature alloy. Straight line represents the position of the scan on the secondary-electron image (top portion represents cobalt; bottom, chromium).

detected appears as a dot on the CRT, with regions of high concentration characterized by a high dot density.

7.4. Comparison of Wavelength-Dispersive Spectrometers with Conventional Energy-Dispersive Spectrometers

An ideal x-ray detector would be small, inexpensive, easy to operate, collect most of the x-rays emitted from a sample, have a resolution of the natural x-ray linewidth being measured (a few electron volts), and be capable of collecting spectral data rapidly without losing information. Neither wavelength-dispersive spectrometers nor Si(Li) or HPGe detectors individually have all of these characteristics, but when used together the two techniques do, in fact, complement each other. Table 7.2 summarizes a comparison of the major features of both modes of detection. An item-by-item analysis of Table 7.2 follows.

7.4.1. Geometric Collection Efficiency

Geometric collection efficiency refers to the solid angle of spectrometer acceptance, $(\Omega/4\pi) \times 100\%$. As was illustrated in Fig. 7.26, the angle subtended in the plane of the focusing circle of a WDS does not change with λ. However, orthogonal divergence perpendicular to that plane will lead to a decreased collection efficiency with increasing λ for a given crystal. In the EDS case the higher geometric collection efficiency is a result of the greater ease in positioning the detector close to the sample (often less than 1 cm). Furthermore, although the solid angle can be varied for retractable detectors (see Fig. 7.3), the nature of the detection process does not require

Table 7.2. Comparison between Types of X-Ray Spectrometers

Operating characteristic	WDS (crystal diffraction)	EDS (silicon, energy-dispersive)
Geometric collection efficiency	Variable, <0.2%	<2%
Overall quantum efficiency	Variable, <30%	~100% for 2–16 keV
	Detects $Z \geq 4$	Detects $Z \geq 10$ (Be window); detects $Z \geq 4$ (windowless or thin window)
Resolution	Crystal-dependent (5 eV)	Energy-dependent (130 eV at 5.9 keV)
Instantaneous acceptance range	~Spectrometer resolution	Entire useful energy range
Maximum count rate (counts/s)	50,000 on an x-ray line	Resolution-dependent, <4000 over full spectrum for best resolution
Minimum useful probe size (nm)	~200	~5
Typical data-collection time	Tens of minutes	Minutes
Spectral artifacts	Rare	Major ones include escape peaks, pulse pileup, electron-beam scattering, peak overlap, and window absorption effects

that the detector be physically moved to encompass its entire energy range as is the case for a WDS.

7.4.2. Quantum Efficiency

Overall quantum efficiency is a measure of the percentage of the x-rays entering the spectrometer that are counted. At low beam currents EDS systems generally have higher count rate capability than WDS per unit beam current partially due to a higher geometric collection efficiency and partially due to higher inherent detector quantum efficiency. With a detector using a thin window close to 100% of the x-rays in the 1.0- to 15-keV energy range striking the detector will be collected. At higher energies a certain percentage of x-ray photons will be transmitted through the silicon crystal, whereas at low energies a certain percentage of photons will be absorbed in the window. Significant absorption of soft x-rays can also occur in the surface "dead layer" or gold contact layer on the detector crystal. Absorption and detector noise generally limit meaningful light element analysis to $Z \geq 6$.

By comparison, the quantum counting efficiency of crystal spectrometers is generally less than 30%, partially due to transmission losses in the proportional counter tube (see Fig. 7.31) and partially due to losses in the diffraction crystal. Large-d-spacing crystals and thin-window proportional counters make it possible to detect elements down to beryllium. Consideration of the overall quantum counting efficiency leads to the conclusion that the useful energy range for WDS and EDS windowless or thin-window systems is usually about 0.1–15 keV.

7.4.3. Resolution

Resolution for both WDS and EDS systems is usually measured as the FWHM. As already described in Eq. (7.1) and illustrated in Fig. 7.15, EDS resolution is energy-dependent in a predictable way, with values normally determined for Mn $K\alpha$ as measured with a ^{55}Fe source. Even at 130 eV the resolution is still about 30 times poorer than that obtainable with WDS using a good quartz or lithium fluoride crystal. The principal effect of reduced resolution is to lower the peak-to-background ratio P/B at a given energy and hence the sensitivity or minimum detectability limit of a given element (see Chapter 9 for a discussion of trace element detection). The reduction in P/B occurs because it is necessary to sample a wider energy interval containing more background counts to obtain a major fraction of the peak counts P. Figure 7.41, taken from the work of Geller (1977), illustrates this point for the case of a 160-eV detector, used to obtain P and P/B for Fe $K\alpha$ and Si $K\beta$ on pure iron and silicon, respectively. The term P^2/B is often used for comparison of sensitivities for the detection of trace amounts of a particular element. The value of P^2/B goes through a maximum when the energy band chosen for peak integration is approximately equal to the FWHM. Note that the peak-to-background ratio must be obtained by integrating both the peak and background counts in the region of interest and taking the ratio of the sums. This quantity is known as the integrated peak-to-background

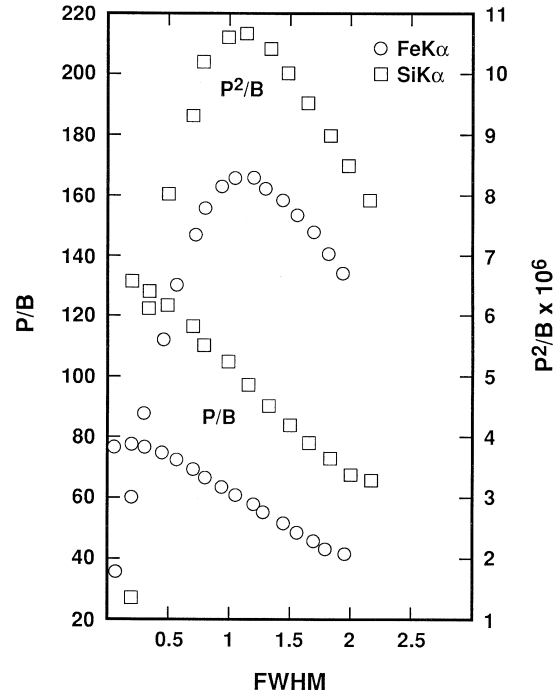


Figure 7.41. Integration limits and their effects on P/B and P^2/B for Fe and Si measurements with an EDS. (From Geller, 1977.)

ratio. The situation is somewhat different in evaluating the performance of a WDS system because optimum performance is obtained by tuning the spectrometer to a diffraction angle corresponding to the maximum peak intensity. Then the spectrometer is detuned to obtain corresponding background levels.

Integrating WDS peaks by scanning through them is basically an exercise in seeing the system's response to a variation of the spectrometer efficiency. Scanning in the vicinity of a peak may have to be used, however, as a means of determining the position of a maximum for those quantitative measurements particularly if the spectrometer is detuned between sample and standard readings. Integrated peak intensities may, in fact, be required in the analysis of the light elements in those cases where the characteristic peaks are broadened. Unlike the EDS, integrating WDS peaks does not add to the precision, but is a means of avoiding the introduction of systematic errors.

The superior energy resolution of the crystal spectrometer results in significantly higher peak-to-background ratios and better spectral dispersion, thereby minimizing the possibility of peak overlap. This can be readily seen by comparing spectra from the same superalloy standard obtained using a crystal spectrometer (Figs. 7.36a and 7.36b) with that obtained from a state-of-the-art (130 eV) Si(Li) detector (Fig. 7.42). Figure 7.36a shows clearly distinguished Ta $L\alpha$, Ni $K\beta$, and W $L\alpha$ lines, whereas these same lines are hardly discernible in Fig. 7.42. Similarly, the Ta, W, and Re $M\alpha$ lines are easily separated using TAP crystal, but remain unresolved with the Si(Li) detector. The inferior resolution of the solid state detector often makes it necessary to establish the presence of a series of spectral lines for a given element when the identification of a particular peak is ambiguous

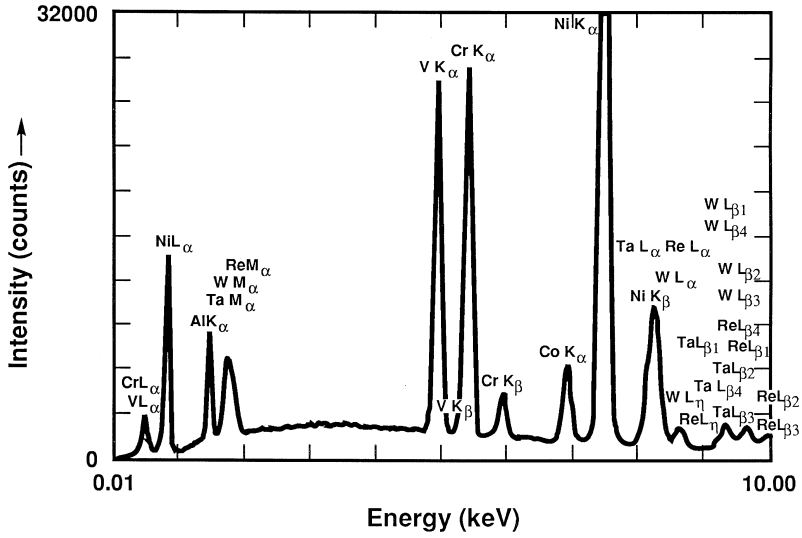


Figure 7.42. EDS spectrum of the same high-temperature superalloy analyzed by WDS in Fig. 7.36.

or the peak of a suspected line is obscured by another element. In such cases it is common practice to use the line markers or collect one or more pure elemental spectra and display them simultaneously with that of the unknown in order to make a direct comparison.

7.4.4. Spectral Acceptance Range

The term "instantaneous acceptance range" refers to that portion of the spectrum that can be measured at any instant of time. For the WDS, we have a serial detection system where only those pulses very close to the selected Bragg angle will be measured while all others are essentially ignored. The EDS, on the other hand, is a parallel detection system with a large acceptance range. Photons of any energy within a broad acceptance range (typically from a few hundred eV to more than 10 keV) will be measured as long as they are sufficiently separated in time from preceding or subsequent pulses so as not to cause overlaps that would interfere with accurate energy measurement.

7.4.5. Maximum Count Rate

As already shown in Fig. 7.9, the maximum useful output count rate for systems operating at optimum resolution is about 2000–3000 counts/s for the total counts collected over the entire energy range of the excited x-rays (0 to E_0). If the composition of a minor component is sought, this number may drop well below 100 counts/s for the element of interest and therefore necessitate long counting times to achieve needed levels of precision. In the WDS with the spectrometer tuned to a specific element, count rates in excess of 50,000 counts/s at the peak energy are possible without loss of energy resolution.

7.4.6. Minimum Probe Size

The previous point leads directly to a discussion of the minimum useful probe size for x-ray analysis. As described in detail in Chapter 2, Fig. 2.26, for each type of electron source and beam voltage there exists a probe current that can be associated with a given probe diameter. For conventional tungsten thermionic sources the probe current varies with the eight-thirds power of the beam diameter [Eq. (2.21)] with typical values at 20 kV ranging from 10 pA for a 20-nm (200-Å) probe diameter, to 2.5 nA for a 100-nm (1000-Å) probe diameter, to 1 μA for a 1000-nm (10,000-Å) probe diameter. For an EDS system utilizing a 4-mm-diameter detector placed 1 cm from a sample of pure nickel, a count rate of about 10^4 counts/s is obtained for a 35° take-off angle, a 20-nm (200Å) probe, and 100% quantum efficiency. According to Fig. 7.9, if the maximum energy resolution (largest time constant) was desired, the input count rate of 10^4 counts/s would give an output count rate of about 3000 counts/s. This would correspond to a 70% dead time, which would be too high to expect a good linear response from the amplifier. Therefore either the energy resolution would have to be

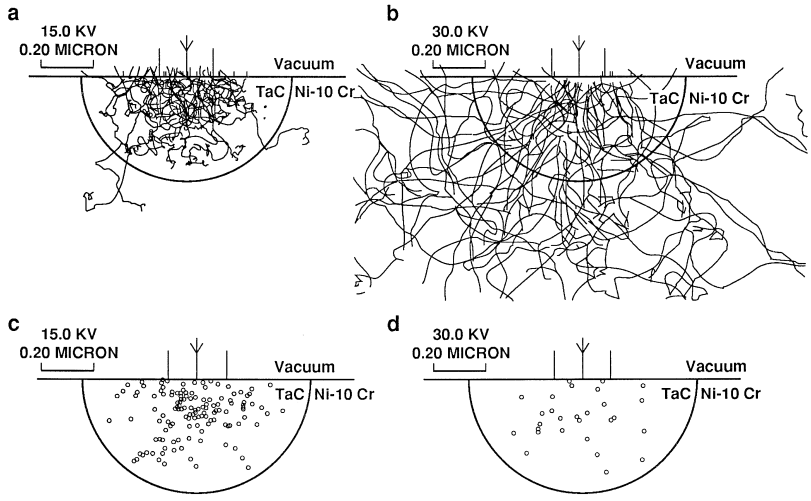


Figure 7.43. Monte Carlo-simulated interaction of a 0.2-μm-diameter beam with a hypothetical 1.0-μm-diameter hemispherical TaC inclusion in a NiCr matrix. (a) Electron trajectories, 15 keV; (b) electron trajectories, 30 keV, (c) Ta $M\alpha$ x-rays at 15 keV, (d) Ta $M\alpha$ x-rays at 30 keV.

decreased (lower the time constant) or the rate would have to be reduced by either retracting the detector or lowering the beam current.

On the other hand, the corresponding WDS count rate for a 10-pA probe current would probably be more like 100 counts/s, which would be too low for practical use. For bulk samples (more than a few micrometers thick) spatial resolution for chemical analysis does not improve for probes much less than 0.5 μm in diameter because the volume of x-ray production is determined by electron beam scattering and penetration rather than by the probe size at beam energies of 10 keV and above. This point is dramatically demonstrated in Fig. 7.43, which shows a series of Monte Carlo simulations of electron beam scattering and x-ray production for a 200-nm (0.2-μm)-diameter probe with a hypothetical l-μm TaC inclusion in a Ni–Cr matrix. The electron trajectories and consequently the region of x-ray production, particularly above 15 keV, can easily exceed 1 μm, or five times the beam diameter. Increasing the probe current by changing the condenser lens setting would also increase the probe diameter, but the net effect for small probes would only be a slight increase in the x-ray excitation volume. A significant increase in the probe current might allow feasible WDS analysis. Section 7.5.3 will discuss other opportunities for WDS analysis at low beam voltages using brighter sources and parallel-beam x-ray spectrometers.

The situation is quite different, however, in the examination of thin foils and biological sections where spatial resolution equal to or even smaller than the foil thickness is possible. In this case the low x-ray yields necessitate having a detector with both the high geometric collection and overall quantum efficiency of the EDS system. It is for this reason that the EDS has been so successful when coupled to both scanning and analytical electron microscopes.

7.4.7. Speed of Analysis

From a practical point of view one of the best features of the EDS system is the speed with which data can be collected and interpreted. The continuous acceptance of a large energy range has distinct advantages for qualitative analysis, which partially offset some of the disadvantages previously described. When a WDS system is mechanically scanned it dwells at each resolvable wavelength for only a short part of the total scan duration. Therefore while looking at one element or even a section of the background it is throwing away information about all of the other elements. Unless the WDS is specifically programmed to go to peak positions, anywhere from only 1/100 to 1/1000 of the total data collection time may be associated with the measurement of each individual peak. In the EDS case a 100-s counting time coupled with a count rate of 2000 counts/s leads to a spectrum containing 200,000 counts. Even if half of those counts are background, most of the measurable constituents present in amount greater than a few tenths of a percent will probably be detected. Furthermore, use of the line markers and other interpretive aids can result in a qualitative analysis in a matter of minutes. In the WDS case several crystals covering different wavelength ranges would be employed with typical data collection and interpretation times being 5–30 min.

7.4.8. Spectral Artifacts

The WDS system is relatively free of spectral artifacts of the type that cause peak position shifts or incorrect elemental assignments. The EDS, on the other hand, is subject to a number of difficulties that can lead the unsuspecting analyst into trouble. Artifacts arise at each stage of the spectral measurement process. Detection artifacts include peak broadening, peak distortion, silicon escape peaks, absorption, and the silicon internal fluorescence peak. Pulse-processing artifacts include pulse pileup, sum peaks, and sensitivity to errors in dead-time correction. Additional artifacts arise from the EDS–microscope environment, including microphonics, ground loops, and oil and ice contamination of the detector components. Both WDS and EDS can be affected by stray radiation (x-rays and electrons) in the sample environment, but because of its much larger solid angle of collection, the EDS is much more readily affected by stray radiation. However, because of the large collection angle, the EDS is less sensitive to spectrometer defocusing effects with sample position.

In summary, this comparison suggests that the strengths of the EDS and WDS systems compensate for each other's weaknesses. The two types of spectrometers are thus seen to be complementary rather than competitive. At the present stage of development of x-ray microanalysis, it is clear that the optimum spectrometry system for analysis with maximum capabilities is a combination of the EDS and WDS systems. As a result of the on-going revolution in the development of laboratory computers, automated systems which efficiently combine WDS and EDS spectrometers are available from several manufacturers.

After many years when the principal focus in spectrometer development concentrated on incremental improvements in existing technologies a number of recent developments have resulted in more significant departures in detector concepts and design. These include x-ray microcalorimetry, silicon drift detectors, and the use of flat-crystal spectrometers in conjunction with capillary optics.

7.5.1. X-Ray Microcalorimetery

The x-ray microcalorimeter is an energy-dispersive detector that works by measuring the temperature increase that results from the absorption of each individual x-ray photon. Because the heat liberated by each interaction is extremely small, and it must be distinguished from all other nearby heat sources, the fabrication and successful operation of such a detector is a great technical challenge that only recently has been realized (Wollman *et al.,* 1997). The major advantage of this approach is that it offers energy resolution equal to or better than WDS, yet it does not require scanning through the spectrum and is therefore similar to EDS. The basic operation of the spectrometer is illustrated in Fig. 7.44. The x-ray absorber is a thin, square film (e.g., a silver patch 250 μm $\times$ 250 μm $\times$ 2 μm thick). It is mounted in contact with a transition-edge sensor (TES). The TES is a superconducting bilayer film (e.g., Al/Ag) held at its critical transition temperature (i.e., the temperature at which the material switches between being superconducting and being a normal conductor). The TES establishes a stable equilibrium in the electrothermal circuit consisting of a voltage source, an inductor, the Ag absorber, and the TES. When an x-ray photon is absorbed the resistance of the TES increases by an amount that can be related to the energy of the incident x-ray photon. As the resistance changes, any current flowing through it will change correspondingly. The change in

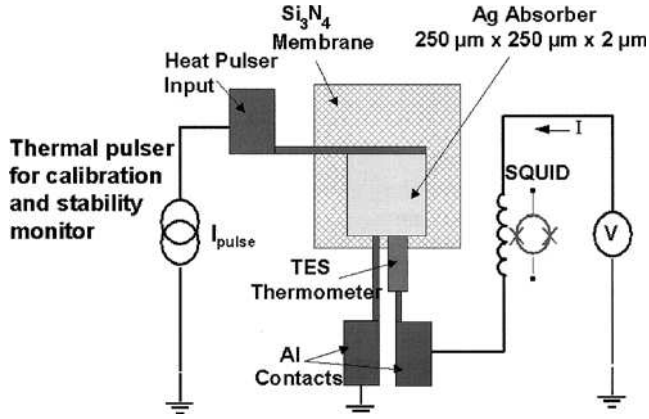


Figure 7.44. Schematic operation of an x-ray microcalorimeter. (Wollman *et al.,* 1997.)

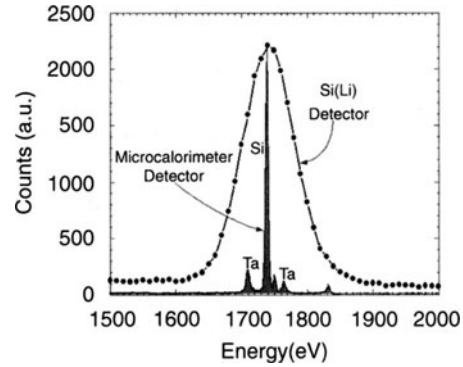


Figure 7.45. Comparison of a conventional silicon EDS spectrometer with a microcalorimeter EDS spectrometer for a thin film of TaSiN.

current is measured at the inductor as a changing magnetic field with a superconducting quantum interference device (SQUID), which is a very sensitive current amplifier. Subsequent circuitry is then used to relate the current pulse to the x-ray energy and the results can be stored in a CXA similar to what was described for EDS.

Figure 7.45 shows a comparison of a conventional silicon EDS spectrometer with a microcalorimeter EDS spectrometer for a thin film of TaSiN. Because the energy resolution of the microcalorimeter is better than 10 eV, the Ta and Si peaks are easily separated, whereas the peaks all appear as one in the case of the conventional EDS. Recent data indicate that energy resolutions of greater than 2 eV are possible with microcalorimeters and therefore future availability of this method is highly desirable.

There are, however, a number of development challenges that must be met before they will be in broad use. First, they must operate at temperatures that can only be reached with very sophisticated refrigeration systems. The approach used by the National Institutes of Standards and Technology (NIST) team involves a combination of liquid nitrogen and helium dewars with an adiabatic demagnetization refrigerator (ADR). Another approach in use is a pulse tube based refrigerator. A second issue is the development of window systems that will allow x-ray pulses to enter the detector, but provide infrared blocking. A third and very critical issue is count rate capability. As with all detectors, the system must fully process a pulse before accepting another. In this case this means that the detector must return to its initial operating temperature, which is done by providing a thermal conductance path for the heat flow from the absorber. The time required for this process is a limiting factor and the current overall system count rates possible are only about 500 counts/s. This problem may eventually be overcome with the use of arrays of detectors. The use of arrays may also help deal with another problem, that of collection efficiency. Because the absorbers are small and cannot be located too close to the sample, the solid angle for collection can be quite small. This problem can be

overcome, in part, by the use of capillary optics; however, the use of such optics adds complexity to the alignment process and restricts the x-rays that can easily be detected to 10 keV and preferably lower. Another concern is the physical size and cost of microcalorimeters, which currently are much larger, heavier, and costlier than silicon or HPGe EDS systems. The technology is still in its infancy and the attractive features it offers should spur the research necessary to overcome the limitations just described.

7.5.2. Silicon Drift Detectors

Although the improved energy resolution of the microcalorimeter over the conventional silicon EDS is a very desirable feature, there are often situations where high-count-rate capability can be equally important. X-ray mapping is a good example. Struder *et al.* (1999) describe in detail advances in a silicon-based detector technology that were first introduced more than a decade earlier. These detectors are referred to as silicon drift detectors (SDDs) and the basic concept of how they work is shown schematically in Fig. 7.46. The device consists of an *n*-type silicon wafer that has concentric shallow rings of *p*-type material starting at an anode ring close to the center of the wafer. An electric field gradient is produced by applying a potential difference between ring 1 and the outermost ring. When an x-ray photon enters the detector (from the bottom, i.e., non-patterned, surface as shown) electron–hole pairs are produced as in the case of the conventional silicon detector and the electrons are attracted to the anode. The charge on the anode is then coupled to the gate of an FET physically fabricated in the center of the detector, where the initial stage of pulse amplification occurs. The major advantage of this type of detector over the conventional silicon detector is the small value of the anode capacitance, which gives rise to faster rise time, larger amplitude pulses. What this means in practice is that much higher count rates are possible and the detector performs well even at room temperature, where the energy

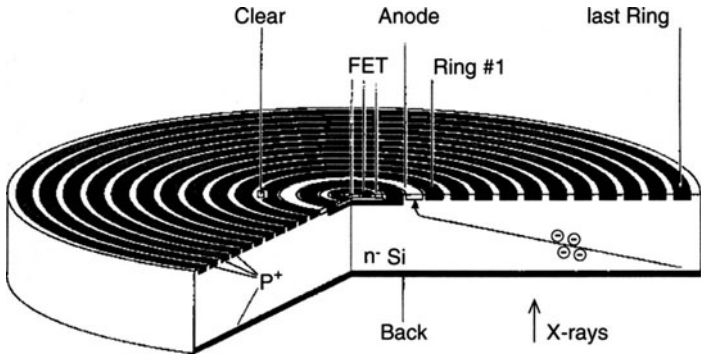


Figure 7.46. Schematic operation of a silicon drift detector with integrated FET preamplifier; active area = 5–10 mm^2. (Courtesy of Rontec Instruments.)

resolution is about 170 eV at Mn $K\alpha$. Use of a Peltier cooler to -30°C can improve the resolution to 130–135 eV at low count rates. Although there is a loss of resolution at higher count rates, if peaks are not so close as to cause overlap problems, very high count rates have been achieved. For example, input count rates for Mn $K\alpha$ of 250,000 counts/s have been measured with a detector with 215-eV resolution. The corresponding output count rate was still about 100,000 counts/s. As shown in Fig. 7.47, the spectra obtained with these detectors are very similar to those obtained with conventional EDS systems. The main limitation of the technique is the limited sensitivity for elements with low atomic numbers ($Z < 11$) due to low-energy noise. The value of high count rates for x-ray mapping is clearly illustrated in Fig. 7.48, which shows the extremely detailed maps that are possible when large numbers of counts are collected. Even higher count rates, in excess of a 1,000,000 counts/s for extremely fast mapping, will also be possible in the near future as arrays of detectors are developed (The collection area of current individual detectors are about 5–10 mm^2.).

7.5.3. Parallel Optic Diffraction-Based Spectrometers

One of the principal disadvantages of the curved-crystal WDS spectrometers described previously is that they have a very small collection angle because of design constraints. Therefore it may be difficult or impossible to get a statistically meaningful x-ray signal when beam currents of 1 nA or less are used, particularly at low electron beam voltages. Low voltages are essential for optimizing spatial resolution by decreasing the volume of x-ray excitation. This is unfortunate because x-ray production efficiency goes up with increasing beam voltage. Gains in collection efficiency can be made, however, with the use of capillary and reflective optical techniques. The devices, which essentially act as a light pipe for x-rays, can be used to collect x-rays from the point source (electron beam impact point) and then steer the collected x-rays such that they emerge as a parallel beam from the x-ray optics. This beam can then be reflected from flat analyzing crystals because the angle of incidence is constant across the crystal. The crystal can also be located at a fixed distance many centimeters from the source (see Fig. 7.49). The crystal set is similar to that used with any WDS system except that the crystals are flat. Conventional tungsten, LaB_6, and cold field emitters are generally not capable of putting more than 1 nA in a focused probe of 100 nm or less at voltages below 5 keV. Therefore, brighter high current sources such as Shottky field emission systems may be essential to providing the electron fluxes and stability necessary for high-spatial-resolution *in situ* analysis. Figure 7.50 demonstrates an x-ray scan with a parallel diffraction spectrometer run over a low-energy portion of the spectrum for a specimen of AlGaAs at 5 kV and 5 nA collected in about 10 min. Under these conditions the probe size is about 100 nm, and the excitation volume is still several times larger, but still small compared to the >1000-nm resolution expected when higher beam voltages are used.

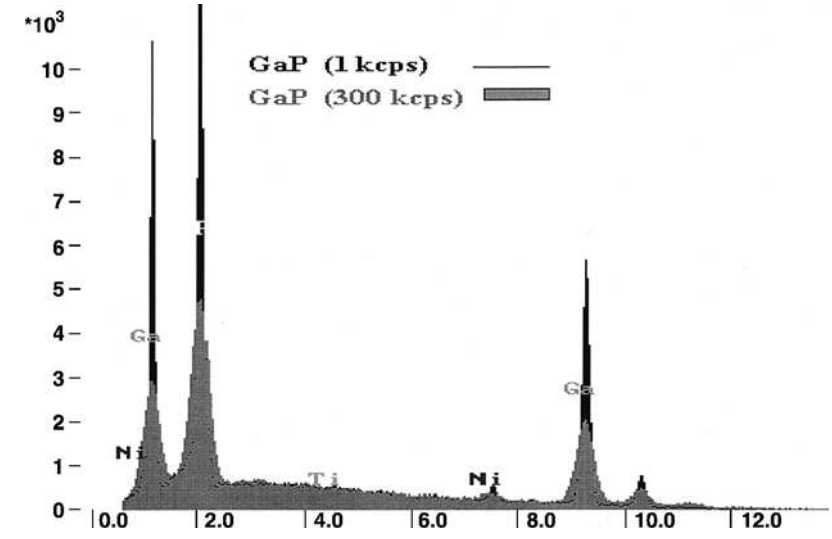


Figure 7.47. Spectra obtained with a silicon drift detector (300 kcps) and a Si (Li) detector. (1 kcps). (Courtesy of Rontec Instruments.)

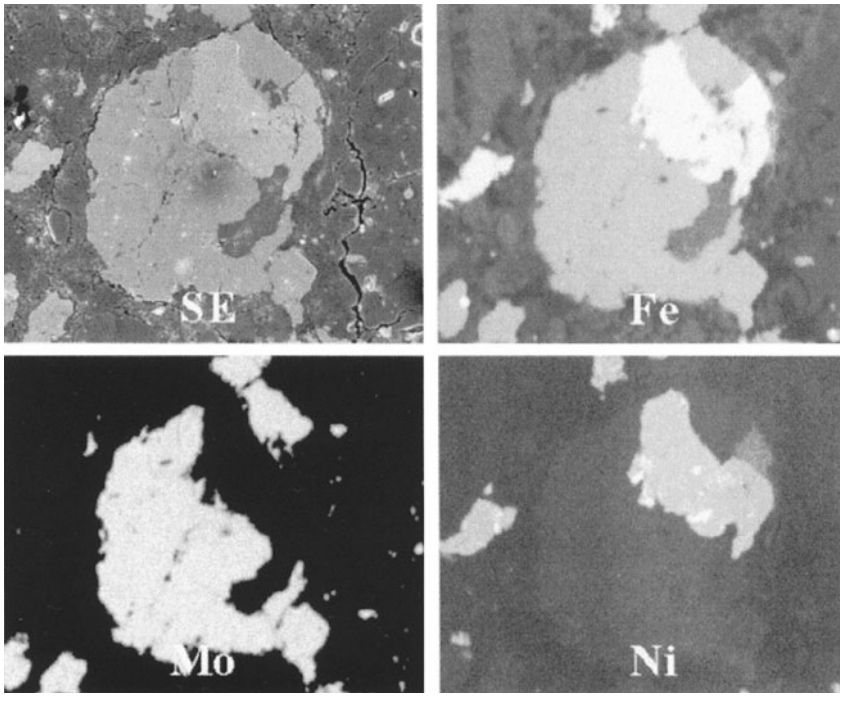


Figure 7.48. X-ray map of iron alloy obtained at very high count rate using a silicon drift detector. (Courtesy of Rontec Instruments.)

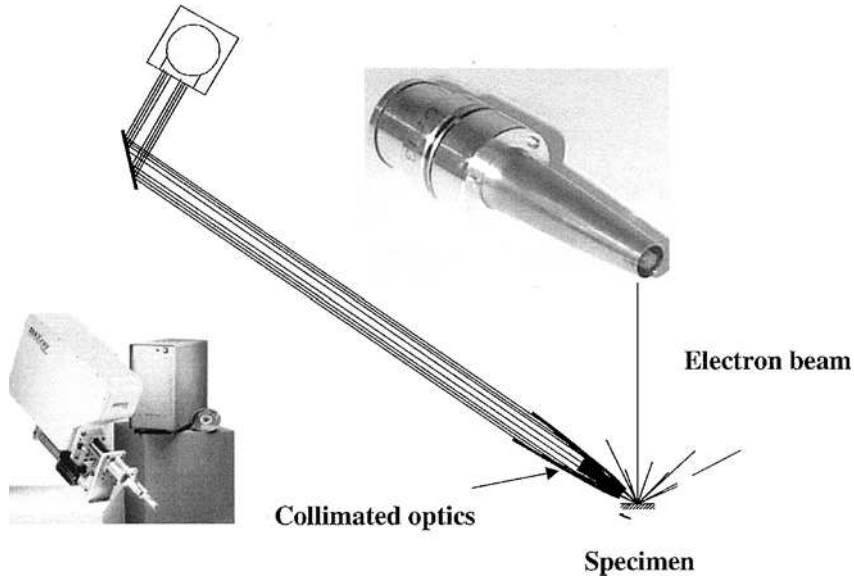


Figure 7.49. Schematic representation of parallel-beam x-ray spectrometer. (Courtesy of ThemoNORAN Instruments.)

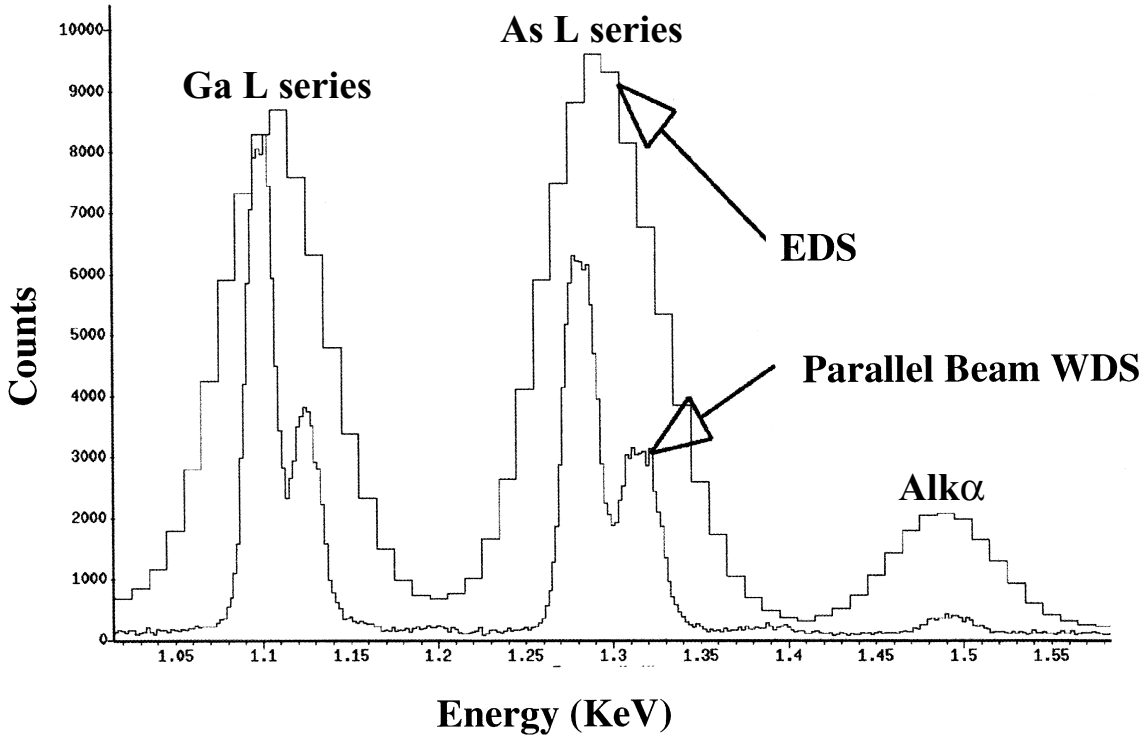


Figure 7.50. Spectral scans of AlGaAs comparing a parallel-beam x-ray spectrometer with results from a conventional silicon EDS system.

Blodgett, K. B., and I. Langmuir (1937). *Phys. Rev*. **51,** 964.
Elad, E., C. N. Inskeep, R. A. Sareen, and P. Nestor (1973). *IEEE Trans. Nucl. Sci.* **20,** 354.
Fiori, C. E., and D. E. Newbury (1978). In *SEM/1978/I,* SEM, Inc., AMF O'Hare, Illinois, p. 401.
Fitzgerald, R., K. Keil, and K. F. J. Heinrich (1968). *Science* **159,** 528.
Freund, H. U., J. S. Hansen, S. J. Karttunen, and R. W. Fink (1972). In *Proceedings 1969 International Conference on Radioactivity and Nuclear Spectroscopy* (J. H. Hamilton and J. C. Manthuruthil, eds.), Gordon & Breach, New York, p. 623.
Geller, J. D. (1977). In *SEM/1977/I, IIT Research Institute,* Chicago, p. 281.
Goulding, G. S. (1972). *Nucl. Instr. Meth.* **100,** 493.
Holliday, J. E. (1963). In *The Electron Microprobe* (T. D. McKinley, K. F. J. Heinrich and D. B. Wittry, eds.), Wiley, New York, p. 3.
Lifshin, E., and M. F. Ciccarelli (1973). In SEM/1973 IIT Research Institute, Chicago, Illinois., p. 89.
Mott, R. B., and J. J. Friel (1995). In *X-ray Spectrometry in Electron Beam Instruments* (D. B. Williams, J. I. Goldstein, and D. E. Newbury, eds.), Plenum Press, New York, p. 127.
Pell, E. M. (1960). *J. Appl. Phys*. **31,** 291.
Reed, S. J. B. (1975). *Electron Microprobe Analysis,* Cambridge university Press, Cambridge.
Struder, L., N. Meidinger, D. Stotter, J. Kemmer, P. Lechner, P. Leutenegger, H. Soltau, F. Eggert, M. Rohde, and T. Schulein. (1999) *Microsc. Microanal*. **4,** 622.
Woldseth, R. (1973). *X-ray Energy Spectrometry,* Kevex Corp., Foster City, California.
Wollman, D. A., K. D. Irwin, G. C. Hilton, L. L. Dulcie, D. E. Newbury, and J. M. Martinis (1997). *J. Microsc*. **188,** 196.

8

Qualitative X-Ray Analysis

8.1. Introduction

The first stage in the analysis of an unknown is the identification of the elements present, a process referred to as qualitative analysis. Qualitative analysis is itself a powerful tool in microanalysis. When a simple understanding of the nature of a chemical microstructure is sought, often the elemental identification accompanied by a broad classification of concentration into categories (such as "major," "minor," or "trace") is sufficient to solve many problems. Note that the terms "major," "minor," and "trace" as applied to the concentration of constituents of a sample are not strictly defined and are therefore somewhat subjective. Each analytical method tends to define "trace" differently based upon its degree of fractional sensitivity. Because the energy dispersive x-ray spectrometer (EDS) limit of detection in bulk materials is about 0.1 wt%, the following arbitrary working definitions will be used in this text:

- *Major:* more than 10 wt% ($C > 0.1$ mass fraction)
- *Minor:* 1–10 wt% ($0.01 \leq C \leq 0.1$)
- *Trace:* less than 1 wt% ($C < 0.01$)

Note that the definition of trace has no bottom limit. Generally, wavelength-dispersive x-ray spectrometry (WDS) can achieve limits of detection of 100 ppm in favorable cases, with 10 ppm in ideal situations where there are no peak interferences and negligible matrix absorption (see Chapter 9, Section 9.8.4).

Qualitative x-ray analysis is often regarded as straightforward, meriting little attention. The reader will find far more references to quantitative analysis than to qualitative analysis, which has been relatively neglected in the literature, with a few exceptions (e.g., Fiori and Newbury, 1978). It is clear that the accuracy of the final quantitative analysis is meaningless if the elemental constituents of a sample have been misidentified. As a

general observation, the identification of the major constituents of a sample can usually be done with a high degree of confidence, but, when minor or trace-level elements are considered, errors in peak assignments can arise unless careful attention is paid to the problems of spectral interferences, artifacts, and the multiplicity of spectral peaks observed for elements of intermediate and high atomic number. Because of the differences in approach to qualitative EDS and WDS analysis, these techniques will be treated separately.

Virtually all commercial x-ray microanalysis software systems include an automatic qualitative analysis function (further discussion below). The careful analyst will certainly make use of this automatic function, but will always confirm the identifications suggested by the automatic qualitative analysis, particularly for low-intensity peaks likely to represent minor and trace-level constituents. To achieve this degree of skill, the analyst needs to learn the strategy for performing a qualitative analysis manually. By carefully examining spectra, first of simple and then increasingly complex mixtures of elements, and making manual peak identifications, the analyst can develop a high degree of skill in qualitative analysis.

An accurate qualitative x-ray analysis depends critically upon several different types of information. First, because the energy of the peak position is the critical information needed to identify the element, the EDS must be calibrated to deliver peak positions within 10 eV for photon energies above 1 keV. Below 1 keV, even a well-calibrated EDS may show some nonlinearity in peak position due to incomplete charge collection and photon coincidence.

With a properly calibrated EDS, the specific energy of each characteristic x-ray peak is measured and entered into a database. Each measured peak energy is compared with reference peak information most likely provided to the analyst as a software function on the computer-based x-ray spectrum analyzer (CXA), where the output is likely given by a visual display marker ("*KLM* markers"). For very old systems, the analyst may find an "energy slide rule" or a graphical presentation of peak energies. When employing any of these aids, the analyst should be aware of some of the potential shortcomings. The graphs, energy slide rules, and even some computer "*KLM*" markers may not list all x-ray peaks observed in practice. (Some software systems inexplicably permit the user to exclude some potentially detectable x-ray peaks from consideration. This is an extremely dangerous practice with no obvious benefit. If the analyst is in a multi-user environment and this exclusion option exists, it should be defeated at the start of the work session as part of a personal measurement quality assurance plan.) Frequently, the $L\mathrm{l}$, $L\eta$, $M\zeta$, and $M_{\mathrm{II}}N_{\mathrm{IV}}$ peaks are neglected, but these peaks are easily detected when the parent is a major constituent. Without knowledge of the existence of these peaks in the reference database, errors may be made in ascribing these peaks to other elements which are not actually present in the specimen. A comprehensive table of x-ray peaks, such as Bearden's compilation (Bearden, 1967), is the ultimate reference, but for EDS qualitative analysis, this compilation

is often too detailed. For example, Bearden lists 25 *L*-family peaks. Even when the integrated spectral count is as high as 5,000,000, only 9 of these *L*-family peaks will usually be observed in high-quality EDS spectra. The remaining 16 *L*-family peaks are lost because of low relative intensity or the overlap of other members of the family. Note, however, that the higher resolution of WDS systems will increase the number of peaks that can be detected, requiring reference to a detailed x-ray compilation.

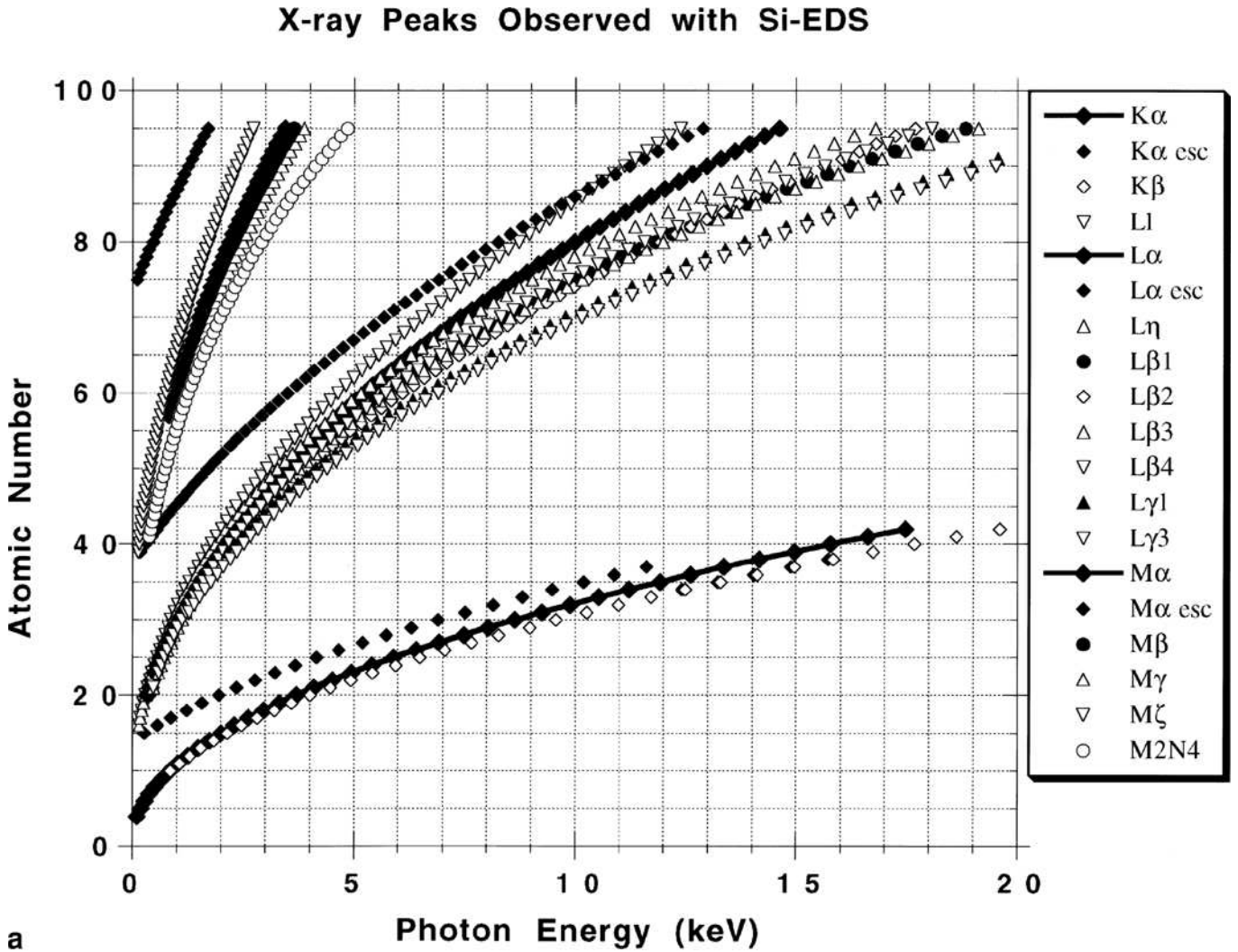
As an aid to qualitative EDS analysis, Fiori and Newbury (1978) published a graphical representation of all x-ray peaks observed in high-quality (5,000,000 counts integrated intensity) EDS spectra. This dataset is plotted Fig. 8.1a for the photon energy range from 0.1 to 20 keV. The range from 0.1 to 5 keV, appropriate for low-voltage microscopy and microanalysis, is expanded in Fig. 8.1b. These plots provide a convenient compilation of x-ray energies and also allow for a rapid visual evaluation of possible interferences. Figure 8.1 is intended as an aid to qualitative analysis in conjunction with computer-generated *KLM* markers for x-ray energies. The dataset used for Fig. 8.1 is available on the accompanying Enhancements CD in the Database Chapter.

The second important piece of information with which the analyst must be familiar is the concept of a family of x-ray peaks. When the beam energy exceeds the critical x-ray ionization energy for a shell or subshell so that it can be ionized, all possible transitions involving that shell vacancy will take place, producing a family of peaks except for the simplest of atoms ($Z \leq 9$, fluorine). These family structures will become more complicated as the electronic structure of the atom increases in complexity with increasing atomic number. With a beam energy of 20 keV or more, all possible peaks of an element in the range 0.1 to approximately 15 keV (depending on the value of the critical ionization potential for a particular element) will be excited with increasing efficiency as the photon energy decreases. Detecting the presence in the spectrum of all possible members of a family of peaks increases the confidence that can be placed in the identification of that element. Because the family members must all exist, the absence of a particular peak should immediately raise suspicion in the analyst's mind that a misidentification may have been made and other possible elements in that energy range should be considered. Most importantly, all members of the x-ray family that can occur must be carefully located and noted or suitably marked to avoid misidentification as another element.

8.2. EDS Qualitative Analysis

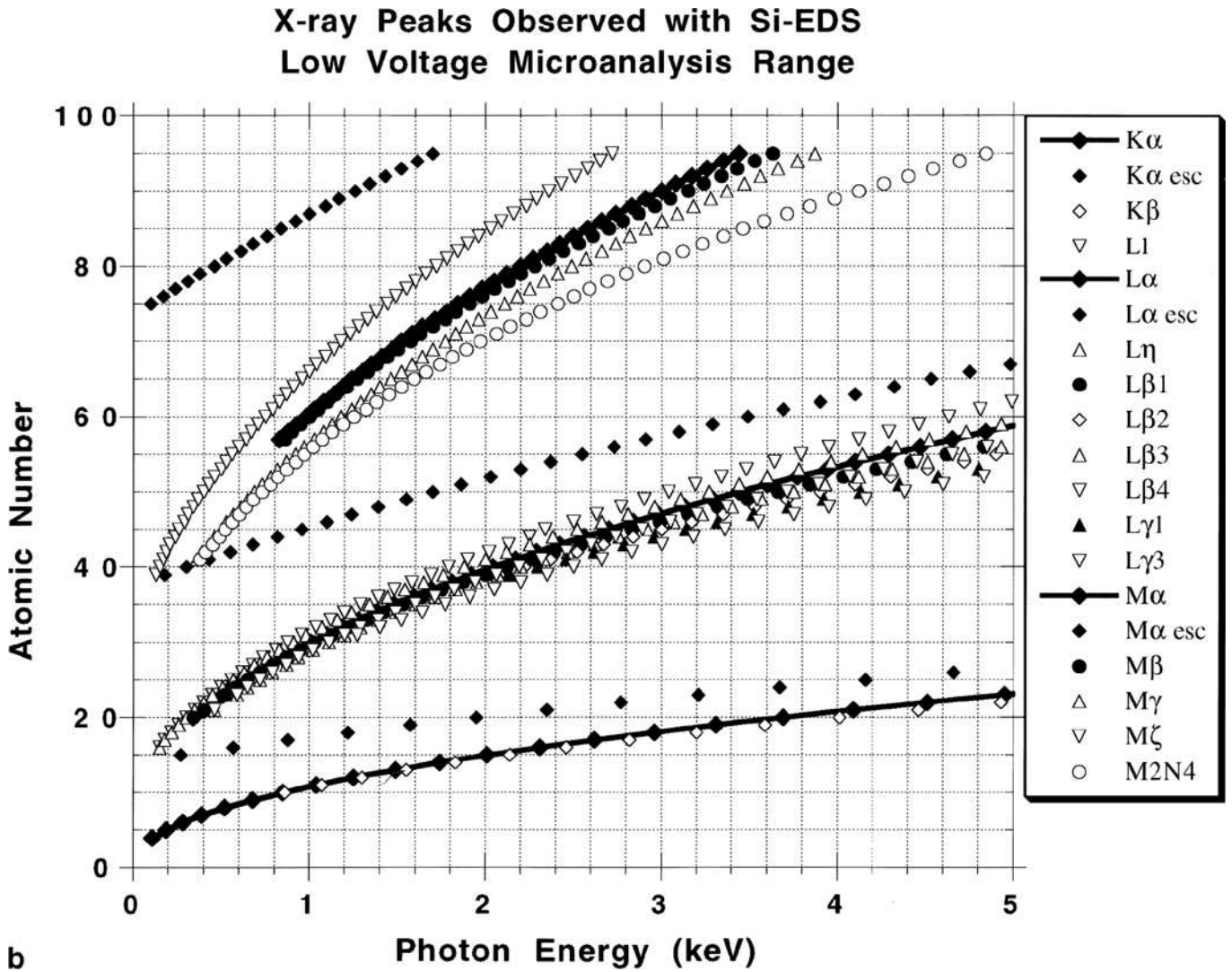
8.2.1. X-Ray Peaks

The energy-dispersive x-ray spectrometer is a powerful tool for qualitative x-ray microanalysis. The fact that the total spectrum of interest, from 0.1 to the beam energy (e.g., 20 keV) can be acquired in a short time (10–100 s) allows for a rapid evaluation of the specimen constituents.



a

Figure 8.1. Plot of all x-ray peaks observed by Si EDS (129 eV at Mn $K\alpha$) in high-count spectra. (a) Full analytical range, 0.1–20 keV.



b

Figure 8.1. (*Continued*) (b) Low-voltage microanalysis range, 0.1–5 keV.

Because the entire x-ray spectrum is always acquired at every specimen location analyzed, the information necessary to perform a complete qualitative analysis at every measured location is always available. Every complete EDS spectrum, rather than just intensities from preselected regions of interest, should always be stored in case it is later necessary to review the assignment of peaks. Moreover, by always having the complete EDS spectrum available, the analyst has the opportunity to recognize when an unexpected constituent appears, an invaluable feature when characterizing complex microstructures. Because the EDS detector has virtually constant efficiency (near 100%) in the range 3–10 keV, the relative peak heights observed for the families of x-ray peaks are close to the values expected for the signal as it is emitted from the sample. On the negative side, the relatively poor energy resolution of EDS compared to the WDS leads to frequent spectral interference problems as well as the inability to separate the members of the x-ray families that occur at low energy (<3 keV). Also, the existence of spectral artifacts peculiar to EDS, including the Si-escape peaks and sum peaks, increases the complexity of the spectrum, particularly when low-relative-intensity peaks are considered.

To aid in the identification of unknowns, it is useful to consider the appearance of the *K*, *L*, and *M* families in EDS spectra as a function of position in the energy range 0.1–15 keV. One of the most useful exercises a new entrant into the SEM/x-ray microanalysis field can do is to study x-ray spectra from pure elements and simple compounds. Learning to identify the general shape of families of peaks as a function of photon energy and the degree to which low-peak-intensity family members can be detected is a great help in systematically solving complex unknowns. In fact the shape of resolved and unresolved x-ray families is information that is generally not used by automatic software qualitative analysis systems.

The approximate weights of peak members in a family provide another source of important information in identifying elements. For photon energies above approximately 3 keV, the energy separation of the members of a family of x-ray peaks is large enough that, despite the peak broadening introduced by a Si EDS system, it is possible to recognize more than one peak, even if they are not completely separated. Typical *K*, *L*, and *M* shell peaks in the photon energy range where separation occurs for at least some family members are shown in Figs. 8.2 (Cu *K* family), 8.3 (Sn *L* family), 8.4 (Ta *L* family), and 8.5 (U *M* family), with x-ray intensity as the vertical axis and x-ray energy as the horizontal axis. A comprehensive line marker plot of family members showing the peak positions and approximate weights are also given in the EDS figures. The EDS *K* family consists of two recognizable peaks $K\alpha$ and $K\beta$. The ratio of the intensities of the $K\alpha$ peak to the $K\beta$ peak is approximately 10:1 when these peaks are resolved, and this approximate ratio should be apparent in the identification of an element. Any substantial deviation from this ratio should be viewed with suspicion as originating from a misidentification or the presence of a second element. The *L* series as observed by EDS consists of $L\alpha$ (l), $L\beta_1$ (0.7), $L\beta_2$ (0.2), $L\beta_3$ (0.08), $L\beta_4$ (0.05), $L\gamma_1$ (0.08),

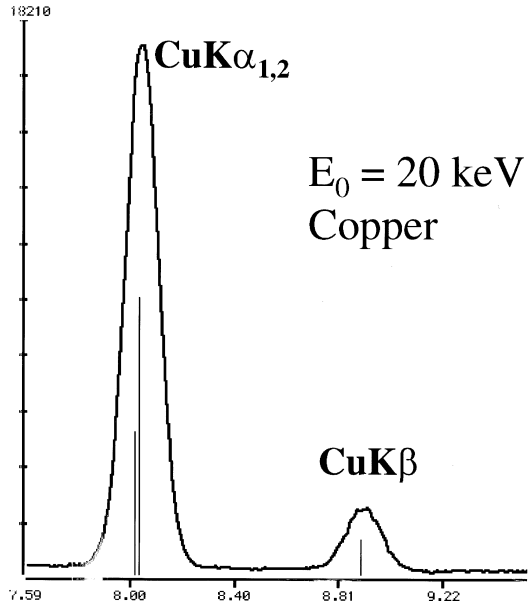


Figure 8.2. Copper *K* family showing separation of the Cu *Kα* and Cu *Kβ* peaks.

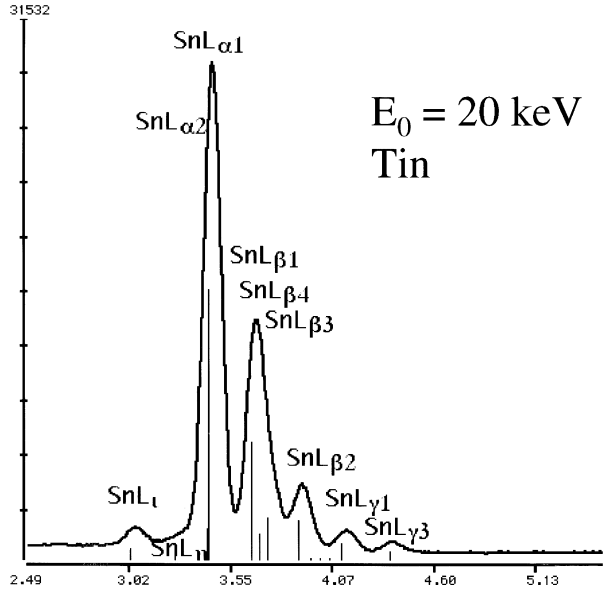


Figure 8.3. Tin *L* family showing incomplete separation of family members.

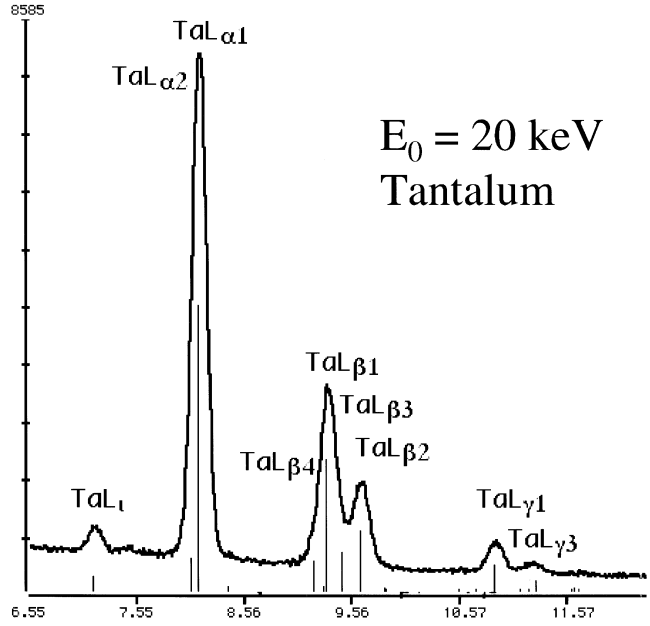


Figure 8.4. Tantalum *L* family showing greater separation of the family members; note that the *Lβ* series is not fully resolved.

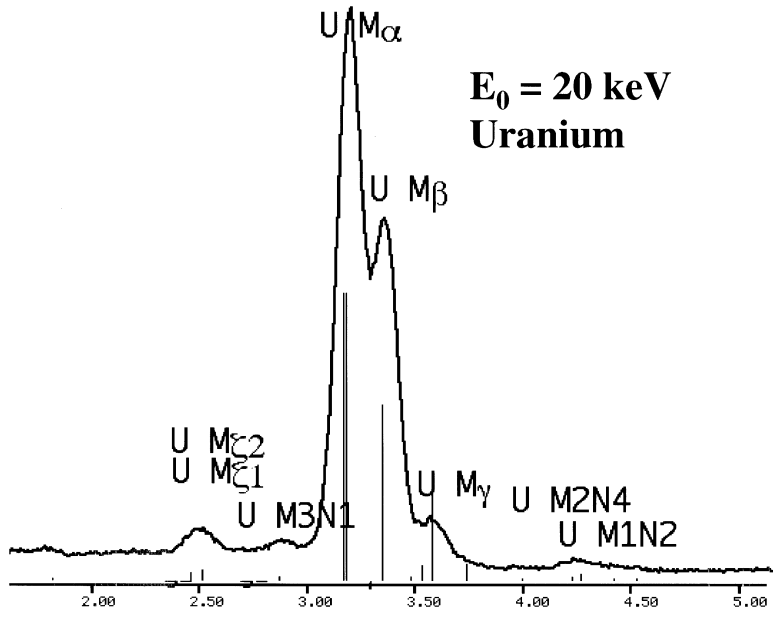


Figure 8.5. Uranium *M* family showing incomplete separation of main components.

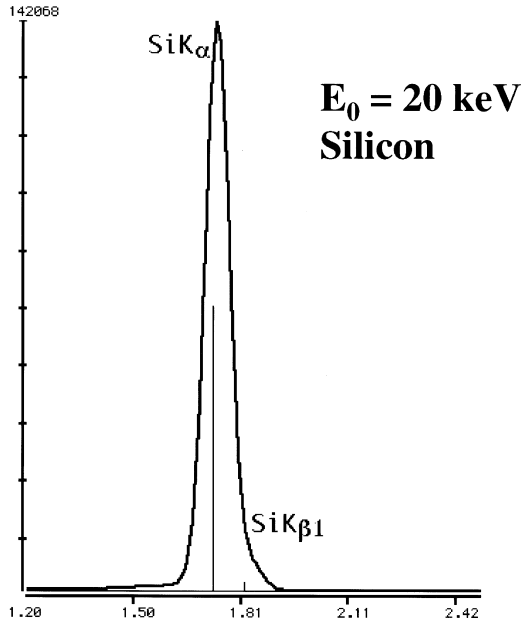


Figure 8.6. Silicon *K* family, showing only one peak with $K\alpha$ and $K\beta$ unresolved.

$L\gamma_3$ (0.03), Ll (0.04), and $L\eta$ (0.01). The observable EDS *M* series consists of $M\alpha$ (l), $M\beta$ (0.6), $M\gamma$ (0.05), $M\zeta$ (0.06), and $M_{II}N_{IV}$ (0.01), with some additional minor *M* peaks for the heaviest elements (Fig. 8.5). The values in parentheses give approximate relative intensities, but because these intensities vary with the element in question, with the overvoltage, and possibly also with chemical state, the weights must be considered only approximate.

Between 2 and 3 keV, depending on the particular element and the EDS resolution, the main family members α and β collapse into just one main peak as the energy separation of the members of the *K*, *L*, or *M* families becomes so small that the peaks fall within the resolution of the EDS system. The appearance of these families below 2 keV is illustrated in Figs. 8.6 (Si *K*, 1.74 keV), 8.7 (Y *L,* 1.92 keV), and 8.8 (Ta *M*, 1.71 keV). Note that the unresolved low-energy $K\alpha$ and $K\beta$ peaks appear to be nearly Gaussian (because of the decrease in the relative height of the $K\beta$ peak to a few percent of the height of the $K\alpha$ for Si). The *L* and *M* peaks appear asymmetric because of the presence of the unresolved β peak of intermediate weight near the main peak. This asymmetry can help to make the identification of the intermediate and heavy elements more robust.

All x-ray peaks for which the critical excitation energy is exceeded by the beam energy will be observable. Therefore in a qualitative analysis, all peaks for each element must be located. Considering the 0.1- to 15-keV range, if a high-energy *K* peak is observed [4.5 keV (titanium) and above],

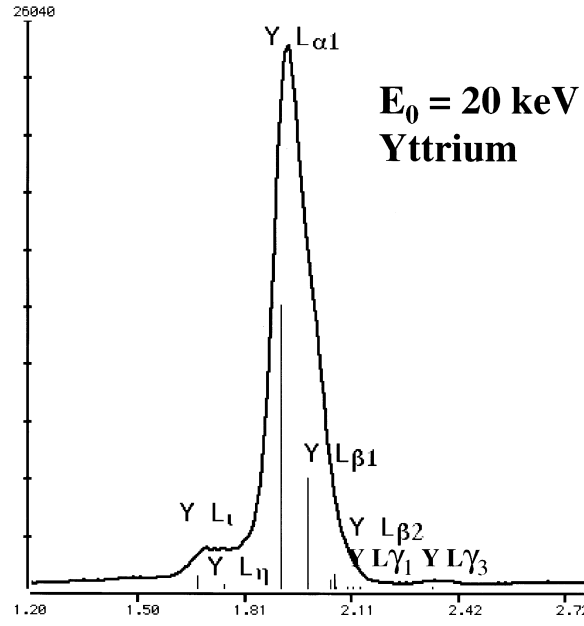


Figure 8.7. Yttrium *L* family, with *Lα* and *Lβ* unresolved. Note the low-intensity peaks associated with Y *Ll*–Y *Lη* below the main peak and Y *Lγ*$_3$ above the main peak.

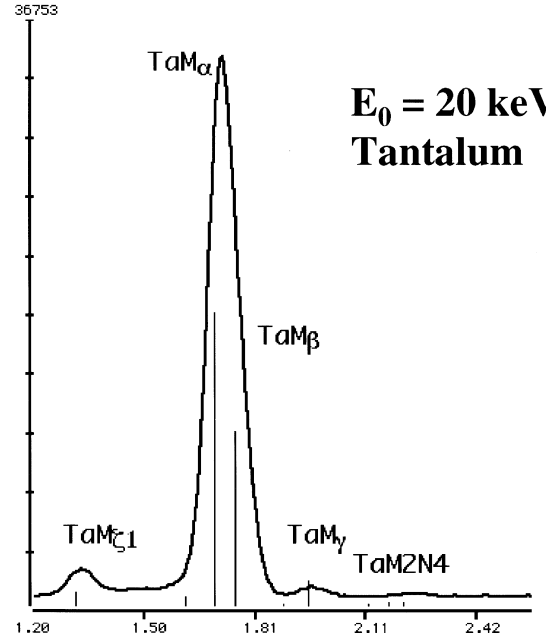


Figure 8.8. Tantalum *M*-family, with *Mα* and *Mβ* unresolved. Note the low-intensity peaks for *Mζ* below the main peak and *Mγ* and M_2N_4 above the main peak.

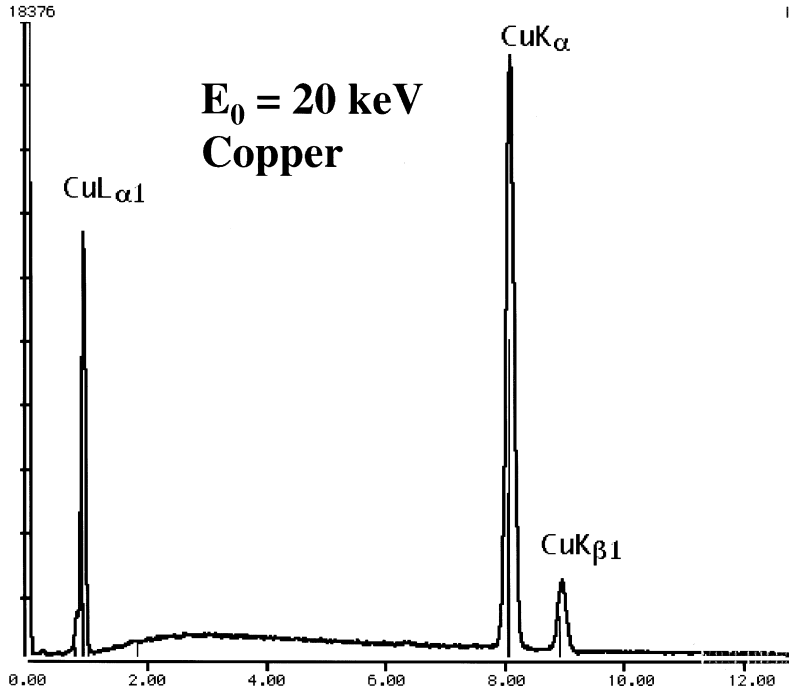


Figure 8.9. Copper *K*-family and *L*-family peaks observed simultaneously. When Cu *K* is excited, the lower energy Cu *L* family must also be present.

then a low-energy *L* peak must also be present for the element. Figure 8.9 shows this situation for copper *K* and *L*. Similarly, if a high-energy *L* peak is observed [4.5 keV (barium) or above], then a low-energy *M* peak must also be present. Figure 8.10 illustrates this situation for tantalum *L* and *M*. Because of large differences in the generation and absorption of low- and high-energy x-rays, it is not generally possible to make use of relative peak heights between *K*, *L*, or *M* families of the same element in qualitative analysis.

The development of reliable low-beam-energy ($E_0 \leq 5$ keV) SEMs has created new interest in low-photon-energy microanalysis. Thin-window or windowless EDS provides access to the x-ray energy range below 1 keV, where the *K* peaks of the light elements beryllium, boron, carbon, nitrogen, oxygen, fluorine, and neon occur. As the x-ray energy decreases below 1 keV, peak shapes tend to deviate somewhat from the Gaussian shape characteristic of higher energy due to incomplete charge (on the low-energy side) and pulse coincidence (count rate-dependent) on the high-energy side), as shown in Fig. 8.11 for carbon. A major consideration in energy-dispersive spectrometry below 1 keV is the problem of interferences between low-Z *K* x-rays and *L*- and *M*-family x-rays of heavier elements. Figure 8.12 shows the situation for the oxygen *K* peak (0.523 keV) in a spectrum of Fe_2O_3. Also shown is the position of the Cr *L* family (0.571 keV).

If oxygen is present at low concentration as a minor or trace constituent while chromium is a major constituent, the oxygen peak would be barely discernible on the shoulder of the chromium peak. If vanadium (V L = 0.510 keV) is added to the mixture, the interference situation becomes untenable without quantitative spectral deconvolution. On the other hand, when oxygen and carbon are present at high concentrations, then the O K peak and C K peak tend to dominate the spectrum and suppress the visibility of L- and M-shell peaks, which have very low fluorescence yields in this energy region. Figure 8.13 shows the wide range of elements for which the principal L- and M-shell peaks suffer interference from C K or O K.

8.2.2. Guidelines for EDS Qualitative Analysis

From the previous discussion, we can construct sets of general and specific guidelines for qualitative analysis.

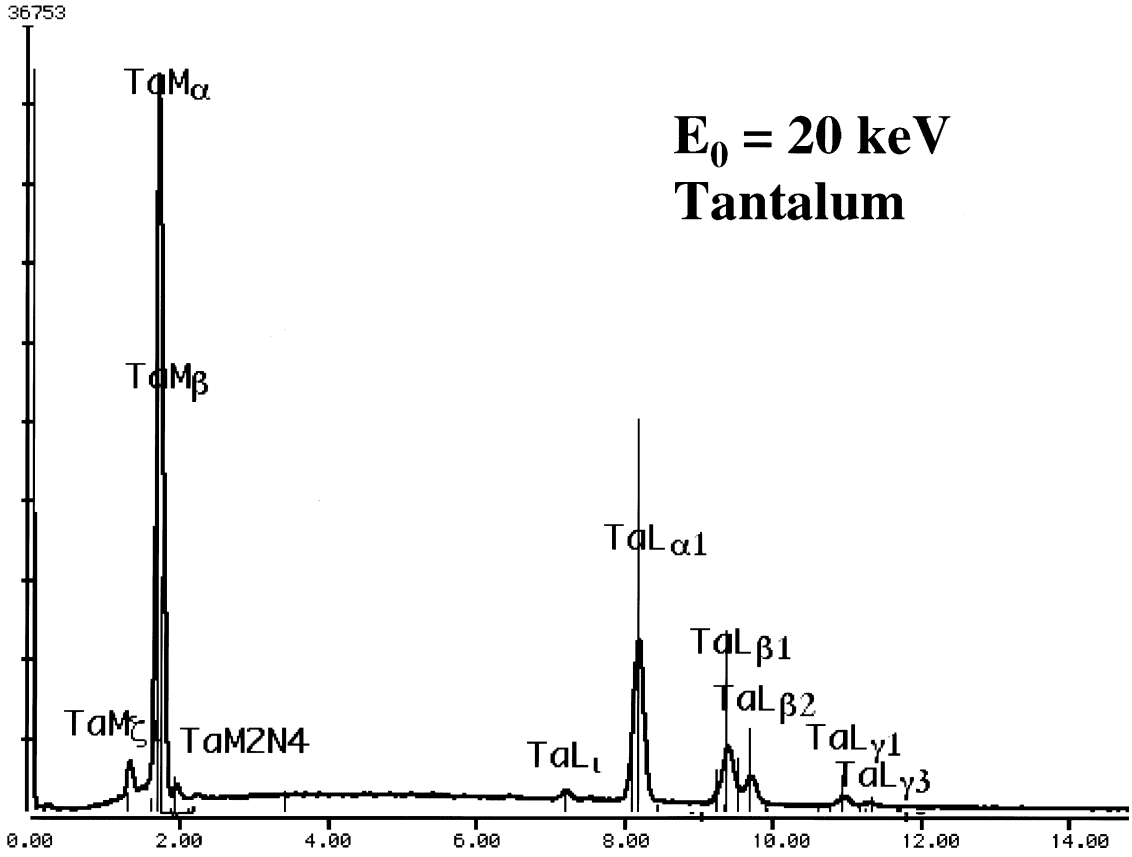


Figure 8.10. Tantalum L family and M family observed simultaneously. When Ta L is excited, the lower energy Ta M family must also be present.

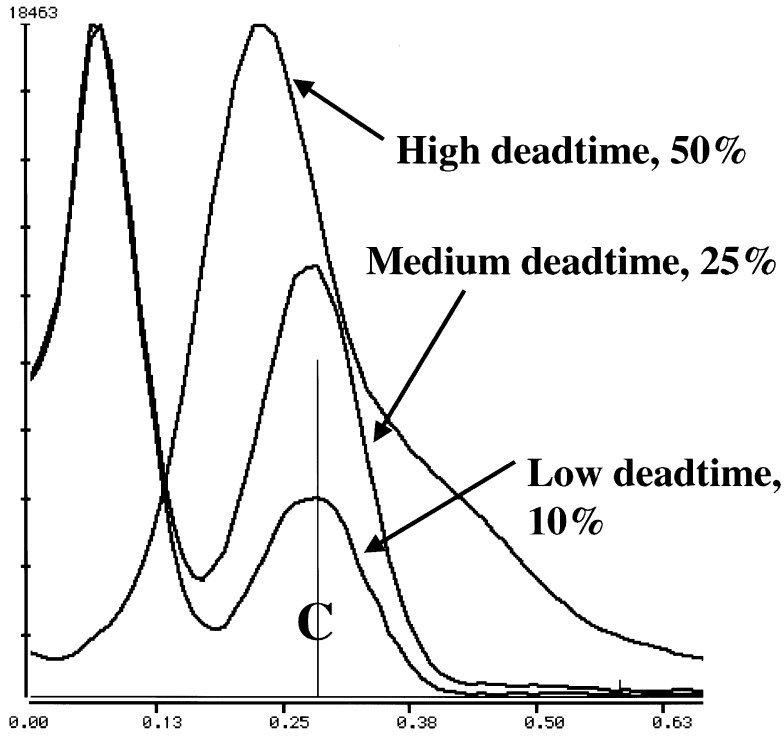


Figure 8.11. Carbon *K* peak: Changes in peak position and shape at low, intermediate, and high dead time.

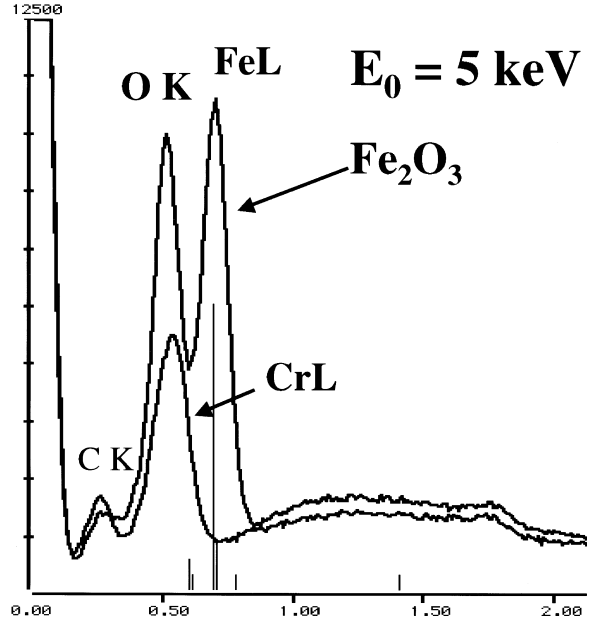


Figure 8.12. Low-photon-energy spectrum of Fe_2O_3 showing O *K* and Fe *L* peaks. The spectrum of pure chromium showing the position of Cr *L* peak is superimposed, showing strong interference between O *K* and Cr *L*.

8.2.2.1. General Guidelines for EDS Qualitative Analysis

1. Only peaks that are statistically significant should be considered for identification. The minimum size of the peak P after background subtraction should be three times the standard deviation of the background at the peak position, that is, $P > 3(N_B)^{1/2}$. This peak height can be approximately estimated directly on the EDS display from the statistical scatter in the background on either side of the peak. When the individual channels are plotted as "points"(using small symbols such as periods) the resulting "thickness" of the background trace is due to statistical fluctuations in the counts of each energy channel and is a measure of $(N_B)^{1/2}$. The peak above the background should be at least three times this thickness. If it is difficult because of statistical fluctuations in the count to decide whether a peak exists above the continuum, then more counts should be accumulated in the spectrum to "develop" the peak. (For more information on detectability limits see Chapter 9, Section 9.9.4.)

2. In order to satisfy guideline 1 and obtain adequate counts in the spectrum, it is tempting to use a high count rate. However, conventional monolithic EDS systems become increasingly susceptible to the introduction of artifacts, such as sum peaks, as the count rate increases. For an EDS system operating at its best energy resolution the maximum total spectrum

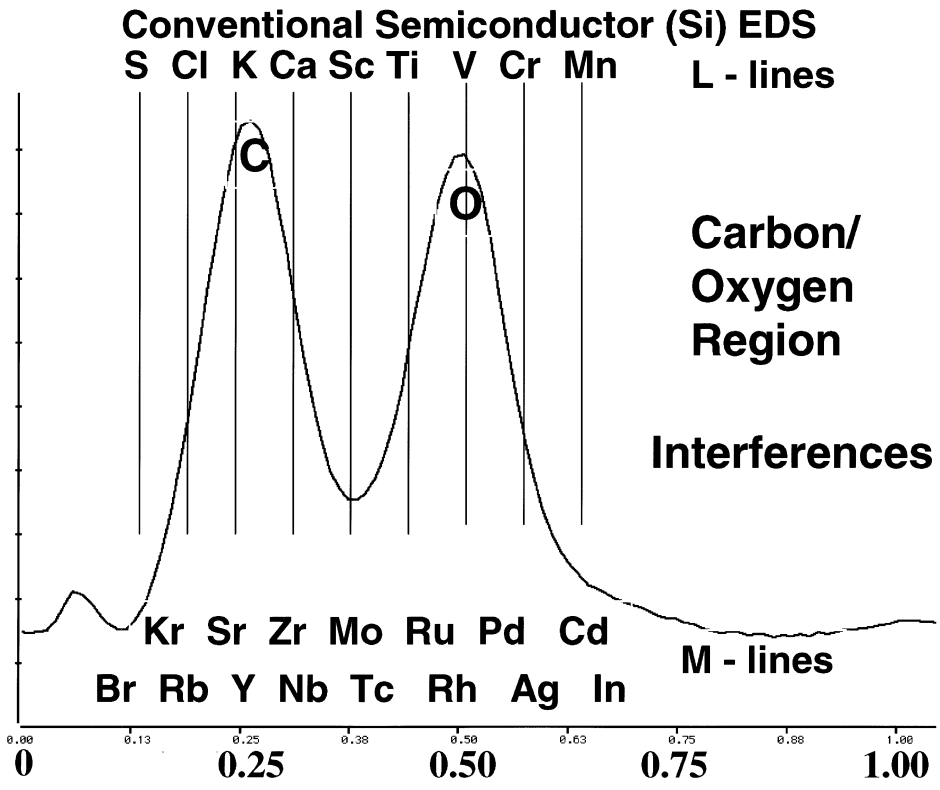


Figure 8.13. The C *K* and O *K* regions showing peak positions of *L*- and *M*-shell peaks that suffer interference.

input count rate should be kept below approximately 2500 counts/s. An alternative criterion is that the dead time should be kept below 30%. Recently developed digital pulse processing systems can operate at higher limiting count rates while retaining the optimum resolution and minimizing coincidence peaks.

3. The EDS spectrometer should be calibrated so that the peak positions are found within 10 eV of the tabulated values. Note that, because of the possible drift of the electronics, the calibration should be checked frequently as part of a measurement quality assurance plan. Further, whereas a linear response is ususally obtained for photon energies above 1 keV, the low range of photon energies below about 750 eV is likely to show significant deviation from calibration, even when the EDS shows near ideal linearity from 1 to 15 keV.

4. Suitable x-ray peaks for identifying the elemental range from beryllium to uranium will be found in the x-ray energy range from 0.1 to 20 keV. To provide an adequate overvoltage to excite x-ray peaks in the upper half of this range a beam energy in the range 20–30 keV should be used. A beam energy of 20 keV is a good compromise between the need to create an adequate overvoltage and the need to minimize absorption in the sample, which increases as the beam energy and depth of penetration increase. However, such a beam energy may mask the presence of light elements that only produce x-ray energies below 2 keV. The very high overvoltage, $U > 10$, for this photon energy range will lead to deep penetration of the specimen by the electron beam, and consequently, very high absorption, such that 50–99% of the low-energy x-rays will be absorbed in the specimen. To avoid missing possible light elements present at low concentration, the spectrum accumulation should be repeated, if possible, with a beam energy in the range 5–10 keV. Occasionally, because of interferences, for example, Mo $L\alpha$ and S $K\alpha$ and $K\beta$, it is necessary to confirm the presence of the heavy element by using x-ray peaks greater than 15 keV in energy, for example, Mo $K\alpha$ at 17.48 keV, which arises from the Mo K-edge at 20 keV. If the beam energy can be increased to give at least $U > 1.5$ ($E_0 = 30$ keV for the Mo K-edge), and long spectrum accumulation times are used, then these high-energy x-ray peaks can prove valuable for establishing the presence or absence of an element.

5. In carrying out accurate qualitative analysis, a conscientious "bookkeeping" method must be followed. When an element is identified, all x-ray peaks in the possible families excited must be marked off, particularly low-relative-intensity members, and also including artifacts such as the Si escape peak and sum peaks. In this way, one can avoid later misidentification of those low-intensity family members as belonging to some other element apparently present as a minor or trace concentration. Peaks arising from stray excitation in the sample chamber ("system peaks"), previously described in Chapter 7, should also be marked off.

8.2.2.2. Specific Guidelines for EDS Qualitative Analysis

1. Set the vertical gain of the CXA so that all peaks are contained in the display field. Begin at the high-energy end of the spectrum and

work downward in energy because the members of the *K*, *L*, and *M* families are more widely separated at high energies and likely to be resolved.

2. Determine the energy of a large peak. If it corresponds closely to a $K\alpha$ peak of an element, immediately look for a $K\beta$ peak with about 10% of the $K\alpha$ peak height. $K\alpha$ and $K\beta$ peaks of elements starting at sulfur (2.31 keV) will be resolved with a high-resolution (129 eV) EDS spectrometer. If the spectrum has too few counts to confidently identify the $K\beta$ peak, then more time should be spent on accumulating a more statistically robust spectrum.

3. If the $K\alpha$ and $K\beta$ pair does not fit the unknown, try the *L* series, noting the multiplicity of high-intensity *L* peaks, $L\alpha$ (l), $L\beta_1$ (0.7), $L\beta_2$ (0.2), $L\beta_3$ (0.08), $L\beta_4$ (0.05), $L\gamma_1$ (0.08), $L\gamma_3$ (0.03), Ll (0.04), and $L\eta$ (0.0l), which must be found to confirm the identification. These peak weights are only approximate.

4. *M*-family peaks can be observed for elements starting at barium (approximately 1 keV) up to uranium (U$M\alpha = 3.5$ keV). Whereas $M\alpha(1)$ and $M\beta$ (0.6) are poorly resolved, the lower intensity $M\zeta$ (0.06), $M\gamma$ (0.05), and $M_{\mathrm{II}}N_{\mathrm{IV}}$ (0.01) are separated from the main peak and usually visible. These low-intensity *M*-family members can be very easily misidentified. Note that even for uranium, the *M* peaks occur at an energy of 3.5 keV or less.

5. When a candidate element is identified, all peaks of all families (*K*, *L* or *L*, *M*) of that element should be marked off before proceeding. Next, escape peaks (parent − 1.74 keV) and sum peaks associated with the major peaks of the element should be located and marked. The magnitude of the escape peak is a constant fraction of the parent peak for each element, ranging from about l% for P $K\alpha$ to 0.01% for Zn $K\alpha$. The sum peak magnitude depends on the count rate and will generally only be visible for elements present at very high concentration with efficiently excited x-ray peaks that will yield the necessary count rate for coincidence to become significant. Figure 8.14 shows this effect for Al $K\alpha$ (1.49 keV), which produces a sum peak at 2.98 keV that is often mistaken for Ag $L\alpha$ (2.98 keV) or Ar $K\alpha$ (2.96 keV). Many automatic qualitative analysis software tools correctly deal with escape peaks, but misidentify sum peaks, as shown in Fig. 8.14. Coincidence events producing sum peaks are strongly count-rate-dependent. To test for the presence of a sum peak, the spectrum can be recorded a second time at a substantially reduced count rate (lower dead time). Sum peaks will be lost at low dead time.

6. At low x-ray energy (less than 3 keV), peak separation and the limited resolution of the EDS spectrometer will likely restrict element identification to only one peak. Note that low-energy *L* or *M* peaks will be accompanied by high-energy *K* or *L* peaks in the 4- to 20-keV range, which can aid identification. Moreover, low-energy *L* and *M* parent peaks are accompanied by low-level family members such as Ll, $L\gamma_3$ and $M\zeta$, $M2$N4, but these will likely only be useful for confirming qualitative analysis when the element is present at high concentration. When the element is present as a minor or trace constituent, only a spectrum collected with extremely

high counting statistics (total spectrum > 1 million counts) is likely to show these minor family peaks to aid identification.

7. When all of the high-intensity peaks in the spectrum have been identified and all family members and spectral artifacts have been located, the analyst is ready to proceed to identify the remaining low-intensity peaks. Any low-intensity peaks that are unidentified after the above procedure should belong to elements present at minor or trace concentrations. Note that for such minor or trace elements, only the principal α peak in a family may be visible. The peaks at lower relative intensity in a family will probably be lost in the statistical fluctuations of the background. As a result, the confidence with which minor or trace elements can be identified is necessarily poorer than for major elements. If the positive identification of minor and trace elements is critical to the result, then a spectrum containing a greater number of counts will be needed. In many cases it may be necessary to resort to WDS qualitative analysis for the added spectral resolution to increase the analyst's confidence in identifying minor or trace elements.

8. Below 1 keV, qualitative EDS analysis becomes more problematic. If the specimen contains elements of intermediate and high atomic number, interferences may arise from the *L*- and *M*-family x-rays of these elements

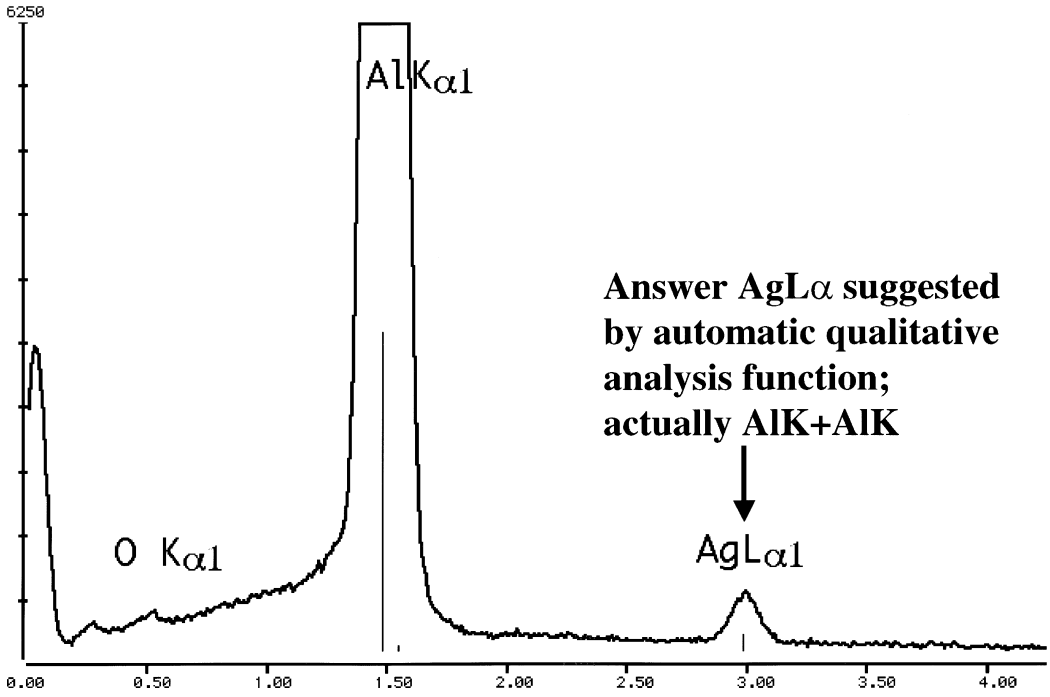


Figure 8.14. The Si EDS spectrum of aluminum taken at high dead time (40%) so that a significant coincidence (sum) peak occurs for Al $K+$ Al K. Note that the automatic qualitative analysis algorithm in this CXA suggested that the sum peak was Ag L, an incorrect assignment.

upon the *K* peaks of the light elements (for example, see Figs. 8.12 and 8.13). If the qualitative analysis procedure has been carefully followed to this point, any *L*- and *M*-family x-rays in the region below 1 keV and escape peaks originating from parent peaks in the range 2–3.75 keV will have been identified based upon higher energy *K*- and *L*-family peaks, so that possible interferences in this region will have been noted. The analyst can therefore proceed to identify the remaining peaks in this region as low-energy *K* x-rays of the light elements. The interference situation of *L*- and *M*-shell x-rays with the light element *K* peaks is often severe when the heavy elements are present at high concentration. Unambiguous identification of light elements should then be performed with the wavelength-dispersive spectrometer because of its much higher spectral resolution. Alternatively, peak deconvolution can be performed on the EDS spectra, but the poor counting statistics often found in the low-energy region will limit the accuracy of deconvolution of peaks when the light-element peak is less than 10% of the overlapping *L* or *M* peak. In reality, the fluorescence yield of *L*- and especially *M*-family x-rays is low compared to *K*-shell x-rays, so that often the *K* peak can be recognized.

9. As a final step, by using Fig. 8.1 and the Enhancement CD database, the analyst should consider what peaks may be hidden by interferences. If it is important to know of the presence of those elements, it will be necessary to resort to careful peak deconvolution and/or WDS analysis.

8.2.3. Examples of Manual EDS Qualitative Analysis

1. *Incisor of Rat*. The spectrum shown in Fig. 8.15 was obtained from the region of mineralized tissue formation in a rat incisor. Following the above procedure, we first find that the peaks at 3.69 and 4.01 keV correspond to Ca $K\alpha$ and $K\beta$, and also locate Ca *L*, Fig. 8.15a. At lower energy, the peaks at 2.015 and 2.310 keV appear to also be a $K\alpha$, $K\beta$ pair, but whereas P $K\alpha$ is found to correspond to the 2.015-keV peak, P $K\beta$ is unresolved, Fig. 8.15a. With additional spectrum accumulation and vertical expansion of the intensity scale, further study reveals that the 2.310-keV peak is S $K\alpha$, Fig. 8.15b. The remaining well-resolved low-energy peak is the O *K* at 0.523 keV. In Fig. 8.15a there is an interesting bump in the background just below Ca $K\alpha$, which may be a peak, but which is so close to the $3(N_B)^{1/2}$ limit that more spectral accumulation time must be spent before attacking this feature. Neither the escape peak of Ca $K\alpha$ nor the sum peak of P $K\alpha$ occurs in this energy region. Further accumulation of the spectrum reveals that the peak is *K* $K\alpha$, with K $K\beta$ lost under Ca $K\alpha$. The possible interferences of the main peaks (Ca $K\alpha$: Sb, Te; P $K\alpha$: Y, Zr) do not correspond to elements likely to exist in this system.

2. *Multielement Glass*. The EDS spectrum of a complicated, multielement glass (NIST glass K309) is shown in Fig. 8.16. Beginning at the high-energy end, the peaks at 6.40 and 7.06 keV correspond to Fe $K\alpha$ and

Fe $K\beta$, respectively. Note that Fe $L\alpha$ at 0.704 keV is just barely observed owing to high absorption. The next series of four peaks near 4.5 keV might at first be thought to be Ti $K\alpha$, $K\beta$ and V $K\alpha$, $K\beta$. However, the peak positions are not proper, nor are the relative peak heights. This series is the Ba L family: Ba $L\alpha$ (4.47 keV), Lβ_1 (4.83), $L\beta_2$ (5.16), $L\gamma$ (5.53), with others in the family unresolved. The peaks at 3.69 and 4.01 keV are Ca $K\alpha$ and Ca $K\beta$. At low energy, the peak at 1.74 keV is Si $K\alpha$ and the peak at 1.49 keV is Al $K\alpha$. The O K peak is found at 0.523 keV. Expansion of the vertical gain reveals no low-intensity peaks except those associated with artifacts. Note that in this example, if it were important to detect minor/trace titanium or vanadium, the interference of the Ba L family would be so severe that EDS analysis would not resolve the question and WDS analysis would certainly be needed.

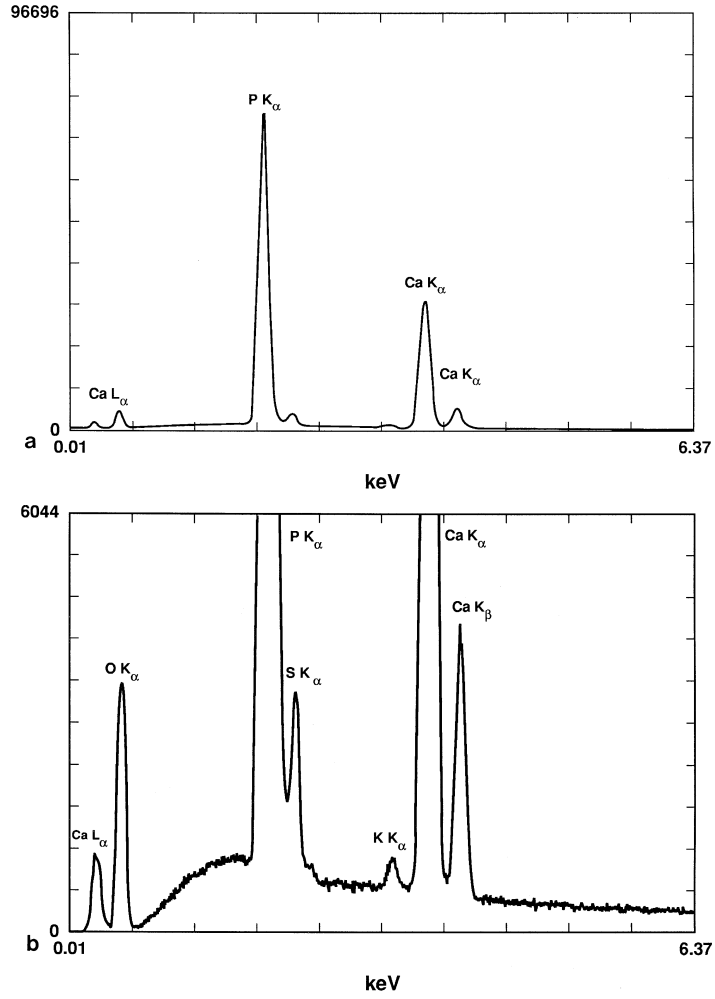


Figure 8.15. The Si EDS spectrum (diamond ultrathin window, resolution 145 eV at Mn $K\alpha$) of the enamel-forming region of a rat ("Nigel") incisor; $E_0 = 20$ keV. (a) Identification of the major constituents, calcium (including Ca L) and phosphorus. (b) Further accumulation and vertical expansion to show identification of sulfur, oxygen, and potassium.

8.2.4. Pathological Overlaps in EDS Qualitative Analysis

The limited energy resolution of the Si EDS frequently causes the analyst to be confronted with serious peak overlap problems. In some cases, the overlaps are so severe that an analysis for an element of interest can not be carried out with the EDS. An example is shown in Fig. 8.17, where the interference of Ti *K* and Ba *L* x-ray peaks is illustrated in the superimposed spectra. Problems with overlaps fall into two general classes: (1) the misidentification of peaks and (2) the problem of separating two overlapping peaks even if the analyst knows both are present. It is difficult to define the degree of overlap that can be successfully solved without careful consideration of the relative peak heights and the spectrum counting statistics. The greater the investment in spectrum accumulation, the more successful will be the deconvolution algorithms. In general, however, it will be very difficult to unravel two peaks of substantially unequal intensity (peak ratio greater than 3:1) when separated by less than 50 eV, no matter what peak stripping method is used. The analyst should check for the possibility of overlaps within 100 eV of a peak of interest. When the problem involves identifying and measuring a peak of a minor or trace constituent in the neighborhood of a main peak of a major constituent, the

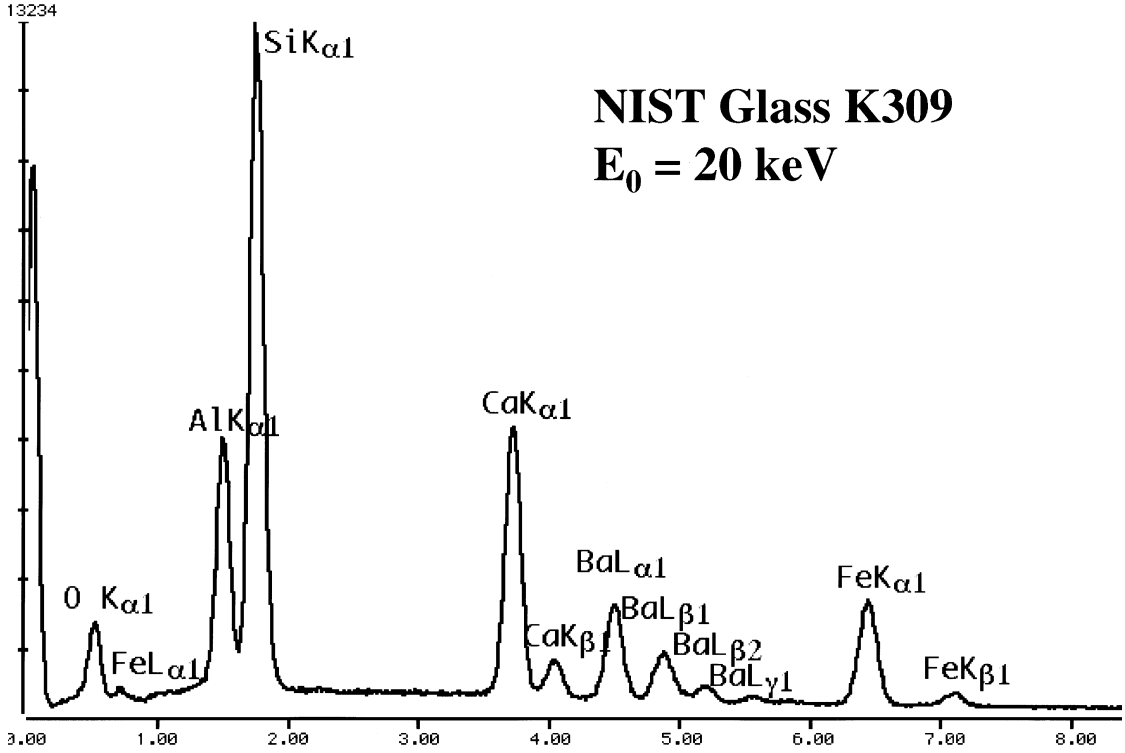


Figure 8.16. The Si EDS spectrum of NIST glass K309, showing the assignment of peaks by qualitative analysis.

problem is further exacerbated, and overlaps may be significant even with 200-eV separation in the case of major/minor constituents. When peaks are only partially resolved, the overlap will actually cause the peak channels for both peaks to shift by as much as 10–20 eV from the expected value. This phenomenon is encountered in the copper *L*-family spectrum, Fig. 8.9, where the unresolved Cu $L\alpha$ and Cu $L\beta$ peaks form a single peak whose peak channel is 10 eV higher than the Cu $L\alpha$ energy. Because copper *K* and *L* peaks are often used as a set of calibration peaks for EDS, the peak overlap effect in the *L* series can lead to a 10-eV calibration error at the low-energy end of the spectrum if the Cu $L\alpha$ energy is used directly. It is better to calibrate the system on a low-energy and a high-energy *K* peak, for example, pure Mg and pure Zn.

Because of the multiplicity of x-ray peaks and the number of elements, it is not practical to list all significant interferences. Figure 8.1 can be conveniently used to assess possible interferences at any energy of interest by striking vertical line from the element and peak of interest and noting all peaks within a given energy width on either side. Short lists of frequently encountered elemental interferences that arise from peak overlaps in biological, geological, and materials science EDS analysis are presented in Tables 8.1 and 8.2.

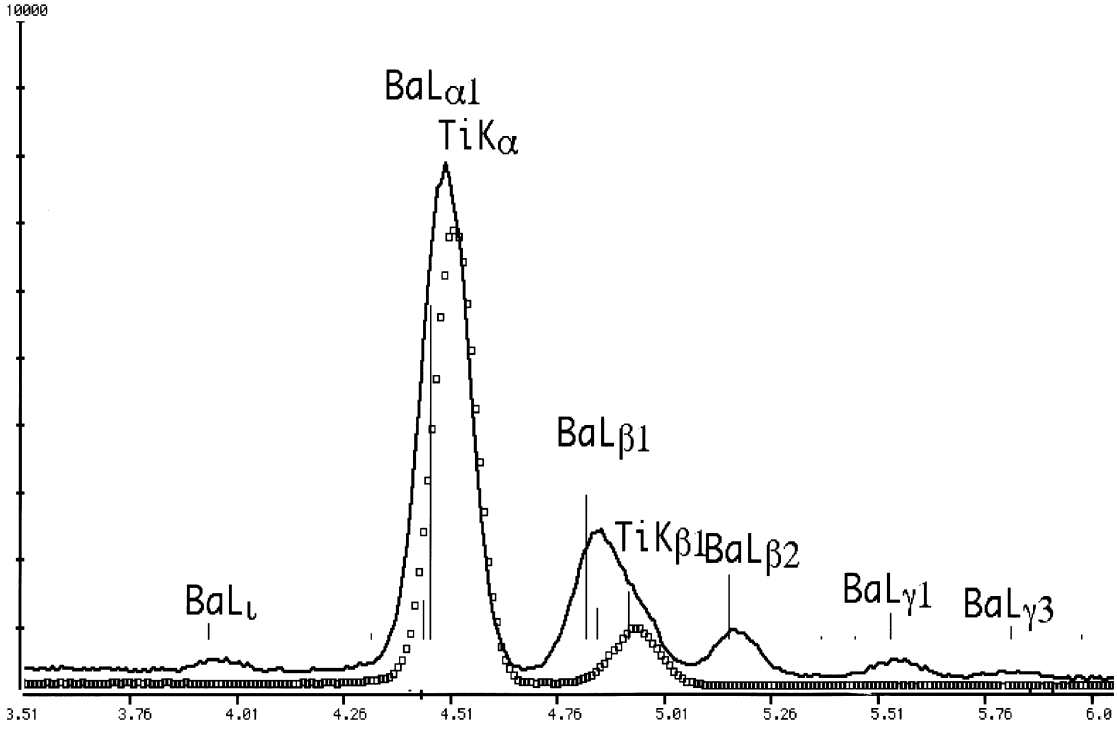


Figure 8.17. Superimposed Si EDS spectra of Ti $K\alpha$ and Ti $K\beta$ (points) and the Ba *L*-family peaks (continuous line) illustrating a high degree of peak interference.

Table 8.1. Common Interferences in Biological X-Ray Microanalysis

Element in stain or fixative	Interfering x-ray peak	Interferes with
U	*M*	K *Kα*
Os	*M*	Al *Kα*
Ru	*L*	S *Kα*
Ag	*L*	Cl *Kα*
As	*L*	Na *Kα*
Cu (grid)	*L*	Na *Kα*

1. *Biological EDS Analysis*. Many interferences that are encountered in biological x-ray microanalysis arise from the use of heavy metal fixatives and stains in sample preparation. The heavy metals have *L*- and *M*-family x-ray peaks in the 1- to 5-keV range that can interfere with the *K* peaks of low-atomic-number materials, Na to Ca, which are important biologically. Table 8.1 lists the typical heavy metal stains and the elements for which significant peak overlap occurs.

2. *Materials Science EDS Analysis*. Because of the large number of elements encountered in materials science analysis, the number of possible interferences is much greater than in biological analysis, and so the analyst must constantly check his or her work to avoid errors. Particularly insidious is the systematic interference across the first transition metal series, in which the *Kβ* peak of an element interferes with the *Kα* peak of the next higher atomic number, as indicated in Table 8.2. Quantitative EDS analytical systems can correct for these interferences through peak deconvolution methods, such as multiple linear least squares fitting (see Chapter 9). In

Table 8.2. Common Interferences in Geological and Materials Science X-Ray Microanalysis

Element	Interferes with	Where found
Ti *Kβ*	V *Kα*	Steels, Fe–Ti oxides
V *Kβ*	Cr *Kα*	Steels
Cr *Kβ*	Mn *Kα*	Steels
Mn *Kβ*	Fe *Kα*	Steels
Fe *Kβ*	Co *Kα*	Steels, magnetic alloys
Co *Kβ*	Ni *Kα*	Steels, hard surfacing alloys
S *Kα*, *β*	Mo *Lα*, Pb *Mα*	Minerals, lubricants, sulfides, sulfates
W *Mα*, *β*	Si *Kα*, *β*	Semiconductor processing
Ta *Mα*, *β*	Si *Kα*, *β*	Semiconductor processing
Ti *Kα*	Ba *Lα*	Optoelectronics, silicates
As *Kα*	Pb *Lα*	Pigments
Sr *Lα*	Si *Kα*	Silicates (feldspars in particular)
Y *Lα*	P *Kα*	Phosphates

qualitative analysis, however, when a minor element is interfered with by a major-element x-ray peak, it is often very difficult to detect the presence of the minor element.

An example of a classic problem is the mutual interference of the S *K* (2.308 keV), Mo *L* (2.293 keV), and Pb *M* (2.346 keV) peaks, which lie within an energy range of only 50 eV. The resolution of the Si EDS system becomes an important consideration when trying to identify which of these elements is present. As shown in Figure 8.18, by moving from a 152-eV (Mn $K\alpha$) detector, which is typical of large-area (e.g., 60 mm^2) detectors, to a 129-eV (Mn $K\alpha$) detector, the S $K\alpha$–$K\beta$ pair is nearly resolved, and the β members of the Mo *L* and Pb *M* families can be seen. Moreover, the minor peak components are more prominent at 129-eV resolution because of increased peak-to-background. With this improved resolution, it is possible to recognize which element is present. However, when mixtures occur, such as MoS_2 and PbS, the spectral interference is so severe that to accurately identify the relative amounts, advanced

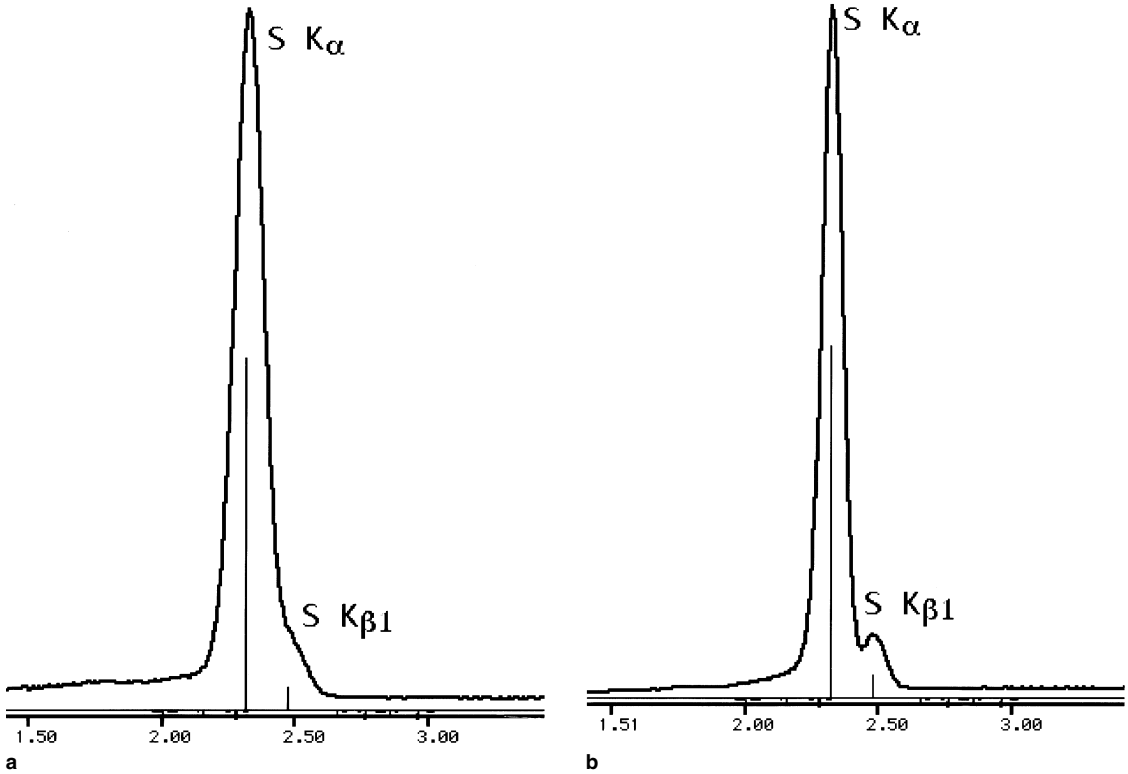
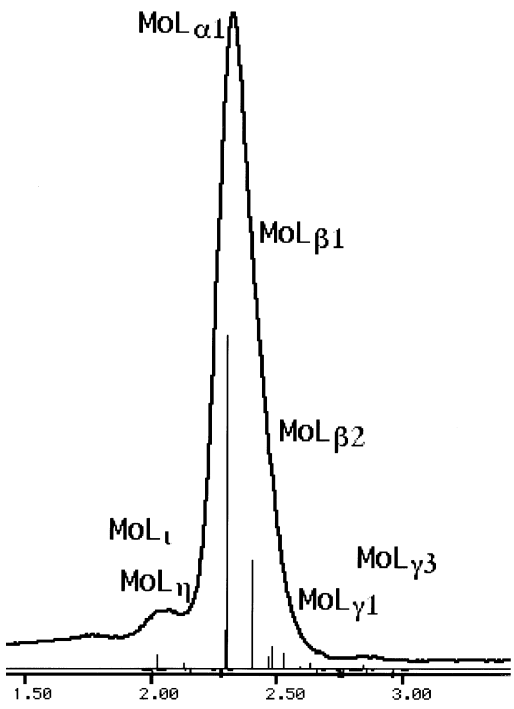
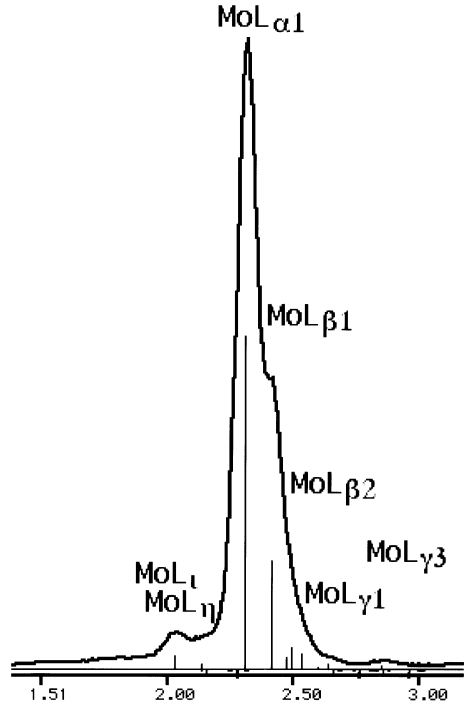


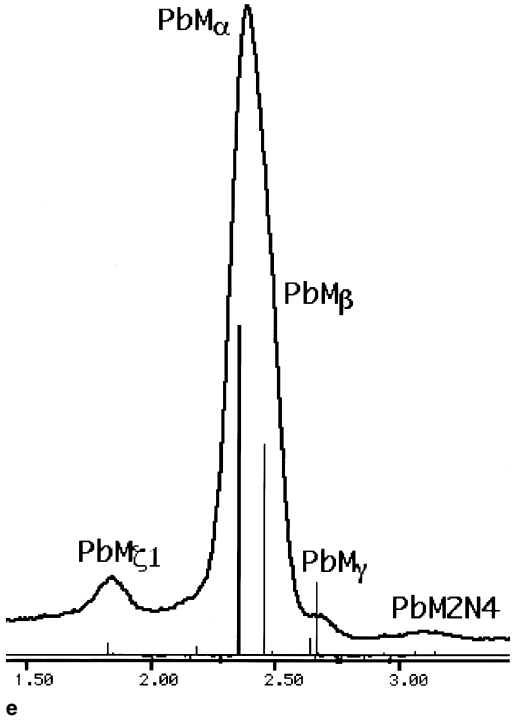
Figure 8.18. Effect of spectrometer resolution, 152 eV (Mn $K\alpha$) versus 129 eV (Mn $K\alpha$): (a) S *K* family, 152 eV, (b) S *K*, 129 eV; (c) Mo *L* family, 152 eV; (d) Mo *L*, 129 eV; (e) Pb *M* family, 152 eV; (f) Pb *M*, 129 eV. In the spectra measured at 129-eV (Mn $K\alpha$) resolution, note the detection of the β member of each family as well as the more prominent appearance of the minor family peaks.



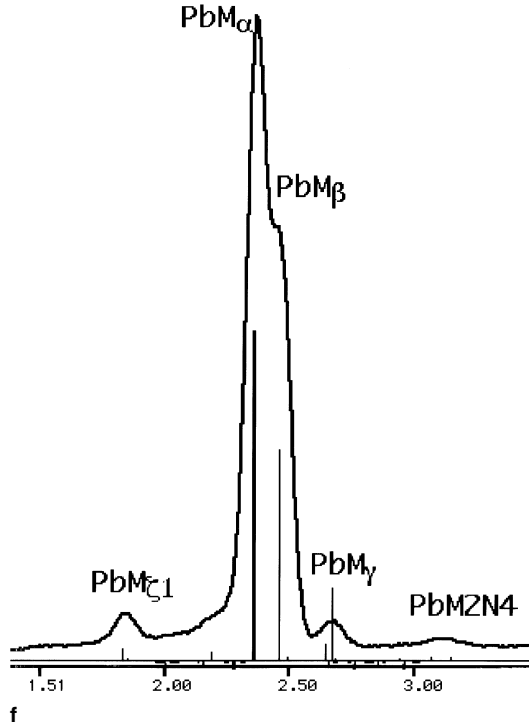
c



d



e



f

Figure 8.18. (*Continued*)

spectral processing or recourse to wavelength-dispersive spectrometry is necessary.

8.2.5. Advanced Qualitative Analysis: Peak Stripping

Having asked the question raised in step 9 in Section 8.2.2.2, what can be done to determine whether peak interference prevents the analyst from identifying an element of interest? Given the limitations considered in Section 8.2.4 concerning pathological overlap situations, does the analyst using EDS have any recourse in the "gray" area between fully resolved peaks and peaks that are so close together that there are no practical means of separating them and therefore requiring that a high-resolution spectrometric method such as WDS must be used? Digital spectral processing is highly flexible and presents several useful tools to the analyst for manipulating EDS spectra. "Peak fitting" and "peak stripping" are general terms for a number of mathematical processes that allow the analyst to selectively remove specific peak structures from the experimental spectrum in question by comparison to purely mathematical descriptions of peaks or to experimentally measured peak structures. Such manual processes are particularly effective at revealing hidden peaks if there exists a member of a family that is not suffering interference. By scaling the reference family to that clear peak, the complete family intensity can be subtracted on a channel-by-channel basis from the experimental spectrum until the resulting background matches the observed experimental background of nearby regions. Any hidden peaks should then be evident.

An example of this type of manual fitting is shown in Fig. 8.19. Figure 8.19a is an EDS spectrum of a maraging steel expanded to show the Fe K and Ni K region of the spectrum. The peak associated with the Fe $K\beta$ is anomalously broad and should arouse the analyst's suspicion. The analyst can examine this anomalous peak by selecting a peak-stripping function from the CXA scaled to the Fe $K\alpha$ peak, but simultaneously removing the complete K family, that is, Fe $K\alpha$ *and* Fe $K\beta$. After progressively removing the Fe $K\alpha$ peak intensity until the background is reached (Fig. 8.19b), the analyst finds that there is indeed a strong feature remaining from the anomalously wide Fe $K\beta$ structure after the Fe $K\beta$ intensity has been removed. This feature is a peak for Co $K\alpha$, and its corresponding Co $K\beta$ peak is recognizable as a distortion on the side of Ni $K\alpha$.

Detailed spectral fitting processes involving *all* spectral peaks are a fundamental part of quantitative analysis and will be presented in Chapter 9. After fitting the EDS spectrum to extract characteristic peak intensities for the quantitative procedure, by examining the residuals when the fitted spectrum is subtracted from the original, it is often possible to recognize peaks that were not detected in the original qualitative analysis. Such comparisons enable the analyst to make final corrections to the list of elements and repeat the quantitative procedure to assure a complete analysis. This issue will be considered more fully in Chapter 9, Section 9.3.

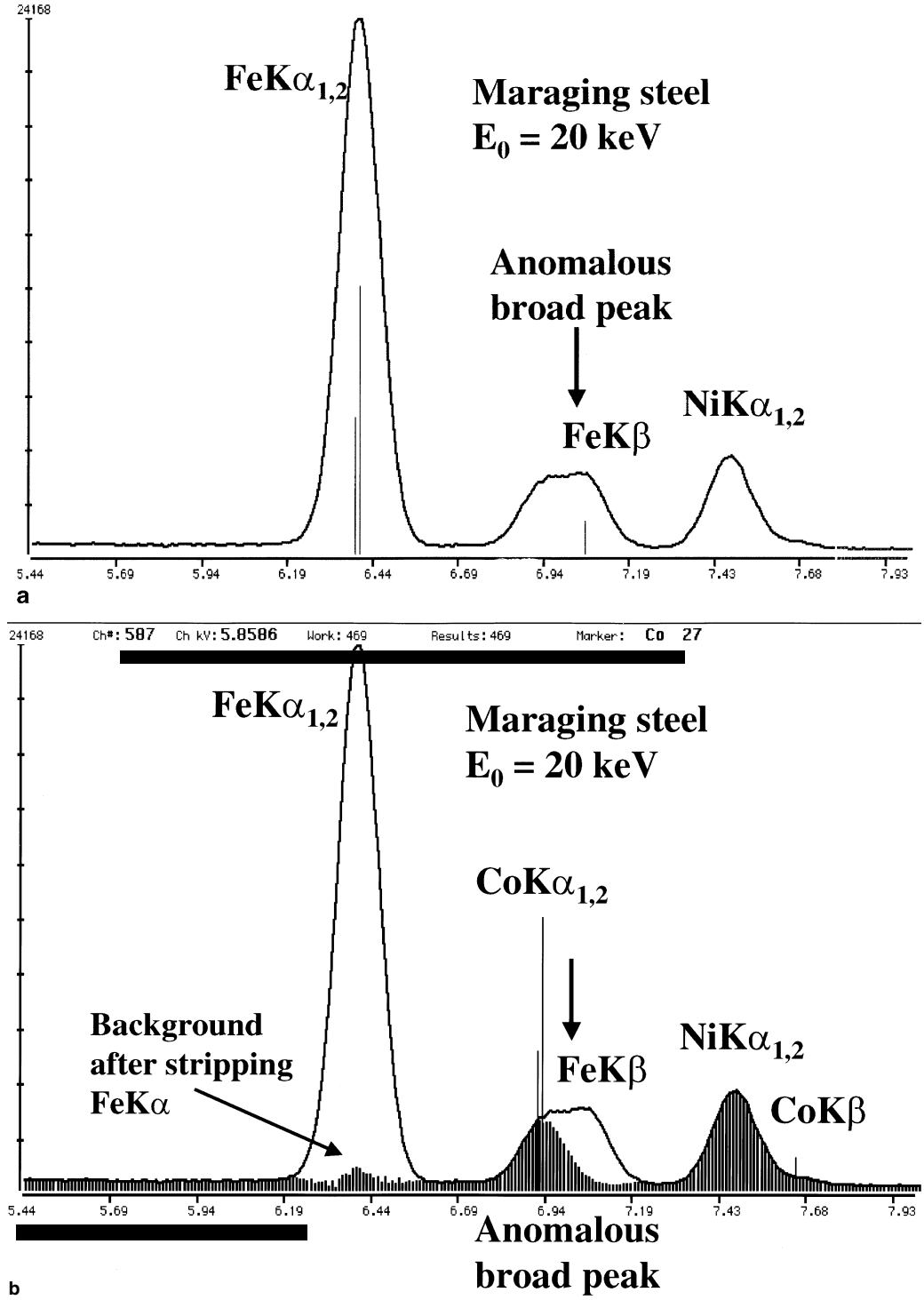


Figure 8.19. (a) Si EDS spectrum of maraging steel. (b) Appearance of the spectrum after subtraction of the Fe *Kα*–Fe *Kβ* family scaled to the Fe *Kα* peak, which does not suffer interference, showing the Co *Kα* peak.

Most modern computer-based analytical systems for energy-dispersive x-ray spectrometry include a routine for automatic qualitative analysis. Such a routine represents an attempt to create an expert system in which the guidelines described in the preceding sections for manual qualitative analysis are expressed as a series of conditional tests for recognizing and classifying peaks. Several different approaches to automatic qualitative analysis are possible, but in general the basic steps are as follows:

1. Background removal, by spectral filtering or background modeling
2. Peak searching, in which peak positions and amplitudes are determined
3. Peak recognition, in which possible candidate elements for assignment to the peaks are cataloged from look-up tables
4. Peak stripping, in which the intensity for a recognized element, including minor and artifact peaks such as the Si-escape peak, is stripped from a candidate peak to determine if any other elements are present in that energy region.

The success with which such an expert system operates depends on several factors:

1. Has the analyst accumulated a statistically valid spectrum prior to applying the automated qualitative analysis procedure?
2. The completeness of the x-ray database: Are the complete x-ray families included in the look-up tables, such as the Ll and $L\eta$ members of the L family and the $M\zeta$ and $M_{\mathrm{II}}N_{\mathrm{IV}}$ members of the M family?
3. Are x-ray artifacts such as the escape peaks and sum peaks properly accounted for?

Automatic qualitative analysis systems range from the naive to the sophisticated. In the naive approach, all possible elemental peaks within a specified energy range, for example, ± 25 eV, of a candidate peak are reported without application of any rules such as families of peaks and weights of peaks to sort out the invalid assignments. Peak stripping is not usually incorporated in such a system. The analyst is left with the task of applying the guidelines for manual qualitative analysis. In the most sophisticated systems, a true expert system with a rigorous codification of the qualitative analysis guidelines is used to reject misidentifications. The identified constituents are often listed in order of decreasing confidence.

Can the results reported by an automatic qualitative analysis system be trusted? Generally it is difficult to assign a quantitative measure of the degree of confidence with which a qualitative identification is made. The measurement of the statistics of the peak relative to the local background provides a measure of the confidence that a peak actually exists, although there is no numerical value for the likelihood of an identification being correct. For major constituents, the automatically determined answer is

virtually always correct, but for minor and trace constituents, the best advice is *caveat emptor*. The "report table" generated by a sophisticated computer-assisted analytical system may look extremely impressive, but that should not be taken as an excuse to accept its elemental identifications without question. The example of the misidentification of the Al sum peak as Ag $L\alpha$ in Fig. 8.14 should be taken as a caution against blindly accepting the computer's interpretation. The careful, experienced analyst really is smarter than the computer! Just as in manual qualitative analysis, the analyst must contribute common sense to solving the problem. Common sense is one of the most difficult concepts to incorporate in an expert system! It is therefore really the responsibility of the analyst to examine each putative identification, major constituents included, and determine if it is reasonable when other possibilities are considered. As always, an excellent procedure in learning the limitations of an automatic system is to test it against known standards of increasing complexity. Even after successful performance has been demonstrated on selected complex standards, the careful analyst will habitually check the suggested results on unknowns; suggestions of the presence of scandium, promethium, or technetium should always be treated with healthy skepticism!

8.3. WDS Qualitative Analysis

8.3.1. Wavelength-Dispersive Spectrometry of X-Ray Peaks

For qualitative analysis with wavelength-dispersive spectrometers, one seeks to identify the elements in the sample by determining the angles at which Bragg's law is satisfied as the spectrometer is scanned through a range of angles. Peaks are observed at those angles θ at which the following condition is satisfied:

$$n\lambda = 2d \sin\theta, \tag{8.1}$$

where d is the spacing of the WDS diffractor (e.g., LiF), λ is the wavelength of the x-rays, and n is the integer order of the reflection ($n = 1, 2, 3$, etc.). WDS spectra are typically read out in terms of wavelength units, $\sin\theta$ units, or millimeters of crystal movement. Because of the inherent comfort analysts have in dealing with x-ray spectra plotted the EDS fashion (intensity versus photon energy), many CXA systems permit the transformation of WDS spectra from $\sin\theta$ or wavelength units into a photon energy axis. This chapter will include examples of both types of WDS presentation. Figure 8.20 shows an example of the transformation of a WDS spectrum from its raw form in Fig. 8.20a, in which the units correspond to a plot width of 50-mm displacement. That same spectrum after conversion into energy units (at 10 eV/channel) is shown in Fig. 8.20b. Note the use of the *KLM* markers from the CXA database to identify the peaks. Because of the nonlinear nature of the λ-to-E conversion process, resulting from the relation $E \sim 1/\lambda$, the transformation of a peak that is symmetric in

wavelength space will not yield a similar symmetric peak in energy space. Thus some fine details of the WDS spectrum are inevitably lost or distorted upon conversion to energy space.

The strategy for qualitative WDS analysis is distinctly different from qualitative EDS analysis. The resolution of the WDS spectrometer is much better, typically <10 eV compared to 130 eV for EDS, which leads to a peak-to-background ratio at least 10 times higher for WDS. As a result of the improved resolution and higher peak-to-background ratio, more members of the family of x-ray peaks for a given element can be detected and must therefore be accounted for in a logical fashion to avoid subsequent misidentification of minor peaks. This is illustrated in Fig. 8.20c, where the EDS and WDS spectra for the cadmium *L* family are compared. Many of the peaks observed in the WDS spectrum are not detected in the EDS spectrum.

The analyst must recognize the consequences of all of the terms in Bragg's law, Eq. (8.1). For a particular value of λ (and therefore a specific photon energy) and a particular crystal d spacing, the consequence of the term n (which can take on any integer value from $n = 1, 2, 3, \ldots$, i.e., the order of the reflection changes) is to generate a series of θ_B values where photons of the same x-ray energy appear in the spectrum. If the parent peak ($n = 1$) is obtained on a given crystal, there may be another value of θ_B at which peaks corresponding to this same λ and energy will be found for

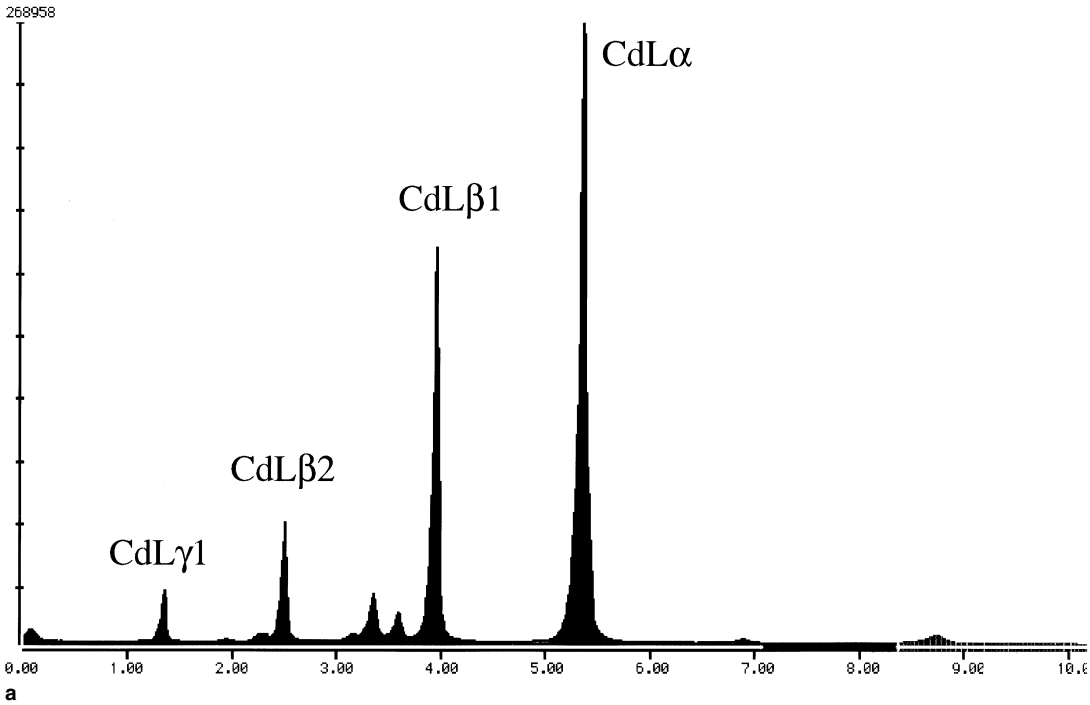
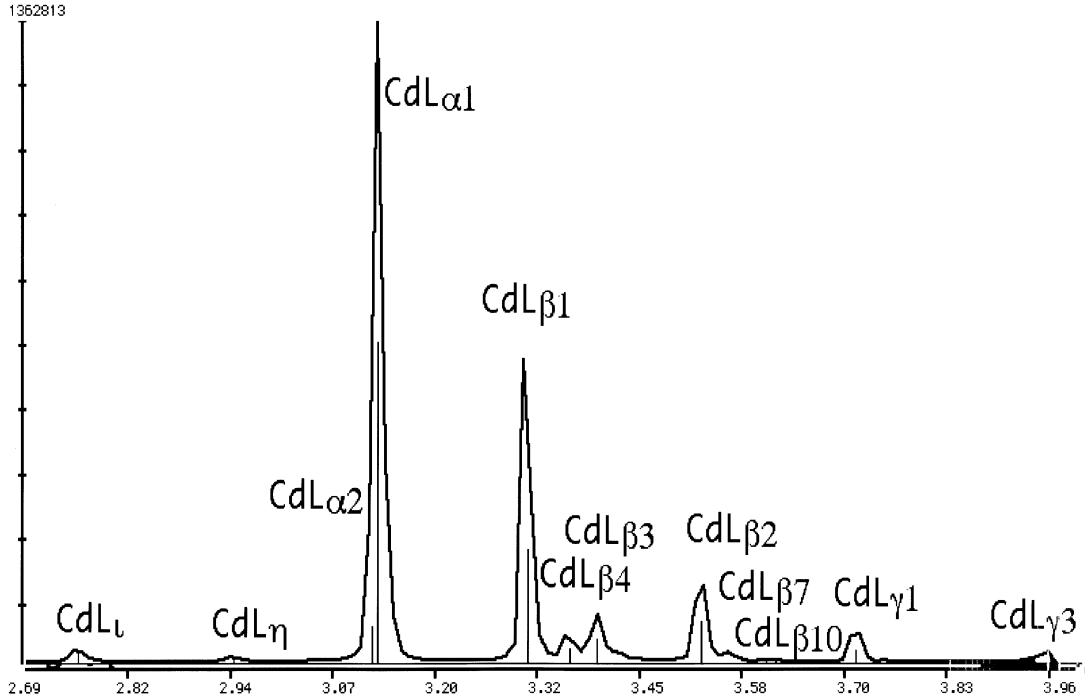
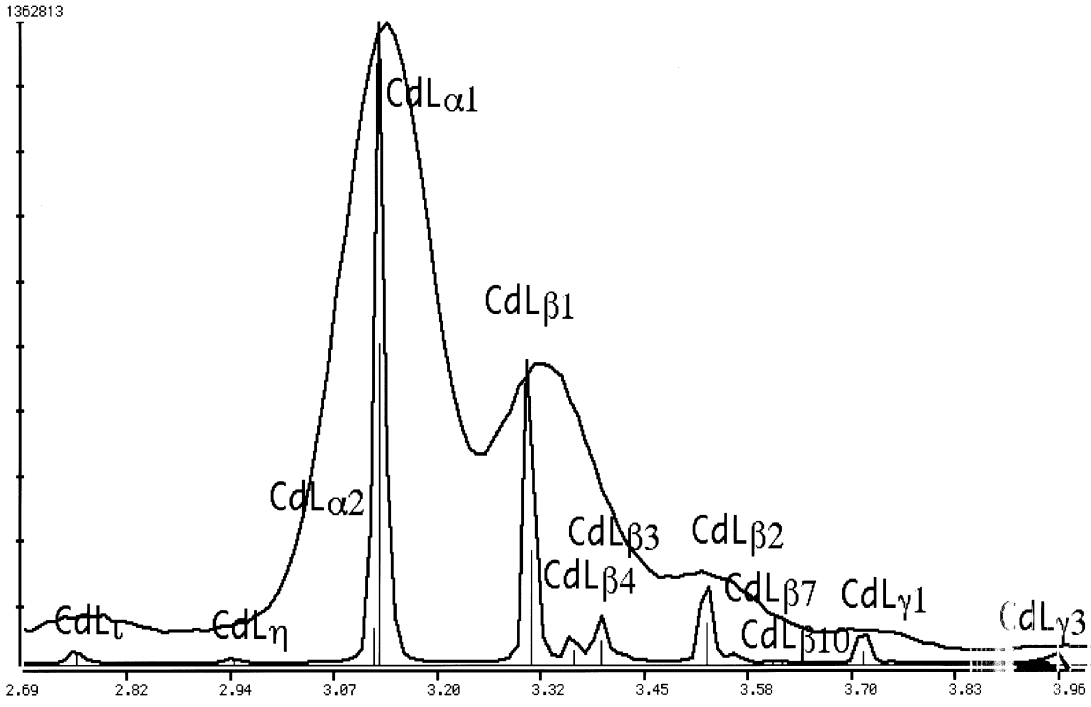


Figure 8.20. (a) Raw data from a WDS scan of Cd *L*-family peaks. Horizontal scale in crystal displacement. (b) conversion to an energy scale, with the Cd *L* peaks labeled; (c) superposition of a Si EDS spectrum of Cd on the converted WDS spectrum.



b



c

Figure 8.20. (*Continued*)

$n = 2$, 3, etc. If other crystals, that is, other d spacings, are considered, additional reflections may be detected, for example, corresponding to $n = 4$, 5, etc. Thus, the complete collection of peaks for an element that must be considered includes not only all of the family of x-ray peaks, but also the higher order reflections associated with each of the parent peaks.

A second consequence is that for a given setting of θ_B on a particular crystal with spacing d, x-rays of different energies that satisfy the same value of the product $n\lambda$ will all be diffracted. Thus, if diffraction occurs for $n = 1$ and $\lambda = \lambda_1$, diffraction will also occur for $n = 2$ and $\lambda = (\lambda_1/2)$ and so on. As an example, consider sulfur in an iron–cobalt alloy. The S $K\alpha$ ($n = 1$) peak occurs at 5.372 Å, whereas Co $K\alpha$ ($n = 1$) occurs at 1.789 Å. The third-order Co $K\alpha$ peak falls at 3×1.789 Å $= 5.367$ Å. These peaks are so close in wavelength that they would not be adequately separated by an ordinary WDS spectrometer. While such interferences are unusual, the analyst must be aware of their possible occurrence. Although the interference cannot be resolved spectroscopically, the S $K\alpha$ and Co $K\alpha$ x-rays are of different energies and can be separated electronically. This method takes advantage of the fact that the x-rays of different energies create voltage pulses of different magnitude. In the sulfur–cobalt example, the voltage pulses produced by the Co $K\alpha$ are about three times larger than those of the S $K\alpha$. By using the pulse height discriminator previously described (Chapter 7), one can set a voltage window around the S $K\alpha$ pulse distribution that excludes the Co $K\alpha$ pulse distribution. It should be noted that once a discriminator window has been established, the spectrometer should not be tuned to another peak without removal of the discriminator window or resetting to the appropriate value for another peak.

From these considerations, it is obvious that WDS qualitative analysis is not as straightforward as EDS qualitative analysis. Nevertheless, WDS

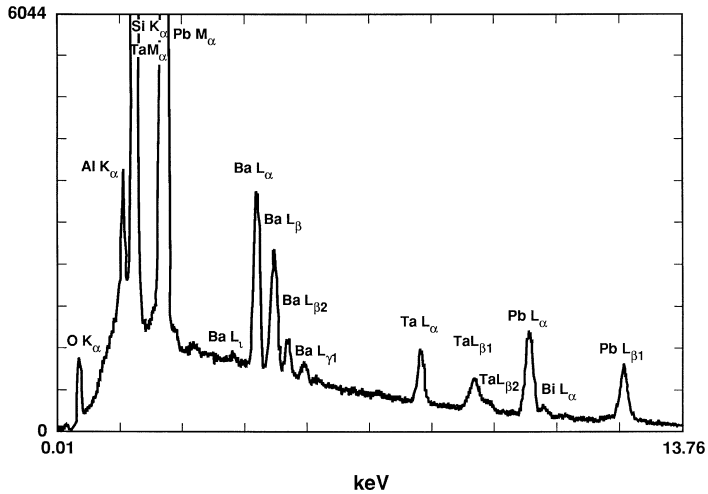


Figure 8.21. The Si-EDS spectrum of NIST glass K–251 (composition listed in Table 8.3) containing heavy elements.

qualitative analysis offers a number of valuable capabilities not possessed by EDS, as illustrated in Fig. 8.21 and 8.22, which compare EDS and WDS spectra on a heavy element glass (NIST K-251, the composition of which is given in Table 8.3).

1. The high resolution of the WDS allows the separation of almost all overlaps that plague the EDS spectrum. Thus in Fig. 8.21, bismuth is extremely difficult to detect in the presence of lead in the EDS spectrum, but is readily apparent in the WDS spectrum (Fig. 8.22a). Further, the multiplicity of peaks which can be observed aids in a positive identification of each element. However, this multiplicity makes the bookkeeping

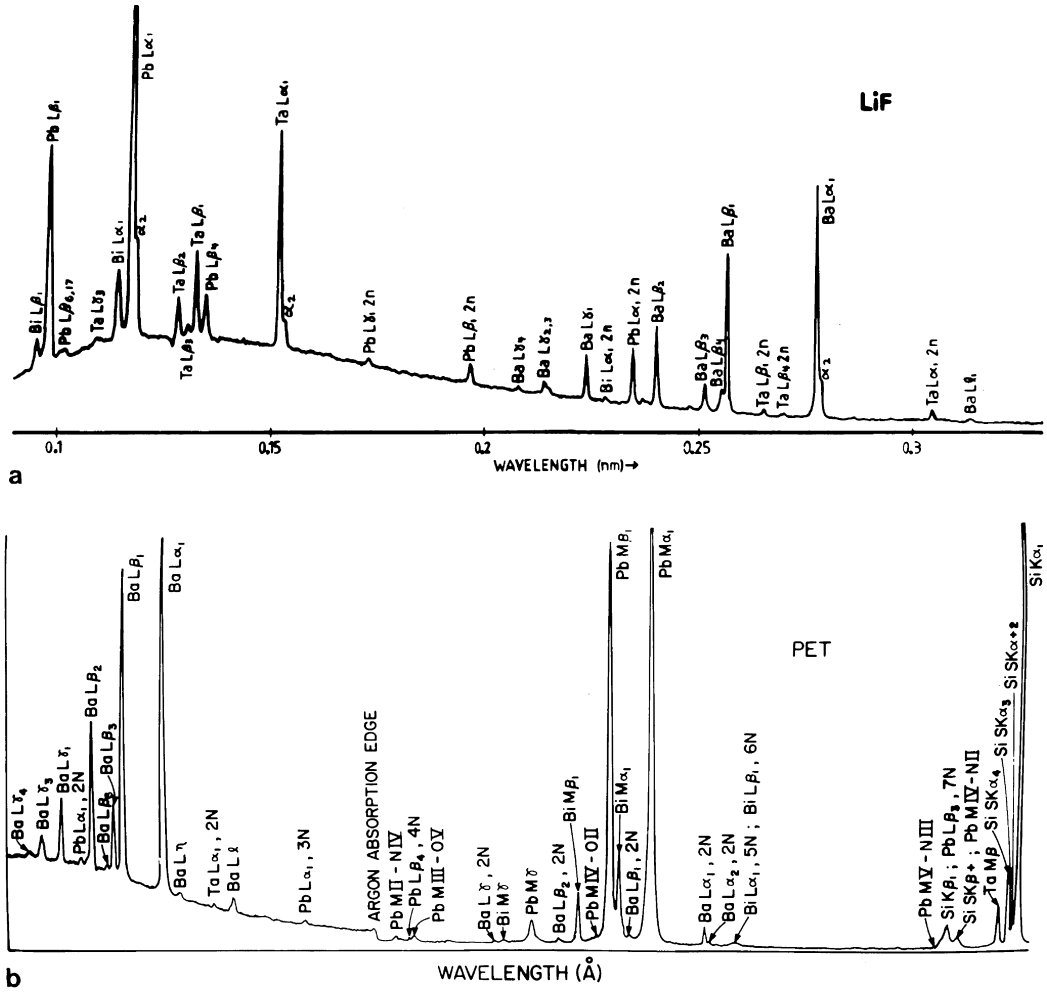
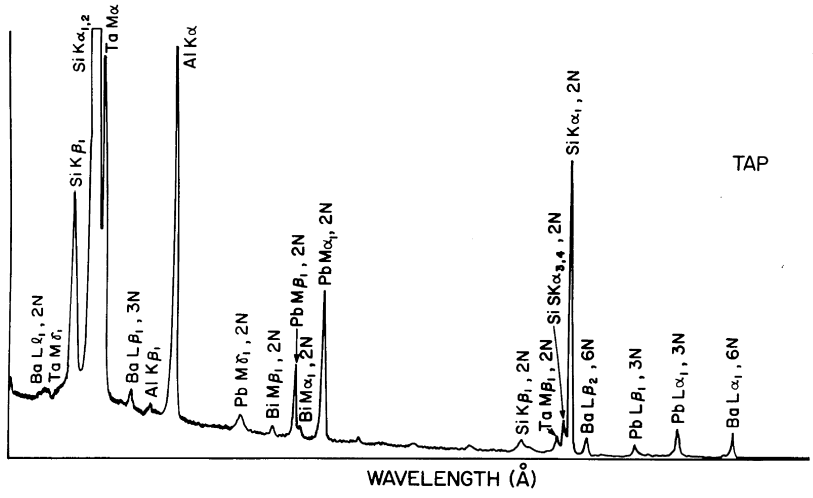
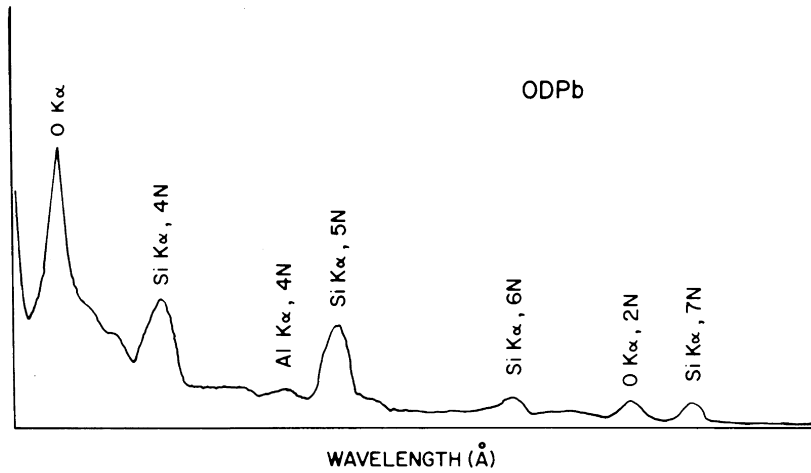


Figure 8.22. WDS spectra of NIST glass K–251 for (a) LiF diffractor, (b) PET diffractor, (c) TAP diffractor, and (d) lead stearate diffractor.



c



d

Figure 8.22. (*Continued*)

that much more difficult and important in order to avoid misassignment of low-intensity peaks.

2. The improved peak-to-background ratio of the WDS spectrometer provides the capability of detection and identification at a concentration about a factor of 10 lower than the EDS (roughly 100 ppm for the WDS compared to 1000 ppm for EDS).

3. The WDS spectrometer can detect elements with atomic numbers in the range $4 < Z < 11$. Windowless or thin-window EDS spectrometers can also detect elements as low as $Z = 4$. However, the detection limits of EDS are much poorer than WDS for the light elements, and the interference situation for EDS in the low-photon-energy range from intermediate- and heavy-element L and M peaks is significant.

Table 8.3. Composition of NIST K-251

Element	Wt%
O	23.0
Al	1.3
Si	14.0
Ba	9.0
Ta	4.1
Pb	44.1
Bi	4.5

The principal drawbacks of WDS qualitative analysis in addition to the multiplicity of peaks and higher order reflections noted above are the following.

1. Qualitative analysis is much slower than in EDS, with scan times of 30 min per crystal and beam currents of 100 nA or more required for a full-wavelength scan at high sensitivity. With automated WDS, the process can be speeded up by just scanning those portions of the spectrum where peaks occur, and more rapidly stepping between peak positions. A full qualitative analysis (for elements with $Z \geq 4$) requires the use of at least five crystals. If multiple spectrometers are available, these crystals can be scanned in parallel. The multiplicity of peaks requires much more of the analyst's time for a complete solution of the spectrum, especially if it is done manually (utilizing a full x-ray table, e.g., Bearden, 1967).
2. Much higher beam currents are required for use in the WDS compared to the EDS because of the considerably lower geometric and quantum efficiencies of the WDS (see Chapter 7). For fragile specimens such as biological targets, these high currents may be unacceptable.

8.3.2. Guidelines for WDS Qualitative Analysis

Given a sample of the complexity of that shown in Figs. 8.22a–8.22d, the following procedure should be used:

1. Because of the possibility of a peak originating from a high-order reflection, it is best to start at the highest photon energy, that is, the shortest wavelength end of the spectrum for each crystal, where the probability is highest for finding first-order, $n = 1$, peaks.
2. The highest intensity peak should be selected in the short-wavelength end of the LiF crystal scan and its wavelength determined. From a complete x-ray reference such as that of Bearden (1967), the possible elements that could produce the peak in question, such as a $K\alpha_{1,2}$ or $L\alpha_{1,2}$, should be noted. In parallel with the concept of a family of peaks introduced in the description of EDS qualitative analysis, when a candidate element is tentatively associated with a peak designated $K\alpha_{1,2}(n = 1)$, the analyst should immediately locate the associated $K\beta_1$ peak. Again, the ratio of $K\alpha$ to $K\beta$ should be roughly 10 : 1. However, because of changes in crystal and detector efficiency, the expected ratios may not always be

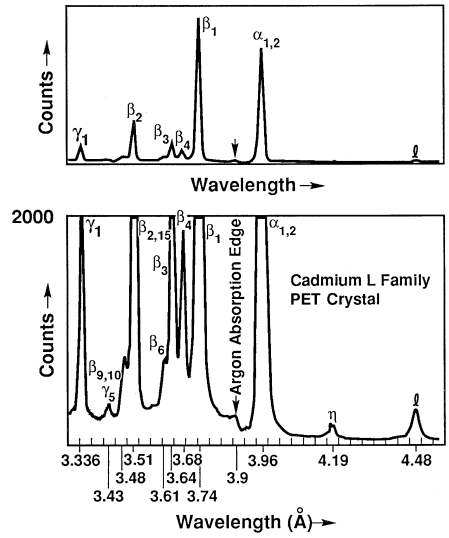


Figure 8.23. WDS scans of Cd showing change in gain at the Ar *K* edge.

found. For example, in the Cd spectrum in Fig. 8.23 recorded with a flow detector operating with P-10 gas (90% argon, 10% methane) the efficiency of the detector approximately doubles on the low-wavelength side of the argon *K* absorption edge (*K* edge = 3.203 keV). Hence the $L\beta_1$ peak, which should be approximately 60% as big as the $L\alpha$, peak, is actually larger. The efficiency doubles at the argon *K* edge because the flow-proportional x-ray detector of this spectrometer utilizes P-10 gas (90% Ar–10% methane). The dimensions of the detector and the pressure of the P-10 gas permit a certain fraction of x-rays of wavelength longer than the edge wavelength to pass through the gas without interaction. For those x-rays with wavelengths

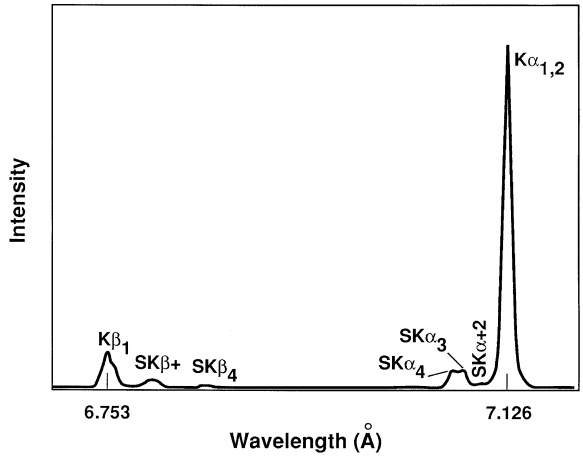


Figure 8.24. WDS spectrum of pure Si taken with a PET crystal at a beam energy of 20 keV. Note the presence of satellite lines labeled S.

shorter than the edge wavelength a greater fraction (approximately twice as many) will interact with the gas and hence be detected. This effect is not seen in the WDS scan of the Cd L family in Fig. 8.20 because a sealed xenon (L_{III} edge = 4.782 keV) counter tube was used.

Note also that the resolution of the WDS with some crystals such as LiF and quartz is sufficient to show at least some separation of $K\alpha$ into $K\alpha_1$ and $K\alpha_2$, which have a ratio of $K\alpha_1 : K\alpha_2$ of 2 : 1. Similarly, if an $L\alpha$ peak is suspected, the full L family should be sought. Note that L peaks in addition to those listed in Fig. 8.1 (i.e., $L\alpha_{1,2}$, $L\beta_1$, $L\beta_2$, $L\beta_3$, $L\beta_4$, $L\gamma_1$, $L\gamma_3$, $L\iota$, $L\eta$) for Si EDS analysis may be found because of the excellent resolution and peak-to-background ratio of WDS. In identifying families of peaks, it is possible that, because of wavelength limitations of the crystals, only the main peak may be found (e.g., Ge $K\alpha$ on LiF, where Ge $K\beta$ lies outside the crystal range). Applying this approach to the LiF spectrum in Fig. 8.22a indicates that the major peak at 0.118 nm could correspond to As $K\alpha$ or Pb $L\alpha$. Inspection for other family members reveals the presence of the Pb L family, but not the As $K\beta$. Thus Pb is identified with certainty and As is excluded.

3. Once the element has been positively identified and all possible first-order members of the family of x-ray peaks have been located, which may require examining spectra from more than one crystal (e.g., Zn $K\alpha$, $K\beta$ on LiF; Zn L family on PET), the analyst should locate all possible higher order peaks associated with each first-order peak throughout the set of crystals. Continuing the heavy element glass example, the higher orders for the Pb L family are noted in the remainder of the LiF spectrum as well as on the PET and TAP spectra in Figs. 8.22b and 8.22c. No higher order Pb L peaks are found in the lead stearate spectrum (Fig. 8.22d). Note that in this case the orders can extend to values as high as 7, as shown in Fig. 8.22d for silicon.

4. Only after all possible family members of the element and all higher order reflections of each member have been identified should the analyst proceed to the next unidentified high-intensity, low-wavelength peak. The procedure is then repeated.

5. Wavelength-dispersive spectrometry is sufficiently complicated that a number of special considerations arise: (a) For some elements, the parent $K\alpha$ peak ($n = 1$) may lie outside the range of the low-wavelength crystal (usually LiF), whereas higher order reflections of that $K\alpha$ peak may still be found. An example is zirconium $K\alpha$ at 0.078 nm (0.785 Å). (b) Satellite peaks will also be observed as low-intensity peaks on the high-energy shoulder of a high-intensity peak. Satellite peaks arise from doubly ionized atoms and are illustrated in Fig. 8.24. Note that several of the satellite peaks of the Si $K\alpha$ peak are nearly as high as the Si $K\beta$ peak.

References

Bearden, J. A. (1967). *Rev. Mod. Phys*. **39,** 78.

Fiori, C. E., and D. E. Newbury (1978). In *SEM/1978/1,* SEM Inc., AMF O'Hare, Illinois, p. 401.

9

Quantitative X-Ray Analysis: The Basics

9.1. Introduction

As discussed in Chapters 6–8, the x-rays emitted from a specimen bombarded with the finely focused electron beam of the scanning electron microscope (SEM) can be used to identify which elements are present (qualitative analysis). Using flat-polished samples and a proper experimental setup and data reduction procedure, one can use the measured x-rays from the identified elements to quantitatively analyze chemical composition with an accuracy and precision approaching 1%. We term the analysis of flat-polished samples "conventional microprobe analysis." The beam energy for conventional microprobe analysis is usually $E_0 \geq 10$ keV.

This chapter provides an overview of the basic principles and techniques used for determining chemical composition, on the micrometer scale, for conventional microprobe analysis. Our intention is to provide the conceptual basis for an understanding of the x-ray microanalytical data reduction procedures for conventional flat-polished samples that today are almost always incorporated into a computer-based, integrated analysis system with which the analyst interacts as a user. It is extremely important to grasp the underlying physical principles of the data reduction procedures to enable the user to apply the best analytical strategy. The x-ray microanalysis software often presents several possible choices of data correction procedures for the user to make, and an optimal choice obviously depends on proper knowledge of the relative merits of each approach. This chapter discusses the theories for x-ray generation and emission and introduces the ZAF and $\phi(\rho z)$ techniques for conventional quantitative analysis of flat-polished specimens. Standardless analysis, special procedures for geological analysis, and precision and sensitivity of analysis are also described. Chapter 10 presents material on x-ray microanalysis of nonideal sample geometries (layered samples, particles, rough surfaces, etc.) and the use of special x-ray techniques such as low-voltage microanalysis and compositional mapping.

9.2. Advantages of Conventional Quantitative X-Ray Microanalysis in the SEM

There are three major advantages of conventional quantitative x-ray microanalysis: (1) small analytical volume, (2) nondestructive analysis capability, and (3) analytical precision and accuracy at the ±1–2% level. These advantages are briefly described in the following paragraphs.

1. Most materials are chemically heterogeneous on the microscopic scale and compositional data are necessary to investigate the properties and behavior of materials at the micrometer spatial scale. As shown in Chapter 6, x-rays can be generated in conventional flat-polished samples, depending on the initial electron beam energy and atomic number, from volumes with linear dimensions as small as 1 μm. This means that a volume as small as 10^{-12} cm^3 can be analyzed. Assuming a typical density of 7 g/cm^3 for a transition metal, the composition of 7×10^{-12} g of material can be determined. Elemental constituents in this small mass of the sample, selected by the electron/x-ray interaction volume, can be determined to concentrations ranging as low as 0.1%, based on the practical limit of energy-dispersive x-ray spectrometry (EDS) in the SEM. The limit of detection in terms of mass for a specific element is then $\sim10^{-14}$–10^{-15} g in the SEM. For reference, a single atom of iron weighs about 10^{-22} g.

2. The focused electron probe of the SEM is largely nondestructive and the specimen can be reexamined using optical microscopy or other techniques. This capability is extremely important when microanalysis or microstructural information must be complemented by other data from the same sample. Virtually all types of hard materials can be analyzed using the SEM. Among these materials are metals and their alloys, minerals, and nonmetals such as nitrides, carbides, oxides, etc. Certain specimen classes are vulnerable to degradation during analysis. Biological materials contain water that will boil off in the vacuum or under electron bombardment. Polymers and biological materials, on the other hand, can chemically decompose due to electron-induced reactions. Both types of materials are poor conductors. Conventional x-ray analysis is often difficult to perform and these materials are considered as special cases (see Chapters 10, 12, and 13).

3. It is important to know how much of an element is present in a given analytical volume, that is, the *accuracy* of the analysis. In addition it is important to know what the statistical variation of the x-ray measurement is at each analysis point, that is, the *precision* of the analysis. Conventional quantitative analysis with flat-polished samples allows for analytical accuracy at the ±1–2% level. Instrument-limited precision can be readily achieved in the 0.1–1% range. This means, for example, that we can analyze a micrometer-sized region of a sample containing 10.0 wt% of an element and know that that analysis has an *accuracy* of ±0.1–0.2 wt%. In addition, we can distinguish this analysis from another analysis point containing 10.1 or 10.2 wt% of that element, a precision of ±1–2%. These levels of accuracy and precision can be achieved by accurately extracting characteristic x-ray peaks from the background and other interfering peaks, by having accurate correction procedures to convert x-ray measurements to composition, and by obtaining statistically sufficient x-ray counts from

the element of interest at each analysis point. An introduction to correction procedures is given in Sections 9.3–9.7. Precision of the analysis is discussed in Section 9.6.4.

9.3. Quantitative Analysis Procedures: Flat-Polished Samples

The procedure for obtaining a chemical analysis of a micrometer-sized region in a sample using x-ray microanalysis involves several steps, the least of which may be the quantitative calculation step itself. Quantitative analysis procedures involve the following steps:

1. Define the problem: What do we want to learn?
2. Perform a qualitative analysis on a sample to determine the elements to be measured.
3. Select the necessary sample(s) and appropriate standards.
4. Prepare the sample and standards for x-ray analysis. Flat-polished samples are used for conventional x-ray analysis.
5. Measure the appropriate x-ray intensities for quantitative analysis.
6. Develop the appropriate x-ray intensity k ratios.
7. Perform quantitative matrix corrections to obtain numerical concentrations.

The following paragraphs outline the quantitative analysis procedure.

Step 1. *Define the problem*. It is very important to decide just what the analyst wishes to learn from the analysis. This decision defines the analysis procedures to be used including the setup for the instrument (keV, beam current, analysis time, etc.). One has to decide whether the analysis involves the measurement of a composition gradient or the measurement of composition at interfaces between phases or precipitates. Analysis volume may be a critical factor here and is controlled primarily by the energy of the electron beam and secondarily by the probe current. The measurement of homogeneity or minimum detectability limits for specific elements involves statistical measurements controlled by peak intensity, x-ray peak-to-background ratios, and counting times. Phase composition and multielement analysis measurements are often complex. The analysis problem will determine how many elements will be measured and perhaps require long counting times for trace or minor elements. Area scans across complex microstructures can yield compositional data if adequate analysis time is available and point-to-point scanning can be done through automation. Finally, routine analysis of multiple samples involves other sample handling and instrumentation capabilities.

Step 2. *Perform a qualitative analysis to determine the elements to be measured.* Before attempting a quantitative analysis, the analyst must be sure to perform a qualitative analysis to identify the constituents present with a high degree of confidence. This topic has been discussed in detail in Chapter 8. This qualitative analysis may be performed on a "raw" piece of the unknown which has not yet undergone full preparation, for

example, metallography. No quantitative analysis, no matter how carefully performed, is useful if the spectral peaks being quantified actually belong to another element! It is useful to estimate the composition range of the elements identified in the sample. Knowing approximate composition ranges allows one to choose the optimum combination of spectrometers for analysis as well as the analysis time(s). Furthermore, if not all the elements can be measured (e.g., C, N, O, etc.), accommodations must be made in the correction schemes for x-ray microanalysis.

Step 3. *Select the necessary sample(s) and appropriate standards.* For quantitative x-ray microanalysis, the sample and standards must be stable under the electron beam. If the sample and standards are not stable under the beam, specific action must be taken to deal with this situation (see Chapters 10–14). In addition, it is necessary to prepare specimens and standards that are conducting. Using conducting standards eliminates charging effects, which can lead to movement of the electron beam from the selected analysis position and/or changes in x-ray intensity.

A standard for x-ray microanalysis is a substance which has a known composition and must be homogeneous at the microscopic level. Most substances are not microscopically homogeneous, however, and they cannot serve as x-ray microanalysis standards. Because one must use the standards several times during a typical analysis session, each micrometer-sized region must have the same composition or analysis errors will be severe. It is also necessary to have a full analysis of the standard so that elements other than the one of interest can be accounted for in the x-ray analysis procedure. If the standard is not well-characterized, element x-ray intensities can be influenced by the presence of the unknown elements. It is most useful to have standards close in composition for each element in the sample. If standards are close in composition, the amount of correction necessary for x-ray microanalysis will be minimized and the accuracy of the analysis will be improved. Fortunately, and this is one of the great strengths of electron-excited x-ray microanalysis, accurate quantitative analysis can be performed with simple standards consisting of pure elements. For those elements whose properties are incompatible with the requirements of measurement in a vacuum or are unstable under electron bombardment (e.g., phosphorus), a simple stoichiometric compound, such as GaP, will suffice.

Step 4. *Prepare the sample and standards for x-ray analysis*. Flat-polished samples are used for conventional x-ray analysis. The sample and standards should be flat-polished with the last polish leaving scratches much less than 0.1 μm in size, that is, scratches which cannot be seen in the optical microscope. A flat polish minimizes height differences at interfaces and minimizes or even eliminates surface layers such as oxide or corrosion. For all samples, it is critical to mount the materials in a conducting medium or to coat the materials so that the beam electrons find a path to ground. For a coated surface, the sample must be connected to a good ground path on the sample holder/stage. Chapter 14 discusses coating and conductivity techniques in some detail.

A flat-polished sample must not be etched or chemically treated once a flat polish at the submicrometer level is completed. The etching or chemical treatment will alter the surface composition and create height differences

at interfaces and between phases, which will result in too much or too little x-ray generation. Given the strict specimen preparation requirements, it is often difficult to locate areas of interest for analysis once the sample is prepared. In most analysis situations, the areas of interest are marked on the sample (scratches, microhardness indentations, etc.) before the final polish is performed. Alternately backscattered electron images can be used to compare areas with those previously identified by optical microscopy or SEM. After the final polish and before analysis, samples should be stored in dry environments or in vacuum to minimize the amount of surface contamination that may build up. This is particularly critical when analysis of oxygen or carbon is performed.

Step 5. *Measure the appropriate x-ray intensities for quantitative analysis*. Now that we have determined the problem to be solved, identified the elements to be analyzed, and secured and prepared the samples and standards, we are ready to begin the quantitative analysis process. This process requires the following steps:

(a) *Obtain the x-ray spectrum of the specimen and standards under defined and reproducible conditions*. Quantitative analysis makes the assumption that all measurements for a given element, in both the specimen and the standards, are performed with identical spectrometer response (efficiency, spectrometer take-off angle, calibration, and resolution). The measurement is performed at the same beam energy and under an accurately known electron dose: beam current multiplied by the spectral accumulation time. The experimental measurement procedure is shown in Fig. 9.1. If possible, E_0 should be chosen to obtain an optimum overvoltage U, the ratio of the electron beam energy to the critical energy to excite x-rays from the elements of interest, in the range 2–3. Normalization can be applied to adjust for differences in dose if spectra are recorded for different times and/or at different beam currents. However, the accuracy of the normalization process is dependent on the accuracy with which the beam current and spectrum accumulation time have been controlled. Hence, an accurate beam-current measuring system (specimen stage Faraday cup and calibrated picoammeter) is necessary and for EDS analysis an accurate dead-time correction by the EDS system must be made.

(b) *Process the spectra of the specimen and standards to remove the background.* To obtain the sample intensity I_i or the standard characteristic intensity $I_{(i)}$ the background continuum must be subtracted from the measured characteristic x-ray intensities (Fig. 9.1). The analyst uses the ratio of

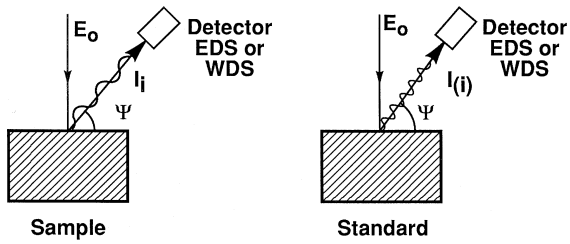


Figure 9.1. General experimental setup for quantitative x-ray microanalysis.

the measured intensities after continuum background is subtracted ($I_i / I_{(i)}$) from sample and standard, respectively, which is called the k ratio (or "k-value"), to obtain the composition of the sample. By measuring the k ratio, instrumental factors and other constant factors related to the specimen and standard will drop out. For minor and trace constituents, the background formed by the x-ray bremsstrahlung or continuum is a significant fraction of the characteristic peak intensity. Incomplete or inaccurate removal will adversely affect the accuracy of quantitation.

Different procedures are used to subtract background x-rays for WDS and EDS spectra. For WDS it is often a relatively easy matter to get a very good estimate of the background under a peak by detuning the spectrometer to wavelengths slightly above and below the tails of the characteristic peak. The value of the background at the peak setting is obtained by linear interpolation (Fig. 9.2). For all measurements, especially for minor or trace constituents (<1 wt%), it is good practice to scan the spectrometer through the peak to inspect the nearby regions for suitable background locations free from other peaks and artifacts.

For EDS background subtraction, the process is more complex because of the relatively poor resolution of the spectrometer. The x-ray intensities at a given characteristic peak in the EDS are determined from the number of counts present in a region of interest (ROI) set around the peak centroid. The size of the ROI is usually equal to or greater than the FWHM of peak (see Chapter 7, Section 7.2.7.1). The contents of the ROI contain counts from both the characteristic line and the background, so removal of the latter from both the sample and standard intensities is necessary to determine the k ratio. One of several approaches is generally used. They are as follows.

Interpolation or Extrapolation. Figure 9.3 shows a portion of a pure chromium spectrum. The background value at the peak channel determined by interpolation is about 100 counts if points B and D are used or about 130 counts if points B and C are used. The latter pair is the better of the two because the chromium absorption edge at 5.989 keV produces an abrupt drop in the background at higher energies, which is obscured by the characteristic Cr peaks. If the spectrum had been collected with a

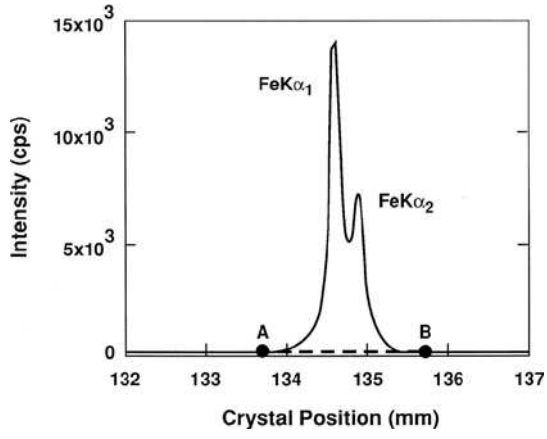


Figure 9.2. Commonly used method of background measurement with a WDS using linear interpolation between measured background positions at A and B. Crystal position is equivalent to wavelength in this figure.

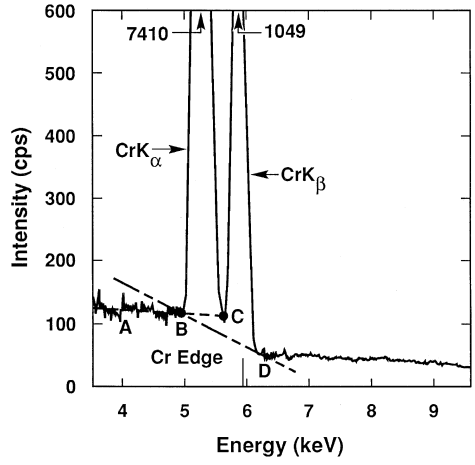


Figure 9.3. EDS background fitting by linear interpolation (points *B–C* or *B–D*) or extrapolation (points *A–B*) in a simple spectrum of chromium (Lifshin, 1975).

detector system with poorer resolution than that used in obtaining Fig. 9.3 (160 eV at 5.89 keV), then point *C* could not be used. In this example, the determination of the background by extrapolation of the low-energy points *A* and *B* to point *C* should be used in preference to interpolation of points *B* and *D*. The error associated with use of this approach would affect the measurement of the pure chromium standard by less than 1%. If the spectrum had been that of a copper–1 wt% chromium alloy, the Cr portion of the spectrum would look substantially the same with a correspondingly lower peak-to-background ratio. The error introduced by this background extrapolation would, however, rise to 100% for the 1% Cr constituent.

For a mixture of elements, the spectrum becomes more complex, and interpolation is consequently less accurate. Figure 9.4 shows a stainless steel spectrum taken under the same operating conditions as the Cr spectrum in Fig. 9.3. There is a sufficient number of $K\alpha$ and $K\beta$ peaks such that the use of extrapolation is questionable for Fe and Ni, whereas interpolation between points *A* and *B* must be performed over a 3-keV range. Because the real background is irregular owing to the presence of Cr and Fe absorption

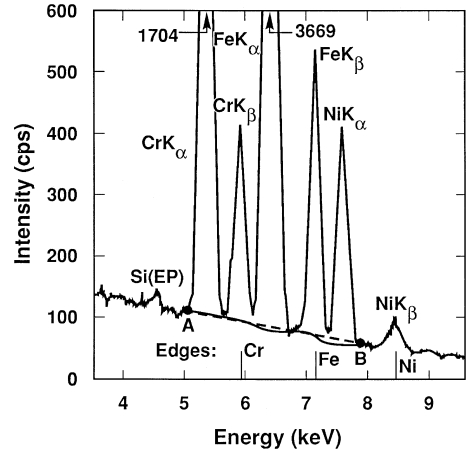


Figure 9.4. EDS background fitting by linear interpolation in a complex spectrum of stainless steel.

edges, an error of about 3% would occur in the Ni, intensity measurement in this specimen using background extrapolation by linear interpolation.

Background Modeling. In this approach the background is calculated from a physical model of the continuum based either on theory or experimental measurements. The first attempts at this type of background subtraction were based on the use of the purely theoretical model of Kramers (1923) to predict the spectral distribution of the continuum generated within the sample or standard. This distribution was modified by sample absorption, absorption in detector windows, and electrical contacts, and then convoluted with the detector response as determined by its FWHM (Fiori *et al.,* 1976). Details of the background modeling approach are given in the Enhancements on the accompanying CD, Section E9.1. An example of applying this technique to fitting the continuum from a complex sample is presented in Figure 9.5. The sample is a mineral, Kakanui hornblende. It is possible to detect a Mn concentration of less than 700 ppm with the background modeling approach.

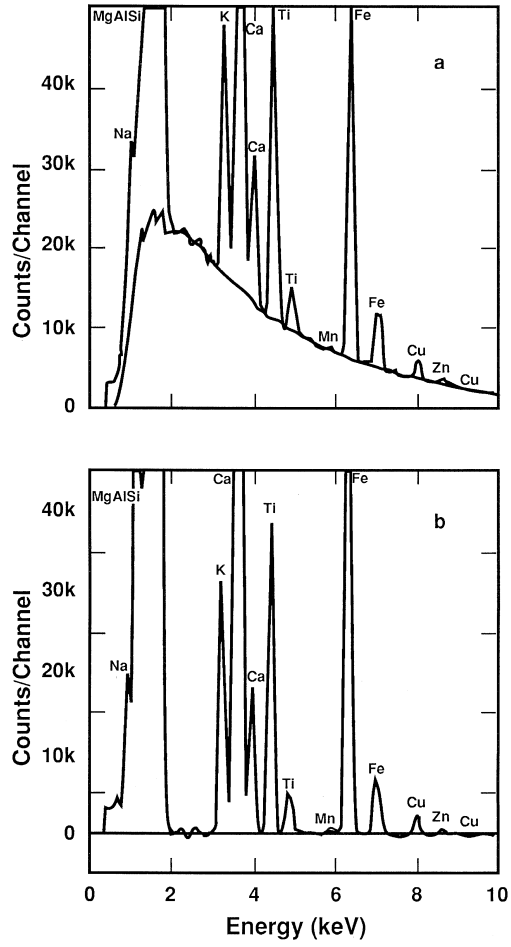


Figure 9.5. Background correction in a Kakanui hornblende. (a) Fitted continuum curve. The observed spectrum is superposed. Note the presence of the Mn $K\alpha$ peak at 5.9 keV. The concentration of Mn is less than 700 ppm. (b) Background subtracted.

Background Filtering. In this method of background calculation, the continuum component of an x-ray spectrum is viewed as an undesirable signal whose effect can be removed by mathematical filtering or modification of the frequency distribution of the spectrum. Knowledge of the physics of x-ray production, emission, and detection is not required.

A filtering technique which is now widely used was developed by Schamber (1978). It has been successfully applied to a wide variety of systems and uses what is known as a "top hat" digital filter. The top hat filter is a simple and elegant algorithm, a fact which can easily be obscured by the mathematical formalism required to describe it. Simply stated, the top hat filter is a special way of averaging a group of adjacent channels of a spectrum, assigning the "average" to the center "channel" of the filter and placing this value in a channel in a new spectrum, which we will call the filtered spectrum. The filter is then moved one channel and a new "average" is obtained. The process is repeated until the entire spectrum has been stepped through. The filter in no way modifies the original spectrum; data are only taken from the original to create a new spectrum. Further detailed information on the "top hat" digital filter can be found in the Enhancements on the CD, Section E9.2.

The effects of passing a top hat digital filter through an x-ray spectrum as recorded by a Si(Li) spectrometer system are to (1) strongly suppress the background and statistical scatter and (2) significantly alter the shape of the spectral peaks. Clear advantages of the method are simplicity and the fact that an explicit model of the continuum is not required. However, because the continuum has been suppressed, the information it carried (i.e., average atomic number, mass-thickness, etc.) is no longer available. The filter method is currently the most widely used procedure to remove background effects. The result of digitally filtering an original, experimentally generated spectrum of a complex geological specimen, Kakanui hornblende, is shown in Fig. 9.6.

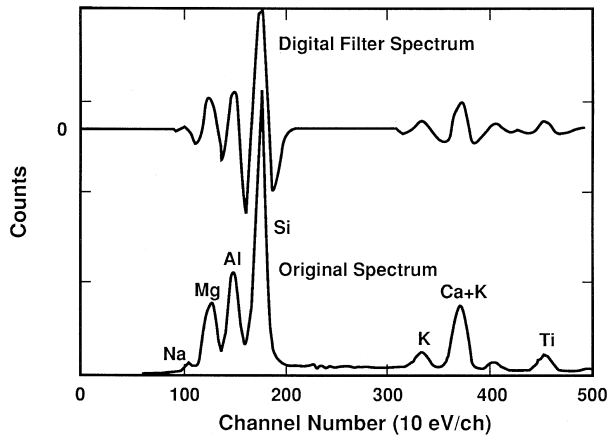


Figure 9.6. Spectrum of Kakanui hornblende (lower plot) and the digital filter of the spectrum (upper plot).

(c) *Process the spectra of the specimen and standards if peak overlaps occur.* Peak intensities used for quantitative WDS calculations are normally based on sample and standard measurements made at the peak centroid followed by background subtraction. Although there are some noteworthy peak overlaps such as Mo $L\alpha$ (2.293 keV) and S $K\alpha$ (2.308 keV), they are relatively rare and can even sometimes be circumvented by the use of other nonoverlapping lines for an analysis.

In the case of EDS, however, poorer energy resolution often leads to characteristic peak overlap situations, particularly at lower energies. A further complication is that in EDS analysis the counts used for the sample and standard are those contained in a region of interest (ROI) selected for each element rather than the peak centroid value used in WDS analysis (see chapter 7, Section 7.4.3). ROIs normally correspond to 1–2 times the FWHM of the detector and are selected to maximize the product of P/B times P as described in Section 9.8.4. Figure 9.7 illustrates what happens

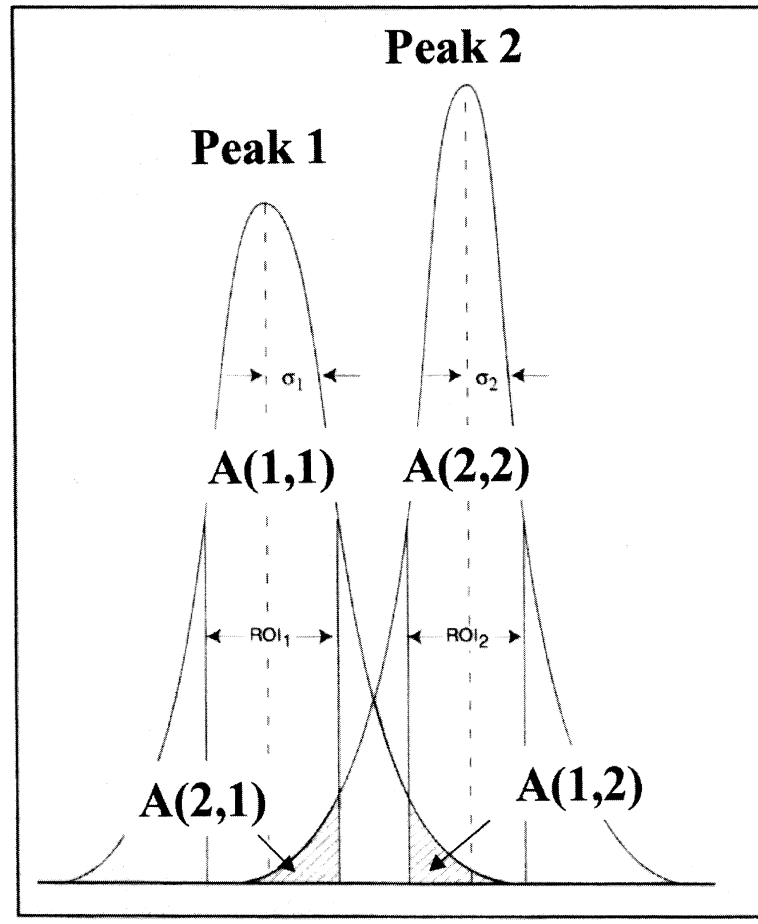


Figure 9.7. Example of peak overlap of peak 1 and peak 2 in EDS analysis. The tails of the two peaks overlap.

when two elemental peaks are so close that one of the tails of one adds counts to the region of interest of the other. For peak 1 the integrated counts within ROI_1 is given by $A(1, 1)$ and the tail of peak 2 contributes $A(2, 1)$. Similarly, for peak 2 the integrated number of counts within ROI_2 is given by $A(2, 2)$ and the tail of peak 1 contributes $A(1, 2)$. The challenge in dealing with real spectra (after background removal) where $A(1, 1)$ and $A(2, 1)$ are combined in ROI_1 is to determine the value of $A(1, 1)$ for peak 1 and in a similar fashion determine $A(2, 2)$ for peak 2. This can be done in practice either by peak deconvolution programs or by the use of overlap coefficients.

With peak deconvolution the measured spectrum (after background removal) is viewed as the sum of weighted spectra of the two individual spectra each containing the pure elemental peaks 1 and 2. A multiple linear least squares fit is then used to determine the weighting factors that give the best match to the measured spectra. If the weighting factors are expressed as fractions of the pure peaks, then those fractions are, in fact, the k ratios needed for the quantitative analysis programs. The trick with this method is to know the exact shape of the pure peaks because they do not always follow a well-defined mathematical function such as a Gaussian. Instead it is usually necessary to work with measured reference spectra determined on the spectrometer actually used for the measurements, hoping that there is no variation in spectral shape and/or position due to count rate or chemical bonding effects.

An alternative approach is to use the method of overlap coefficients. This method does not require an exact knowledge of the peak shape, yet is effective if the peak shape and position do not change significantly between the sample and standard. The first step is to measure pure spectra for each element. The next is to determine the ratio of the integrated number of counts at the setting for the ROI of the overlapping element (even though it is not there) and the value at the ROI for the pure element being measured. This fraction is called an overlap coefficient. Thus the overlap of peak 2 in ROI_1 is given by

$$O(1, 2) = \frac{A(2, 1)}{A(2, 2)}. \tag{9.1a}$$

Similarly,

$$O(2, 1) = \frac{A(1, 2)}{A(1, 1)}. \tag{9.1b}$$

Then to a first approximation

$$A(1, 1) = \mathrm{ROI}_1 - O(1, 2) \cdot \mathrm{ROI}_2 \tag{9.2a}$$

and

$$A(2, 2) = \mathrm{ROI}_2 - O(2, 1) \cdot \mathrm{ROI}_1. \tag{9.2b}$$

In this manner one uses the area of the peak not analyzed to estimate its contribution to the ROI of the peak analyzed. It is stated as a first approximation because the peak not analyzed is enhanced by the one analyzed (e.g., ROI_2 contains some counts from element 1, so its value is slightly

overstated). It is possible to write out an infinite series to demonstrate this effect just using the equations above, but the terms will become smaller and smaller, so it is generally not necessary to use more than one term unless the peaks are very close. Forming a k ratio for subsequent quantitative analysis for element 1 involves taking the value of $A(1, 1)$ and dividing it by the corresponding value of the total counts contained in ROI_1 for the element 1 standard and a similar calculation for element 2.

Step 6. *Develop the x-ray intensity k ratios using the specimen intensity I_i and the standard intensity $I_{(i)}$ after the continuum background is subtracted.* The x-ray intensity ratios k_i, k_j, k_k, etc. are obtained for each element i, j, k, etc. present in the sample (see Fig. 9.1).

Step 7. *Perform quantitative matrix corrections to obtain numerical concentrations.* This critical step is discussed in more detail in the following sections.

9.4. The Approach to X-Ray Quantitation: The Need for Matrix Corrections

Upon initial examination, it would seem that quantitative analysis should be extremely simple. Just form the ratio of the characteristic x-ray intensity for a given element measured from the specimen to that measured from the standard, and that ratio should be equal to the ratio of concentrations for a given element between the specimen and the standard. As was first noted by Castaing (1951), the primary generated intensities are roughly proportional to the respective mass fractions of the emitting element. If other contributions to x-ray generation are very small, the measured intensity ratios between specimen and standard are roughly equal to the ratios of the mass or weight fractions of the emitting element. This assumption is often applied to x-ray quantitation and is called Castaing's "first approximation to quantitative analysis"; it is given by

$$C_i / C_{(i)} = I_i / I_{(i)} = k_i \tag{9.3}$$

The terms C_i and $\mathrm{C}_{(i)}$ are the composition in weight (mass) concentration of element i in the unknown and in the standard, respectively. The ratio of the measured unknown-to-standard intensities after continuum background is subtracted and peak overlaps are accounted for, $I_i / I_{(i)}$, is the basic experimental measurement which underlies all quantitative x-ray microanalysis and is given the special designation the "k ratio."

Careful measurements performed on homogeneous substances of known multielement composition compared to pure-element standards reveal that there are significant systematic deviations between the ratio of measured intensities and the ratio of concentrations. An example of these deviations is shown in Fig. 9.8, which depicts the deviations of measured x-ray intensities in the iron–nickel binary system from the linear behavior predicted by the first approximation to quantitative analysis, Eq. (9.3). Figure 9.8 shows the measurement of $I_i / I_{(i)}$ for Ni $K\alpha$ and Fe $K\alpha$ in nine well-characterized homogeneous Fe–Ni standards (Goldstein *et al.,* 1965).

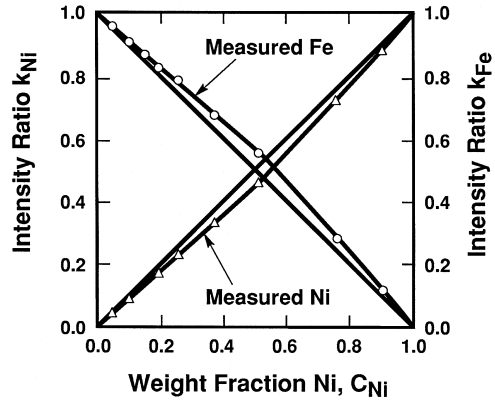


Figure 9.8. Measured Fe–Ni intensity ratios, k ratios, versus weight fraction of Ni for 30 keV. Curves are measured k ratios, straight lines represent ideal behavior.

The data were taken at an initial electron beam energy of 30 keV and a take-off angle ψ of 52.5°. The intensity ratio k_{Ni} or k_{Fe} is the $I_i / I_{(i)}$ measurement for Ni and Fe, respectively. The straight lines plotted between pure Fe and pure Ni indicate the relationship between composition and intensity ratio given in Eq. (9.3). For Ni $K\alpha$, the actual data fall below the first approximation and indicate that there is an absorption effect taking place, that is, more absorption in the sample than in the standard. For Fe $K\alpha$, the measured data fall above the first approximation and indicate that there is a fluorescence effect taking place in the sample. In this alloy the Ni $K\alpha$ radiation is heavily absorbed by the iron and the Fe $K\alpha$ radiation is increased due to x-ray fluorescence by the Ni $K\alpha$ radiation over that generated by the bombarding electrons.

In most quantitative chemical analyses, the measured intensities from specimen and standard need to be corrected for differences in electron backscatter, density, x-ray cross section, and energy loss as well as absorption within the solid in order to arrive at the ratio of generated intensities and hence the value of C_i. These effects are referred to as matrix or interelement effects. The magnitude of the matrix effects can be quite large, exceeding factors of 10 or more in certain systems. Recognition of the complexity of the problem of the analysis of solid samples has led numerous investigators to develop the theoretical treatment of the quantitative analysis scheme first proposed by Castaing (1951).

9.5. The Physical Origin of Matrix Effects

What is the origin of these matrix effects ? The x-ray intensity *generated* for each element in the specimen is proportional to the concentration of that element, the probability of x-ray production (ionization cross section) for that element, the path length of the electrons in the specimen, and the fraction of incident electrons that remain in the specimen and are not backscattered. It is very difficult to calculate the absolute generated intensity for the elements present in a specimen directly. Moreover, the intensity

that the analyst must deal with is the *measured* intensity. The measured intensity is even more difficult to calculate, particularly because absorption and fluorescence of the generated x-rays may occur in the specimen, thus further modifying the measured x-ray intensity from that predicted on the basis of the ionization cross section alone. Instrumental factors such as differing spectrometer efficiency as a function of x-ray energy must also be considered. Many of these factors are dependent on the atomic species involved. Thus, in mixtures of elements, matrix effects arise because of differences in elastic and inelastic scattering processes and in the propagation of x-rays through the specimen to reach the detector. For conceptual as well as calculational reasons, it is convenient to divide the matrix effects into atomic number, Z_i, x-ray absorption, A_i, and x-ray fluorescence, F_i, effects.

Using these matrix effects, the most common form of the correction equation is

$$C_i/C_{(i)} = [ZAF]_i \cdot I_i/I_{(i)} = [ZAF]_i \cdot k_i, \tag{9.4}$$

where C_i is the weight fraction of the element of interest in the sample and $C_{(i)}$ is the weight fraction of i in the standard. This equation must be applied separately for *each* element present in the sample. Equation (9.4) is used to express the matrix effects and is the common basis for x-ray microanalysis in the SEM/EPMA.

It is important for the analyst to develop a good idea of the origin and the importance of each of the three major nonlinear effects on x-ray measurement for quantitative analysis of a large range of specimens.

9.6. ZAF Factors in Microanalysis

The matrix effects Z, A, and F all contribute to the correction for x-ray analysis as given in Eq. (9.4). This section discusses each of the matrix effects individually. The combined effect of ZAF determines the total matrix correction.

9.6.1. Atomic Number Effect, Z

9.6.1.1. Effect of Backscattering (R) and Energy Loss (S)

One approach to the atomic number effect is to consider directly the two different factors backscattering (R) and stopping power (S) which determine the amount of generated x-ray intensity in a sample. Dividing the stopping power S for the sample and standard by the backscattering term R for the sample and standard yields the atomic number effect Z_i for each element i in the sample. A discussion of the R and S factors follows.

1. Backscattering R

The process of elastic scattering in a solid sample leads to backscattering, which results in the premature loss of a significant fraction of the beam electrons from the target before all of the ionizing power of those electrons has been expended in generating x-rays of the various elemental

constituents. From Fig. 3.8, which depicts the backscattering coefficient as a function of atomic number, this effect is seen to be strong, particularly if the elements involved in the unknown and standard have widely differing atomic numbers. For example, consider the analysis of a minor constituent, 1% of aluminum in gold, against a pure aluminum standard. In the aluminum standard, the backscattering coefficient is about 15% at a beam energy of 20 keV, whereas for gold the value is about 50%. When aluminum is measured as a standard, about 85% of the beam electrons completely expend their energy in the target, making the maximum amount of Al $K\alpha$ x-rays. In gold, only 50% are stopped in the target, so by this effect, aluminum dispersed in gold is actually underrepresented in the x-rays generated in the specimen relative to the pure aluminum standard. The energy distribution of backscattered electrons further exacerbates this effect. Not only are more electrons backscattered from high-atomic-number targets, but, as shown in Figs. 3.13a, the backscattered electrons from high-atomic-number targets carry off a higher fraction of their incident energy, further reducing the energy available for ionization of inner shells. The integrated effects shown in Figs. 3.8 and 3.13a form the basis of the "R factor" in the atomic number correction of the "ZAF" formulation of matrix corrections.

2. Energy loss S

The rate of energy loss due to inelastic scattering also depends strongly on the atomic number. For quantitative x-ray calculations, the concept of the stopping power S of the target is used. S is the rate of energy loss given by Eq. (3.3) divided by the density ρ [$S = -(1/\rho)\ dE/ds$]. Using the Bethe formulation for the rate of energy loss dE/ds, Eq. (3.3), Table 3.1, one observes that the stopping power is a decreasing function of atomic number. The low-atomic-number targets actually remove energy from the beam electron more rapidly with mass depth ρz (the product of the density of the sample ρ and the depth dimension z) than do high-atomic-number targets.

An example of the importance of the atomic number effect is shown in Fig. 9.9. This figure shows the measurement of the intensity ratios k_{Au}

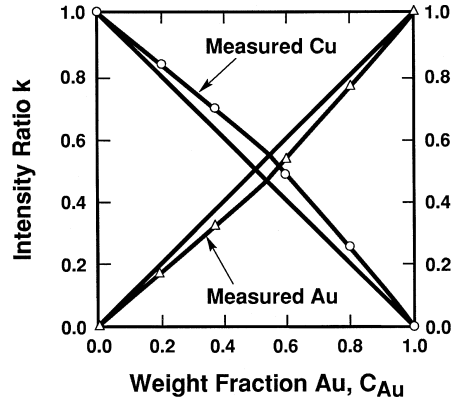


Figure 9.9. Measured Au–Cu intensity ratios k_{Au} and k_{Cu} versus the weight fraction of Au at 25 keV. Curves are measured k ratios, straight lines represent ideal behavior.

and k_{Cu} for Au $L\alpha$ and Cu $K\alpha$ for four well-characterized homogeneous Au–Cu standards (Heinrich *et al.,* 1971). The data were taken at an initial electron beam energy of 15 keV and a take-off angle of 52.5° and pure Au and pure Cu were used as standards. The atomic number difference between these two elements is 50,79 (Au) minus 29 (Cu). The straight lines plotted on Fig. 9.9 between pure Au and pure Cu indicate the relationship between composition and intensity ratio given in Eq. (9.3). For both Au $L\alpha$ and Cu $K\alpha$, the absorption effect A_i is less than 1% and the fluorescence effect F_i is less than 2%. For Cu $K\alpha$, the measured data fall above the first approximation and almost all the deviation is due to the atomic number effect, the difference in atomic number between the Au–Cu alloy and the Cu standard. As an example, for the 40.1-wt% Au specimen, the atomic number effect Z_{Cu} is 0.893, an increase in the Cu $K\alpha$ intensity by 11%. For Au $L\alpha$, the measured data fall below Castaing's first approximation and almost all the deviation is due to the atomic number effect. As an example, for the 40.1-wt% Au specimen, the atomic number effect Z_{Au} is 1.24, a decrease in the Au $L\alpha$ intensity by 24%. In this example, the S factor is larger and the R factor is smaller for the Cu $K\alpha$ x-rays, leading to a larger S/R ratio and hence a larger Z_{Cu} effect. Just the opposite is true for the Au $L\alpha$ x-rays, leading to a smaller Z_{Au} effect. The effects of R and S tend to go in opposite directions and tend to cancel.

9.6.1.2. X-Ray Generation with Depth, $\phi(\rho z)$

A second approach to calculating the atomic number effect is to determine the x-ray generation in depth as a function of atomic number and electron beam energy. As shown in Chapter 3, the paths of beam electrons within the specimen are well represented by Monte Carlo simulations of electron trajectories. In the Monte Carlo simulation technique, the detailed history of an electron trajectory is calculated in a stepwise manner. At each point along the trajectory, both elastic and inelastic scattering events can occur. The production of characteristic x-rays, an inelastic scattering process, can occur along the path of an electron as long as the energy E of the electron is above the critical excitation energy E_c of the characteristic x-ray of interest.

Figure 9.10 displays Monte Carlo simulations of the positions where K-shell x-ray interactions occur for three elements, Al, Ti, and Cu, using an initial electron energy E_0 of 15 keV. The incoming electron beam is assumed to have a zero width and to impact normal to the sample surface. X-ray generation occurs in the lateral direction x and in the depth dimension z. The micrometer marker gives the distance in both the x and the z dimensions. Each dot indicates the generation of an x-ray; the black regions indicate that a large number of x-rays are generated. This figure shows that the x-ray generation volume decreases with increasing atomic number (Al, $Z = 13$ to Cu, $Z = 29$) for the same initial electron energy. The decrease in x-ray generation volume is due to (1) an increase in elastic scattering with atomic number, which causes the electron path to deviate from the initial beam direction, and (2) an increase in critical excitation energy E_c, which gives a corresponding decrease in overvoltage U ($U = E_0/E_c$) with

atomic number. This decreases the fraction of the initial electron energy available for the production of characteristic x-rays. A decrease in overvoltage U decreases the energy range over which x-rays can be produced.

One can observe from Fig. 9.10 that there is a noneven distribution of x-ray generation with depth z for specimens with various atomic numbers and initial electron beam energies. This variation is illustrated by the histograms on the left side of the Monte Carlo simulations. These histograms plot the number of x-rays generated with depth into the specimen. In detail the x-ray generation for most specimens is somewhat higher just below the surface of the specimen and decreases to zero when the electron energy E falls below the critical excitation energy E_c of the characteristic x-ray of interest.

As illustrated from the Monte Carlo simulations, the atomic number of the specimen strongly affects the distribution of x-rays generated in specimens. These effects are even more complex when considering more interesting multielement samples as well as the generation of L- and M-shell x-ray radiation.

Figure 9.10 clearly shows that x-ray generation varies with depth as well as with specimen atomic number. In practice it is very difficult to measure or calculate an absolute value for the x-ray intensity generated with depth. Therefore, we follow the practice first suggested by Castaing (1951)

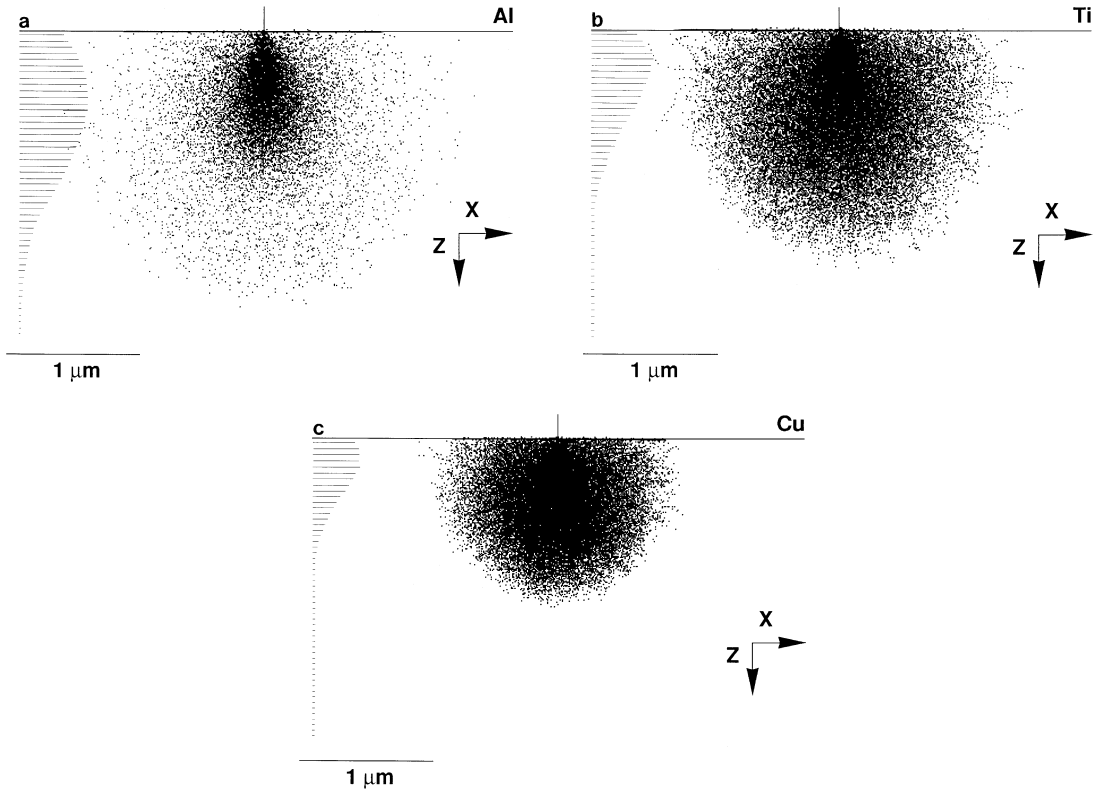


Figure 9.10. Monte Carlo calculations of x-ray generation at 15 keV for (a) Al, (b) Ti, and (c) Cu.

of using a relative or a normalized generated intensity which varies with depth, called $\phi(\rho z)$. The term ρz is called the mass depth and is the product of the density of the sample ρ(g/cm^3) and the linear depth dimension z (cm), usually given in units of g/cm^2. The mass depth ρz is more commonly used than the linear depth term z. The use of the mass depth removes the strong variable of density when comparing specimens of different atomic number. Therefore it is important to recognize the difference between the two terms as the discussion of x-ray generation proceeds.

The general shape of the depth distribution of the generated x-rays, the $\phi(\rho z)$ versus ρz curve, is shown in Fig. 9.11. The amount of x-ray production is related to the amount of elastic scattering, the initial electron beam energy, and the energy of the x-ray line of interest. As the incident beam penetrates the layers of material in depth, the length of the trajectory in each successive layer increases because (1) elastic scattering causes the electron to deviate out of the normal straight line path, requiring a longer path to cross the layer, and (2) backscattering results in electrons that were scattered deeper in the specimen crossing the layer in the opposite direction. Due to these factors, x-ray production increases with depth from the surface, $\rho z = 0$, and goes through a peak ϕ_m at a certain depth ρR_m (see Fig. 9.11). Another consequence of backscattering is that surface layer production ϕ_0 is larger than 1.0 in solid samples because the backscattered electrons excite x-rays as they leave the sample. After the depth ρR_m, x-ray production begins to decrease with depth because the backscattering of the beam electrons reduces the number of electrons available at increasing depth ρz and the remaining electrons lose energy and therefore ionizing power as they scatter at increasing depths. Finally x-ray production goes to zero at $\rho z = \rho R_x$, where the energy of the beam electrons no longer exceeds E_c.

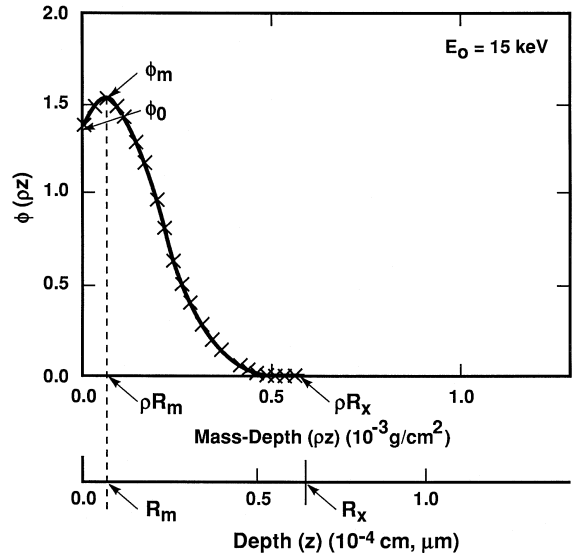


Figure 9.11. Schematic for the measurement of a $\phi(\rho z)$ curve.

Now that we have discussed and described the depth distribution of the production of x-rays using the $\phi(\rho z)$ versus ρz curves, it is important to understand how these curves differ with the type of specimen that is analyzed and the operating conditions of the instrument. The specimen and operating conditions that are most important in this regard are the average atomic number Z of the specimen and the initial electron beam energy E_0 chosen for the analysis. Calculations of $\phi(\rho z)$ versus ρz curves have been made for this chapter using the PROZA program (Bastin and Heijligers, 1990). In Fig. 9.12, the $\phi(\rho z)$ curves for the $K\alpha$ lines of pure Al, Ti, and Cu specimens at 15 keV are displayed. The shapes of the $\phi(\rho z)$ curves are quite different. The ϕ_0 values increase from Al to Ti due to increased backscattering, which produces additional x-ray radiation. The ϕ_0 for Cu is smaller than the value for Ti because the overvoltage U for the Cu $K\alpha$ x-ray line at an E_0 of 15 keV is low (1.67) and the energy of many of the backscattered electrons is not sufficient to excite Cu $K\alpha$ x-rays near the surface. The values of ρR_m and ρR_x decrease with increasing Z and a smaller x-ray excitation volume is produced. This decrease would be much more evident if we plotted $\phi(\rho z)$ versus z, the depth of x-ray excitation, because the use of mass depth includes the density.

Figure 9.13 shows calculated $\phi(\rho z)$ versus ρz curves, using the PROZA program (Bastin and Heijligers, 1990, 1991a, b) at an initial beam energy of 15 keV for Al $K\alpha$ and Cu $K\alpha$ radiation for the pure elements Al and Cu. These curves are compared in Fig. 9.13 with calculated $\phi(\rho z)$ versus ρz curves at 15 keV for Al $K\alpha$ and Cu $K\alpha$ in a binary sample containing Al with 3 wt% Cu. The ϕ_0 value of the Cu $K\alpha$ curve in the alloy is smaller than that of pure Cu because the average atomic number of the Al–3 wt% Cu sample is so much lower, almost the same as pure Al. In this case, less backscattering of the primary high-energy electron beam occurs and fewer Cu $K\alpha$ x-rays are generated. On the other hand the Al $K\alpha$ $\phi(\rho z)$ curves for the alloy and the pure element are essentially the same because the average atomic number of the specimen is so close to that of pure Al. Although the

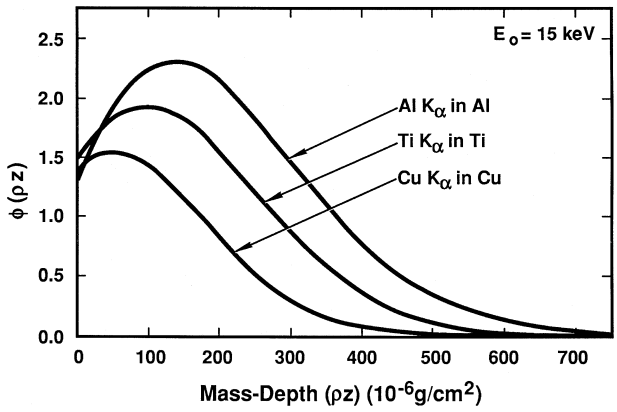


Figure 9.12. Calculated $\phi(\rho z)$ curves for Al $K\alpha$ in Al, Ti $K\alpha$ in Ti, and Cu $K\alpha$ in Cu at 15 keV.

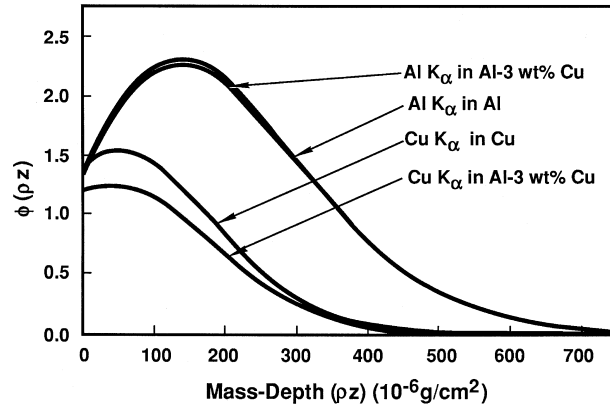


Figure 9.13. Curves of $\phi(\rho z)$ versus (ρz) for Al $K\alpha$ and Cu $K\alpha$ radiation in Al, Cu, and Al–3 wt% Cu at 15 keV, calculated using the PROZA program.

variation of $\phi(\rho z)$ curves with atomic number and initial operating energy is complex, a knowledge of the pertinent x-ray generation curves is critical to understanding what is happening in the specimen and the standard for the element of interest.

The generated x-ray intensity for each element x-ray line i in the specimen $I_{i,\text{gen}}$ can be obtained by taking the area under the $\phi(\rho z)$ versus ρz curve, that is, by summing the values of $\phi(\rho z)$ for all the layers $\Delta(\rho z)$ in mass thickness within the specimen for the x-ray line i of interest. We will call this area $\phi(\rho z)_{i,\text{gen}}$ Area. Table 9.1 lists the calculated values, using the PROZA program, of the $\phi(\rho z)_{\text{gen}}$ Area for the 15-keV $\phi(\rho z)$ curves shown in Fig. 9.13 (Cu $K\alpha$ and Al $K\alpha$ in the pure elements and Cu Kα and Al $K\alpha$ in an alloy of Al–3 wt% Cu) and the corresponding values of ϕ_0. A comparison of the $\phi(\rho z)_{\text{gen}}$ Area values for Al $K\alpha$ in Al and in the Al–3 wt% Cu alloy shows very similar values, whereas a comparison of the $\phi(\rho z)$ Area values for Cu $K\alpha$ in pure Cu and in the Al–3 wt% Cu alloy shows that about 17% fewer Cu $K\alpha$ x-rays are generated in the alloy. The latter variation is due to the different atomic numbers of the pure Cu and the Al–3 wt% Cu alloy specimen. The different atomic number matrices cause a change in ϕ_0 (see Table 9.1) and the height of the $\phi(\rho z)$ curves.

The atomic number correction Z_i for element i can be calculated by taking the ratio of $\phi(\rho z)_{\text{gen}}$ Area for the standard to $\phi(\rho z)_{\text{gen}}$ Area for element i in the specimen. Pure Cu and pure Al are the standards for Cu $K\alpha$ and Al $K\alpha$, respectively. The values of the calculated ratios of generated x-ray

Table 9.1. Generated X-Ray Intensities in Cu, Al, and Al–3 wt% Cu Alloy[a]

Sample	X-ray line	$\phi(\rho z)_{\text{gen}}$ Area (g/cm^2)	Z_i	ϕ_0
Cu	Cu $K\alpha$	3.34×10^{-4}	1.0	1.39
Al	Al $K\alpha$	7.85×10^{-4}	1.0	1.33
Al–3 wt% Cu	Cu $K\alpha$	2.76×10^{-4}	1.21	1.20
Al–3 wt% Cu	Al $K\alpha$	7.89×10^{-4}	0.995	1.34

[a] $E_0 = 15$ keV.

intensities for pure element standard to specimen (atomic number effect, Z_{Al}, Z_{Cu}) are also given in Table 9.1. As discussed in Section 9.6.1, it is expected that the atomic number correction for a heavy element (Cu) in a light element matrix (Al–3 wt% Cu) is greater than 1.0 and the atomic number correction for a light element (Al) in a heavy element matrix (Al–3 wt% Cu) is less than 1.0. The calculated data in Table 9.1 also show this relationship.

In summary, the atomic number correction, Z_i for element i is equal to the ratio of Z_i in the standard to Z_i in the specimen. Using appropriate $\phi(\rho z)$ curves, one can calculate Z_i by taking the ratio of I_{gen} for the standard to I_{gen} for the specimen for each element i in the sample. It is important to note that the $\phi(\rho z)$ curves for multielement samples and elemental standards that can be used for the calculation of the atomic number effect inherently contain the R and S factors discussed previously.

9.6.2. X-Ray Absorption Effect, A

Figure 9.14 illustrates the effect of varying the initial electron beam energy using Monte Carlo simulations of the positions where K-shell x-ray interactions occur for Cu at three initial electron energies, 10, 20, and 30 keV. This figure shows that the Cu characteristic x-rays are generated deeper in the specimen and the x-ray generation volume is larger as E_0 increases. From these plots, we can see that the sites of inner shell ionizations

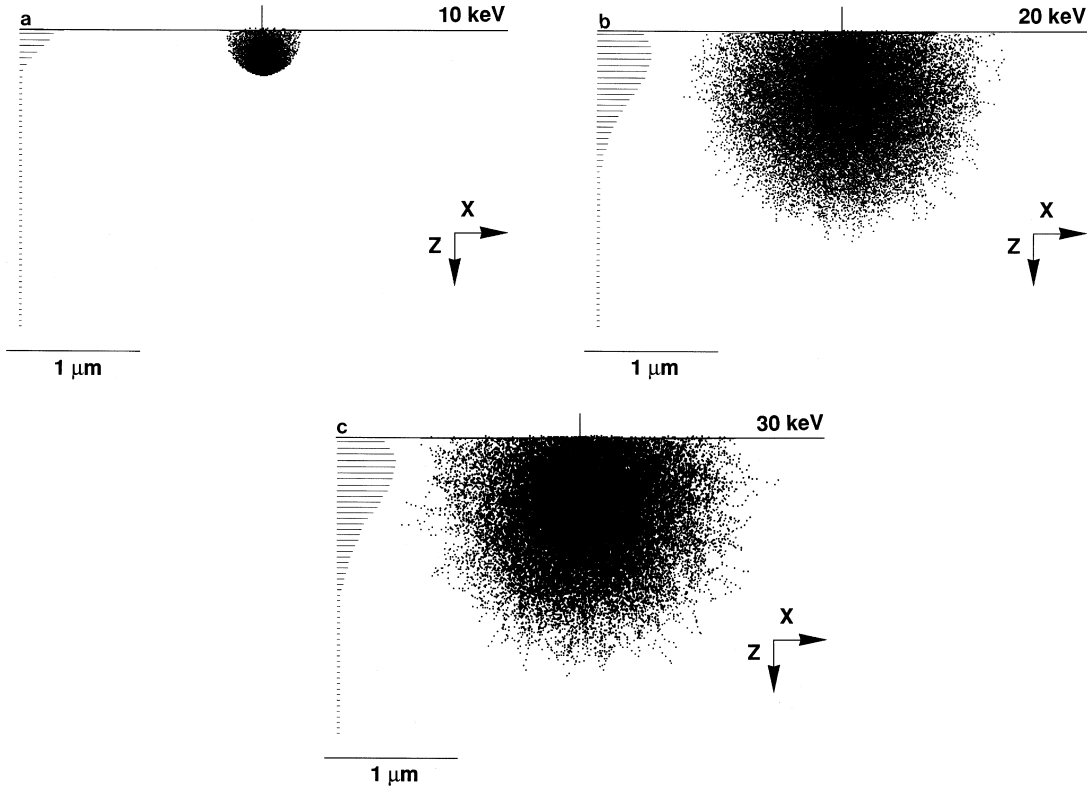


Figure 9.14. Monte Carlo calculations of x-ray generation volume for Cu at (a) 10, (b) 20, and (c) 30 keV.

which give rise to characteristic x-rays are created over a range of depth below the surface of the specimen.

To reach the detector, the x-rays will have to pass through a certain amount of matter, and, as explained in Chapter 6, the photoelectric absorption process will decrease the intensity. It is important to realize that the x-ray photons are either absorbed or else they pass through the specimen with their original energy unchanged, so that they are still characteristic of the atoms that emitted the x-rays. Absorption follows an exponential law, and so, as x-rays are generated deeper in the specimen, a progressively greater fraction is lost to absorption.

From the Monte Carlo plots of Fig. 9.14, one recognizes that the depth distribution of ionization is a complicated function. To quantitatively calculate the effect of x-ray absorption, an accurate description of the x-ray distribution in depth is needed. Fortunately, the complex three-dimensional distribution can be reduced to a one-dimensional problem for the calculation of absorption because the path out of the specimen toward the x-ray detector only depends on depth. The $\phi(\rho z)$ curves, as discussed previously, give the generated x-ray distribution of x-rays in depth (see Figs. 9.11–9.13). Figure 9.15 shows calculated $\phi(\rho z)$ curves for Cu $K\alpha$ x-rays in pure Cu for initial beam energies of 10, 15, and 30 keV. The curves extend deeper (in mass depth or depth) in the sample with increasing E_0. The ϕ_0 values also increase with increasing initial electron beam energies because the increase in the initial beam energy increases U and the energy of the backscattered electrons increases with higher values of E_0.

The x-rays that escape from any depth can be found by placing the appropriate path length in the x-ray absorption equation for the ratio of the measured x-ray intensity I to the generated x-ray intensity at some position in the sample I_0,

$$I/I_0 = \exp[-(\mu/\rho)(\rho t)]. \tag{9.5}$$

The terms in the absorption equation are μ/ρ, the mass absorption coefficient: ρ, the specimen density; and t, the path length (PL) that the x-ray

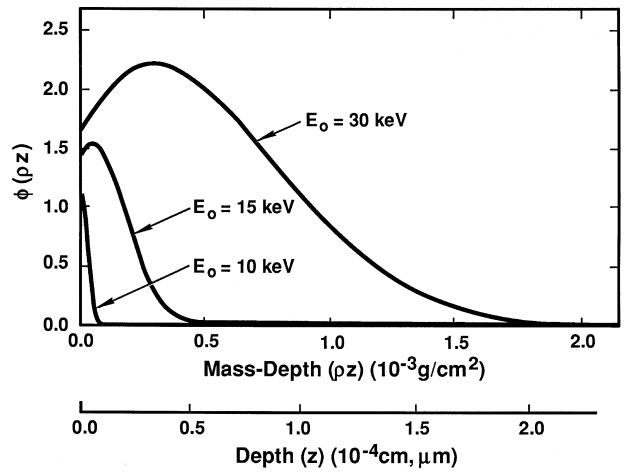


Figure 9.15. Calculated $\phi(\rho z)$ curves for Cu $K\alpha$ in Cu at 10, 15, and 30 keV.

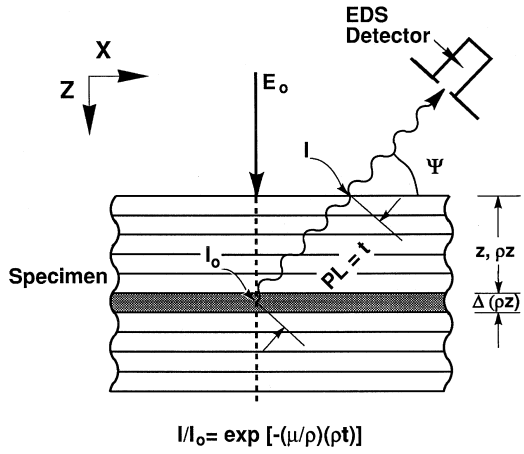


Figure 9.16. Schematic diagram of absorption in the measurement or calculation of $\phi(\rho z)$ emitted. PL = path length, ψ = take-off angle.

traverses within the specimen before it reaches the surface, $z = \rho z = 0$. For the purpose of our interests, I represents the x-ray intensity that leaves the surface of the sample and I_0 represents the x-ray intensity generated at some position within the x-ray generation volume. Because the x-ray spectrometer is usually placed at an acute angle from the specimen surface, the so-called take-off angle ψ, the PL from a given depth z is given by PL = z csc ψ as shown in Fig. 9.16. When this correction for absorption is applied to each of the many layers $\Delta(\rho z)$ in the $\phi(\rho z)$ curve, a new curve results, which gives the depth distribution of emitted x-rays. An example of the generated and emitted depth distribution curves for Al $K\alpha$ at an initial electron beam energy of 15 keV (calculated using the PROZA program; Bastin and Heijligers, 1990, 1991a, b) is shown in Fig. 9.17 for a trace amount (0.1 wt%) of Al in a pure copper matrix. The area under the $\phi(\rho z)$ curve represents the x-ray intensity. The difference in the integrated area between the generated and emitted $\phi(\rho z)$ curves represents the total

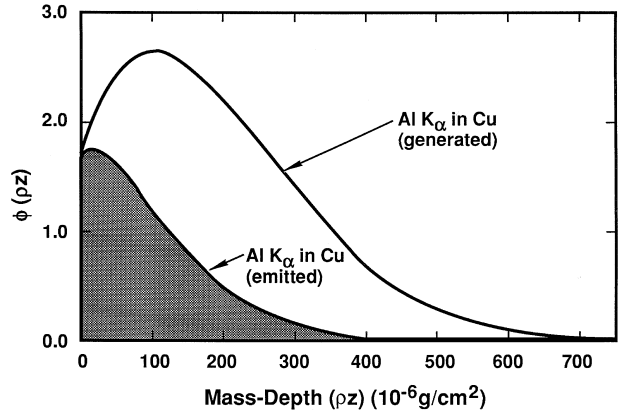


Figure 9.17. Calculated generated and emitted $\phi(\rho z)$ curves for Al $K\alpha$ in a Cu matrix at 20 keV.

x-ray loss due to absorption. The absorption correction factor in quantitative matrix corrections is calculated on the basis of the $\phi(\rho z)$ distribution. Figure 9.8, for example, illustrates the large amount of Ni $K\alpha$ absorbed in an Fe–Ni sample.

X-ray absorption is usually the largest correction factor that must be considered in the measurement of composition by x-ray microanalysis. For a given x-ray path length, the mass absorption coefficient μ/ρ for each measured x-ray line controls the amount of absorption. The μ/ρ varies greatly from one x-ray line to another and is dependent on the matrix elements of the specimen (see Chapter 6). For example, the mass absorption coefficient for Fe $K\alpha$ radiation in Ni is 90.0 cm^2/g, whereas the mass absorption coefficient for Al $K\alpha$ radiation in Ni is 4837.5 cm^2/g. Using Eq. (9.5) and a nominal path length of 1 μm in a Ni sample containing small amounts of Fe and Al, the ratio of x-rays emitted at the sample surface to the x-rays generated in the sample I/I_0 is 0.923 for Fe $K\alpha$ radiation, but only 0.0135 for Al $K\alpha$ radiation. In this example, Al $K\alpha$ radiation is very heavily absorbed with respect to Fe $K\alpha$ radiation in the Ni sample. Such a large amount of absorption must be taken account of in any quantitative x-ray analysis scheme. Even more serious effects of absorption occur when considering the measurement of the light elements, C, N, O, etc. For example, the mass absorption coefficient for C $K\alpha$ radiation in Ni is 17,270 cm^2/g, so large that in most practical analyses, no C $K\alpha$ radiation can be measured if the absorption path length is 1 μm. Significant amounts of C $K\alpha$ radiation can only be measured in a Ni sample within 0.1 μm of the surface. In such an analysis situation, the initial electron beam energy should be held below 10 keV so that the C $K\alpha$ x-ray source is produced close to the sample surface.

As shown in Fig. 9.15, x-rays are generated up to several micrometers into the specimen. Therefore the x-ray path length (PL $= t$) and the relative amount of x-rays available to the x-ray detection system after absorption (I/I_0) vary with the depth at which each x-ray is generated in the specimen. In addition to the position, ρz or z, at which a given x-ray is generated within the specimen, the relation of that depth to the x-ray detector is also important because a combination of both factors determines the x-ray path length for absorption. Figure 9.18 shows the geometric relationship between the position at which an x-ray is generated and the position of the collimator that allows x-rays into the EDS detector. If the specimen is normal to the electron beam (Fig. 9.18), the angle between the specimen surface and the direction of the x-rays into the detector is ψ. The angle ψ is called the take-off angle. The path length $t =$ PL over which x-rays can be absorbed in the sample is calculated by multiplying the depth in the specimen z where the x-ray is generated by the cosecant (the reciprocal of the sine) of the take-off angle ψ. A larger take-off angle will yield a shorter path length in the specimen and will minimize absorption. The path length can be further minimized by decreasing the depth of x-ray generation R_x, that is, by using the minimum electron beam energy E_0 consistent with the excitation of the x-ray lines used for analysis. Table 9.2 shows the variation of the path length that can occur if one varies the initial electron beam energy for Al Kα x-rays in Al from 10 to 30 keV and the take-off

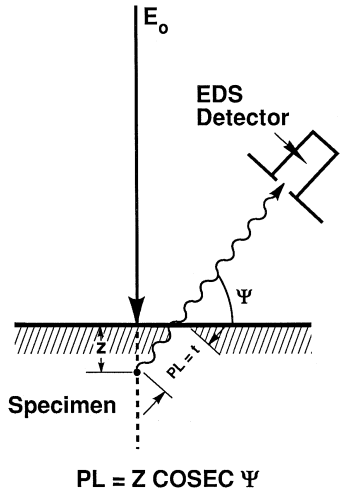


Figure 9.18. Diagram showing the absorption path length in a solid, flat-polished sample. PL = path length, ψ = take-off angle.

angle ψ from 15 to 60 deg. The variation in PL is larger than a factor of 20, from 0.35 μm at the lowest keV and highest take-off angle to 7.7 μm at the highest keV and lowest take-off angle. Clearly the initial electron beam energy and the x-ray take-off angle have a major effect on the path length and therefore the amount of absorption that occurs.

In summary, using appropriate formulations for x-ray generation with depth or $\phi(\rho z)$ curves, one can obtain the effect of absorption by considering absorption of x-rays from element i as they leave the sample. The absorption correction A_i can be calculated by taking the ratio of the effect of absorption for the standard to x-ray absorption for the specimen for each element i in the sample. The effect of absorption can be minimized by decreasing the path length of the x-rays in the specimen.

9.6.3. X-Ray Fluorescence, F

Photoelectric absorption results in the ionization of inner atomic shells, and those ionizations can also cause the emission of characteristic x-rays. For fluorescence to occur, an atom species must be present in the target that has a critical excitation energy less than the energy of the characteristic x-rays being absorbed. In such a case, the measured x-ray intensity from this second element will include both the direct electron-excited intensity

Table 9.2. Path Length PL for Al $K\alpha$ X-Rays in Al

E_0	Take-off angle ψ (deg)	R_x (μm)	PL (μm)
10	15	0.3	1.16
10	60	0.3	0.35
30	15	2.0	7.7
30	60	2.0	2.3

as well as the additional intensity generated by the fluorescence effect. Generally, the fluorescence effect can be ignored unless the photon energy is less than 5 keV greater than the edge energy.

The significance of the fluorescence correction F_i can be illustrated by considering the binary system Fe–Ni. In this system, the Ni $K\alpha$ characteristic energy at 7.478 keV is greater than the energy for excitation of Fe K radiation, $E_c = 7.11$ keV. Therefore an additional amount of Fe $K\alpha$ radiation is produced beyond that due to the direct beam on Fe. Figure 9.8 shows the effect of fluorescence in the Fe–Ni system at an initial electron beam energy of 30 keV and a take-off angle ψ of 52.5°. Under these conditions, the atomic number effect Z_{Fe} and the absorption effect A_{Fe} for Fe $K\alpha$ are very close to 1.0. The measured k_{Fe} ratio lies well above the first-approximation straight-line relationship. The additional intensity is given by the effect of fluorescence. As an example, for a 10 wt% Fe–90 wt% Ni alloy, the amount of iron fluorescence is about 25%.

The quantitative calculation of the fluorescence effect requires a knowledge of the depth distribution over which the characteristic x-rays are absorbed. The $\phi(\rho z)$ curve of electron-generated x-rays is the starting point for the fluorescence calculation, and a new $\phi(\rho z)$ curve for x-ray generated x-rays is determined. The electron-generated x-rays are emitted isotropically. From the x-ray intensity generated in each of the layers $\Delta(\rho z)$ of the $\phi(\rho z)$ distribution, the calculation next considers the propagation of that radiation over a spherical volume centered on the depth ρz of that layer, calculating the absorption based on the radial distance from the starting layer and determining the contributions of absorption to each layer $\Delta(\rho z)$ in the x-ray-induced $\phi(\rho z)$ distribution. Because of the longer range of x-rays than electrons in materials, the x-ray-induced $\phi(\rho z)$ distribution covers a much greater depth, generally an order of magnitude or more than the electron-induced $\phi(\rho z)$ distribution. Once the x-ray-induced $\phi(\rho z)$ generated distribution is determined, the absorption of the outgoing x-ray-induced fluorescence x-rays must be calculated with the absorption path length calculated as above.

The fluorescence factor F_i is usually the least important factor in the calculation of composition by evaluating the [ZAF] term in Eq. (9.4). In most analytical cases secondary fluorescence may not occur or the concentration of the element that causes fluorescence may be small. Of the three effects Z, A, and F that control x-ray microanalysis calculations, the fluorescence effect F_i can be calculated (Reed, 1965) with sufficient accuracy so that it rarely limits the development of an accurate analysis.

9.7. Calculation of ZAF Factors

It is very difficult to present a thorough discussion of the various methods for calculating the ZAF correction factors without going into a great deal of detail. The calculation schemes have been developed over a period of some 50 years and in many cases involve some very rigorous physics and mathematics. In this section, we discuss the methodology for

formulating the *ZAF* correction factors and discuss some of the key advances that have been made.

9.7.1. Atomic Number Effect, Z

There are two approaches to obtaining the atomic number effect, Z. One approach involves the calculation of the backscattering (R) and stopping power (S) factors; see Section 9.6.1.1 for a discussion of these factors. The other approach is to determine the x-ray generation in depth as a function of atomic number and electron beam energy; see Section 9.6.1.2. We will briefly describe both approaches to the calculation of the atomic number effect.

The most popular calculation method for the atomic number correction is that given by Duncumb and Reed (1968). In this approach R and S for each element i are calculated separately for sample and standard. The scheme for obtaining R involves the calculation of the fraction of ionization remaining in a target of a certain atomic number after the loss due to the backscattering of beam electrons. The value of R also varies with the initial electron beam energy and the energy necessary to excite the element of interest i. The scheme for obtaining S involves calculating the stopping power S. The major problem in evaluating S is determining the mean ionization potential J and the average energy along the path length of the electron E_{m}. A detailed discussion of the atomic number matrix correction Z is given in the Enhancements on the CD, Section E9.3.

The second approach to calculating the atomic number correction is to determine x-ray generation in depth, $\phi(\rho z)$, as a function of atomic number and electron beam energy. Fortunately, many $\phi(\rho z)$ curves have been measured experimentally. With available $\phi(\rho z)$ curves, numerous researchers attempted to accurately model these curves for element-i x-rays as a function of matrix composition and beam energy. Packwood and Brown (1981) developed one of the most useful formulations of the $\phi(\rho z)$ curves. The most popular calculation methods for the $\phi(\rho z)$ approach to the atomic number correction are PROZA, formulated by Bastin *et al.* (1986), using the Packwood–Brown $\phi(\rho z)$ approach, and PAP, due to Pouchou and Pichoir (1984). A detailed discussion of the $\phi(\rho z)$ curves and the atomic number correction is given in the Enhancements, Sections E9.4 and E9.5.

9.7.2. Absorption Correction, A

Two approaches can be taken to calculate the absorption effect, A_i. One approach calls for the direct calculation of x-ray generation and the amount of x-ray absorption in sample and standard for each element i. The most commonly used absorption correction for this approach is that of Philibert–Duncumb–Heinrich (Philibert, 1963; Duncumb and Shields, 1966; Heinrich, 1969). This absorption correction was developed using an empirical expression for $\phi(\rho z)$. It accounts for effects of the mass absorption coefficient μ/ρ for each measured x-ray line, the excitation energy E_{c}, the beam energy E_0, and the take off-angle ψ. A detailed discussion

of the absorption matrix correction A is given in the Enhancements Section E9.6.

The second approach uses the $\phi(\rho z)$ technique. The $\phi(\rho z)$ absorption correction first considers x-ray production by calculating $\phi(\rho z)$ versus (ρz) and then sums the effect of absorption of element i from various depths in the sample to the surface of the sample. The $\phi(\rho z)$ absorption correction takes into account the measured x-ray line, the sample matrix, the excitation energy E_c, the beam energy E_0, and the take-off angle ψ. The most popular calculation methods for the $\phi(\rho z)$ approach to the absorption correction were formulated by Bastin and Heijligers (1984) using the Packwood–Brown $\phi(\rho z)$ approach and by Pouchou and Pichoir (1984). A detailed discussion of the absorption correction obtained by using $\phi(\rho z)$ curves is given in the Enhancements, Section E9.7. Both of the major absorption correction schemes are formulated for both the sample and standard in order to calculate A_i.

9.7.3. Characteristic Fluorescence Correction, F

The most popular calculation method for the fluorescence correction F was formulated by Reed (1965), from the original equations of Castaing (1951). The correction takes into account (1) the generation of element-j x-rays in the sample from the electron beam, (2) the absorption of element-j x-rays as they propagate through the sample and produce fluorescent x-rays of element i, and (3) the absorption of element i as element-i x-rays leave the sample. The additional fraction of x-rays produced by fluorescence of element i by element j represents the fluorescence correction F. A detailed discussion of the fluorescence correction F is given in the Enhancements Section E9.8.

9.7.4. Calculation of ZAF

As discussed in Sections 9.4 and 9.5, the calculation of the composition of a multielement sample using the electron probe microanalyzer involves the measurement of the x-ray intensity ratios of sample to standard of all the elements in the sample. To calculate the composition, the atomic number correction Z, the absorption correction A, and the fluorescence correction F are evaluated. The most popular ZAF correction formulations used for electron probe microanalysis have been summarized in the preceding sections. One approach, the traditional ZAF formulation, involves calculations of Z, A, and F from fundamental equations describing the physical phenomena. The other approach uses the $\phi(\rho z)$ formulation for ZA and includes the fundamental equations describing characteristic fluorescence F. A more detailed discussion of the ZAF formulations is given in the Enhancements, Section E9.9. These correction schemes and others are incorporated in computer programs supplied with EDS units or with the automation packages of the EPMA instruments.

It is of interest to determine the accuracy of the various correction schemes for the calculation of ZAF effects. One way to assess the accuracy of the ZAF technique is to compare *calculated* k' values with *measured*

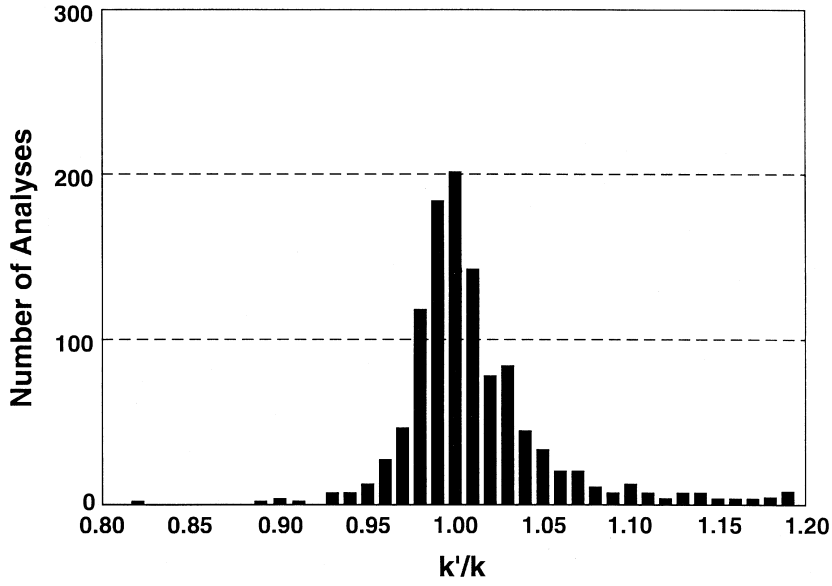


Figure 9.19. Error histogram of *ZAF* analyses versus k'/k (Bastin and Hiejligers, 1997). The Duncumb and Reed atomic number correction, the Philibert–Duncumb–Heinrich absorption correction, and the Reed fluorescence correction were used for calculations of the k ratio. The database contained 1113 analyses of medium to heavy elements ($Z > 11$). The number of analyses is plotted versus the ratio between the calculated (k') and the measured k ratio (k) for the given composition and the energy of the electron beam. Courtesy of G. F. Bastin.

k values from samples of the same composition. Calculated k' ratios can be obtained from a number of calculation schemes for *ZAF*. Measured k ratios can be obtained over a range of electron beam energies for a given sample. The results of calculated versus measured intensity ratios are typically displayed in an "error" histogram by plotting the number of analyses versus the ratio of k'/k. When the ratio k'/k equals 1.0, there is perfect agreement between calculated and measured k ratios for a given sample. An error histogram is developed when the k ratios for a large number of multielement samples of known composition, treated as unknowns and analyzed against pure-element standards, are compared to k' ratios calculated for samples of the same composition.

Figure 9.19 shows an error histogram using the Duncumb and Reed atomic number correction, the Philibert–Duncumb–Heinrich absorption correction, and the Reed fluorescence correction. The data base of 1113 analyses used in the figure include elements of atomic number greater than 12. This database of measured k values was drawn from Bastin *et al.* (1986) and updated by Bastin and Heijligers (1997). The so called *ZAF* histogram of Fig. 9.19 develops a very long tail toward higher k'/k ratios, which is entirely due to the very poor performance on Al and Si measurements in highly absorbing matrices (Ni, Fe) at higher voltages. This is a demonstration of the failure of the absorption correction of the most popular *ZAF* calculation program. Figure 9.20 shows an error histogram using the $\phi(\rho z)$ technique and calculated with the PROZA96 matrix correction program

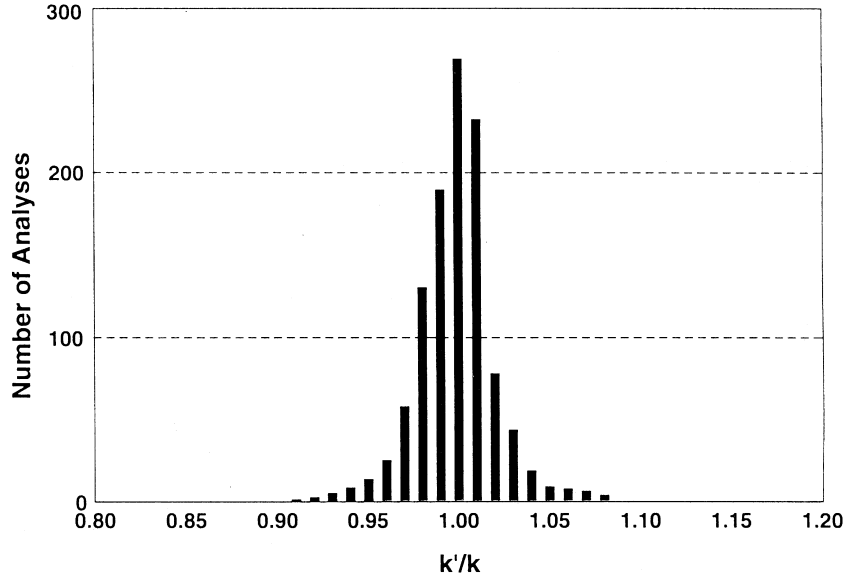


Figure 9.20. Error histogram of $\phi(\rho z)$ analyses versus k'/k (Bastin and Hiejligers, 1997). The database contained 1113 analyses of medium to heavy elements ($Z > 11$). The number of analyses is plotted versus the ratio between the calculated (k') and the measured k ratio (k) for the given composition and the energy of the electron beam. Approximately 97% of the analyses fall within ±5% relative error. Courtesy of G. F. Bastin.

(Bastin *et al.,* 1997). The same database of 1113 analyses was used as in Fig. 9.19. These calculated *ZAF* values offer a very good performance. Approximately 97% of all the data fall within a ±5% relative error.

In the following section, a number of examples illustrate the effect of elements of differing atomic number in a sample and the effect of varying electron beam energy on the absorption effect in particular. The effect of using elemental standards versus standards with compositions close to the sample of interest, the effect of measuring low-energy x-rays, and the effect of fluorescence will also be illustrated. These examples show how important it is to calculate the magnitude of the *ZAF* effects before the analysis process is undertaken. Adjustments in beam energy, analyzed x-ray lines, elemental standards, and, in some cases, take-off angle can be made to optimize the accuracy of the analysis.

9.7.5. The Analytical Total

When all of the elemental constituents have been calculated with equation 9.4, the sum should equal 1.0 or 100% but in actual experience it rarely does. The deviation from unity is a direct consequence of the reality of measurement errors, both random (statistics) and systematic (imperfections in the matrix corrections as well as the intensity measurements). The analytical total actually conveys information. For example, if an elemental constituent is missed during qualitative analysis, this situation will be

manifest as an anomalously low analytical total. If an oxide inclusion in a metallic alloy is measured but no oxygen calculation is made, either directly against an oxygen containing standard such as FeO or indirectly by assumed stoichiometry, then the analytical total will typically fall into the range 0.6 to 0.8, (60% to 80%). This large deviation should alert the analyst to an unexpected situation. With good measurement control and *all* constituents calculated, the analytical total should fall such that $0.98 \leq \text{total} \leq 1.02$ (98% to 102%). Outside this range, the analyst should endeavor to understand why this happened. It is very important that the analyst avoid normalized totals or an element by difference.

9.8. Practical Analysis

9.8.1. Examples of Quantitative Analysis

We now consider the way in which the individual *ZAF* correction procedures are utilized to calculate the compositions of a series of specimens.

9.8.1.1. Al–Cu Alloys

Table 9.3 summarizes the *ZAF* factors that have been calculated using the $\phi(\rho z)$ versus ρz generated and emitted curves as well as the conventional Philibert–Duncumb–Heinrich (PDH) calculation scheme for the Al–3 wt% Cu and the Cu–0.1 wt% Al alloys discussed in Sections 9.6.1 and 9.6.2. For Cu $K\alpha$ and Al $K\alpha$, pure-element standards are used. The Cu $K\alpha$ x-ray has more than sufficient energy to excite Al $K\alpha$ x-rays (more than 5 keV greater than E_c for Al $K\alpha$ radiation). However, the fluorescence excitation process is inefficient. Therefore the fluorescence factors F_{Cu} and F_{Al} are 1.0 and no fluorescence correction is needed.

1) Al–3 wt% Cu Alloy. The analysis of this alloy illustrates the importance of the atomic number correction Z. The atomic number effect Z_i for

Table 9.3. X-Ray Microanalysis of Two Al–Cu Alloys[a]

Sample	X-ray line	Z_i	A_i	F_i	ZAF_i	k_i
Al–3 wt% Cu						
$\phi(\rho z)$ calculation	Cu $K\alpha$	1.21	1.0	1.0	1.210	0.0248
	Al $K\alpha$	0.995	1.039	1.0	1.0345	0.938
PDH calculation	Cu $K\alpha$	1.253	1.0	1.0	1.253	0.0240
	Al $K\alpha$	0.995	1.046	1.0	1.041	0.932
Cu–0.1 wt% Al						
$\phi(\rho z)$ calculation	Cu $K\alpha$	0.999	1.0	1.0	1.00	0.999
	Al $K\alpha$	0.857	2.42	1.0	2.071	0.00048
PDH calculation	Cu $K\alpha$	1.0	1.0	1.0	1.000	0.999
	Al $K\alpha$	0.806	2.58	1.0	2.076	0.00048

[a] $E_0 = 15$ keV, $\psi = 40°$. Pure-element standards.

Al is very small, 0.995, because there is a very small difference in atomic number between specimen and standard (see Fig. 9.13). The atomic number effect Z_i for Cu, however, is significant, 1.21 (Table 9.3). The Cu $K\alpha$ radiation in the Al–3 wt% Cu alloy is decreased primarily due to the smaller number of backscattered electrons available in the Al matrix (see Fig. 9.13).

The absorption correction for Cu $K\alpha$ in the Al–Cu alloy, A_{Cu}, is 1.0 (Table 9.3) at 15 keV and $\psi = 40^\circ$ because the mass absorption coefficients are very small; $(\mu/\rho)_{\mathrm{Cu}}^{\mathrm{Cu}K\alpha} = 53.7\ \mathrm{cm^2/g}$, $(\mu/\rho)_{\mathrm{Al}}^{\mathrm{Cu}K\alpha} = 49.6\ \mathrm{cm^2/g}$, No absorption correction is needed. On the other hand, Al $K\alpha$ in the Al–Cu alloy is highly absorbed by Cu and the mass absorption coefficients for Al $K\alpha$ are much larger than those of Cu $K\alpha$, $(\mu/\rho)_{\mathrm{Al}}^{\mathrm{Al}K\alpha} = 385.7\ \mathrm{cm^2/g}$, $(\mu/\rho)_{\mathrm{Cu}}^{\mathrm{Al}K\alpha} = 5377$. However, the absorption effect A_i for Al $K\alpha$ in the Al–3 wt% Cu alloy is small, 1.039, because only 3 wt% of the highly absorbing element Cu is present in the sample.

Applying Eq. (9.2) in Section 9.5 and using pure-element standards, we find

$$\begin{aligned} C_{\mathrm{Cu}} &= [ZAF]_{\mathrm{Cu}} \cdot k_{\mathrm{Cu}} \\ C_{\mathrm{Al}} &= [ZAF]_{\mathrm{Al}} \cdot k_{\mathrm{Al}}. \end{aligned} \tag{9.6}$$

Using the $\phi(\rho z)$ calculation scheme for the Al–3 wt% Cu specimen, we find that ZAF_{Cu} is 1.21 (Table 9.3). As discussed above, the correction is due primarily to a large atomic number effect. For Al, ZAF_{Al} is 1.0345. As discussed above, the correction is due primarily to the absorption correction. For this example, the measurement of k_{Cu} must be multiplied by 1.21 to obtain the concentration of Cu, C_{Cu}, and the measurement of k_{Al} must be multiplied by 1.0345 to obtain the concentration of Al, C_{Al}. Values of ZAF from the PDH calculation are similar to those of the $\phi(\rho z)$ calculation. However, the difference in composition that is calculated using Eq. (9.4) and the two ZAF procedures gives an error in the resultant alloy composition of about 3% relative in the Cu content (± 0.1 wt%) and about 3% relative in the Al content (± 3 wt%) of the alloy. This variation in concentration exceeds our goal of analytical accuracy at the ± 1–2% level.

2) Cu–0.1 wt% Al Alloy. The analysis of this alloy illustrates the importance of the absorption correction A. In the Cu–0.1 wt% Al alloy, the ZAF factor for Cu $K\alpha$ is 0.999, very close to 1.0 (Table 9.3). This result is not surprising because the sample is 99.9 wt% Cu. On the other hand, the Al $K\alpha$ radiation shows a significant atomic number correction Z_{Al} of 0.857 because the average atomic number of the sample is very close to that of Cu, which is much higher than the atomic number of Al. A very significant Al absorption correction A_{Al} of 2.42 using the $\phi(\rho z)$ calculation scheme (Table 9.3) is calculated. This large absorption correction was discussed in Section 9.6.2 and illustrated in Fig. 9.12. A large amount of generated Al $K\alpha$ radiation is absorbed in the sample before reaching the sample surface. For this alloy, and for analysis conditions of $E_0 = 15$ keV and $\psi = 40^\circ$, measurements of k_{Al} must be multiplied by a ZAF factor of 2.07 to obtain the aluminum concentration C_{Al}. Corrections of 100% are

Table 9.4. X-Ray Microanalysis of a Ni–10 wt% Fe Alloy[a]

	X-ray line	Z_i	A_i	F_i	ZAF_i	k_i
15 keV						
$\phi(\rho z)$ calculation	Fe $K\alpha$	1.011	1.002	0.784	0.795	0.126
	Ni $K\alpha$	1.0	1.004	1.0	1.004	0.897
PDH calculation	Fe $K\alpha$	1.008	1.003	0.784	0.793	0.126
	Ni $K\alpha$	1.0	1.004	1.0	1.005	0.896
30 keV						
$\phi(\rho z)$ calculation	Fe $K\alpha$	1.015	1.008	0.722	0.739	0.135
	Ni $K\alpha$	1.0	1.018	1.0	1.018	0.884
PDH calculation	Fe $K\alpha$	1.013	1.011	0.722	0.739	0.135
	Ni $K\alpha$	1.0	1.022	1.0	1.022	0.881

[a] $\psi = 52.5°$. Pure-element standards.

quite large, but fortunately the ZAF factors calculated by the $\phi(\rho z)$ and the PDH calculation schemes are the same (Table 9.3).

9.8.1.2. Ni–10 wt% Fe Alloy

The analysis of this alloy illustrates the importance of the fluorescence correction F for measuring the Fe content. As discussed in Section 9.6.3, the Ni $K\alpha$ characteristic energy (7.478 keV) is greater than the energy for the excitation of Fe $K\alpha$ radiation ($E_c = 7.11$ keV). Therefore an additional amount of Fe $K\alpha$ radiation is obtained beyond that produced by the electron beam itself. The fluorescence effect is quite large in the Ni–10 wt% Fe alloy considered here. At an initial electron beam energy of 15 keV, F_{Fe} is 0.784 (see Table 9.4). In effect, an additional 27.5% of Fe $K\alpha$ radiation is produced. At an initial electron beam energy of 30 keV, the fluorescence effect is increased; F_{Fe} is 0.722. The effect of fluorescence is the same in both the $\phi(\rho z)$ calculation and the PDH calculation because both schemes use the Reed (1965) correction. As with the absorption correction, the fluorescence correction can be minimized (brought closer to 1.0) as the initial beam energy is decreased. Figure 9.8 illustrates the major effect of Fe $K\alpha$ fluorescence for a range of Fe–Ni alloys.

The atomic numbers of Fe and Ni are very close and therefore the atomic number corrections for Fe $K\alpha$ and Ni $K\alpha$ are essentially zero (Z_{Fe}, $Z_{Ni} \approx 1.0$) as shown in Table 9.4. The absorption corrections for Fe $K\alpha$ and Ni $K\alpha$ are also very small (A_{Fe}, $A_{Ni} \approx 1.0$). The Ni $K\alpha$ is more highly absorbed in this alloy because of the fluorescence effect and is observed by the lower value of A_{Ni} versus A_{Fe} (Table 9.4). Figure 9.8 also illustrates the major effect of Ni $K\alpha$ absorption for a range of Fe–Ni alloys.

9.8.1.3. Ni–38.5 wt% Cr–3.0 wt% Al Alloy

Multielement analyses are of much more interest in x-ray microanalysis. All the matrix elements must be taken into account in the analysis and more often than not there is a range of atomic numbers among the

Table 9.5. X-Ray Microanalysis of a 58.5 wt% Ni–38.5 wt% Cr–3.0 wt% Al Alloy[a]

	X-ray line	Z_i	A_i	F_i	ZAF_i	k_i
15 keV						
$\phi(\rho z)$ calculation	Ni $K\alpha$	1.002	1.02	1.0	1.021	0.573
	Cr $K\alpha$	1.01	1.007	0.952	0.968	0.398
	Al $K\alpha$	0.896	1.976	1.0	1.769	0.017
PDH calculation	Ni $K\alpha$	1.006	1.02	1.0	1.027	0.570
	Cr $K\alpha$	1.006	1.008	0.952	0.966	0.399
	Al $K\alpha$	0.852	2.12	1.0	1.81	0.0166
30 keV						
$\phi(\rho z)$ calculation	Ni $K\alpha$	0.997	1.08	1.0	1.071	0.546
	Cr $K\alpha$	1.014	1.024	0.922	0.957	0.402
	Al $K\alpha$	0.916	3.97	1.0	3.631	0.00826
PDH calculation	Ni $K\alpha$	1.0	1.09	1.0	1.086	0.538
	Cr $K\alpha$	1.01	1.03	0.922	0.960	0.401
	Al $K\alpha$	0.887	4.42	1.0	3.92	0.00765

[a] $\psi = 40^\circ$. Pure-element standards.

elements present in the sample. The analysis of a 58.5 wt% Ni–38.5 wt% Cr–3.0 wt% Al alloy illustrates the importance of the atomic number and absorption effect for Al and the fluorescence effect for Cr. The importance of operation at low beam energies is also illustrated.

Table 9.5 gives the results of *ZAF* calculations for the Ni–Cr–Al alloy obtained for 15- and 30-keV operation with a take-off angle of 40°. For Ni $K\alpha$, the *ZAF* correction is quite small at 15 keV and about 7–9% relative at 30 keV. For Cr $K\alpha$, there is a significant fluorescence correction of about 5% relative at 15 keV and 8% relative at 30 keV. The large corrections, however, occur for the measurement of Al $K\alpha$. The atomic number correction is about 10% relative because the average atomic number of the alloy is about 25, whereas the atomic number of Al is 13. The absorption correction is very large, about a factor of 2.0 at 15 keV and about a factor of 4 at 30 keV. The significant absorption correction A_{Al} is due to the large mass absorption coefficient for Al in Ni (4545 cm^2/g). The mass absorption coefficient for Ni in Al, for example, is much smaller (61.3 cm^2/g). As shown in Table 9.5, the *ZAF* correction for Al $K\alpha$ is about 1.8 at 15 keV and 3.6–3.9 at 30 keV. There is no question that the use of a lower electron beam energy will decrease the absorption distance of x-rays in the specimen. Lower-keV operation is recommended for analysis when the mass absorption coefficient is large, particularly for x-ray energies less than 1.5 keV.

It is useful to compare that *ZAF* corrections calculated with the $\phi(\rho z)$ and the PDH calculation schemes for this Ni–Cr–Al alloy. Differences of less than 1% relative for *ZAF* are calculated for Ni $K\alpha$ and Cr $K\alpha$ from the two calculation schemes. Absorption and atomic number corrections for Ni $K\alpha$ and Cr $K\alpha$ are small. However, for Al $K\alpha$ differences are as large as 10% relative, particularly at 30 keV, where the *ZAF* correction is as large as 3.92 (Table 9.5). In summary, for multielement analyses it is best to operate under conditions where the *ZAF* corrections are as close to 1.0 as possible.

Table 9.6. Microanalysis Calculation for Pyroxene for $E_0 = 15$ keV and $\psi = 52.5^{\circ}$[a]

Element	Wt%	Oxide	Wt%	Z_i	A_i	F_i	ZAF_i	k_i
	(a) Pure-element standards: calculation of element concentration							
Si	24.9	SiO_2	53.3	1.023	1.215	0.999	1.241	0.201
Al	0.59	Al_2O_3	1.11	1.047	1.360	0.992	1.412	0.0042
Cr	0.42	Cr_2O_3	0.62	1.201	1.016	0.985	1.201	0.0035
Fe	7.38	FeO	9.5	1.212	1.006	1.00	1.220	0.0605
Mg	8.5	MgO	14.1	1.010	1.475	0.996	1.482	0.0574
Ca	15.2	CaO	21.2	1.077	1.032	0.997	1.108	0.137
O[b]	43.0			1.118	1.804	1.00	2.02	0.2129
	(b) Pure oxide standards: calculation of oxide concentration							
Si	24.9	SiO_2	53.3	0.973	1.106	0.999	1.075	0.496
Al	0.59	Al_2O_3	1.11	0.989	1.180	0.992	1.158	0.00959
Cr	0.42	Cr_2O_3	0.62	1.101	1.019	0.985	1.105	0.00561
Fe	7.38	FeO	9.5	1.136	1.008	1.00	1.145	0.0830
Mg	8.5	MgO	14.1	0.977	1.208	0.996	1.175	0.120
Ca	15.2	CaO	21.2	1.034	1.033	0.997	1.065	0.199
O[c]	43.0							

[a] Calculations made using the PROZA x-ray quantitation program (Bastin and Heijligers, 1990, 1991a, b).

[b] FeO used as the O $K\alpha$ standard.

[c] Measured by stoichiometry.

9.8.1.4. Pyroxene: 53.5 wt% SiO_2, 1.11 wt% Al_2O_3, 0.62 wt% Cr_2O_3, 9.5 wt% FeO, 14.1 wt% MgO, and 21.2 wt% CaO

The analysis of a complex oxide sample illustrates the importance of matrix effects. In this example different standards can be employed as well as different operating conditions. In this silicate pyroxene, $K\alpha$ lines are used for all elements. One has a choice of standards, either pure elements or pure oxide standards. For Si, Al, Cr, Fe, Mg, and Ca one can use pure elements plus a standard such as FeO for pure oxygen. The other choice is to use oxides such as SiO_2, Al_2O_3, Cr_2O_3, FeO, MgO, and CaO. The stoichiometry of the cation–anion ratio of the oxides is used to obtain the oxygen content of the sample.

Tables 9.6 and 9.7 summarize calculated *ZAF* factors for the seven-element pyroxene for a take-off angle ψ of 52.5°. Two values of initial beam energy E_0-15 and 30 keV are chosen in order to compare the values of the *ZAF* terms. Tables 9.6 and 9.7 contain *ZAF* factors calculated when pure elements are used as standards and when simple oxides are used as standards. For oxygen analysis, a FeO standard is used because a pure-element standard is not available. *ZAF* calculations are made using the $\phi(\rho z)$ calculation scheme.

For the pyroxene microanalysis, using elemental standards and an initial electron beam energy of 15 keV (Table 9.6a), the atomic number effect Z_i for all the elements is greater than 1.0. The atomic number correction is needed because the average atomic number of the specimen is less than that of each elemental standard and FeO, the standard for oxygen. The

Table 9.7. Microanalysis Calculation for Pyroxene for $E_0 = 30\,\text{keV}$ and $\psi = 52.5^{\circ a}$

Element	Wt%	Oxide	Wt%	Z_i	A_i	F_i	ZAF_i	k_i
	(a) Pure-element standards: calculation of element concentration							
Si	24.9	SiO_2	53.3	1.019	1.714	0.998	1.74	0.143
Al	0.59	Al_2O_3	1.11	1.045	2.215	0.989	2.29	0.0026
Cr	0.42	Cr_2O_3	0.62	1.167	1.059	0.98	1.21	0.0035
Fe	7.38	FeO	9.5	1.168	1.026	1.00	1.20	0.0616
Mg	8.5	MgO	14.1	1.010	2.585	0.994	2.60	0.0328
Ca	15.2	CaO	21.2	1.058	1.111	0.995	1.17	0.130
O[b]	43.0			1.10	2.448	1.00	2.70	0.1597
	(b) Pure oxide standards: calculation of oxide concentration							
Si	24.9	SiO_2	53.3	0.977	1.304	0.998	1.272	0.421
Al	0.59	Al_2O_3	1.11	0.994	1.488	0.989	1.463	0.00759
Cr	0.42	Cr_2O_3	0.62	1.087	1.070	0.98	1.140	0.0054
Fe	7.38	FeO	9.5	1.111	1.033	1.00	1.148	0.0828
Mg	8.5	MgO	14.1	0.982	1.521	0.994	1.484	0.0950
Ca	15.2	CaO	21.2	1.024	1.114	0.995	1.135	0.187
O[c]	43.0							

[a] Calculations made using the PROZA x-ray quantitation program (Bastin and Heijligers, 1990, 1991a, b).

[b] FeO used as the O $K\alpha$ standard.

[c] Measured by stoichiometry.

absorption effect A_i is high, 20–50%, for Si, Al, and Mg. The absorption effect for oxygen is quite large, more than 80%, primarily due to the high O $K\alpha$ mass absorption coefficients in the metallic elements.

The fluorescence effect F_i is small, very close to 1.0, except for Cr. X-rays of Cr $K\alpha$ are excited by Fe $K\alpha$ radiation and F_{Cr} is 0.985, a 1.5% effect. The atomic number effect Z_{Cr}, on the other hand, is 1.201, a 20% effect. The atomic number effect is much larger and therefore the fluorescence effect for Cr is relatively unimportant.

The *ZAF* correction for this pyroxene varies from a minimum of 1.1 for Ca to 1.5 for Mg, 10–50% corrections for the metallic elements in the pyroxene. The *ZAF* correction for oxygen is very significant, a factor of 2.02, over 100%. Therefore, even a small error in calculating *ZAF* will cause a significant error in the chemical analysis of the mineral.

Geologists usually take advantage of natural and synthetic mineral standards to reduce matrix effects even further. Using oxide standards, the measurement of oxygen is unnecessary. However, one must assume stoichiometry in the oxides to take advantage of these standards. Because the atomic number of oxide standards is much closer to the average atomic number of the specimen, the magnitude of the atomic number effect is closer to 1.0 and a smaller correction is needed. In this example (compare Table 9.6a and 9.6b), Z_{Cr} is reduced from 1.20 to 1.10 and Z_{Al} from 1.047 to 0.989. The amount of the absorption effect is also decreased because the amount of absorption of a characteristic x-ray line in the oxide standard is much closer to the amount of absorption in the specimen. For Si $K\alpha$, Al $K\alpha$, and Mg $K\alpha$, which are most heavily absorbed, the value of the absorption

effect A_i is halved (compare Table 9.6a and 9.6b). The fluorescence effect F_i is unaffected because fluorescence does not occur either in the pure oxide or in the metal standards. In this analysis situation, the ZAF correction is reduced from a minumum of 1.07 for Ca to a maximum correction of ~1.18 for Mg. Small errors in the ZAF calculation will not affect the analysis in a major way and oxygen is calculated by stoichiometry. The analysis of ceramic oxides as well as carbides and nitrides can be accomplished more accurately by using compound standards as long as stoichiometry is preserved. If this is not the case, oxygen, carbon, and/or nitrogen must be measured separately.

One can compare the calculation of pyroxene composition at an initial beam energy of 15 keV with a higher initial beam energy of 30 keV. Because the path length for absorption is longer, there will be a larger absorption correction. In comparing A_i values at 15 and 30 keV (Tables 9.6a and 9.7a) for pure-element standards, one sees that A_i for Si, Al, and Mg increases from a 20–50% correction at 15 keV to a 70–150% correction at 30 keV. The oxygen absorption correction effect increases as well from 80% to 140%. At 30 keV, the atomic number effect Z_i is decreased as much as 4%, not nearly enough to compensate for the increase in the absorption effect. Therefore, as discussed earlier, a lower electron beam energy is preferable for microanalysis measurements.

The use of pure oxide standards at 30 keV (Table 9.7b) can minimize the effect of Z_i and A_i with regard to pure-element standards. However, the absorption effect for Si, Al, and Mg is still twice that obtained at an initial electron beam energy of 15 keV. As in almost all samples, x-ray microanalysis should be performed (1) at as low an initial electron beam energy as possible and (2) by using standards for each element with an atomic number as close to that of the average atomic number of the specimen as possible.

It is of interest to compare the results of the ZAF and the $\phi(\rho z)$ calculations for the pyroxene sample run under similar conditions, 15 and 30 keV and $\psi = 52.5^\circ$, using oxide standards. The calculated ZAF_i values are very similar, within 1% relative, at 15 and 30 keV, except for Mg $K\alpha$. Manganese is the lowest atomic number element analyzed and the atomic number correction calculated by the $\phi(\rho z)$ method is about 5% higher than that calculated by the ZAF method.

9.8.2. Standardless Analysis

In the preceding sections of this chapter we have presented the traditional method of quantitative x-ray microanalysis, which relies on careful measurements under identical conditions of the unknown and standards. The standards required for this procedure can be as simple as pure elements, for example, Fe, Ni, Cr, etc., whereas for those elements that are not stable in a vacuum or under electron bombardment, for example, S, Cl, F, etc., simple binary compounds can be used. The success of this procedure is established by the error histograms shown in Figs. 9.19 and 9.20, which demonstrate that for major elements (mass fraction >0.1), the results

are within ±5% relative for 97% of the analyses (Bastin and Heijligers, 1997).

This traditional method of quantitative x-ray microanalysis was developed to support x-ray intensity measurements made with wavelength-dispersive x-ray spectrometers. By measuring the intensity ratio (*k* ratio) for the same elemental x-ray peak in the unknown and in the standard, the spectrometer efficiency is identical for both the unknown and standard intensities and quantitatively divides out in the *k* ratio. In a careful analysis of sources of error, Heinrich and Yakowitz (1968) showed additional advantages to the *k* ratio in terms of relaxed requirements for accuracy of other physical parameters used in the matrix corrections.

The energy dispersive x-ray spectrometer (EDS) became popular in part because the analyst could always observe the complete x-ray spectrum with every measurement, so that qualitative analysis, at least of major constituents, should always be robust. The traditional *k*-ratio method of quantitative EPMA was demonstrated to be readily adaptable to EDS measurements with an accuracy shown to be similar to that of WDS for major constituents for which peak interferences were not significant. Moreover, the EDS was found to be reasonably stable over long periods of time. With a careful measurement quality assurance plan in operation, quantitative EDS measurements could be performed with stored standard spectra, especially for photon energies above 1 keV. However, EDS measurements are subject to variability in the SEM because of the added degrees of freedom for sample positioning. Because the SEM does not normally include an optical microscope for highly precise positioning relative to the EDS, the sequential measurement of unknowns and standards under identical conditions became more problematic. This situation was further exacerbated in the important class of SEMs based upon cold-field-emission electron guns, where the continual change in the beam current due to gas contamination of the field emission tip following high-voltage conditioning made quantitative analysis subject to error introduced by beam variability.

To respond to these challenges, methods of "standardless" quantitative x-ray microanalysis have been developed, which rely only on the measured EDS spectrum of the unknown. The standard intensity that forms the denominator of the *k* ratio is provided by calculation rather than direct measurement. Moreover, standardless methods are not dependent on the absolute electron dose, so that beam intensity variations can be ignored, a real advantage for the cold FEG SEMs. Only accurate values of the beam energy, x-ray take-off angle, and spectrometer efficiency are required. By eliminating the measurement of standards and the need for quantitative control over sample position and many instrument parameters, this approach clearly simplifies the overall analysis procedure. There is, however, a significant price to be paid for this convenience, and that is a marked loss in accuracy that has been documented with the current generation of standardless analysis methods compared to the traditional method of standards measurement (Newbury *et al.,* 1995; Newbury, 1999).

Standardless methods fall into two broad classes depending on how the required standard intensities are calculated:

1. "True" standardless, in which the appropriate standard intensities are calculated from first principles, considering all aspects of x-ray generation, propagation through the solid target, and detection
2. "Fitted standards standardless," in which the standard intensities are derived from a suite of experimental standards measurements performed on an instrument at the software manufacturer's laboratory; these remotely measured standards intensities are adjusted for the characteristics of the local instrument actually being used to measure the unknowns

With either route to "standard intensities" the resulting standardless "k ratios" (unknown/calculated standard) are subjected to matrix corrections with one of the usual approaches, ZAF, $\phi(\rho z)$, etc. Because all connections to electron dose and spectrometer geometric efficiency (solid angle) are lost, standardless analysis totals are always normalized to unity (100%), with oxygen either directly measured or else included in the analytical total by means of stoichiometry. The analyst must realize that this 100% total implies nothing about accuracy.

9.8.2.1. First-Principles Standardless Analysis

Calculating an equivalent standard intensity I_{std} from first principles requires solution of the equation

$$I_{\text{std}} = \left(\frac{\rho N_0}{A} \omega \int_{E_0}^{E_c} \frac{Q}{dE/ds} \, dE \right) R f(\chi) \varepsilon(E_v). \tag{9.7}$$

The material terms are the density ρ, Avogadro's number N_0, the atomic weight A, and the fluorescence yield ω. The terms in large parentheses represent the excitation function: Q is the ionization cross section, dE/ds is the stopping power, and E is the instantaneous electron energy, which for x-ray generation varies between E_0, the incident beam energy, and E_c, the critical excitation energy. The other terms correct for the loss of x-ray production due to electron backscattering (R), the self-absorption of x-rays propagating through the solid [$f(\chi)$], and the efficiency of the detector [$\varepsilon(E_v)$] as a function of photon energy E_v.

The accuracy with which the intensity can be calculated depends on our knowledge of the critical physical parameters required for Eq. (9.7), especially the ionization cross section, the fluorescence yield, the stopping power, the backscatter loss, the x-ray self absorption within the target, and the detector efficiency. Let us briefly consider each term (Newbury, 1999):

1. Examination of the published K-shell *ionization cross sections* reveals a variation among leading forms that exceeds 25%, especially in the low-overvoltage range, $1 \leq U \leq 3$, which is the main operational range of electron-excited x-ray spectrometry (Powell, 1976), The situation for L-shell and M-shell cross sections is substantially poorer.
2. Published experimental determinations of the K-shell *fluorescence yield* show a variation of more than 25% for many elements. The situation for L- and M-shell transitions is again much more unsatisfactory.

3. The classic Bethe formulation of the *stopping power* becomes inaccurate at low beam energies (<5 keV), and eventually with decreasing energy becomes physically unrealistic with a sign change. The accuracy of the stopping power at low electron energies matters for calculating Eq. (9.7) because the cross section must be integrated down to E_c, which for low-energy x-rays (e.g., C, N, O, F) involves electron energies in this regime. Several authors (e.g., Joy and Luo, 1989) have suggested modifications to the Bethe formulation to correct for the low-beam-energy regime. Unfortunately, the base of experimental measurements necessary to select the best choice is just being developed.

4. The *backscatter loss* correction factor R was initially formulated based upon experimental measurements and has been refined with advanced Monte Carlo simulations that enable the rigorous calculation of R over all scattering angles and energy losses. The R factor is probably known to an accuracy within a few percent across the periodic table and the energy range of interest.

5. The *self-absorption* of x-rays in the hypothetical standard is calculated based upon the distribution in depth of x-ray production, a parameter that has been extensively studied by experimental measurement of layered targets and by Monte Carlo electron trajectory simulation. Fortunately, the absorption correction is generally small for the x-rays of a pure element, so that at least for higher energy characteristic x-rays, for example, photon energies greater than 3 keV, there is little error in this factor. However, the self-absorption increases both as photon energy decreases and as incident electron energy increases, so that the error in calculating the intensity emitted from an element emitting low-energy photons, for example, carbon, can be significant.

6. The EDS *detector efficiency* can present a major source of error in standardless analysis because we are effectively comparing characteristic x-rays of different energies. Detector efficiency is mainly controlled by absorption losses in the window(s) and detector structure as discussed in Chapter 7. The expression for detector efficiency, Eq. (9.7) (see Enhancements, Section E9.10), consists of a multiplicative series of absorption terms for each component: detector window, gold surface electrode, semiconductor (Si or Ge) dead layer (a partially active layer below the electrode and the source of incomplete charge phenomena), and a transmission term for the detector thickness. Additionally, for most practical measurement situations there may be pathological absorption contributions from contamination on the detector crystal, usually arising from ice buildup due to pinhole leaks in the window or support, and from contamination on the detector window usually deposited as hydrocarbons from the microscope environment.

Examples of efficiency curves are shown in Fig. 9.21 for one detector configuration with various vacuum isolation windows. Whereas the efficiency approaches unity for photons in the range 4–10 keV, the efficiency varies strongly for photon energies below 4 keV, and is especially complex in behavior below 1 keV due to absorption edges from window

components. This region is also quite sensitive to detector "aging" effects that arise from the buildup of ice, as illustrated in Fig. 9.22, where the efficiency for photons with energies just above the oxygen K edge varies sharply with aging.

The ranges of the six critical physical parameters discussed above lead to considerable uncertainty, ranging over a factor of ± 2 or more, in predicting standard intensities. Thus, creating a standardless analysis protocol based upon first-principles prediction of standard intensities is likely to be subject to large errors. To assess the magnitude of these errors, a test of the "first-principles" standardless analysis procedure embedded in NIST–NIH Desktop Spectrum Analyzer yielded the error distribution shown in Fig. 9.23 (Newbury *et al.,* 1995). To prepare this error distribution, EDS spectra were recorded from an extensive series of microhomogeneous NIST Standard Reference Materials and Research Materials, binary compounds, minerals, and glasses of known composition. All possible combinations of K-, L-, and M-shell x-rays above 0.9 keV were tested in these analyses. Oxygen was analyzed by the method of assumed stoichiometry, but the oxygen results were not included in the error distribution. Comparing the error distribution of Fig. 9.23 with that of the traditional method

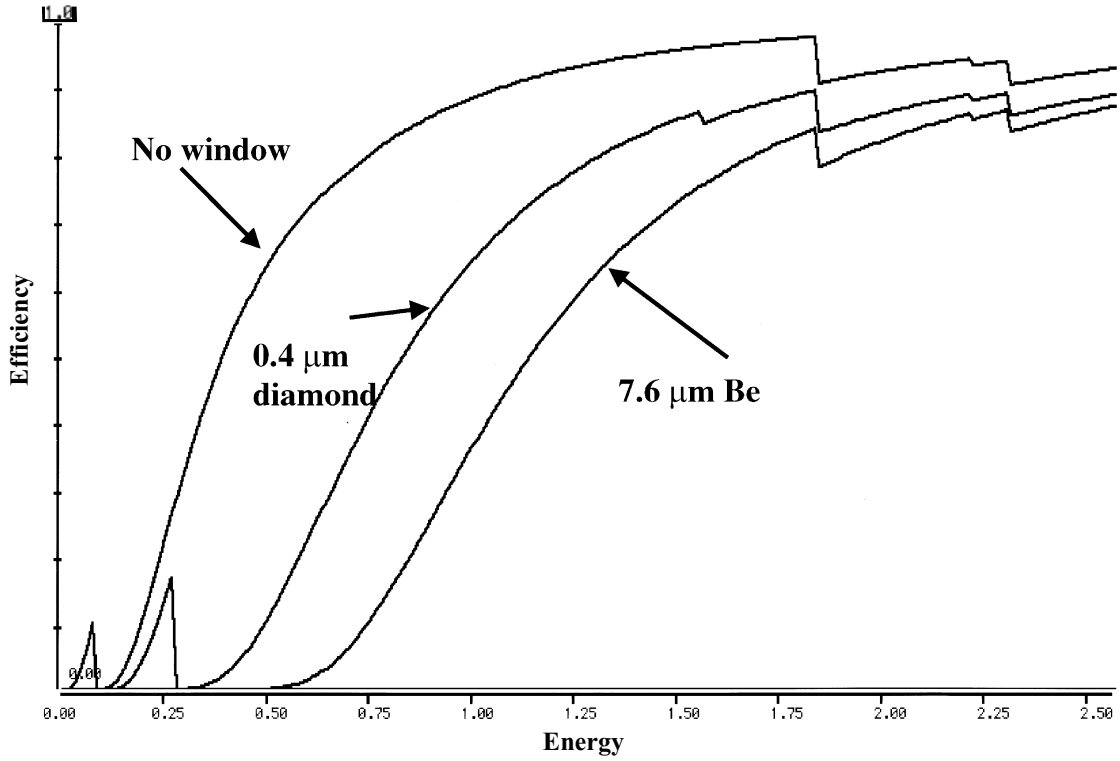


Figure 9.21. EDS detector efficiency plots for a detector structure with a 10-nm Au front surface electrode, a 100-nm silicon "dead" layer, as simulated with the NIST–NIH Desktop Spectrum Analyzer, and either no window, a 0.4-μm diamond window with a 20-nm Al light-reflective coating, or a 7.6-μm beryllium window.

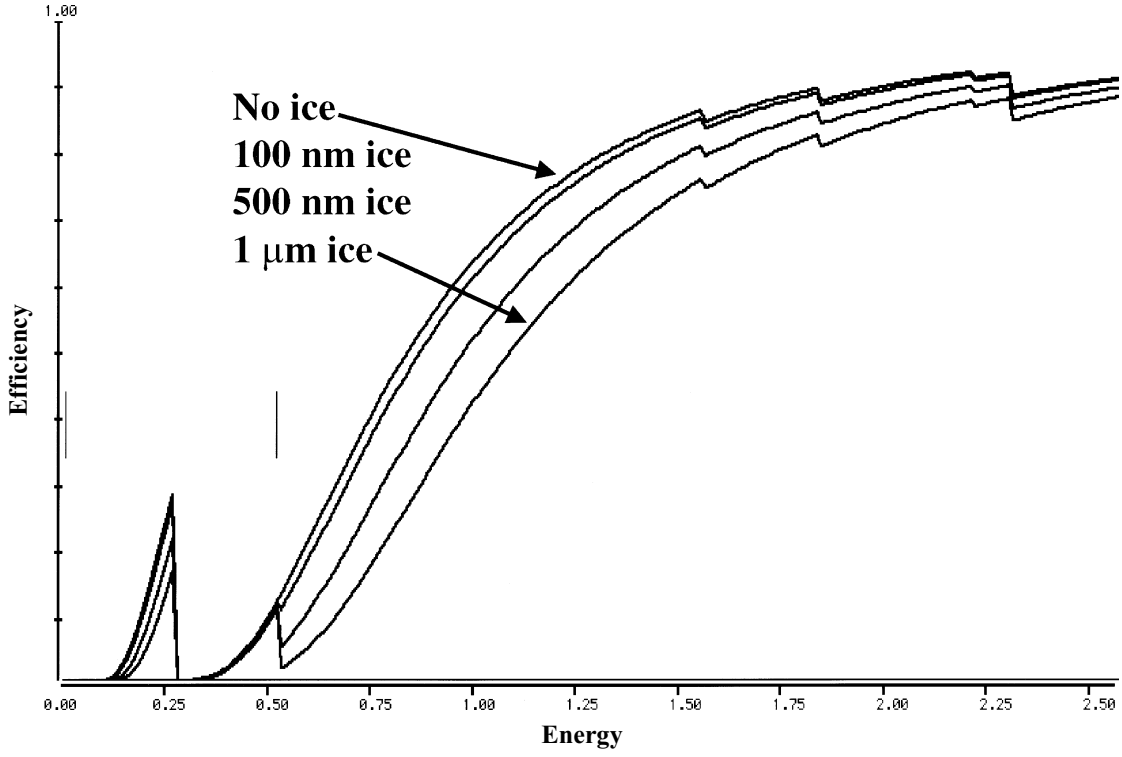


Figure 9.22. EDS detector efficiency plots showing the effect of the buildup of ice for a detector structure with a 0.4-μm diamond window with a 20-nm Al light-reflective coating, a 10-nm Au front surface electrode, and a 100-nm silicon "dead" layer, as simulated with NIST–NIH Desktop Spectrum Analyzer, for no ice, 0.1 μm of ice, 0.5 μm of ice, and 1 μm of ice.

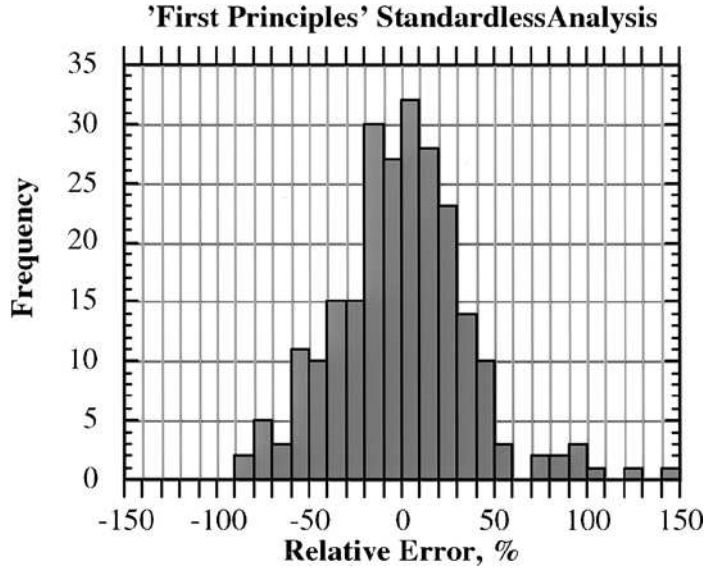


Figure 9.23. Distribution of relative errors observed for broad-based analysis of microhomogeneous materials with the first-principles standardless analysis procedure embedded in the NIST–NIH Desktop Spectrum Analyzer (Newbury, 1999).

(Figs. 9.19 and 9.20) shows the error bins to be an order of magnitude broader, so that 95% of the analyses are found to lie within ±50% relative. When the analytical errors become this large, the compositional values that result from the procedure are often so far from the correct values that it is difficult to make meaningful use of the data for all but the coarsest levels of compositional comparison.

9.8.2.2. "Fitted-Standards" Standardless Analysis

Experimental measurements of suites of pure-element standard intensities, such as those shown for the *K*-shell peaks for Ti to Ge in Fig. 9.24, often show sufficiently regular behavior that mathematical fitting procedures can be developed to describe the relationship of x-ray intensity with atomic number. Such mathematical fits can be used to interpolate for missing elements, for example, Ga in the series shown in Fig. 9.24. Moreover, the intensity of characteristic x-rays for a single element as a function of overvoltage was shown in Chapter 6, Eq. (6.8), to also follow a predictable

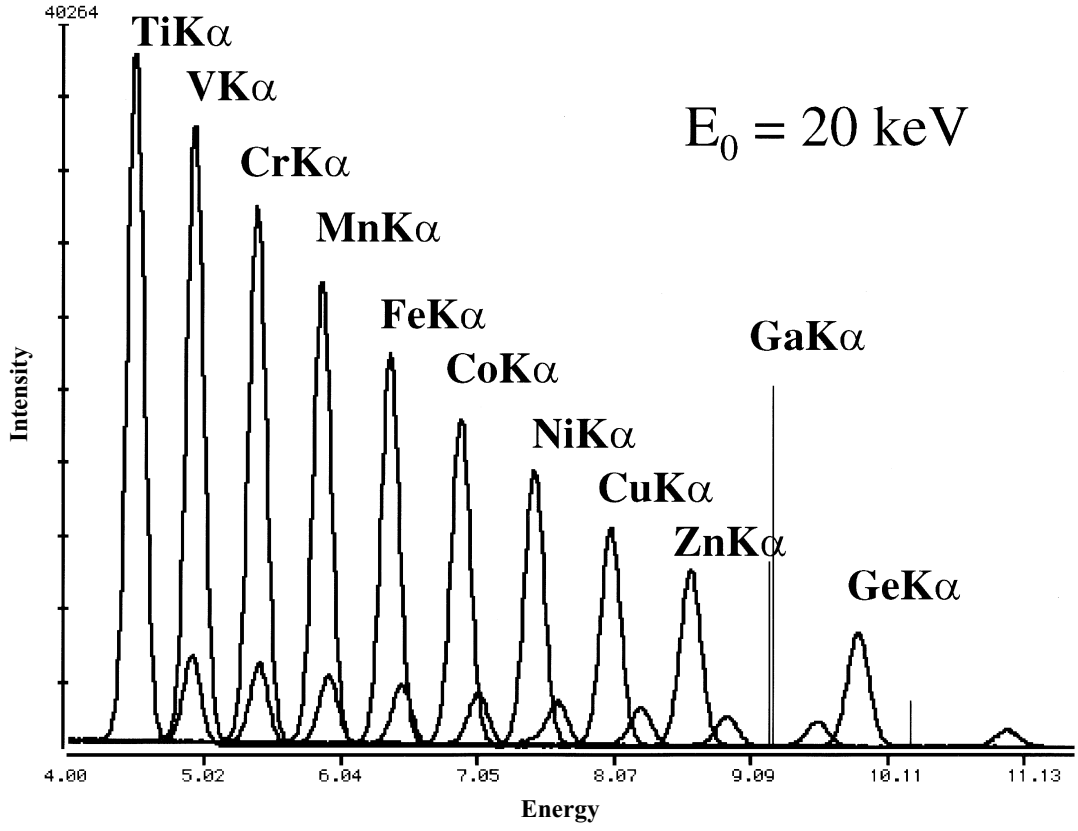


Figure 9.24. Pure-element intensities at 20 keV and constant dose for Ti through Ge, with the exception of Ga. Note the regular behavior.

relationship. Thus, by measuring a suite of standard intensities for the K, L, and M families at one or more beam energies, a database can be developed which can be used with the known dependence on overvoltage to calculate a new suite of intensities appropriate to another beam energy or to interpolate to determine the intensities of missing elements from the experimental suite. In order to relate such measurements to the local SEM/EDS actually involved in measuring the unknown, the detector efficiency must be known for both the original and the local EDS units. The efficiency can be calculated [using Eq. (7.9) in the Enhancements, Section E9.10] if the various parameters are known with sufficient accuracy. Alternatively, spectra of the x-ray bremsstrahlung measured for the same element (e.g., B or C) at the same beam energy on both the remote and local SEM/EDS instruments can be ratioed on a channel-by-channel basis to give a correction factor as a function of photon energy to adjust each standard spectrum to the value appropriate for the local EDS detector efficiency.

Procedures such as that described in the previous paragraph form the basis for the "fitted-standards" approach to standardless analysis. Note that this method is not really standardless because it is ultimately tied to a database of experimentally measured pure element intensities. Moreover, because the procedure is based on actual measurements and only a few physical parameters are needed to adjust the intensities for different conditions, we can expect that the fitted-standards approach should be more accurate than the "first-principles" approach. Indeed, when "fitted-standards" standardless quantitative procedures are tested against the same extensive sample under the same conditions, error histograms of the sort shown in Figs. 9.25 and 9.26 are obtained (Newbury, 1999). The width of the

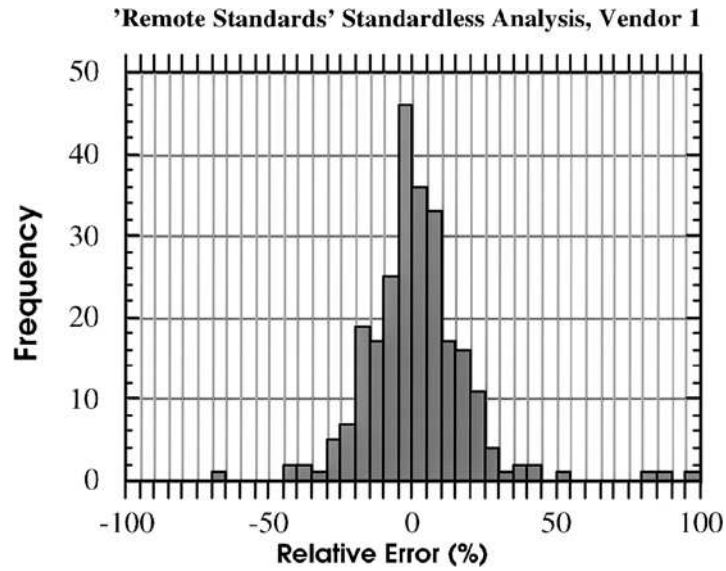


Figure 9.25. Distribution of relative errors observed for broad-based analysis of microhomogeneous materials with a commercial fitted-standards standardless analysis procedure (Newbury, 1999).

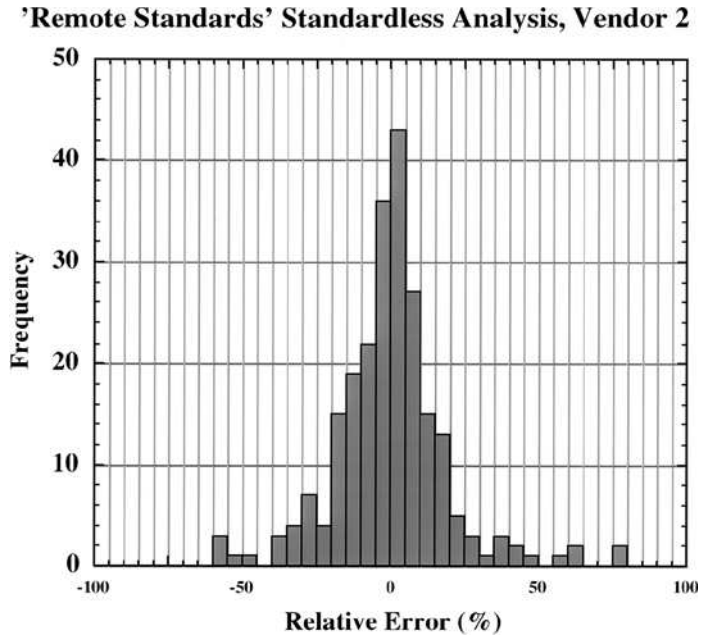


Figure 9.26. Distribution of relative errors observed for broad-based analysis of microhomogeneous materials with a second commercial fitted-standards standardless analysis procedure (Newbury, 1999).

distribution is a factor of two better than the "first-principles" approach. Compared with the error distribution of the traditional method, Figs. 9.19 and 9.20, the error distribution for "fitted-standards" standardless analysis is still a factor of five broader, so that 95% of the results lie within ±25% relative to the correct value. The origin of this degraded performance rests primarily with the *L*- and *M*-shell x-ray intensities, which are not as well behaved as the *K*-shell systematics seen in Fig. 9.24.

In general, "fitted-standards" standardless analysis achieves its best accuracy for situations in which all constituents are measured with *K*-shell x-rays, especially when those characteristic peaks are of similar photon energy, for example, Cr–Fe–Ni in stainless steels. When the photon energy spread increases, for example, Al and Si in a steel, or characteristic peaks from different shells must be used in the same analysis (e.g., $YBa_2Cu_3O_7$), the accuracy tends to degrade. The analytical total, including oxygen by direct or stoichiometric calculation, is always normalized to unity (100%), so that the analyst will not be alerted to missing constituents through an anomalous value of the analytical total.

If the analyst chooses to employ any standardless analysis procedure, he or she should not be misled by the accuracy suggested by numerical values of the concentration that are often reported to four (or more!) apparently significant figures. Note that a high degree of precision, given by the counting statistics and often below 1% in realistic SEM/EDS measurements, does not imply anything about the accuracy of a measurement.

Quantitative analysis software should be accompanied from the supplier by a realistic report of its accuracy performance on known test materials. If not, it is the analyst's responsibility to perform tests on suitable standards containing the elements of interest; for example, if S in Fe is of interest, test the analytical procedure on the mineral pyrite, FeS_2. Otherwise, it is incumbent on the analyst to report a conservative estimate of analytical performance of the level seen in Figs. 9.25 and 9.26.

It is also necessary to recognize the utility and popularity of standardless analysis. There is no question that the accuracy of standardless analysis will improve and may eventually approach the accuracy of the traditional k-ratio approach. Significant research will be necessary to accomplish this goal.

9.8.3. Special Procedures for Geological Analysis

9.8.3.1. Introduction

In the early 1960s, the *ZAF* method was not well developed and analysis errors were often quite significant. Furthermore, computer data reduction was much less generally available. In response to this state of affairs, Ziebold and Ogilvie (1964) developed a hyperbolic or empirical correction method for quantitative analysis of binary metal alloys. They expressed this relationship for element 1 in the sample in the hyperbolic form

$$(1 - k_1)/k_1 = a_{12}(1 - C_1)/C_1 \tag{9.8}$$

or

$$C_1/k_1 = a_{12} + (1 - a_{12})C_1, \tag{9.9}$$

where k_1 is the k ratio, sample to standard, of element 1 [Eq. (9.3)] and C_1 is the concentration of element 1 in the sample. The term a_{12} is the a factor, which is a constant for element 1 in a binary system of elements 1 and 2 for a given value of electron beam energy E_0 and take-off angle ψ. If this relationship is correct, a plot of C_1/k_1 versus C_1 is a straight line with a slope $(1 - a_{12})$. Such a hyperbolic relationship between C_1 and k_1 has been shown to be correct for several alloys (e.g., see Figs. 9.8 and 9.9) and oxide systems. However, it is difficult to obtain standards to determine a factors for a large variety of binary systems.

Nearly 25 years ago, Bence and Albee (1968) proposed the use of empirical a factors to correct electron microprobe analyses of multielement oxide and silicate minerals, similar to procedures proposed by Ziebold and Ogilvie (1964). This paper has been one of the most cited articles in the geological sciences and the Bence–Albee analytical technique has played an important role in x-ray microanalysis in geology. In the intervening years since the original Bence–Albee paper, new empirical a factors have been developed (Bence and Holzworth, 1977; Laguitton *et al.,* 1975) by replacing the constant a factor with a polynomial function of concentration. Other studies have indicated that the original a factors produced some systematic

errors in processing geological data (Armstrong, 1988). Although *ZAF* and $\phi(\rho z)$ procedures have been developed which yield accurate compositional values in oxide and silicate systems (see Sections 9.2 and 9.3), many analysts continue to use the Bence–Albee procedure for geological samples. The following sections outline the method and give an application of the method for quantitative analysis.

9.8.3.2. Formulation of the Bence–Albee Procedure

Ziebold and Ogilvie (1964) also developed an analysis procedure for a ternary system containing elements 1, 2, and 3 with compositions C_1, C_2, and C_3 and intensity ratios k_1, k_2, and k_3, respectively. For element 1

$$(1 - k_1)/k_1 = a_{123}(1 - C_1)/C_1, \tag{9.10}$$

where

$$a_{123} = (a_{12}C_2 + a_{13}C_3)/(C_2 + C_3). \tag{9.11}$$

Alternatively, Eq. (9.10) can be given as

$$C_1/k_1 = a_{123} + (1 - a_{123})C_1. \tag{9.12}$$

Similar relationships can be written for elements 2 and 3. In this procedure, the ternary a factors are determined from each of the individual binary a factors a_{12}, a_{13}, and a_{23}. In such an analysis, a factors must be measured or calculated for three binary systems.

Extending this formulation, Bence and Albee (1968) showed that, in a system of n components, for the nth component

$$C_n/k_n = \beta_n, \tag{9.13}$$

where

$$\beta_n = \frac{k_1a_{n1} + k_2a_{n2} + k_3a_{n3} + \cdots + k_na_{nn}}{k_1 + k_2 + k_3 + \cdots + k_n}. \tag{9.14}$$

In this equation a_{n1} is the a factor for the $n1$ binary, that is, for the determination of element n in a binary consisting of element n and of element 1. Therefore a_{n2} is the a factor for a binary consisting of element n and of element 2, and so on. The value of a_{nn}, which is the a factor for n in n is unity. Similarly, the notation a_{1n} represents the a value for element 1 in a binary of elements 1 and n. The value of k_1 is the relative intensity ratio found for element 1 in the specimen to the standard used for element 1. This notation can be used for all the elements in the specimen and, for example, k_n is the relative intensity ratio for element n to the standard used for element n. It is important to note that the a factor values must be known for the experimental conditions E_0 and ψ used for analysis.

To solve completely a system of n components, one needs $n(n-1)$ values for a. The a value for each of the n pure elements is unity as noted above. Thus $n(n-1)$ standards are required. Bence and Albee (1968) obtained a number of suitable standards of geological interest and determined their respective a values at various initial electron beam energies and at one

take-off angle. The standards used were binary oxides. As an example, a_{NiFe} represents the a factor for Ni in the NiO–FeO binary and is used for the determination of the element Ni in a binary consisting of NiO and FeO.

Given the necessary matrix of a values, the analyst chooses analytical standards and determines $k_1, k_2, \ldots, k_n$. Then $\beta_1, \beta_2, \ldots, \beta_n$ can be calculated. From these values, a first approximation to $C_1, C_2, \ldots, C_n$ is found. Iteration continues until the differences between calculated values of β are arbitrarily small. In practice, the result really depends on how good the standard is; in geological work, standards of nearly the same average atomic number as the specimen are usually chosen. Hence, atomic number effects are reduced. Fluorescence effects are usually small in silicate minerals and oxides. Hence, most of the final correction is for absorption.

9.8.3.3. Application of the Bence–Albee Procedure

Tests of the validity of this procedure indicate that it yields results comparable to those obtainable with the *ZAF* method. The chief obstacle to its general use is the need for many homogeneous, well-characterized standard materials. For this reason, the a-factor method is sometimes combined with the *ZAF* or the $\phi(\rho z)$ method. In this case, one calculates the necessary a-factor matrix for n elements by assuming a concentration value C and using the *ZAF* or the $\phi(\rho z)$ method to compute the corresponding k values. The value of the a factor can be calculated from Eq. (9.8). For example, Albee and Ray (1970) calculated correction factors for 36 elements relative to their simple oxides for $E_0 = 15$ and 20 keV and for several take-off angles using the *ZAF* technique. Nevertheless, a calculated a factor is subject to all of the uncertainties associated with the *ZAF* or the $\phi(\rho z)$ methods as discussed in Sections 9.2 and 9.3. The use of such corrections, however, permits x-ray microanalysis using a small suite of simple crystalline oxides as standards and the calculation procedure is much simpler than the *ZAF* or the $\phi(\rho z)$ corrections.

As an example of the use of the empirical method, one can consider the microanalysis of the seven-component pyroxene discussed previously for the *ZAF* and the $\phi(\rho z)$ analysis (Tables 9.6 and 9.7). The $K\alpha$ lines for Si, Al, Cr, Fe, Mg, and Ca are used and the analysis was calculated for $E_0 = 15$ keV and $\psi = 52.5^\circ$. The appropriate a factors $[n(n-1) = 30]$ for these operating conditions taken from Albee and Ray (1970) are listed in Table 9.8. The β_i factors and intensity ratios k_i calculated for each element i from Eq. (9.14) for the meteoritic pyroxene are given in Table 9.9. In addition the calculated ZAF_i factors and k_i values from the *ZAF* and the $\phi(\rho z)$ procedures are listed in Table 9.9. The β and *ZAF* factors differ by no more than 2% relative and all three techniques give comparable results.

The Bence–Albee method has the special attributes that it works directly with the oxide standards that are most appropriate for the analysis of geological unknowns and that because the calculations involve only simple arithmetic, final results are obtained in the shortest time of any of the correction methods.

Table 9.8. Empirical a Factors for a Meteoritic Pyroxene Analysis[a,b]

Element	MgO	Al_2O_3	SiO_2	CaO	Cr_2O_3	FeO
Mg	1.0	1.03	1.09	1.26	1.50	1.76
Al	1.62	1.0	1.01[b]	1.14	1.28	1.37[c]
Si	1.29[c]	1.34[c]	1.0	1.05	1.12	1.19[c]
Ca	1.08	1.06	1.08	1.0	0.91	0.92
Cr	1.13	1.11	1.12	1.14	1.0	0.88
Fe	1.15	1.13	1.15	1.14	1.08	1.0

[a] Albee and Ray (1970).
[b] $\psi = 52.5^\circ$, $E_0 = 15$ keV.
[c] Modified by T. Bence, personal communication (1978).

9.8.3.4. Specimen Conductivity

Almost all specimens for geological analysis are coated with a thin layer of C or metal to prevent surface charging effects (see Chapter 15). However, x-rays are generated in depth in the specimen and the surface conducting layer has no influence on electrons that are absorbed micrometers into the specimen. Electrical fields may be built up deep in the specimen, which can ultimately affect the shape of the $\phi(\rho z)$ curves.

One problem to consider is the potential of electron energy loss in the conducting layer. Such a loss can happen due to the quality of the conducting layer and the current in the electron beam. Such an effect can be observed by measuring the short-wavelength limit of the continuum (see Chapters 6,15). If the value of the highest energy continuum x-rays is lower than E_0, measured separately on a good conducting material, the coating is a major problem. If not corrected, high-energy x-ray lines may suffer diminished intensity by the loss of the overvoltage for x-ray production.

A second problem lies in the possibility that the nonconducting specimen beneath the surface coating will undergo an electrical field buildup, which in some way will influence x-ray production so that the present ZAF or $\phi(\rho z)$ corrections do not describe the x-ray generation and absorption in the specimen. Although no simple solution has been offered, the potential problem must be considered.

Table 9.9. Calculation Using the a-Factor Technique for a Meteoritic Pyroxene[a]

Concentration (wt%)	a factor		ZAF method		$\phi(\rho z)$ method	
	β_i	k_i	ZAF_i	k_i	ZAF_i	k_i
53.3 SiO_2	1.074	0.496	1.091	0.489	1.075	0.496
1.11 Al_2O_3	1.16	0.00957	1.17	0.00947	1.158	0.00959
0.62 Cr_2O_3	1.102	0.00563	1.106	0.0056	1.105	0.00561
9.5 FeO	1.133	0.0838	1.15	0.0828	1.145	0.0830
14.1 MgO	1.179	0.120	1.16	0.122	1.175	0.120
21.2 CaO	1.047	0.203	1.06	0.200	1.065	0.199

[a] $\psi = 52.5^\circ$, $E_0 = 15$ keV.

As discussed in Section 9.8.1, most geologists do not measure oxygen directly, but get their oxygen values by stoichiometry. If oxygen analysis is desired, then C coatings are not optimal because they are highly absorbing for both O and N $K\alpha$ x-rays (see Section 9.10 for an in-depth discussion of light element analysis). Nevertheless, it is important to be aware of the subtle differences in metallic element analysis which may occur in conducting versus nonconducting silicate minerals, oxides, sulfides, etc.

9.8.4. Precision and Sensitivity in X-Ray Analysis

Up to this point we have discussed the corrections for quantitative x-ray microanalysis and the errors associated with the calculation procedure. In addition one must consider the precision and sensitivity of a given analysis. By the precision of an analysis we mean the scatter of the results due to the very nature of the x-ray measurement process. We can take advantage of statistical equations to describe the precision associated with composition determination expressed as confidence limits. This allows us to determine whether a given sample is homogeneous or if groups of samples are in fact different. This approach can also be used to tell if a measurement is statistically different from the background level and thus can be used to determine sensitivity (detection limits) or minimum detectable film thickness.

9.8.4.1. Statistical Basis for Calculating Precision and Sensitivity

Data collection procedures are different for EDS and WDS measurements; however, in each case the goal is the same: to accurately and precisely determine the intensity of the characteristic X-rays of each element emitted from a specimen and corresponding standard at defined energy settings. This process requires both the determination of corresponding background levels and correction for any dead-time effects. Because X-ray emission is a random process over time, each measurement is a sample of a parent distribution consisting of all possible repeated measurements under identical operating conditions. The expected distribution can be described by Poisson statistics, but for large numbers of counts can be approximated with a normal distribution. If the distribution of counts has a mean value of $\overline{N}$, then the standard deviation of that distribution σ is given by

$$\sigma = \sqrt{\overline{N}}. \tag{9.15}$$

Figure 9.27 shows such a distribution with a reminder that, if successive readings are made, 68.3% will be within $+1$ or -1 σ and 95.4% will be within ± 2 σ, and so on. If N only equals 100, then the relative standard deviation σ/N would be 0.1, whereas if N equals 10,000, then the relative standard deviation is 0.01. The smaller the relative standard deviation, the higher is the precision. Clearly, the more counts the better, and it is not uncommon to select operating conditions that will ensure greater than 50,000 counts in any peak measurement of importance. Although larger numbers of counts are sometimes possible by increasing the beam current or counting time, it may not always be practical to do so.

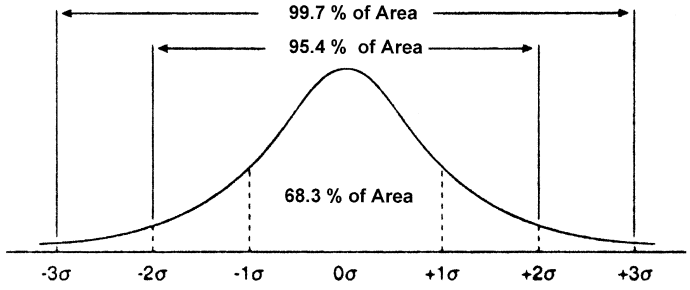


Figure 9.27. Precision x-ray counting statistics.

A common error is to misapply this simple fact to counting samples that have been "background-corrected." When a constant (the background) is subtracted from a counting sample the result no longer obeys Poisson statistics and other, more complicated, relations must be used (Fiori *et al.,* 1984; Small *et al.,* 1987). As the magnitude of the peak diminishes and becomes numerically comparable to the background on which it rides, it is extremely important not to make this mistake because large errors can result.

As Liebhafsky *et al.* (1955) pointed out, the real standard deviation of the experiment S_c is given by

$$S_c = \left[\sum_{i=1}^{n} \frac{(N_i - \overline{N}_i)^2}{n-1} \right]^{1/2}, \tag{9.16}$$

where N_i is the number of x-ray counts for each determination i and

$$\overline{N}_i = \left(\sum_{i=1}^{n} N_i \right) \Big/ n, \tag{9.17}$$

where n is the number of measurements of N_i. The standard deviation S_c equals σ_c only when operating conditions have been optimized. In some SEM instruments, drift of electronic components and of specimen position (mechanical stage shifts) can create operating conditions that are not necessarily ideal. The high-voltage filament supply, the lens supplies, and other associated electronic equipment may drift with time. After a specimen is repositioned under the electron beam, a change in measured x-ray intensity may occur if the specimen is off the focusing circle of a WDS x-ray spectrometer or if the take-off angle ψ of the specimen varies when using an energy-dispersive detector. In practice, for typical counting times of 10–l00 s/point, the actual standard deviation S_c is often at least twice σ_c. If longer counting times are used, S_c/σ_c increases due to instrument drift. Only when counting times are short and the instrument is electronically stable does S_c approach σ_c. Besides the sample signal, sources of variation may also occur if data from reference standards and/or background standards are required (Ziebold, 1967). These, as well as faulty specimen preparation, may also affect the precision of an analysis. Therefore, both instrumental factors and signal variations must be considered when the precision of an analysis is determined.

Before proceeding with the next sections it is important to review what we mean by "peak" (P) and "background" (B) in x-ray microanalysis and how we measure P and B (see Section 9.3, step 5b). For the WDS we measure the intensity at an energy at the center of the peak. To measure the rest of the peak profile it is necessary to "scan" the peak by moving the spectrometer through a range of wavelengths or energies encompassing the peak. We further note that such a WDS spectrometer tuned to the center of a peak is incapable of determining whether an x-ray is a peak or background x-ray. Consequently, what is measured is peak + background (P + B) and there is no way to directly tell one from the other. What is required is to estimate the background that is actually under the peak by one of several methods discussed in Section 9.3, step 5, and subtract the background value from the measured $P + B$. Consequently, when we write P what we really mean is $(P + B) - B$, where the second B is an estimate of the first B. The purpose of the parentheses is to emphasize the fact that the terms $P + B$ are really one quantity. With a WDS spectrometer, it is often an easy matter to get a very good estimate of the background under a peak. The spectrometer is detuned on each side of the peak and measurements are made at specific positions (Section 9.3, step 5), keeping track of how far each measurement is from the peak and how long each measurement took. Because the spectral resolving power of the WDS is $\sim$10 eV, the effects of interfering peaks are "usually" not a problem.

When an EDS is used to make a measurement of P and B, the process is made very much more complicated, and in some cases impossible. As discussed in Section 9.3, step 5, it can be very difficult to get a good estimate of the background under a peak where the background is strongly curved (1–3 keV) or where absorption edges occur. Indeed, in those cases where a "digital filter" is used to remove background, it is important to note that the filter "averages" each channel in a spectrum with its neighbors in such a way that complicates or prevents the use of many of the methods about to be discussed (the channels are no longer independent variates and the ratio of the magnitude of the peaks to the magnitude of the background is altered in the filtering process). Peak overlap is often the rule in EDS (Section 9.3, step 5c) and we must now contend with the statistical effects of the peaks of one element interfering with the background determination of the peaks of another element.

In the following sections it is safe to assume that the methods discussed will work well with data obtained with the WDS. The methods can also be applied to EDS data, provided the background information is obtainable and has not been modified. Furthermore, EDS data are normally processed as the integrated counts within a specified region of interest (ROI) from a statistical point of view. From a statistical viewpoint, if the ROI for an element measured contains counts from an overlapping peak, then those counts, in effect, contribute to the background in addition to the continuum.

9.8.4.2. Precision of Composition

In electron microprobe analysis, what is of ultimate importance is the precision of the composition rather than just that of an individual intensity

measurement. This point has been discussed in detail by a few authors (Ziebold, 1967; Lifshin *et al.,* 1999). Note first that a k ratio consists actually of the averages of four measurements: $\overline{N}$, the intensity measured on the sample; $\overline{N}(B)$, the corresponding background at the same energy; $\overline{N}_s$, the intensity measured on the standard; and $\overline{N}_s(B)$, the corresponding background for the standard:

$$k = \frac{\overline{N} - \overline{N}(B)}{\overline{N}_s - \overline{N}_s(B)}. \tag{9.18}$$

In reality a single measurement of each of the four terms in Eq. (9.18) results in only a single estimate of k and many sets of measurements are required to approach the true value of average value of k. If we accept this approximation, then we can use σ for s. Ziebold (1967) showed that the corresponding precision in the k ratio (σ_k) and the value of the precision of the measurement of concentration C (σ_c) are given respectively by

$$\sigma_k^2 = k^2 \left[\frac{\overline{N} + \overline{N}(B)}{n[\overline{N} - \overline{N}(B)]^2} + \frac{\overline{N}_s + \overline{N}_s(B)}{n'[\overline{N}_s - \overline{N}_s(B)]^2} \right] \tag{9.19}$$

and

$$\sigma_C^2 = C^2 \left[\frac{N + N(B)}{n[N - N(B)]^2} + \frac{N_s + N_s(B)}{n'[N_s - N_s(B)]^2} \right] \left[1 - \frac{(a-1)C}{a} \right]^2, \tag{9.20}$$

where n and n' refer to the numbers of measurements made on the sample and standard, respectively. The parameter a is used in the following equation (Ziebold and Ogilvie, 1964) relating the k ratio and c [see Eq. (9.8)]:

$$\frac{1-k}{k} = a\left(\frac{1-C}{C}\right).$$

If a is not known, then either the ZAF or the $\phi(\rho z)$ method can be used to calculate a value of k for a given C, and then a can be calculated from the above equation. With suitably long counting times, the standard deviation in a composition determination due to counting statistics can be reduced to less than 1% relative.

Equation (9.20) makes it possible to assess statistical uncertainty in an estimate of composition. For example, it can be used to construct a confidence interval (1.96σ for 95% confidence) for the difference of two samples or to plan how long data need to be collected to be able to estimate differences between two samples at the desired level of precision. The calculation of the confidence interval is based on the normal distribution of the estimate of C for large samples. To make a similar calculation using all of the equations incorporated within the ZAF or the $\phi(\rho z)$ method would be very complicated. An example of the use of Eq. (9.20) to calculate σ_c for the Ni concentration in an alloy with a composition of 58.6 wt% Ni, 38.5 wt% Cr, and 3.0 wt% Al is as follows: First determine the average number of Ni $K\alpha$ counts measured on the sample and

standard as well as the corresponding background counts for each. For example, at 10 kV and 100-s counting time for each measurement which was repeated six times ($n = n' = 6$), $\bar{N}_s = 11, 116, \bar{N}_s(B) = 1482, \bar{N} = 6882$, and $\bar{N}(B) = 1263$. Substituting these values in Eq. (9.18) gives $k = 0.583$. Using either the *ZAF* or the $\phi(\rho z)$ technique, this k value corresponds to $C = 0.586$ (58.6%). With k and C known, Eq. (9.8) can be used to show that $a = 1.011$. Therefore all of the quantities are known to evaluate Eq. (9.20) and to obtain $\sigma_c = 0.00472$. Because a 95% confidence interval corresponds to $C \pm 1.96\sigma_c$, the 95% confidence interval for Ni lies between 57.7% and 59.5%. This range is quite reasonable because longer counting times may not be practical due to contamination, drift, sample damage, or simply the availability of instrument time if many points are to be measured. Shorter counting times such as 10 seconds might increase the 95% confidence interval to an unacceptable level if the goal of the measurement is to compare the composition to similar alloys.

Furthermore, it cannot be over emphasized that counting statistics is only one factor that affects precision and that the real precision factor may often be at least a factor of two greater. Estimates of these additional sources of variation can be made in reproducibility experiments. Here the analyst positions a standard under a static electron beam, takes a reading, then mechanically moves to another location as would be the case when measuring a sample and returns to the standard to take another reading, repeating the process several times. Because the position of the standard relative to the x-ray detector is never exactly reproduced nor is it exactly in the same area analyzed, the actual standard deviation S_c calculated with Eq. (9.16) may be greater than σ_c.

9.8.4.3. Sample Homogeneity

An analyst is often asked if a sample and/or a phase is homogeneous. In order to answer this question, the x-ray data must be obtained so that it can be treated statistically. One can either set up criteria for homogeneity and apply them or one can measure the range of composition variation of a sample at a certain confidence level and report that number. Either method allows a more quantitative statement to be made than just a simple "yes" or "no" to questions concerning homogeneity. This section discusses homogeneity criteria that can be used and how to calculate the range and level of homogeneity.

A simplified criterion has been used to establish the homogeneity of a phase or a sample. This criterion states that for a homogeneous sample all the data points n, each taken at a different position on the sample, fall within the $\bar{N} \pm 3\bar{N}^{1/2}$ limits, where N is the raw, uncorrected intensity from the specimen (Yakowitz *et al.,* 1965). If this criterion is satisfied, one assumes that the sample is homogeneous. The variation

$$(\pm 3\bar{N}^{1/2}/\bar{N}) \cdot 100\% \tag{9.21}$$

for the element of interest in the sample represents the level of homogeneity in percent that is measured for the sample, remembering that there must

be an irreducible minimum level due to the fact that x-ray production is statistical in nature. If 100,000 counts are accumulated at each point in a sample and all these points fall within the limits $N \pm 3N^{1/2}$, the sample is homogeneous and according to Eq. (9.21) the level of homogeneity is $\pm 0.95\%$. A level of homogeneity of $\leq \pm 1\%$ is often desired. If the concentration in the sample C is l0 wt%, the range of homogeneity, that is, the minimum variation of concentration that can be validly measured, is ± 0.1 wt%. This equation is only valid when the P/B ratio is high enough to use $\overline{N}$ directly without a background correction.

A more exacting determination of the range (wt%) and level (%) of homogeneity involves the use of (1) the standard deviation S_c of the measured values and (2) the degree of statistical confidence in the determination of $\overline{N}$. The details can be found in Enhancements, Section E9.11

9.8.4.4. Analytical Sensitivity

Analytical sensitivity is the ability to distinguish, for a given element, between two concentrations C and C' that are nearly equal. X-ray counts N and N' for both concentrations therefore have a similar statistical variation. The value of $C - C'$, ΔC, represents the concentration difference that can be measured for a specific level of x-ray counts. The specific details for calculating analytical sensitivity using a statistics approach are given in the Enhancements Section E9.12.

If $\overline{N}$, the average x-ray count for an element of interest, is much larger than $\overline{N}_B$, the average background counts, we can approximate ΔC, the concentration difference that can be measured, by

$$\Delta C = C - C' \geq 2.33C/(n\overline{N})^{1/2} \tag{9.22}$$

and the analytical sensitivity in percent that can be achieved as

$$\Delta C/C(\%) = 2.33(100)/(n\overline{N})^{1/2}. \tag{9.23}$$

For an analytical sensitivity of l%, 54,290 accumulated counts, $n\overline{N}$ from Eq. (9.23), must be obtained from the sample. If the concentration C is 25 wt%, $\Delta C = 0.25$ wt%, and if the concentration C is 5 wt%, $\Delta C = 0.05$ wt%. Although the analytical sensitivity improves with decreasing concentration, it should be pointed out that the x-ray intensity from the sample also decreases directly with the reduced concentration. Therefore longer counting times will become necessary to maintain the 1% sensitivity level.

Equation (9.23) is particularly useful for predicting the necessary procedures for obtaining the sensitivity desired in a given analysis. If a concentration gradient is to be monitored over a given distance in a sample, it is important to predict how many data points should be taken and how many x-ray counts should be obtained at each point. For example, if a gradient from 5 to 4 wt% occurs over a 25-μm region, and 25 1-μm steps are taken across the gradient, the change in concentration per step is 0.04 wt%. Therefore ΔC, the analytical sensitivity at a 95% degree of confidence, must be 0.04 wt%. Using Equation (9.22) and noting that $\overline{N}$ is much larger

than $\overline{N}_B$, we find that $n\overline{N}$ must be at least 85,000 accumulated counts per step. If only 10 2.5-μm steps are used across the gradient, the change in concentration per step is 0.1 wt% and $n\overline{N}$ need only be 13,600 accumulated counts per step. By measuring 10 rather than 25 steps, the analysis time is cut down much more than the obvious factor of 2.5 because the number of required accumulated counts per step, due to sensitivity requirements, also decreases. Furthermore, if the data are plotted, the eye will "integrate" neighboring points and permit the unambiguous determination of even smaller differences.

9.8.4.5. Trace Element Analysis

As the elemental concentration C approaches the order of 1.0 wt% in x-ray microanalysis, $\overline{N}$ is no longer much larger than $\overline{N}_B$. This concentration range, below 1.0 wt%, is often referred to as the trace element analysis range for x-ray microanalysis. The analysis requirement in trace element analysis is to detect significant differences between the sample and the continuum background generated from the sample.

To develop a useful procedure for trace detection we need a criterion that will guarantee that a given element is present in a sample. This criterion can be called the "detectability limit" DL. The detectability limit is governed by the minimum value of the difference $\overline{N} - \overline{N}_B$ that can be measured with statistical significance. Leibhafsky *et al*. (1960) discussed the calculation of detectability limits. They suggest that an element can be considered present if the value of $\overline{N}$ exceeds the background $\overline{N}_B$ by $3(\overline{N}_B)^{1/2}$. A more sophisticated analysis must consider the measured standard deviation, the number of analyses, etc. Such a statistical approach for trace element analysis is given in the Enhancements, Section E9.13.

If we assume for trace analysis that the x-ray calibration curve of intensity versus composition is expressed as a linear function, then C, the unknown composition, can be related to $\overline{N}$ by the equation

$$C = \frac{\overline{N} - \overline{N}_B}{\overline{N}_S - \overline{N}_{SB}} C_S, \tag{9.24}$$

where $\overline{N}_S$ and $\overline{N}_{SB}$ are the mean counts for the standard and the continuum background for the standard, respectively, and C_S is the concentration in wt% of the element of interest in the standard. If we assume a detectability limit of three times the standard deviation of the background, then equation (9.24) gives us

$$C_{DL} = \frac{3\sqrt{\overline{N}(B)}}{\overline{N}_S - \overline{N}(B)} C_S. \tag{9.25}$$

The relative error or precision in a trace element analysis is equal to $(C_{DL}/C_S)\cdot 100\%$ and approaches $\pm 100\%$ as C approaches C_{DL}.

The background intensity $N(B)$ must be obtained accurately so that the unknown concentration C can be measured, as shown by Eq. (9.24). It is usually best to measure the continuum background intensity directly on the

sample of interest. Other background standards may have different alloying elements or a different composition, which will create changes in absorption with respect to the actual sample. Also, such background standards may have different amounts of residual contamination on the surface, which is particularly bad when measuring x-ray lines with energies $\leq$1 keV. The background intensity using a WDS is obtained after a careful wavelength scan is made of the major peak to establish precisely the intensity of the continuum on either side of the peak. Spectrometer scans must be made to establish that these background wavelengths are free of interference from other peaks in all samples to be analyzed.

Using the energy-dispersive spectrometer, it is difficult to measure backgrounds with the accuracy needed to do trace element analysis. Measurements of continuum background with the EDS are discussed in Section 9.3, step 5. If no overlapping peaks or detector artifacts appear in the energy range of interest and the continuum background can be determined, trace element analysis can be accomplished using the EDS detector. Equation (9.25) can also be used to predict C_{DL} when no standards are available. Lifshin *et al*. (1999) showed this can be accomplished using Monte Carlo calculations to predict the spectrum for a standard of known composition and then extracting the needed intensity data from that spectrum.

Another method for calculating C_{DL} was proposed by Ziebold (1967), who suggested using

$$C_{\mathrm{DL}} \geq 3.29a/(n\tau P \cdot P/B)^{1/2}, \tag{9.26}$$

where τ is the time of each measurement, n is the number of repetitions of each measurement, P is the pure element counting rate, P/B is the peak-to-background ratio of the pure element, that is, the ratio of the counting rate of the pure element to the background counting rate of the pure element, and a relates composition and intensity of the element of interest through the Ziebold and Ogilvie (1964) empirical relation [Eq. (9.8)]. To illustrate the use of this relation, the following values of the operating conditions were used for calculating the detectability limit for Ge in iron meteorites (Goldstein, 1967) using a WDS system:

- Beam energy = 35 keV
- τ = 100 s
- Specimen current = 0.2 μA
- n = 16
- P = 150,000 counts
- α = 1.0
- P/B = 200

With these numbers, Eq. (9.26) gives $C_{\mathrm{DL}} >$ 15 ppm (0.0015 wt%). The actual detectability limit obtained after using more sophisticated statistical equations (Enhancements, Section E9.12) is 20 ppm (0.0020 wt%) (Goldstein, 1967). Counting times of the order of 30 min were necessary to achieve this detectability limit. In measuring the carbon content in steels, detectability limits of the order of 300 ppm are more typical with a WDS system if one uses counting times of the order of 30 min, with the

Table 9.10. Comparison of the Minimum Detectability Limit of Various Elements Using EDS and WDS Detection Systems on the Basis of Optimized Operating Conditions[a,b]

Analysis	Element	P (cps)	B (cps)	P/B	Wet chem. (wt%)	C_{DL} (wt%)
EDS	Na $K\alpha$	32.2	11.5	2.8	3.97	0.195
	Mg $K\alpha$	111.6	17.3	6.4	7.30	0.102
	Al $K\alpha$	103.9	18.2	5.7	4.67	0.069
	Si $K\alpha$	623.5	27.3	22.8	26.69	0.072
	Ca $K\alpha$	169.5	19.9	8.5	12.03	0.085
WDS	Na $K\alpha$	549	6.6	83	3.97	0.021
	Mg $K\alpha$	2,183	8.9	135	7.30	0.012
	Al $K\alpha$	2,063	16.1	128	4.67	0.008
	Si $K\alpha$	13,390	37.0	362	26.69	0.009
	Ca $K\alpha$	2,415	8.2	295	12.03	0.009

[a] Geller (1977).

[b] Analysis of Dl-JD-35: EDS data collected at 2000 counts/s (cps) for 180 s, dead-time-corrected (25%) (1.75-nA probe current at 15 keV). WDS data collected for 30 s for each element. Total analysis time 180 s (30-nA probe current at 15 keV).

instrument operating at 10 keV and 0.05 μA specimen current. If higher specimen current can be used without increasing specimen contamination, the detectability limit will be lower.

A criterion which is often used to compare the sensitivity of wavelength and energy-dispersive detectors is the product $P \cdot P/B$ in the Ziebold relation, Eq. (9.26). Geller (1977) compared WDS and EDS detectors in which each detector was optimized for maximum P^2/B. The specimen used in the analysis was a synthetic mineral D1-JD-35 containing Na, Mg, Si, and Ca as the oxides. The EDS data were taken at a specimen current of 1.75 nA at 15 keV. An optimum count rate of 2000 counts/s on the specimen was obtained so that summed peaks would not be significant in the analysis. The WDS data were taken at a specimen current of 30 nA at 15 keV. An optimum count rate was set at 13,000 counts/s on Si, so that the maximum count rate caused less than 1% dead time in the proportional counter. Table 9.10 lists the C_{DL} for the elements present in the sample. For optimized operating conditions, the minimum detectability limit using the WDS is almost 10 times better than the EDS for all elements. Therefore if the solid specimen is not affected by the higher specimen current, the WDS can offer a factor of 10 increase in sensitivity over the EDS. Typical C_{DL} values for the WDS and EDS detectors are 0.1 and 0.01 wt%, respectively.

9.8.4.6. Trace Element Analysis Geochronologic Applications

Recently, there has been considerable interest and development in the application of the electron microprobe to geochronology (Cocherie and Albarede, 2001; Suzuki and Adachi, 1991, 1998; Williams *et al.,* 1999). This involves the precise measurement of the trace elements of U, Th, and Pb in minerals such as monazite (REE, Th phosphate) zircon (Zr silicate), and xenotime (Y phosphate). Because monazite apparently does not incorporate

Pb readily into the structure, all the Pb measured can be assumed to be radiogenic. Knowledge of the decay constants for the relevant isotopes of Th and U permit an age estimation because the length of time necessary to generate the measured Pb concentration can be calculated. Application of this technique in geology adds a powerful perspective to geochronologic studies as the analyses are done *in situ,* preserving rock textures and allowing the correlation of time with the growth, metamorphism, and deformation of rock-forming minerals. Micromapping of crystals prior to analysis reveals compositional domains, which commonly correspond to age domains, indicating multiple growth episodes. The exceptionally high spatial resolution, nondestructive nature, and mapping capabilities of the electron microprobe allow the acquisition of age information currently unattainable by other techniques. This in turn offers insight into metamorphic and deformation histories and into complex data obtained via traditional isotopic methods (Williams *et al.,* 1999; Williams and Jercinovic, 2002). The importance of this technique has led to an interest in improving the sensitivity of the electron probe microanalyzer and has inspired new developments in electron optics, x-ray optics, and software design. The primary hardware improvements concentrate on increasing P, B, and t by improving beam brightness, beam stability at the sample, and x-ray collection efficiency, primarily through the development of ultralarge WDS diffracting crystals.

9.8.4.7. Biological and Organic Specimens

It is now possible to obtain flat polished specimens of biological and organic material. Polymers may be planed smooth at ambient temperatures using a diamond knife. Soft biological specimens may be fixed, dehydrated and infiltrated with liquid resin which may be polymerized to provide a hard sample. Sections are cut at ambient temperatures using a diamond knife. The sections are discarded to reveal a flat polished specimen. (See Chapter 13 for details). Alternatively, quench frozen specimens may be cryoplaned at 140K using a diamond knife to reveal a flat polished surface (See Chapter 14 for details.).

Such flat polished surfaces have been analysed using the ZAF and the P/B method. Marshall and Xu (1998) have used the cryoplaning method to prepare flat polished specimens in which they were able to quantitively map light elements, including oxygen and carbon. Echlin (1999) and Echlin (2001) have used both room temperature and low temperature cryoplane processes on polymers and dry and frozen hydrated biological material to obtain flat polished specimens. Quantitative analysis was carried out with both the ZAF and the peak to local background algorithms using the L-line for Br and the K-line for O, Na, Mg, Al, Si, P, Cl, K, Ca, and Cr.

References

Albee, A. L., and L. Ray (1970). *Anal. Chem*. **42,** 1408.

Armstrong, J. T. (1988). In *Microbeam Analysis—1988* (D. E. Newbury, ed.), San Francisco Press, San Francisco, p. 469.

Bastin, G. F., and H. J. M. Heijligers (1984). In *Microbeam Analysis*—1984 (A. D. Romig, Jr., and J. I. Goldstein, eds.), San Francisco Press, San Francisco, p. 291.

Bastin, G. F., and H. J. M. Heijligers (1990). In *Proceedings 12th International Congress on Electron Microscopy* (L. Peachey and D. B. Williams, eds.), San Francisco Press, San Francisco, Vol. 11, p. 216.

Bastin, G. F., and H. J. M. Heijligers (1991a). In *Electron Probe Quantitation* (K. F. J. Heinrich and D. E. Newbury, eds.), Plenum Press, New York, p. 145.

Bastin, G. F., and H. J. M. Heijligers (1991b). In *Electron Probe Quantitation* (K. F. J. Heinrich and D. E. Newbury, eds.), Plenum Press, New York, p. 163.

Bastin, G. F., and H. J. M. Heijligers (1997). Quantitative Electron Probe Microanalysis of Boron in Binary Borides, Internal Report, Laboratory for Physical Chemistry, Eindhoven University of Technology, Eindhoven, The Netherlands.

Bastin, G. F., H. J. M. Heijligers, and F. J. J. van Loo (1986). *Scanning* **8,** 45.

Bastin, G. F., J. M. Dijkstra, and H. J. M. Heijligers (1997). *X-ray Spectrometry*.

Bence, A. E., and A. Albee (1968). *J. Geol*. **76,** 382.

Bence, A. E., and W. Holzwarth (1977). In *Proceedings 12th MAS Congress—1977,* Boston, MA p. 38.

Castaing, R. (1951). Ph.D. Thesis, University of Paris, Paris.

Cocherie, A., and F. Albarede (2001). *Geochim. Cosmochim. Acta* **65,** 4509.

Duncumb, P., and S. J. B. Reed (1968). In *Quantitative Electron Probe Microanalysis* (K. F. J. Heinrich, ed.), National Bureau of Standards Special Publication 298, U.S. Government Printing Office, Washington, D.C., p. 133.

Duncumb, P., and P. K. Shields (1966). In *The Electron Microprobe* (T. D. McKinley, K. F. J. Heinrich, and D. B. Wittry, eds.), Wiley, New York, p. 284.

Echlin, P. (1999). Microscopy Microanalysis **4,** 577.

Echlin, P. (2001). Microscopy Microanalysis **7,** 211.

Fiori, C. E., R. L. Myklebust, K F. J. Heinrich, and H. Yakowitz (1976). *Anal. Chem*. **48,** 172.

Fiori, C. E., C. R. Swyt, and K. E. Gorlen (1984). *Microbeam Analysis—1984* (A. D. Romig, Jr., and J. I. Goldstein, eds.), San Francisco Press, San Francisco, p. 179.

Geller, J. D. (1997). In *SEM/1977/I,* IIT Research Institute, Chicago, p. 281.

Goldstein, J. I. (1967). *J. Geophys. Res*. **72,** 4689.

Goldstein, J. I., R. E. Hanneman, and R. E. Ogilvie (1965). *Trans. Met. Soc. AIME* **233,** 812.

Heinrich, K. F. J. (1969). Present State of the Classical Theory of Quantitative Electron Probe Microanalysis, National Bureau of Standards Technical Note 521.

Heinrich, K. F. J., and H. Yakowitz (1968). Mikrochim. Acta **5,** 905.

Heinrich, K. F. J., R. L. Myklebust, S. O. Rasberry, and R. E. Michaelis (1971). National Bureau of Standards Special Publication 260. (Washington).

Joy, D. C., and Luo, S. (1989). *Scanning* **11,** 176.

Kramers, H. A. (1923). *Philos. Mag*. **46,** 836.

Laguitton, D., R. Rousseau, and F. Claisse (1975). *Anal. Chem*. **47,** 2174.

Liebhafsky, H. A., H. G. Pfeiffer, and P. D. Zemany (1955). *Anal. Chem*. **27,** 1257.

Liebhafsky, H. A., H. G. Pfeiffer, and P. D. Zemany (1960). In *X-ray Microscopy and X-ray Microanalysis* (A. Engstrom, V. Cosslett, and H. Pattee, eds.), Elsevier/North-Holland, Amsterdam, p. 321.

Lifshin, E. (1975). In Advances in X-ray Analysis, Vol. 19 (R. W. Gould, C. S. Barrett, J. B. Newkirk, and C. O. Roud, eds.) (Kendall/Hunt, Dubugue, Iowa), p. 113.

Lifshin, E., N. Doganaksoy, J. Sirois, and R. Gauvin (1999). *Microsc. Microanal*. **4,** 598.

Marshall, A. T. and W. Xu (1998). *J. Microsc.* **190,** 305.

Newbury, D. E. (1999). *Microsc. Microanal.* **4,** 585.

Newbury, D. E., Swyt, C. R., and Myklebust, R. L. (1995) *Anal. Chem* **67,** 1866.

Packwood, R. H., and J. D. Brown (1981). *X-Ray Spectrom*. **10,** 138.

Philibert, J. (1963). In *Proceedings 34th International Symposium on X-ray Optics and X-Ray Microanalysis* (H. H. Pattee, V. E. Cosslett, and A. Engstrom, eds.), Academic Press, New York, p. 379.

Pouchou, J. L., and R. Pichoir (1984). *Rech. Aerosp*. **3,** 13.

Powell, C. J. (1976). Rev. Mod. Phys. **48,** 33.
Reed, S. J. B. (1965). *Br. J. Appl. Phys*. **16,** 913.
Schamber, F. H. (1978). In *Proceedings 13th National Conference of the Microbeam Analysis Society*, Ann Arbor, Michigan, p. 50.
Small, J. A., S. D. Leigh, D. E. Newbury, and R. L. Myklebust (1987). *J. Appl. Phys.* **61,** 459.
Suzuki, K., and M. Adachi (1998). *J. Metamorphic Geol*. **16,** 23.
Suzuki, K., and M. Adachi (1991). *J. Geochem.* **25,** 357.
Williams, M. L., and M. J. Jercinovic (2002). *J. Struct. Geol.* **24,** 1013.
Williams, M. L., M. J. Jercinovic, and M. P. Terry (1999). *Geology* **27,** 1023.
Yakowitz, H., D. L. Vieth, K. F. J. Heinrich, and R. E. Michaelis (1965). Homogeneity Characterization of NBS Spectroscopic Standards II: Cartridge Brass and Low Alloy Steel, National Bureau of Standards Publication 260–10.
Ziebold, T. O. (1967). *Anal. Chem*. **39,** 858.
Ziebold, T. O., and R. E. Ogilvie (1964). *Anal. Chem*. **36,** 322.

10

Special Topics in Electron Beam X-Ray Microanalysis

10.1. Introduction

Chapter 9 presented the procedures for performing quantitative electron probe x-ray microanalysis for the case of an ideal specimen. The ideal specimen surface is flat and highly polished to reduce surface roughness to a negligible level so that electron and x-ray interactions are unaffected by geometric effects. Such a highly polished surface has a short-range surface topography (sampled at distances less than 1 μm) that is reduced to an amplitude of a few nanometers and the long-range topography (sampled at distances greater than 100 μm) that is reduced to 100 nm or less. These ideal specimens satisfy three "zeroth" assumptions that underlie the conventional EPMA technique:

1. The only reason that the x-ray intensities measured on the unknown differ from those measured on the standards is that the compositions of specimen and standard are different. Specifically, no other factors such as surface roughness, size, shape, and thickness, which can be generally grouped together as "geometric" factors, act to affect the intensities measured on the unknown.
2. The specimen is homogeneous over the full extent of the interaction volume excited by the primary electron beam and sampled by the primary and secondary x-rays. Because x-rays of different excitation energies are generated with different distributions within the interaction volume, it is critical that the specimen has a uniform composition over the full region. If a thin surface layer of different composition than the underlying bulk material is present, this discontinuity is not properly considered in the conventional matrix correction analysis procedure.
3. The specimen is stable under the electron beam. That is, the interaction volume is not modified through loss of one or more atomic or molecular species by the electron beam over the time period necessary to collect the x-ray spectrum (EDS) or peak intensities (WDS). Biological

and polymer specimens are likely to alter composition under electron bombardment.

This chapter will deal with a series of special analytical situations that violate one or more of these basic assumptions. A general caution can be stated to alert the analyst to a possible special analysis situation that is related to geometric factors: If any specimen feature approaches the scale of the interaction volume, a nonstandard analytical condition will almost certainly result. The x-ray range equation (6.12) gives an estimate of the dimensions of the interaction volume and the region within which x-rays from a particular excitation edge can be generated. When the specimen dimension(s) decreases to a value within a factor of 10 of the value given by Eq. (6.12), the analyst should expect to see the effects of geometric factors on the measured x-ray intensities. Indeed, for large, rough objects, the influence of geometric factors is manifested at all dimensions.

Another powerful indicator of the impact of geometric factors on an analysis is the analytical total. If all elemental constituents within the interaction volume are measured and quantified with standards and matrix correction factors (i.e., no element is determined by difference, and oxygen is calculated either directly against a standard or indirectly by the method of assumed stoichiometry of the cations), then the analytical total will typically range between 98% and 102%. The deviation from 100% represents the cumulative error in the determination of each constituent, both from random (i.e., counting statistics) and systematic (e.g., instrumental factors and inaccuracies in matrix correction factors) error sources. When the total moves outside this range, the source of the error should be carefully checked. Possible errors due to instrumental factors such as variations in beam current or specimen factors such as local inclination should be considered first. Specimen geometric effects can lead to excursions both above and below 100%. An analytical total below this range can also indicate the presence of an unmeasured constituent. Specimens that undergo beam damage or migration may also show a deviation in the analytical total.

10.2. Thin Film on a Substrate

A film on a substrate represents a case in which one dimension, the specimen thickness along the beam direction, becomes an important factor in the analysis. The lateral dimensions of the film are typically much greater than the interaction volume. When the film thickness becomes less than the interaction volume dimension for its particular composition, the beam will penetrate through the film into the substrate, generating x-rays in both film and substrate. If the substrate x-rays are not completely absorbed while passing through the film, the analyst will measure a composite spectrum containing contributions from the compositions of both regions. In thin-film analysis, the typical analytical request is to determine both the thickness

and the composition of the film. Some films are even more complex, with two or more layers of different compositions.

Analysis of thin films should begin with a careful qualitative analysis. If the film and substrate are both of unknown composition, the analyst should endeavor to analyze the uncoated substrate separately, including qualitative and quantitative analysis. With knowledge of the elements contained within the substrate, the analyst can proceed to consider the composite spectrum of the film plus substrate. Elements that appear in the composite spectrum that are not contained in the substrate must be the contribution of the film. The analyst should be careful to consider if a constituent is common to both film and substrate. This is the most difficult case to resolve, and it may require careful application of thin-film strategy 1 described below.

Strategy 1. Convert the film-on-substrate situation back into a bulk analysis case. Before applying the advanced procedures described below for the analysis of films, the analyst should first consider the possibility of altering the measurement strategy to convert the film-on-substrate situation back into a bulk analysis case. Because of the strong dependence of the x-ray range on beam energy, $R_x \sim (E_0^{1.68} - E_c^{1.68})$, it may be possible to lower the beam energy sufficiently to contain the interaction volume for the x-rays of interest within the film. By avoiding beam penetration into the substrate, as indicated by elimination of substrate element contributions to the spectrum, the film can be effectively analyzed as a bulk material using the conventional matrix correction procedure. For this strategy to be successful, the reduced beam energy must still be sufficient to excite all of the characteristic x-ray peaks of interest, which requires that $U = E_0/E_c > 1$. For most situations, if $U > 1.25$, a major constituent ($C > 0.1$ mass fraction) will produce an adequate peak above the background. However, as the concentration decreases, a low-overvoltage situation may result in poor detectability.

When applying this "back-to-bulk-analysis" strategy, it may also be necessary to consider changing the atomic shell with which an element is measured. Thus, for a film of aluminum and zinc, it is possible to measure Al $K\alpha$ and Zn $L\alpha$ instead of Al $K\alpha$ and Zn $K\alpha$, which allows a much lower beam energy to be used because for zinc, $E_L = 1.02$ keV and $E_K = 9.66$ keV, whereas for aluminum, $E_K = 1.56$ keV. Thus a beam energy as low as 2.5 keV will give adequate overvoltage for both the Al $K\alpha$ and Zn $L\alpha$ peaks.

Another possible strategy is to tilt the specimen to a high value, $\theta \geq 60^\circ$. At high tilt, the penetration of the beam electrons is reduced by approximately $\cos\theta$, or 50% at $\theta = 60^\circ$, compared to normal incidence. For quantitative analysis, it is vital that the polished bulk standards be measured at the same tilt angle and x-ray detector take-off angle. Clearly, the success of this "back-to-bulk-analysis" approach depends very much on the exact composition and thickness of the film, but it is a useful strategy with which to begin.

Strategy 2. Thin-film analysis. If the back-to-bulk approach is unsuccessful or impractical, the analyst must attack the problem with an advanced film-on-substrate procedure. A series of spectra should be recorded over a

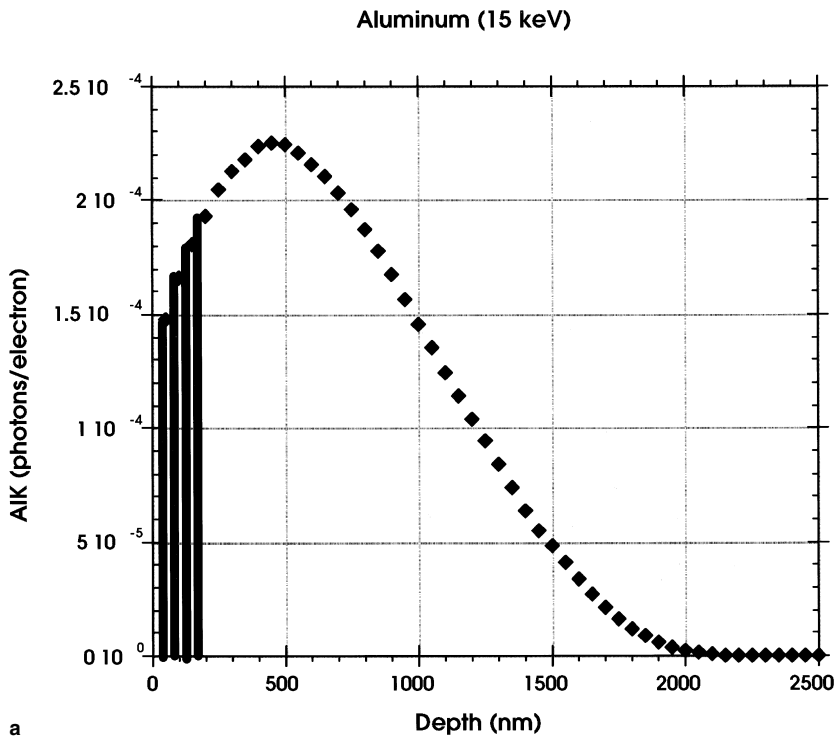
progression of beam energies, for example, 5, 10, 15, 20 keV. By carefully interpreting these spectra, aided, if possible, with knowledge of the substrate composition separately obtained, quantitative analysis of the film composition and its thickness is possible with a high degree of accuracy (e.g., Pochou and Pichoir, 1991).

The $\phi(\rho z)$ description of the depth distribution of x-ray production provides a powerful basis for understanding the origin of the specimen geometry effects of a thin film and how modified analytical procedures are developed to correct for them. Consider the $\phi(\rho z)$ curve for a flat, bulk target of pure element A, as calculated with a Monte Carlo simulation and shown in Fig. 10.1a. The integral of the complete area under this curve represents the total radiation generated from this pure element, and after correction for self-absorption, the integral represents the total emitted radiation, that is, the standard intensity of pure A. Now consider the specimen to consist of a film of A starting at the surface and extending to a depth t, which defines the thickness. Below this film the specimen continues to be pure A. The film is represented by the shaded area in Fig. 10.1a. The area under this portion of the $\phi(\rho z)$ curve represents the x-ray intensity generated within the film and, after correction for self-absorption, the x-ray intensity emitted from the film. We can define a k value for thin-film measurement (k_{film}) as the ratio of these two quantities:

$$k_{i,\text{film}} = \int_0^{\rho t} \phi(\rho z) \exp\left(-\frac{\mu}{\rho} \csc \varphi \rho z\right) d\rho z \Bigg/ \int_0^{\rho z_{\max}} \phi(\rho z) \times \exp\left(-\frac{\mu}{\rho} \csc \varphi \rho z\right) d\rho z, \tag{10.1}$$

where the mass absorption coefficient (μ/ρ) refers to the radiation of interest in the appropriate matrix composition, ψ is the spectrometer take-off angle, ρz represents the depth, and $\rho z_{\max}$ is the maximum range for the x-ray of interest. (For more background information on the $\phi(\rho z)$ technique see the accompanying CD, Enhancements, Chapter 9, Section E9.4.)

This is a trivial case because we cannot distinguish the radiation from the hypothetical layer from the same radiation produced by the rest of the thickness of the bulk target because it is the same material. In reality, a film of one composition will be placed upon a substrate of another composition, so that an abrupt discontinuity will occur in the $\phi(\rho z)$ curve at the depth t (although it is possible that one or more elemental constituents may be shared in both the film and substrate, so that an abrupt discontinuity will not be observed for those shared elements, only a change in slope as the electron scattering and x-ray absorption change at the interface). This situation is illustrated in Fig. 10.1b for the case of a 300-nm-thick aluminum layer on various substrates, where the incident beam energy (15 keV) is sufficiently high that the film is penetrated and the underlying substrate is also excited. The Al intensity is found to vary systematically with the atomic number of the substrate. This behavior can be understood in terms of the increased backscattering that occurs below the film when $Z_{\text{substrate}} > Z_{\text{film}}$. The flux of



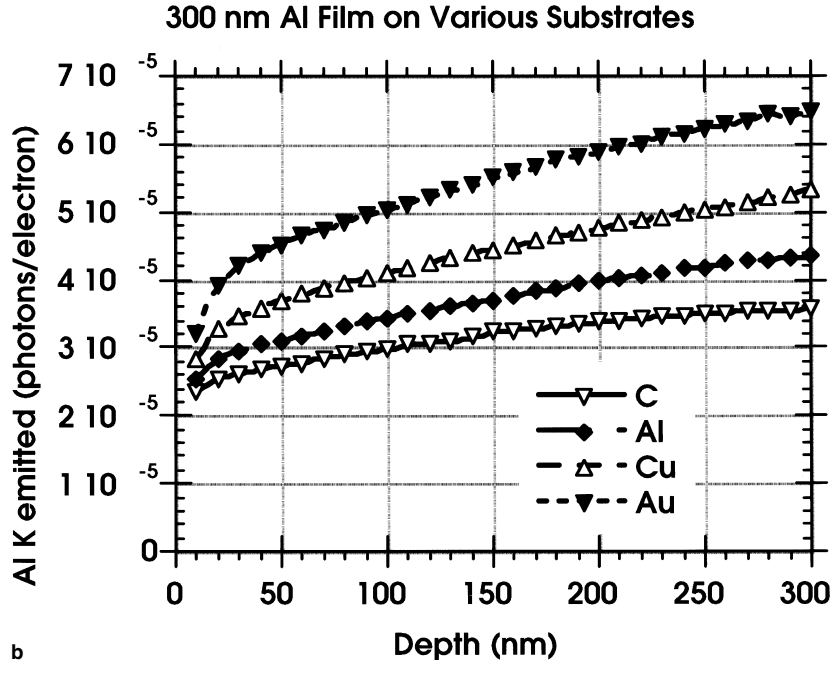


Figure 10.1. (a) Example of a $\phi(\rho z)$ curve for a solid target. A "film" of the same element is shown as a portion of the distribution starting at the surface. (b) $\phi(\rho z)$ curves for 300 nm of Al on various substrates (C, Al, Cu, and Au) at 15 keV calculated by Monte Carlo simulation.

backscattered electrons passing back through the film creates additional x-rays compared to the case where the film is placed on a substrate of similar atomic number. When $Z_{substrate} < Z_{film}$, the reduction in backscattering results in a lower intensity of x-ray production in the film compared to a substrate of similar atomic number to the film. To describe the $\phi(\rho z)$ curve for the supported thin film, the bulk pure-element $\phi(\rho z)$ curve must be modified to include the change in the electron scattering at the interface.

When the beam penetrates the film and also excites the substrate, we can measure two distinct k values, k_{film} and $k_{substrate}$ (if the substrate radiation is not completely absorbed in passing through the film) both relative to bulk pure-element standards, for example, Al $K\alpha$, Cu $K\alpha$, and Cu $L\alpha$ for an Al film on Cu. Again, a modified $\phi(\rho z)$ curve can be calculated for the thick substrate to take into account the electron scattering in the overlying film. The calculation of the emitted x-radiation from the substrate must consider not only the absorption in the substrate, but that in the overlying film as well, where the composition is different.

Such k values for constituents of the film and substrate measured relative to bulk pure-element standards (or in the case of compound standards, the measured standard intensity is recalculated to yield the equivalent pure element intensity) form the measurement basis for quantification of thin films. For very complex situations involving multiple films or when elements are partitioned between the film and substrate, the experimental k values must be measured as a function of the beam energy. By systematically increasing the beam energy and therefore progressively altering the sampling of the film/substrate, a multidimensional variable dataspace is created. This k-value dataspace can be compared to theoretical predictions based upon assumed models of the film thickness and composition and the substrate composition to find the best fit. In general, the more independent information about the system that can be supplied, such as substrate composition, and the more independent experimental k values measured, the more accurate will be the final result of the fitting procedure.

Two theoretical methods, the Monte Carlo simulation and the film-modified $\phi(\rho z)$ curve, are in general use to support quantitative analysis of films. A model of the film composition and thickness is assumed, and the theoretical methods are used to predict k values for comparison with measured k values.

1. The Monte Carlo simulation can be directly adapted to calculate any specified film thickness and film/substrate compositions. Thus, Fig. 10.1b shows an example of a series of Monte Carlo calculations for Al films on various substrates excited with a 15-keV beam. Both generated and emitted x-ray intensities can be calculated. By using the same Monte Carlo simulation to calculate bulk target intensities at the same beam energy and spectrometer take-off angle for the pure elements of the film and substrate elements, values of k_{film} and $k_{substrate}$ can be determined. Because of the small dimensions to be studied, the electron position is followed on a very fine scale by using a "single-scattering" Monte Carlo formulation, which uses a step length based on the elastic mean free path. The crossing of the film/substrate interface and all other electron aspects of the interaction must

be accurately simulated. Moreover, x-rays produced in the film undergo absorption in the film composition only, whereas x-rays produced in the substrate suffer absorption both in the substrate itself as well as in the overlying film through which they must pass to reach the detector. The same Monte Carlo simulation is then used to calculate x-ray emission from bulk substrates so that theoretical k values can be calculated. The chief limitation on the Monte Carlo approach is its lack of treatment of secondary x-ray production in the film and substrate through characteristic and continuum-induced secondary fluorescence, although characteristic fluorescence could be calculated once the distribution of primary radiation has been calculated with the Monte Carlo calculation.

2. The film-modified $\phi(\rho z)$ curve approach starts with the mathematical description of the pure-element $\phi(\rho z)$ curves and systematically alters those expressions to account for the effect of the film/substrate discontinuity. One of the most important factors is the difference in the film and substrate electron scattering, which is necessary to accurately describe the effect seen in Fig. 10.1b. This factor is introduced by means of an electron penetration model, information for which can be obtained from experiments involving foils as well as from Monte Carlo simulation. Once the film-modified $\phi(\rho z)$ expression for the generated intensity is calculated, absorption is treated in a similar fashion to the Monte Carlo simulation. The film-modified $\phi(\rho z)$ method has the advantage that, because the $\phi(\rho z)$ curve is closely tied to experimental data, the effect of continuum-induced secondary fluorescence is automatically embedded in the experimental measurements, at least to a first order. Characteristic secondary fluorescence must still be treated from the primary characteristic distribution if a complete description is to be achieved. Moreover, the film-modified $\phi(\rho z)$ method has a significant advantage in terms of calculation speed because the expressions, although complicated, are determined and can be executed rapidly on a computer, making extensive testing of the variable space possible.

Example 1: Thickness of a pure-element film. This is the simplest film-on-substrate problem, and it is approached by systematically calculating the film and substrate k values for various film-thicknesses with the Monte Carlo simulation or with the film-modified $\phi(\rho z)$ equations to develop k value versus thickness curves at fixed beam energy, as shown in Figs. 10.2a and 10.2b. With these curves, the analyst can immediately convert an experimental k value measured under the same beam and detector efficiency conditions into a corresponding film thickness. If the k value for the substrate radiation is also measured, two independent estimates of the film thickness can be obtained, one each from k_{film} and $k_{\text{substrate}}$.

Example 2: Thickness and composition of a multielement film. When both the thickness and the film composition are unknown, both the Monte Carlo method and the film-modified $\phi(\rho z)$ equation method must be used to explore a multiple-variable dataspace consisting of thickness and each compositional variable. To speed up this process, a good initial estimate of thickness and composition is needed to reduce the search range. The

experimental k values provide the basis for an effective initial estimate. If the characteristic x-rays of the film can be isolated from those of the substrate (i.e., no element is shared in both), then the raw k values for the film can be normalized by the raw sum to give an initial estimate of the concentrations for the film, exactly how the first concentration estimate is obtained in the conventional bulk analysis matrix correction method:

$$C_{i,\text{film}} = k_{i,\text{film}} \Big/ \sum_j k_{j,\text{film}} \tag{10.2}$$

where the summation is taken over all film components j, with oxygen calculated by the method of assumed stoichiometry, if desired.

To estimate the total film thickness, the individual film k values for each element are compared with the corresponding pure-element k value versus thickness plots. For each k_{film} the equivalent mass thickness (ρt, where ρ is the density and t is the linear thickness) is determined as that element's contribution to the overall film (normalized by the pure standard). The separate mass thicknesses (not linear-dimension thicknesses) are then added to give an initial estimate of the mass thickness of the multielement thin film:

$$\rho t_{\text{film}} = \rho t_{1,\text{film}} + \rho t_{2,\text{film}} + \rho t_{3,\text{film}} + \ldots, \tag{10.3}$$

where the subscripted numbers represent the elemental constituents of the film.

With these initial estimates of $C_{i,\text{film}}$ and ρt_{film}, a concentration range $\pm\Delta C_{i,\text{film}}$ and a thickness range $\pm\Delta\rho t_{\text{film}}$ can be explored with the Monte Carlo simulation or the film-modified $\phi(\rho z)$ method to generate k values to compare with the experimentally measured values. The solution is determined when an appropriate statistical measure of the fit of theoretically predicted k value to experimentally measured k value is satisfied.

Software implementations of the film-modified $\phi(\rho z)$ method are sufficiently advanced that the user can enter the estimated film/substrate

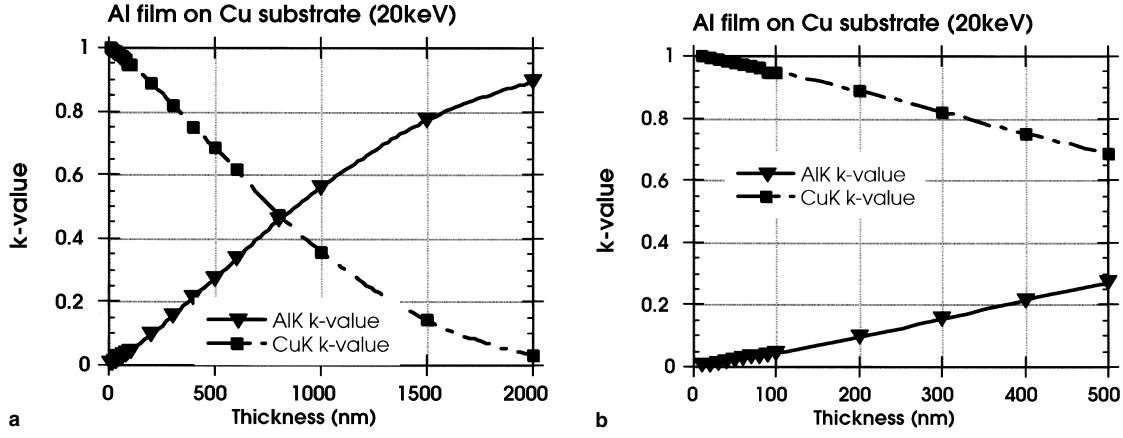


Figure 10.2. Monte Carlo calculations of k_{film} and $k_{\text{substrate}}$ for Al films of various thicknesses on a Cu substrate for an incident beam energy of 20 keV. (a) 0–2000 nm; (b) expanded view of the initial 0–500 nm.

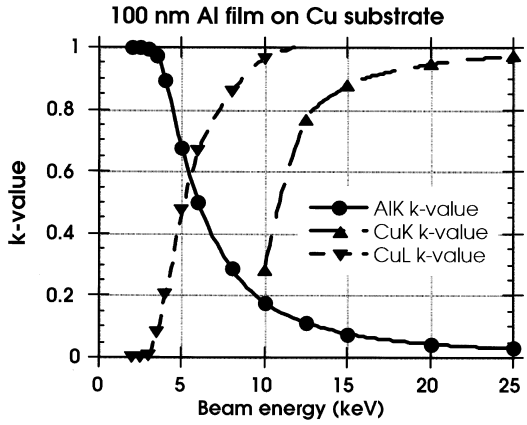


Figure 10.3. Monte Carlo calculations of k_{film} and $k_{substrate}$ for a 100-nm Al film on a Cu substrate at various beam energies.

configuration along with the measured film and substrate k values, and the k value/thickness space will be automatically searched to give the best fit. Such a procedure generally benefits by having as much data as possible to constrain the search, such as k values from both film and substrate as well as k-values versus beam energy. Figure 10.3 shows such a k-value curve for a 100-nm Al film on Cu as a function of beam energy. For the most complicated situations, such as multiple films or buried films, k-value data versus beam energy must be measured to provide sufficient constraints.

Examples of quantitative thin-film analyses performed with the Monte Carlo simulation and with the film-modified $\phi(\rho z)$ method are listed in Tables 10.1 [$\phi(\rho z)$ method] and 10.2 (Monte Carlo simulation). The analyst needs to be aware of the difficulty in rigorously testing these thin-film procedures due to the problem in obtaining independent measurements with which to verify film thicknesses and film/substrate compositions. Where testing has been possible, these quantitative EPMA thin-film procedures have usually shown good accuracy, generally providing results within $\pm 10\%$ for thickness and composition when compared with independent methods such as Rutherford backscattering (RBS) of high-energy ions (Pouchou and Pichoir, 1991).

Table 10.1. Analysis of a Thin Film on a Substrate (Cu–Pd–Au on SiO_2)[a]

			C	
Element	Line	k ratio	PAP Model	RBS
Cu	K	0.191	0.317	0.319
Pd	L	0.157	0.350	0.347
Au	M	0.157	0.333	0.334
Thickness (μg/cm^2)			278	282

[a] PAP, the $\phi(\rho z)$ method of Pouchou and Pichoir (1991). RBS, Rutherford backscattering (high-energy ions).

Table 10.2. Analysis of a Thin Film on a Substrate (Co–Pt on SiO_2)[a]

			C	
Element	Line	k ratio	MC	RBS
Co	K	0.0177	0.185	0.194
Pt	M	0.0581	0.810	0.806
Thickness (μg/cm^2)			53	53

[a] MC, Monte Carlo method of Kyser and Murata (1974). RBS, Rutherford backscattering (high-energy ions).

10.3. Particle Analysis

For films on substrates and foils, only the depth dimension is comparable to the electron range, and hence the modifications to conventional analytical procedures need only be parametrized in that one dimension. When microscopic particles are considered, however, all dimensions of the specimen are important because any or all of the particles may be comparable to the electron range, so that electron scattering and x-ray absorption geometric effects can occur in all dimensions. An example of the complexity of the situation is illustrated in Fig. 10.4 for a wide (500 nm) beam incident on a 2-μm-diameter aluminum sphere. At $E_0 = 5$ keV, which gives adequate overvoltage for Al ($E_c = 1.55$ keV), the electron trajectories are well-contained within the particle. Despite the surface curvature, the electron scattering and x-ray absorption situations are nearly equivalent to a flat, bulk specimen. As the beam energy is increased above 5 keV, the electrons spread to completely fill the particle and escape from all surfaces at 17 keV, and penetrate mostly through the bottom of the particle at 30 keV. The particle scattering and x-ray propagation situations depend on particle size and composition and on the incident beam energy. This complex situation gives rise to three geometry effects on the x-ray intensities, which have parallel physical origins to the *ZAF* factors: the mass effect, the absorption effect, and the fluorescence effect (Small *et al.,* 1978, 1979).

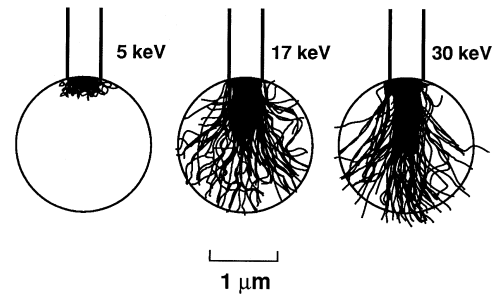


Figure 10.4. Monte Carlo electron trajectory calculations for the interaction of electrons in Al particles as a function of beam energy (5, 17, and 30 keV). The beam diameter is 0.5 μm and the particle diameter is 2 μm. (Newbury *et al.,* 1980.)

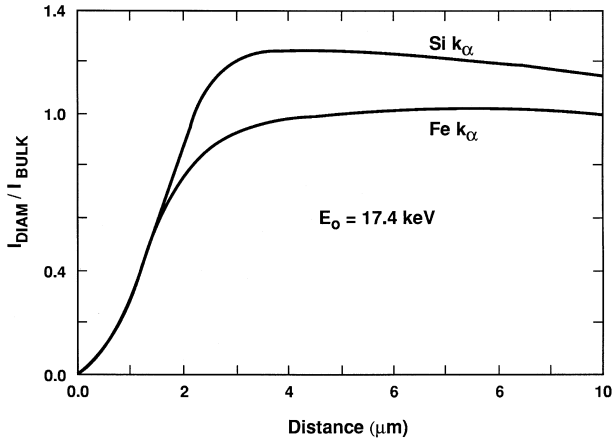


Figure 10.5. Monte Carlo electron trajectory calculations of the normalized intensity I_{diam}/I_{bulk} as a function of diameter for spherical particles of NIST glass K-309 (15% Al_2O_3, 40% SiO_2, 15% CaO, 15% Fe_2O_3, 15% BaO).

10.3.1. Particle Mass Effect

When the particle dimensions approach those of the interaction volume in a bulk solid of the same composition, as estimated with the x-ray range equation, the electrons escaping through the sides and bottom of the particle carry away energy. In the bulk case, these electrons would have remained in the specimen to contribute to the generation of x-rays. The generation of x-rays in particles thus decreases relative to a bulk standard. The intensity response from particles, normalized to a bulk standard of the same composition, as a function of diameter for spherical particles is shown in Fig. 10.5 for two different characteristic x-rays, Fe $K\alpha$ (6.40 keV) and Si $K\alpha$ (1.74 keV). The mass effect *always* leads to a lower intensity measured in the particle compared to the standard. The effect becomes significant for particle diameters below 5 μm, although the onset size depends on the average atomic number and the beam energy. Because the electron range increases as the 1.68 power of the beam energy (Chapter 3), a decrease in the beam energy can significantly reduce the particle size at which the mass effect becomes significant. Since x-rays with high E_c are generated close to the beam impact (shallow x-ray range), the mass effect is most pronounced for photons with low E_c.

10.3.2. Particle Absorption Effect

For a flat, bulk specimen placed at specific angles to the incident beam and the x-ray detector, there will be a well-defined, calculable path along which the x-rays travel through the specimen to reach the detector, as shown in Fig. 10.6. Because x-ray absorption follows an exponential relation, any changes in this path length will have a strong effect on the fraction of x-rays that escape the specimen. As shown in Fig. 10.6, the curved surface of a spherical particle can substantially shorten the x-ray absorption

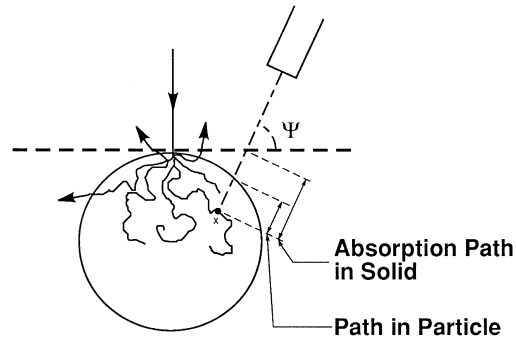


Figure 10.6. Schematic diagram comparing the absorption path in a bulk target with that in a spherical particle.

path for a significant portion of the interaction volume on the side of the particle facing the detector. X-rays generated in this region are more likely to escape from the particle than from a bulk target of the same composition, increasing the measured intensity above what is expected from a bulk target. This situation is illustrated for Si *K* in Fig. 10.5 (where *K* Si exceeds unity) and forms the basis of the particle absorption effect. The particle absorption effect is only significant for low-energy photons, where the threshold for "low" photon energy is approximately 4 keV (but the exact value is highly dependent on the specifics of the composition, particle size, and beam energy). For high-energy x-rays above 4 keV, there will only be slight absorption in the bulk case, so that the increase in emission due to the particle absorption effect will be negligible (e.g., Fe $K\alpha$ in Fig. 10.5, where the ratio approaches unity). For photon energies below 4 keV, the absorption effect generally increases as the photon energy decreases. A "soft" x-ray such as O *K* may suffer absorption of 90% or higher for a bulk target, so that when the particle absorption effect occurs, oxygen will appear at a significantly higher concentration than expected.

10.3.3. Particle Fluorescence Effect

As described in Chapter 6, when an atom absorbs an x-ray through the photoelectric effect, that atom is left ionized in a shell. The subsequent decay of this excitation state will produce either an Auger electron or a characteristic x-ray photon of the absorbing atom, a phenomenon referred to as secondary fluorescence to distinguish it from direct electron ionization that produces the primary fluorescence. Secondary fluorescence from a particular atomic shell can be induced by characteristic or continuum x-rays as long as the photon energy is greater than the edge energy. The efficiency of ionizing that specific edge is greatest for photons with energies just above the edge. The ionization efficiency decreases rapidly as the photon energy increases beyond the edge energy. Thus, only characteristic photons within about 1 keV of an edge can create significant secondary fluorescence. Because continuum photons occur at all energies up to the Duane–Hunt limit, there will always be some secondary fluorescence of

all edges with $E_c < E_0$, the incident beam energy. The range of x-ray absorption and the subsequent production of characteristic x-rays is much greater than the range of electrons. It is therefore possible for a particle to contain virtually all of the electron interaction volume (primary radiation production), but still be small enough compared to the x-ray range so that most of the secondary fluorescence from characteristic and/or continuum x-rays is not produced. This loss of possible intensity is known as the particle fluorescence effect. The particle fluorescence effect lowers the intensity for x-rays measured relative to bulk standards, which are large enough to contain all of the secondary fluorescence. Compared to the particle mass and particle absorption effects, the particle fluorescence effect generally contributes only small errors.

10.3.4. Particle Geometric Effects

When particles are analyzed relative to bulk, pure-element standards (or the equivalent pure-element intensities calculated from compound standards), the most obvious consequence of the particle geometric effects can be observed in the analytical total. If all elemental constituents are measured, or for the case of oxygen included by means of stoichiometric calculations, and no constituents are calculated by difference, then the analytical total carries information. The analytical total rarely equals exactly 1.0 mass fraction (e.g., 100%) because of the analytical error associated with each separate elemental determination. Considering the accuracy normally achieved with standards and matrix corrections for flat, bulk specimens, the analytical total can be expected to range between 0.98 and 1.02 mass fraction. Particle effects, depending on the size and specific elements involved, can cause the analytical total to deviate substantially outside this range. It is therefore important in conducting particle analysis to insist upon selecting the unnormalized analytical total as a particle diagnostic. Table 10.3 contains a series of EDS analyses of glass particles (NIST glass K-411) of various diameters with a focused probe of 50 nm placed at the top center of the particle. The x-ray take-off angle was 40° above the horizontal (50° from the beam). Analytical totals from a larger set of data on this material are plotted in Fig. 10.7. The analytical total is seen to systematically decrease as the particle diameter decreases because of the action of the particle mass effect as electrons penetrate through the sides and bottom of the particle.

Table 10.3. Analysis of K-411 Particles on Bulk Carbon

Diameter (μm)	Mg	Si	Ca	Fe	O	Total
18	0.0959	0.255	0.111	0.106	0.426	0.995
5	0.0986	0.255	0.108	0.100	0.427	0.990
4	0.0966	0.244	0.102	0.0993	0.410	0.951
3	0.0858	0.219	0.0892	0.0894	0.367	0.851
2	0.0786	0.189	0.0753	0.0792	0.320	0.743
1.3	0.0463	0.113	0.0418	0.0453	0.189	0.435

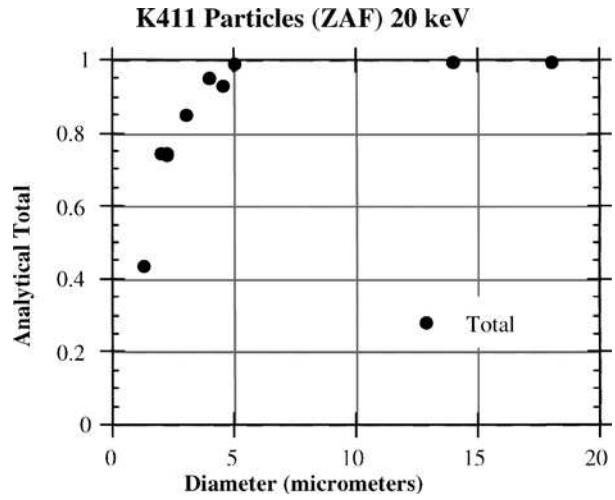


Figure 10.7. Analytical total for K-411 (analyzed against bulk standards with NIST *ZAF* matrix corrections) versus particle diameter.

10.3.5. Corrections for Particle Geometric Effects

10.3.5.1. The Consequences of Ignoring Particle Effects

The concentrations of the individual elements reported in Table 10.3 deviate by progressively greater factors as the particle diameter and the analytical total decrease. For example, for the particle of 1.3 μm diameter, the relative error is approximately 50%, calculated as

$$\text{Relative error} = 100\% \cdot (\text{measured} - \text{true})/\text{true}. \tag{10.4}$$

Errors of such magnitude are so severe that they generally render the analytical results meaningless for any realistic applications. Methods for correcting particle geometric effects follow.

10.3.5.2. Normalization

Simple normalization of the data, performed *after* the unnormalized analytical total has been inspected, is the simplest correction to apply for particle effects. Normalization consists in dividing each concentration value in the set C_i by the raw analytical total of all of the elements detected in that region:

$${}_nC_i = C_i \Big/ \sum_j C_j. \tag{10.5}$$

In order for normalization to yield sensible results, all elements present in the region of the specimen that is sampled by the measured x-ray spectrum must be considered in the raw analytical total. If oxygen is present, its concentration must be calculated either by direct means against an oxygen-containing standard or indirectly by the method of assumed cation stoichiometry. Similarly, other low-atomic-number elements such as carbon must be considered in the analytical total when present. Table 10.4 presents

Table 10.4. Errors Observed in Raw and Normalized Particle Analysis

	Mg	Si	Ca	Fe	O	Total
Bulk	0.0885	0.254	0.111	0.112	0.424	0.990
1.3-μm diam, raw	0.0463	0.113	0.0418	0.0453	0.189	0.435
Relative error (%)	−48	−56	−62	−60	−55	
1.3-μm diam, normalized	0.106	0.260	0.0961	0.104	0.434	
Relative error (%)	+20	+2.2	−13	−7	+2.5	

an example of a particle analysis from the data of Table 10.3 after normalization has been applied to the individual concentrations. The analytical errors are greatly reduced by simple normalization so that the resulting values are much closer to the true concentrations.

The effect of normalization is to force the analysis to be equivalent to a bulk target, that is, the particle mass effect, which generally reduces the intensity of all characteristic x-rays, is eliminated, at least to a first order. Normalization cannot compensate for the particle absorption and fluorescence effects, which are strongly energy-dependent, so that errors due to these effects are still embedded in the data. The actions of these errors can be recognized in Fig. 10.8, which shows a plot of the relative errors in the normalized concentrations as a function of particle diameter for the K-411 particle dataset. The errors are segregated according to the x-ray energy, with magnesium, which is measured with a low-energy x-ray (Mg $K\alpha = 1.25$ keV), showing positive relative errors (i.e., the concentration is overestimated) that increase as the particle diameter decreases. This overestimation occurs because excess Mg $K\alpha$ intensity escapes small-diameter particles due to the particle absorption effect. Elements in this system measured with high-energy photons, iron (Fe $K\alpha = 6.40$ keV) and calcium (Ca $K\alpha = 3.69$ keV), show negative relative errors (i.e., the concentration is underestimated). The errors increase as the particle diameter decreases because these photon energies are affected by the particle mass

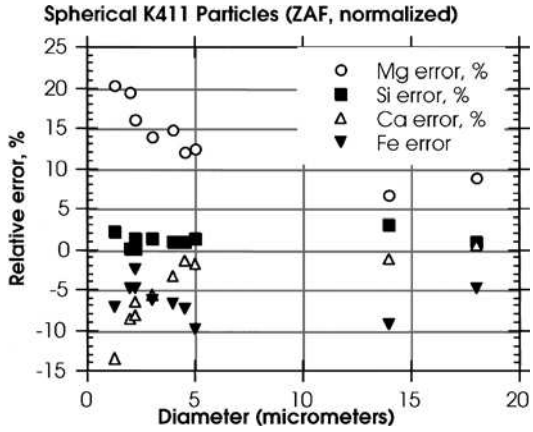


Figure 10.8. Analytical errors in normalized results versus particle diameter.

effect, but are not significantly affected by the particle absorption effect. In the normalization process, these elements are driven to negative relative errors in reaction to the positive relative errors of the low-photon-energy elements. The raw analytical total provides a rough guide to the magnitude of the relative errors, although the range of the relative errors will depend upon the specific details of the composition, the particle size, and the beam energy.

10.3.5.3. Critical Measurement Issues for Particles

Electron beam x-ray microanalysis of particles requires additional consideration of the measurement strategy beyond that necessary for flat, bulk specimen analysis. Specifically, the particle absorption effect can be so strong that it can severely compromise x-ray spectra measured on particles unless special care is taken. Consider the situation illustrated schematically in Fig. 10.9, which shows the dramatic differences in the absorption path when the beam is placed (1) on the front face (i.e., facing the EDS), (2) on top of the particle, and (3) on the back face of the particle. The effect of these differences in the absorption path can be seen directly in EDS spectra recorded on a 25-μm-diameter particle of IN-100 alloy, Fig. 10.10. The spectrum recorded on top of the particle and that recorded on the side of the particle facing the detector contain strong peaks for all of the constituents, but the spectrum recorded on the back side shows severe attenuation of the low-energy photons, with the Ni *L*, Al *K*, and Mo *L* peaks nearly completely eliminated, due to the long absorption path through the material. Note that, despite the loss of the low-energy peaks, x-ray continuum is still found in this same energy range, although it is sharply reduced in intensity compared to the top-position spectrum. The continuum in the low-photon-energy range actually does not arise from the particle, but rather represents electron scattering off the curved side of the particle onto the surrounding bulk carbon substrate. These remote scattered electrons generate C *K* characteristic peak and x-ray continuum photons. Despite the use of a collimator, the EDS has a wide acceptance area at the specimen plane, so that the remotely generated photons can be collected to a distance of hundreds

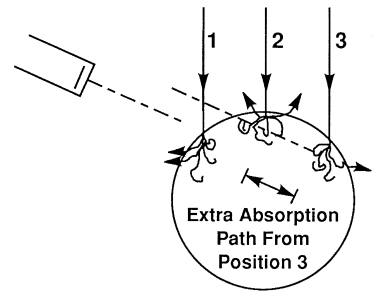


Figure 10.9. Schematic diagram showing the dramatic differences in the absorption path when the beam is placed (1) on the front face (i.e., facing the EDS), (2) on top of the particle, and (3) on the back face of the particle.

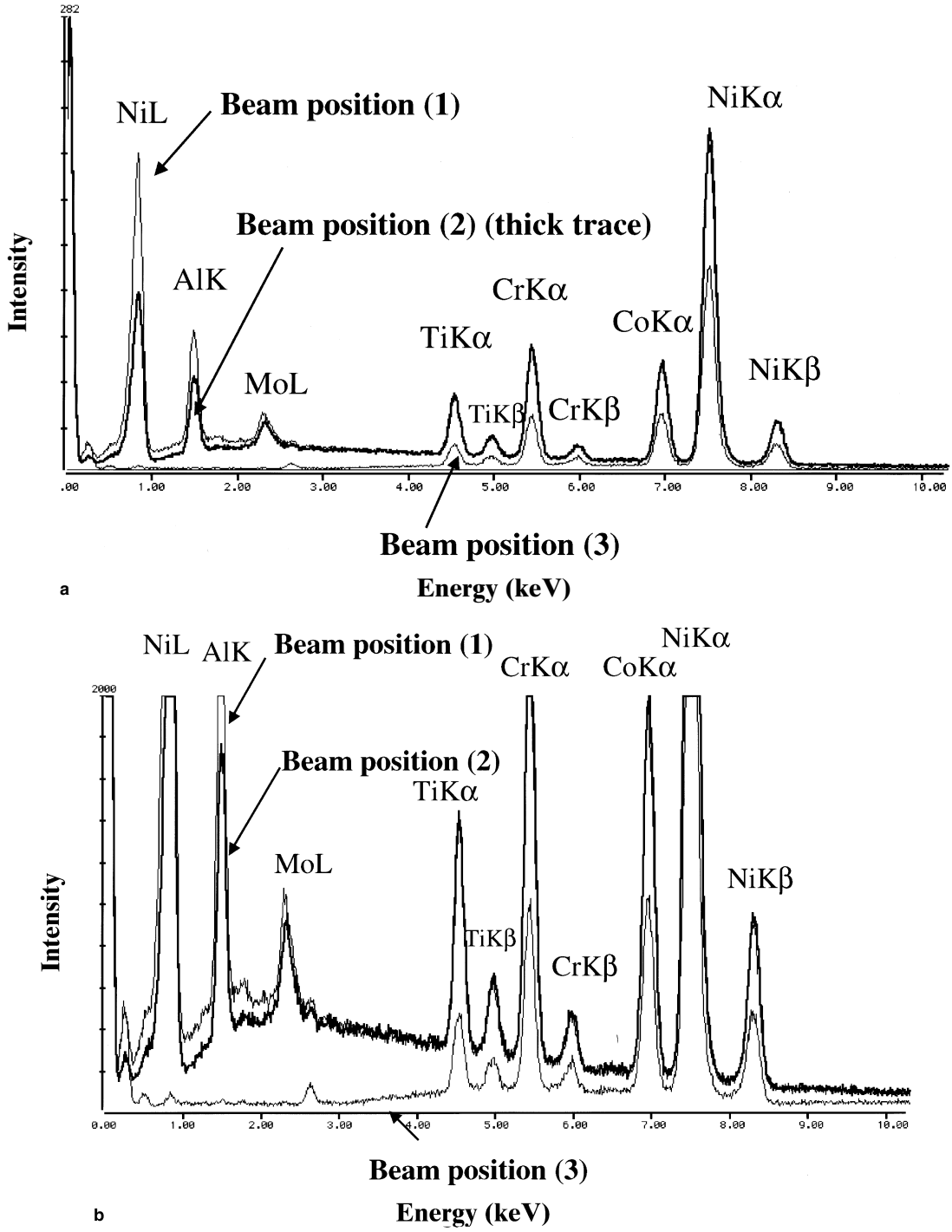


Figure 10.10. (a) Spectra of a 25-μm-diameter particle of alloy IN-100 obtained in the three positions shown in Fig. 10.9; (b) expansion of the vertical axis to show the shape of the low-photon-energy region of position 3, on the side opposite the EDS.

of micrometers to several millimeters from the beam impact point, creating a composite spectrum. The presence of the x-ray continuum in the low-energy portion of the spectrum of a small particle on a bulk substrate therefore must not be taken as a reliable indication that the absence of characteristic peaks in this region is valid.

To avoid collecting spectra on the back side of a particle, the analyst must always be aware of the geometric relationship of the EDS line-of-sight to the particle. A schematic diagram (top view) should be prepared showing the SEM chamber with the relative placement of the Everhart–Thornley (E-T) electron detector and the EDS x-ray detector, as depicted in Fig. 10.11a. This diagram should also include what motion of the mechanical stage, for example, *x*-axis drive, *y*-axis drive, or a combination of *x* and *y* drives, moves the specimen toward the EDS. When the SEM image prepared with the E-T detector is examined, the absolute position of the E-T detector can be immediately determined by changing the bias from positive to negative so as to collect only the highly directional backscattered electrons, as shown in Fig. 10.11b. Because the apparent illumination of an SEM image comes from an external detector, the highly directional negatively biased E-T position is obvious and this direction can be marked with high accuracy. With this direction established, the EDS position can be marked based upon its rigid angular relationship to the E-T detector. Note that there are actually two possibilities for the EDS position, 180 deg apart, because the scan may be a mirror of the actual situation. To correctly select the position, operate the mechanical stage motion that drives the specimen stage toward the EDS, and this will reveal which of the two positions is valid. Finally, in the normal SEM image (Fig. 10.11c), the E-T and EDS positions can be marked because the relative positions of the E-T and EDS detectors are fixed by the geometric relationship imposed by their placement on the SEM. If the SEM image is rotated to achieve a condition of apparent top illumination (see Chapter 4), the EDS position on the image must be relocated (Fig. 10.11d). For an independent confirmation, spectra can be recorded on top of the particle and on the sides facing toward the EDS and away. The shape of the low-energy portion of the EDS spectrum is a strong diagnostic, as shown in Fig. 10.10, to confirm that EDS spectra are being obtained with the minimum absorption path possible.

10.3.5.4. Advanced Quantitative Methods for Particles

Peak-to-Background Particle Method. The peak-to-background method (Small *et al.,* 1978, 1979; Statham and Pawley, 1978; Statham, 1979) is based on the observation that although the intensity of characteristic x-rays emitted from a particle is highly dependent on geometric effects, the peak-to-background ratio taken between the characteristic x-rays and the continuum x-rays *of the same energy* is much less sensitive. Although the two types of radiation are produced by different processes (inner shell ionization versus deceleration in a Coulombic field), both are generated in nearly the same volume. Both forms of radiation will scale similarly with the geometric mass effect, at least to a first order. Both types of radiation have a similar, although not identical, depth distribution, so the

absorption paths are similar. Because the energy of both forms of radiation is identical by definition, the geometric absorption effect will also scale for both. Therefore, the continuum intensity I_B can be used as an automatic internal normalization factor for the geometric effects. Consider the k value for a particle measured relative to a flat, bulk standard of the same compositon, $k_{\text{particle}} = I_{\text{particle}}/I_{\text{bulk}}$. Here I is the characteristic peak intensity corrected for continuum background at the same energy, $I = P - B$. The measured k_{particle} is a strong function of the particle size, but the ratio $(I_{\text{particle}}/I_{\text{B,particle}})/(I_{\text{bulk}}/I_{\text{B,bulk}})$ involving the background at the same photon energy is nearly independent of particle size, except for very small particles, where anisotropy of the continuum emission becomes significant (Newbury *et al.,* 1980). This experimental observation, which has been confirmed by theoretical calculations and Monte Carlo simulations, can be employed in several ways (Small *et al.,* 1978; Statham, 1979). One useful approach is to incorporate the following correction scheme into a conventional *ZAF* method (Small *et al.,* 1978, 1979). Given that

$$\frac{I_{\text{particle}}}{I_{\text{B,particle}}} = \frac{I_{\text{bulk}}}{I_{\text{B,bulk}}}, \tag{10.6}$$

a modified particle intensity I^*_{particle} can be calculated which is equivalent to the intensity that would be measured from a flat, bulk target of the same composition as the particle:

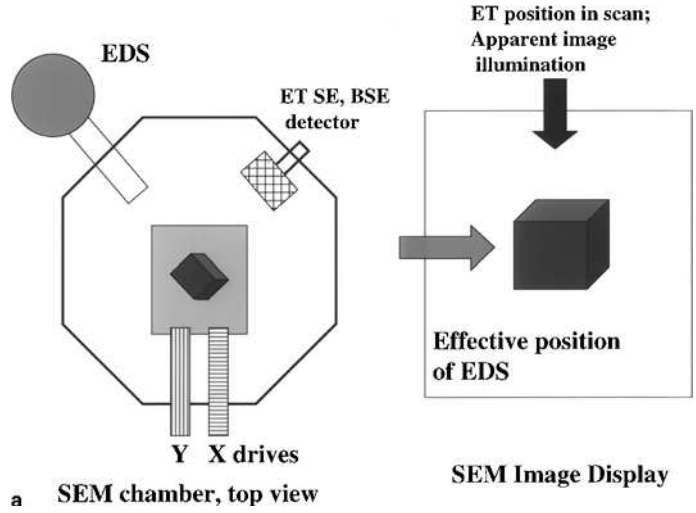
$$I^*_{\text{particle}} = I_{\text{bulk}} = I_{\text{particle}} \frac{I_{\text{B,bulk}}}{I_{\text{B,particle}}}. \tag{10.7}$$

To apply Eq. (10.7) for the analysis of an unknown particle, the quantities I_{particle} and $I_{\text{B,particle}}$ can be measured directly from the energy-dispersive x-ray spectrum. The term $I_{\text{B,bulk}}$ will in general not be available because a standard identical in composition is generally not available. However, as described in Chapter 9, an *estimate* of the concentrations of elements in the unknown is always available in the *ZAF* procedure, including the first step, where $C_i = k_i / \Sigma\, k$. The value of $I_{\text{B,bulk}}$ can therefore be estimated from the background measured on pure-element standards:

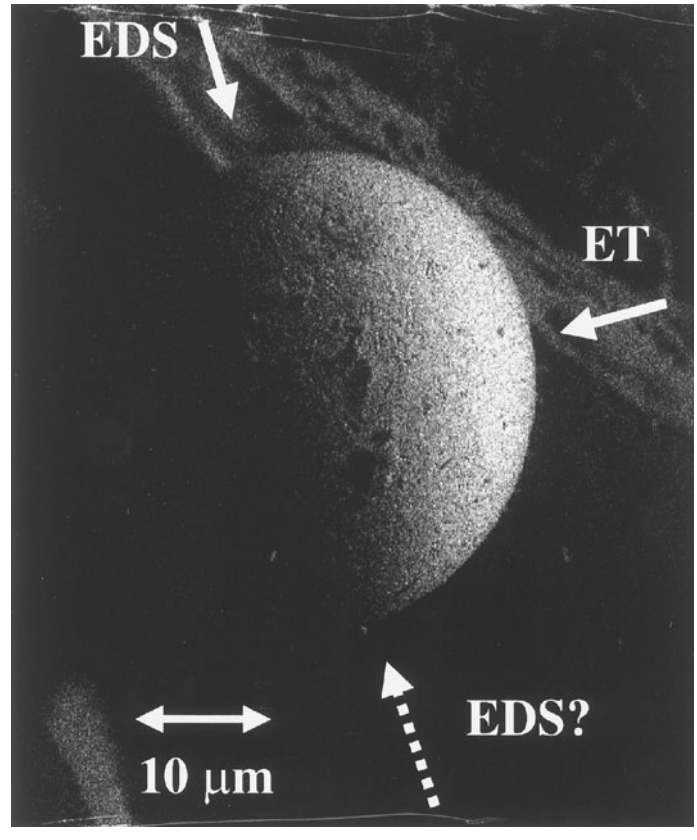
$$I_{\text{B,bulk}} = \sum_j C_j I_{j,\text{B,standard}}, \tag{10.8}$$

where $I_{j,\text{B,standard}}$ is the pure-element bremsstrahlung at the energy of interest and C_j is the concentration of that element. An example of an analysis of an IN-100 particle using the peak-to-background method is given in Table 10.5. The errors are reduced compared to the simple normalization procedure, especially for Al, which is highly absorbed when measured on the back side of the particle.

The special advantage of the peak-to-background method is that it can be applied to spectra obtained with a focused probe directed to a specific location on a particle. Thus, particles that have a chemically heterogeneous substructure can be directly studied. To be effective, the peak-to-background method requires spectra with high counts. Because the ratio of background intensities is used to scale the particle peak intensities, the

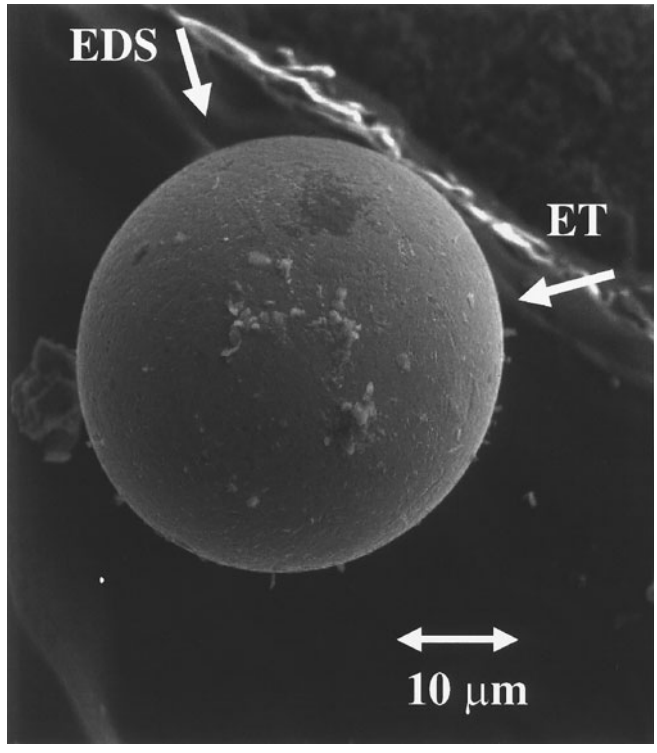


a

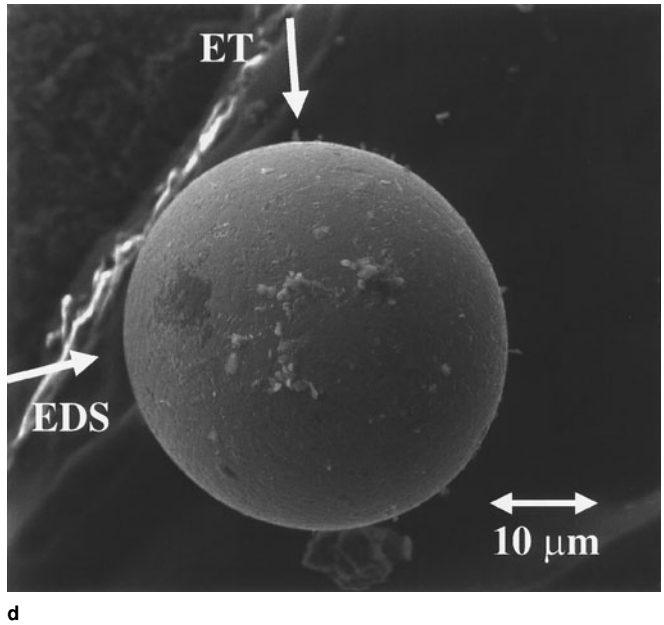


b

Figure 10.11. (a) Schematic diagram of detector placements in the SEM. (b) SEM image (negatively biased ET detector) of a particle revealing the absolute position of the ET detector. The EDS position is marked with the aid of (a); see text.



c



d

Figure 10.11. (*Continued*) (c) Normal SEM image (positively biased ET detector) with EDS position indicated. (d) Normal SEM image rotated to produce top lighting, with consequent necessary relocation of EDS axis (line-of-sight).

Table 10.5. Errors Observed with Peak-to-Background Corrections[a,b]

	C_{bulk}	C_{raw}	RE (%)	C_N	RE (%)	$C_{P/B}$	RE (%)
Al	0.0603	0.0201	−67	0.0241	−60	0.0552	−8
Mo	0.0353	0.0194	−45	0.0233	−34	0.0437	+24
Ti	0.0519	0.0406	−22	0.0487	−6	0.0480	−7
Cr	0.0965	0.0788	−18	0.0945	−2	0.0996	+3
Co	0.155	0.139	−11	0.166	+7	0.156	+1
Ni	0.601	0.536	−11	0.643	+7	0.598	−0.5

[a] Spherical particle: IN-100 alloy, 88 μm diameter, with beam placed 22 μm from the top center on the back side of particle.
[b] C_{raw}, raw concentration; C_N, concentration after normalization; RE, relative error.

statistical uncertainty in the background ratio propagates into the error in each concentration value in addition to the statistics of the characteristic peak. Even more importantly, the peak-to-background method depends on the background radiation originating in the excited volume of the specimen only, and not in the surrounding substrate. When particles become small, the penetration of the beam into the substrate means that continuum will continue to be produced, even if the substrate is a low-atomic-number element such as carbon. As noted above, the EDS has a large acceptance area at the specimen, even when fitted with a collimator. To minimize the extraneous background contributions, the particles should be mounted on a thin (approximately 10–20 nm) carbon film supported on a metal grid (typically copper, as used in the transmission electron microscope) and mounted over a blind hole drilled into a carbon block. The continuum contribution from such a thin film is negligible relative to particles as small as about 250 nm in diameter.

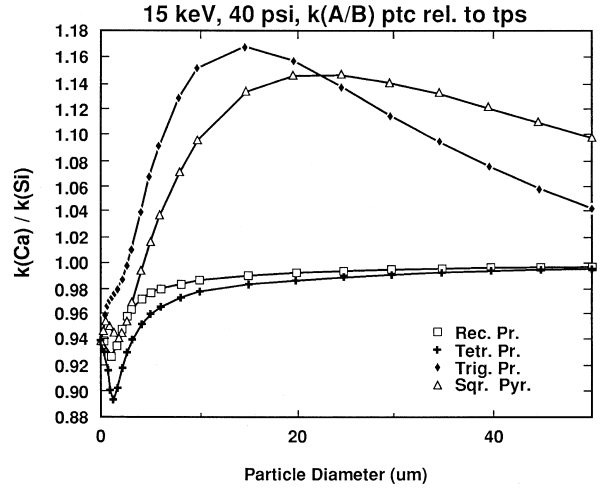


Figure 10.12. Particle shape factor analysis for the *k*-value ratio for Ca/Si as a function of particle size for four different particle shapes (Armstrong, 1991).

Table 10.6. Errors Observed in the Analysis of Particles by Analytic Shape Modeling Compared to Conventional Methods with Normalization (Averages of 140 Particles)[a]

Constituent	Actual concentration	Conventional analysis, normalized	Percentage relative error	Analytic particle analysis	Percentage relative error
Pyrite					
S	53.45	54.8	+2.5	53.5	+0.1
Fe	46.55	45.2	−2.9	46.5	−0.1
Olivine					
MgO	43.94	50.7	+15	43.8	−0.3
SiO_2	38.72	39.3	+1.5	38.9	+0.5
FeO	17.11	9.7	−43	17.1	0.

[a] Armstrong (1991).

Analytic Shape Factors for Particles. Armstrong and Buseck (1975) developed an analytic approach for calculating correction factors for particle geometric effects. Their method is based on bracketing the particle and overscanning during collection of the x-ray spectrum. Correction factors are developed by rigorously calculating the modification of the generation of x-rays by analytical modeling of the penetration of electrons into and propagation of x-rays through specific particle shapes. A suite of particle shapes was considered to approximate real particles, including the sphere, hemisphere, rectangular, tetragonal, cylindrical, and right triangular prisms, and the squared pyramid. The correction factors R_{AB} are calculated in terms of the ratio of predicted k factors for two particular elements k_A/k_B versus particle thickness along the electron beam. An example of a set of calculated factors for several particle shapes and for calcium and silicon is shown in Fig. 10.12. In practice, the analyst must judge which shape most closely corresponds to the unknown particle, and what its thickness is along the beam. The correction factors for all of the constituents are then determined from the appropriate curve for the shape most similar to the

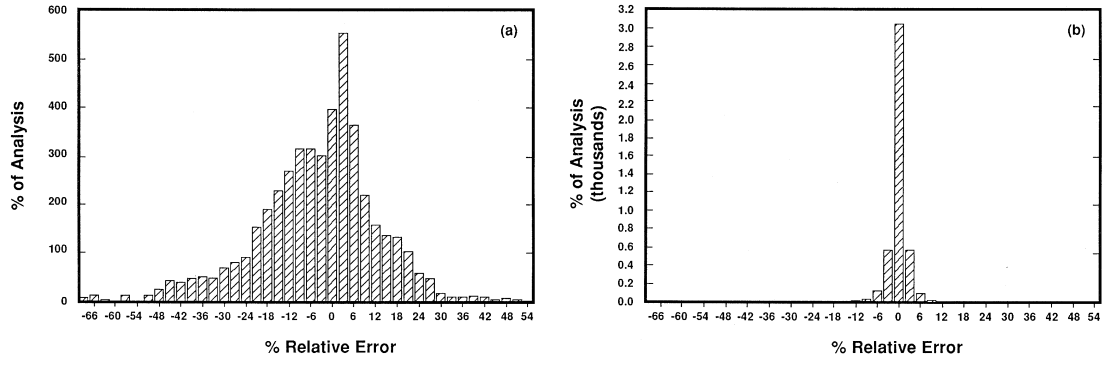


Figure 10.13. Comparison of error histograms of particle analyses with (a) conventional bulk *ZAF* procedures (with simple normalization) and (b) analytic modeling correction factors (Armstrong, 1991).

unknown. As listed in Table 10.6, and shown in the error distributions of Figs. 10.13a and 10.13b, the particle analytic modeling method can produce highly accurate results. The method has two main caveats. The use of overscanning obviously precludes the recognition of inhomogeneous particles, and the analyst's judgement is important in deciding on the appropriate shape/correction factor.

10.4. Rough Surfaces

10.4.1. Introduction

Rough surfaces represent the greatest challenge to quantitative x-ray microanalysis because of the poorly defined measurement conditions that frequently prevail. Similar to the situation for particles, there exist geometric effects that arise from alterations in the scattering of electrons and absorption of x-rays for rough surfaces as compared to flat, bulk standards:

1. A geometric mass effect originates from the increase in electron backscattering with the local tilt angle, which leads to a corresponding decrease in x-ray production, as shown in Fig. 10.14 for a flat iron surface at various tilts and a beam energy of 20 keV.
2. A geometric absorption effect arises because of the extreme variation in the x-ray path length that can be encountered in measuring an arbitrarily oriented surface. With complex surfaces, it is very easy to choose a location for analysis that is on the back side of a feature facing away from the EDS. As in the case of particles, it is even more critical to understand

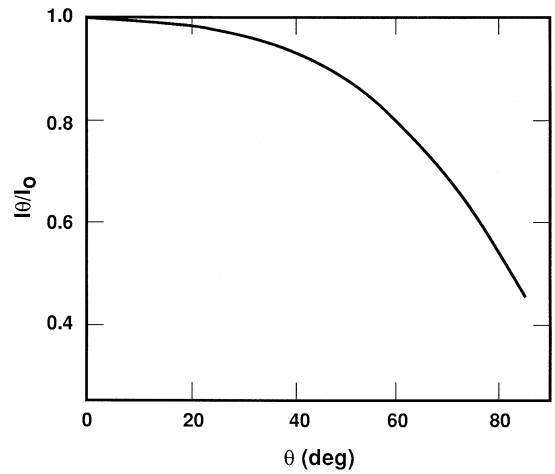


Figure 10.14. Monte Carlo electron trajectory simulation of the emission of Fe *K* x-rays from a tilted specimen of iron, with the intensity normalized to the value at 0° tilt. Beam energy = 20 keV.

the relation between the apparent illumination of the electron image and the relative position of the EDS so that suitable analytical locations can be selected.

3. Because we are dealing with bulk materials rather than small particles, the situation for secondary fluorescence is very similar to the ideal flat, bulk specimen case.

Moreover, the analyst must be much more concerned with the problem of stray radiation when analyzing rough surfaces. The EDS detector, even with a well-designed collimator, typically accepts x-rays from a region around the beam with a diameter of several millimeters. When isolated particles are analyzed, electrons scattering off the particle most likely encounter the substrate. With a well-chosen substrate material, typically carbon, the remote contribution to the spectrum will only be the C *K* peak and x-ray continuum. For a rough surface, the situation is much more severe, as illustrated in Fig. 10.15. Electrons that scatter off the analytical region

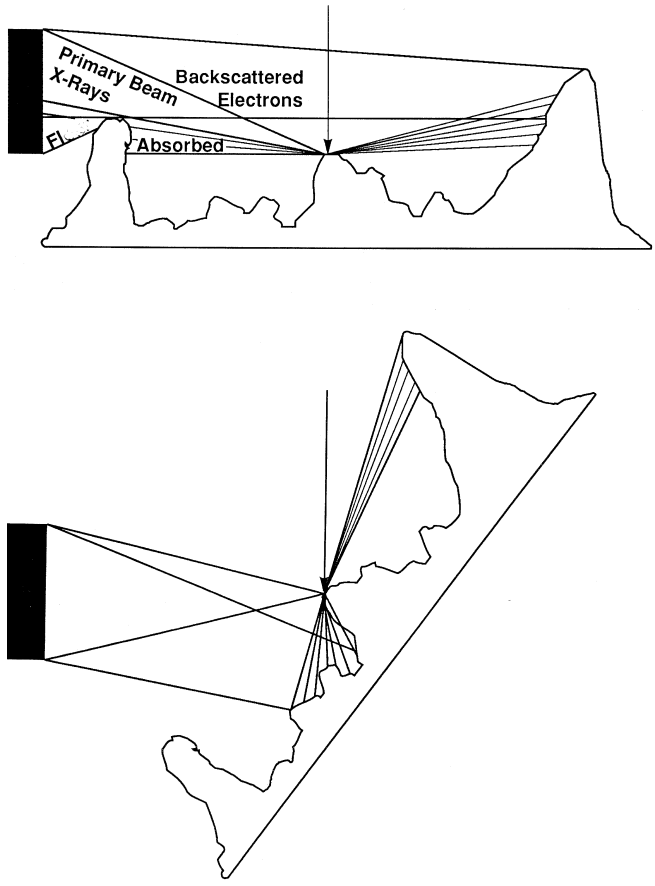


Figure 10.15. Schematic illustration of remote scattering effects from rough surfaces. Electron scattering paths are indicated. Fl, Remote fluorescence.

of interest as well as characteristic and continuum x-rays emitted from the analytical region have a high probability of striking nearby material, which may be of a different composition from the location selected with the beam. Because of this ill-defined scattering situation, the composite spectrum that results will be difficult to understand, even qualitatively. Although the major constituents represented in the spectrum are quite likely to be generated within the primary electron interaction volume, minor and trace constituents must always be considered suspect because such low-level peaks may easily arise from remote excitation. Furthermore, the analyzed region may be screened from the EDS by local specimen topography. Under such conditions, the low-energy photons will suffer relatively more absorption, in a manner similar to the situation illustrated for the large particle in Figs. 10.9 and 10.10. In extreme absorption conditions, major constituents with low atomic number (and consequently low photon energy) may be reduced to the apparent level of minor or even trace constituents. Careful inspection of the shape of the low-energy x-ray continuum is a valuable tool for assessing such pathological conditions. [As a further aid for the analysis of rough targets and large particles, a semiconductor backscattered electron detector can be mounted on or near the EDS detector (Hayashi and Bolon 1979). Because the apparent source of illumination in an SEM image comes from the detector, an image prepared with this BSE detector associated closely with the EDS will be subject to the same shadowing as the EDS x-ray collection. The rapid image collection possible with a BSE

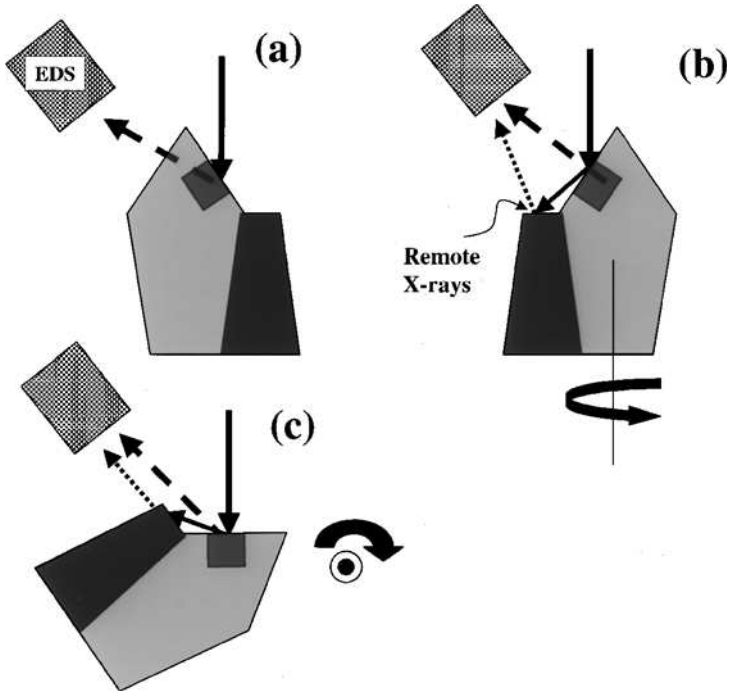


Figure 10.16. Reorientation of a rough specimen to optimize the spectrum quality: (a) Poor orientation relative to the EDS because of a long absorption path, which will selectively diminish low-energy photons. (b) Specimen rotated to bring the feature to face the EDS; the tilt still permits significant remote scattering to contribute to the spectrum. (c) Specimen tilted to bring the feature to the normal beam incidence condition; remote scattering is reduced because backscattering is minimized, but, depending on the exact surroundings, remote contributions to the spectrum are still possible.

detector makes this an efficient tool for assessing the x-ray measurement situation.

10.4.2. Rough Specimen Analysis Strategy

10.4.2.1. Reorientation

Because of the difficulties imposed by the problem of remote excitation, the analyst's first consideration for a suitable strategy should be to reorient the specimen if at all possible, as illustrated in Fig. 10.16, to optimize the spectrum quality. In Fig. 10.16a, the location selected by the beam placement for analysis is located on the back side of the specimen relative to the EDS, which has the dangerous consequence of creating a long absorption path that will selectively diminish low-energy photons. In Fig. 10.16b, the specimen has been rotated to bring the feature to face the EDS, giving a short absorption path. The high degree of tilt still permits significant backscattering to contribute remote excitation to the spectrum. Finally, in Fig. 10.16c, the specimen has been tilted to bring the feature to a normal beam incidence condition. This final orientation minimizes the backscattering coefficient, especially the BSE emerging at low angles to the surface, and consequently diminishes remote excitation. Nevertheless, it must be recognized that there is still sufficient backscattering from intermediate- and high-atomic-number materials to contribute to significant remote scattering. The level of significance of these remote contributions will depend on the exact nature of the specimen topography surrounding the analyzed location. Careful examination of the specimen topography, including stereo pairs, will be necessary for understanding the situation.

10.4.2.2. Normalization

When rough surfaces are analyzed by conventional *ZAF* analysis and flat, bulk standards, the analytical total will often deviate significantly from unity (100%), as illustrated in Table 10.7 (first column). The geometric effects are so severe in these examples, especially the absorption effect, that large errors still remain after simple normalization. The errors can be so large that even the sense of the ratio of the two constituents may be reversed!

10.4.2.3. Peak-to-Background Method

As in the case of particles, the peak-to-background method can compensate for geometric effects, at least to a first order. The examples given in Table 10.7 show quite remarkable improvements in the analytical errors compared to the raw errors and the errors after simple normalization. The P/B method is not a panacea because absorption can so severely attenuate the low-energy photons that they can be effectively lost from the

Table 10.7. Errors Observed in the Analysis of Rough Objects (Fracture Surfaces of Homogeneous Au–Cu Alloys) by Peak-to-Background Modified *ZAF* Method[a]

Constituent	Actual concentration	Conventional *ZAF*	Relative error (%)	Nomalized *ZAF*	Relative error (%)	*P/B ZAF*	Relative error (%)
Analysis 1							
Au	60.3	28.2	−53	49.6	18	58.0	−4
Cu	39.6	28.7	28	50.4	+27	44.0	+11
Analysis 2							
Au	60.3	1.08	−98	29.1	52	52.0	−14
Cu	39.6	2.63	−93	70.8	+79	41.0	+3
Analysis 3							
Au	80.1	95.8	+20	73.8	8	76.9	−4
Cu	19.8	34.0	+72	26.2	+32	19.1	−3

[a] Small *et al*. (1979).

spectrum, even when they represent a major constituent. Moreover, the P/B method is based upon measuring the local background produced in the same electron-excited region as the characteristic x-rays. If there is significant remote scattering, then the P/B method will be invalidated by the composite nature of the spectrum. Thus, specimen reorientation and careful study of the topography surrounding the analyzed region is the best strategy for optimizing the spectrum quality.

10.5. Beam-Sensitive Specimens (Biological, Polymeric)

10.5.1. Thin-Section Analysis

The x-ray microanalysis of biological and polymeric specimens is made difficult, and sometimes impossible, by several forms of radiation damage that are directly caused by the electron beam. At the beam energies used in the SEM (0.5–30 keV), it is possible for the kinetic energy of individual beam electrons to break and/or rearrange chemical bonds. At the beam currents used, particularly in the EPMA with WDS (typically 10–1000 nA), it is also possible to cause highly damaging temperature elevations. Indeed, when analyzing this class of specimen it should be assumed that significant mass loss will occur during the measurement at each point of the specimen. If all constituents were lost at the same rate, then simply normalizing the result would compensate. Unfortunately, the matrix constituents (principally carbon compounds and water) can be selectively lost, whereas the heavy elements of interest in biological microanalysis (e.g., P, S, Ca) remain in the bombarded region of the specimen and appear to be present at effectively higher concentration than existed in the original specimen. What is then required of any analytical procedure for biological and polymeric specimens is a mechanism to provide a meaningful

analysis under these conditions of a specimen that undergoes continuous change.

Marshall and Hall (1966) and Hall (1968) made the original suggestion that the x-ray continuum could serve as an internal standard for monitoring specimen changes. This assumption permitted development of the key procedure for beam-sensitive specimens that is used extensively in the biological community and is also applicable in many types of polymer analysis. This application marks the earliest use of the x-ray continuum as a tool (rather than simply a hindrance) for analysis, and that work forms the basis for the development of the peak-to-local background method described above for particles and rough surfaces. The technique was initially developed for applications in the high-beam-current EPMA, but the procedure works well in the SEM environment. A detailed derivation of the Marshall–Hall method from first principles will be found in the Enrichments, Chapter 10, Marshall–Hall, on the accompanying CD-ROM.

The Marshall–Hall method requires that several key conditions be met:

1. The specimen must be in the form of a thin section, where the condition of "thin" is satisfied when the incident beam penetrates with negligible energy loss. For an analytical beam energy of 10–30 keV, the energy loss passing through a section consisting of carbon approximately 100–200 nm in thickness will be less than 500 eV. This condition permits the beam energy to be treated as a constant, which is critical for the development of the correction formula. Biological specimens are thus usually analyzed in the form of thin sections cut to approximately 100-nm thickness by microtome. Polymers may also be analyzed when similarly prepared as thin sections. Such a specimen configuration also has a distinct advantage for improving the spatial resolution of the analysis compared to a bulk specimen. The analytical volume in such thin specimens is approximately the cylinder defined by the incident beam diameter and the section thickness, which is at least a factor of 10–100 smaller in linear dimensions than the equivalent bulk-specimen case at the same energy, shown in the polymer etching experiment in Fig. 3.4a and the Monte Carlo simulation of the interaction volume for carbon in Fig. 3.6a.

2. The matrix composition must be dominated by light elements, for example, C, H, N, O, whose contributions will form nearly all of the x-ray continuum and whose concentrations are reasonably well known for the specimen. Elements of analytical interest such as sulfur or calcium, the concentrations of which are unknown in the specimen, must only be present preferably as trace constituents (<0.01 mass fraction) so that their effect on the x-ray continuum can be neglected. When the concentration rises above the low end of the minor constituent range (~ 0.03 mass fraction or less), the analyte contribution to the continuum can no longer be ignored.

3. A standard must be available with a known concentration of the trace/minor analyte of interest and for which the complete composition of low-Z elements is also known. The closer the low–Z element composition

of the standard is to that of the unknown, the more accurate will be the results.

The detailed derivation yields the following general expression for the Marshall–Hall method:

$$\frac{I_{\text{ch}}}{I_{\text{cm}}} = c\frac{\dfrac{C_{\text{A}}}{A_{\text{A}}}}{\displaystyle\sum_i \left[C_i\left(\frac{Z_i^2}{A_i}\right)\log_e\left(1.166\frac{E_0}{J_i}\right)\right]}. \tag{10.9}$$

In this equation, I_{ch} is the characteristic intensity of the peak of interest, for example, S $K\alpha$ or Fe $K\alpha$, and I_{cm} is the continuum intensity of a continuum window of width ΔE placed somewhere in the high-energy portion of the spectrum, typically above 8 keV, so that absorption effects are negligible and only mass effects are important. C is the mass concentration, Z is the atomic number, and A is the atomic weight. The subscript A identifies a trace or minor analyte of interest (e.g., S, Fe, etc.) in the organic matrix, and the subscript i represents all elements in the electron-excited region. E_0 is the incident beam energy and J is the mean ionization energy, a function only of atomic number, as described in Eq. (3.3b).

Key condition 2 provides that the quantity $(C_i Z_i^2/A_i)$ in Eq. (10.9) for the biological or polymeric specimen to be analyzed is dominated by the low-Z constituents of the matrix. (Some representative values of $C_i Z_i^2/A_i$ are 3.67 for water, 3.01 for nylon, 3.08 for polycarbonate, and 3.28 for protein with S. Typically the range is between 2.8–3.8 for most biological and many polymeric materials.) The unknown contribution of the analyte C_{A} to the sum may be neglected when considering the specimen because C_{A} is low.

To perform a quantitative analysis, Eq. (10.9) is used in the following manner: A standard for which all elemental concentrations are known and which contains the analyte(s) of interest A is prepared as a thin cross section (satisfying key condition 3). This standard is measured under defined beam and spectrometer parameters to yield a characteristic-to-continuum ratio $I_{\text{A}}/I_{\text{cm}}$. This measured ratio $I_{\text{A}}/I_{\text{cm}}$ is set equal to the right side of Eq. (10.9). Because the target being irradiated is a reference standard, the atomic numbers Z_i, atomic weights A_i, and weight fractions C_i are known for all constituents, and the J_i values can be calculated as needed. The only unknown term is then the constant c in Eq. (10.9), which can now be determined by dividing the measured intensity ratio $I_{\text{A}}/I_{\text{cm}}$ by the calculated term. Next, under the same measurement conditions, the characteristic A intensity and the continuum intensity at the chosen energy are determined for the specimen location(s). Providing that the low-Z elements that form the matrix of the specimen are similar to the standard, or in the optimum case these concentrations are actually known for the specimen (or can be estimated from other information about the actual, localized material being irradiated by the electrons, and *not* some bulk property), then this value of c determined from the standard can be used to calculate the weight fraction of the analyte C_{A} for the specimen.

This basic theme can be extended and several analytes A, B, C, etc., can be analyzed simultaneously if a suitable standard or suite of standards containing the analytes is available. The method can be extended to higher concentrations, but the details of this extension are beyond the scope of this book; a full description and derivation can be found in Kitazawa *et al.* (1983). Most modern commercial computer x-ray analyzer systems have the Marshall–Hall procedure included in their suite of analysis tools. The Marshall–Hall procedure works well for thin specimens in the "conventional" analytical energy regime ($E_0 \geq 10$ keV) of the SEM and EPMA. The method will not work for specimens where the average atomic number is expected to vary significantly from one analysis point to another or relative to that of the standard. A bulk specimen where the beam-damaged region is not constrained by the dimensions of the thin section so that the region continues to change during electron bombardment also violates the key conditions. Consequently, many materials science applications for "soft" materials cannot be accommodated by the classic Marshall–Hall procedure procedure.

10.5.2. Bulk Biological and Organic Specimens

The quantitative procedures devised by Statham and Pawley (1978) and Small *et al.* (1978) for the analysis of particles and rough specimens, described earlier in this chapter, have been adapted to the analysis of bulk biological and organic samples (Roomans, 1981, 1988; Echlin, 1999). The method is based on the use of the ratio between the intensity of the characteristic and background x-rays defined as P/B, where P and B are measured at the same photon energy (or range of energies for EDS spectrometry). The rationale behind the development of the method is that because the characteristic and background x-rays are generated within nearly the same depth distribution, they are subject to the same composition-related absorption, secondary fluorescence, and atomic number effects. It is assumed that the percentage of characteristic x-rays absorbed by the sample is the same as the percentage of background x-rays that are absorbed. If we take the ratio P/B, the absorption factor A is no longer relevant because it has the same effect in the numerator as the denominator. Because backscattered electrons are lost due to changes in atomic number Z, there is a similar decrease in the efficiency of production of both peak and background. The reduced production affects both peak and background in a similar (although not identical) way, and this factor is also canceled out to first order when the ratio P/B is measured. Because nearly all biological and/or organic materials are of low atomic number ($Z^{\max} = 20$) the secondary fluorescence effect F is low and can be treated as a second-order correction.

These assumptions only hold true for homogeneous samples because the characteristic and background x-rays will vary with changes in the average atomic number of the sample. However, this is not considered to have any significant effect in cases where the P/B-ratio method is applied to fully hydrated specimens, which contain 85–90% water, or dried organic material containing a small amount of light element salts. The ratio of

peak area to the background immediately beneath the peak is relatively insensitive to small changes in surface geometry. However, the sample surface should be as smooth as is practicable because uneven fracture faces give unreliable x-ray data due to preferential masking and absorption.

Spectra are processed by the following procedure. The peaks in the spectra of the unknown and a standard of similar composition are fit by an appropriate procedure, such as multiple linear least squares, to determine the peak area for element i, P_i. The spectrum after peak fitting and subtraction is then examined again to determine the background intensity remaining in the peak region of interest, giving the corresponding B_i at the same photon energy. Once accurate P/B ratios are obtained, they can be used for quantitation in a number of ways. The P/B value for one element can be compared with the P/B value for another element in the same sample and to a first order:

$$(C_i/C_j) = h_{ij}[(P/B)_i/(P/B)_j], \tag{10.10}$$

where C_i and C_j are the percentage concentrations of elements i and j and h_{ij} is a correction factor, which can be obtained from measurements on a standard(s) of known composition very similar in composition to the unknown. Once h_{ij} has been empirically obtained for the element(s) of interest, measurements of the P/B ratio(s) from the unknown can be converted into concentration ratios. An advantage of taking the double ratio of (P/B) in Eq. (10.10) is the suppression of matrix effects, to first order.

Alternatively, the P/B value for an element in the sample can be compared with the P/B value for the same element in a standard, provided there is no significant difference in the matrix composition between sample and standard. If the mean atomic number of a given sample is always the same, the Kramers relationship shows that the background radiation is proportional to atomic number. If the mean atomic number of the sample is always the same, then

$$C_i = h_i(P/B)_i, \tag{10.11}$$

where h_i is a constant for each element. If it is possible to analyze all the elements and calculate the concentration of elements such as C, H, O, and N by stochiometry, then relative concentrations can be readily converted to absolute concentrations.

If there is a significant change in the composition between the unknown and standard(s), then a correction must be applied based upon the dependence of the continuum upon atomic number, following the original Marshall–Hall thin-section method:

$$(C_i/C_j) = h_{ij}[(P/B)_i/(P/B)_j]\left[\left(Z_i^2/A_i\right)\Big/\left(Z_j^2/A_j\right)\right], \tag{10.12}$$

where Z and A are the atomic number and weight, respectively.

The peak-to-background-ratio method has been found to be as efficient and accurate for biological materials as the more commonly used *ZAF*

algorithms, which have been designed primarily for analyzing nonbiological bulk samples. The paper by Echlin (1999) gives details of the accuracy and precision of the method as applied to hydrated and organic samples.

10.6. X-Ray Mapping

Scanning electron micrographs prepared with the backscattered electron signal provide direct information on compositional heterogeneity through the mechanism of atomic number contrast. Although such images are quite useful for characterizing microstructures, the compositional information that they contain is nonspecific. That is, although the sequence of the average atomic numbers of the phases present can be recognized and classified by the relative gray levels in the image, the image conveys no specific information to identify the elements present or the concentration levels.

The use of the x-ray signal(s) derived from a spectrometer(s) to prepare scanning images that contain element-specific information was developed very early in the history of the electron probe microanalyzer (Cosslett and Duncumb, 1956; Duncumb, 1957). In the method of "x-ray mapping," "area scanning," or "dot mapping," detection of an x-ray photon with the WDS spectrometer set to a peak position [later adapted to EDS with selected energy windows around the peak(s)] triggers the photographic CRT system to record a white "dot" at the position of the scanned beam. Despite the difficulties in obtaining satisfactory x-ray images because of the considerable time penalty imposed by the weak x-ray signal, x-ray mapping rapidly became one of the most popular modes of performing x-ray analysis. The reasons for this popularity arise from the specificity of the elemental information and the power of visual presentation of information. The human visual process is able to rapidly assimilate information presented in the form of images, and often a problem can be solved by merely knowing where in a microstructure a particular constituent is localized without detailed quantitative analysis.

With the advent of computer-controlled collection of digital x-ray data, the classic dot mapping procedure with direct film recording has been virtually eliminated from current usage. [However, because dot mapping may be the only option available with old SEM/EPMA systems (>15 years), a description of the dot mapping procedure will be found in the Enrichments section for Chapter 10 entitled x-ray Dot Mapping, on the CD-ROM.] Modern computer control systems for SEM/EPMA x-ray spectrometry systems provide various methods of x-ray mapping. These methods range from simply recording and displaying the total x-ray count in an energy window as a scaled gray level, to rigorous quantitative compositional mapping in which a complete compositional analysis is performed at each location in a matrix scan (Newbury *et al.,* 1990 a, b, 1991; Ingram *et al.,* 1998). In quantitative compositional mapping, the gray or color scale at a particular pixel displayed on an analog device such as a CRT is related to the actual concentration of each constituent.

10.6.1. Relative Merits of WDS and EDS for Mapping

Both WDS and EDS can be employed to collect data for x-ray mapping. A comparison of the operating characteristics of these spectrometers is presented in Chapter 7. For the task of x-ray mapping, each spectrometer has advantages and disadvantages.

1. *Count rate:* The WDS can operate at much higher photon counting rates than the EDS if the specimen can withstand the higher beam currents necessary to compensate for the relatively poor geometric and quantum efficiencies of the WDS compared to the EDS. If beam current is no problem, the WDS can process x-rays at 100 kHz or even higher. The EDS in high-resolution operation (e.g., 129 eV at Mn $K\alpha$) is limited to about 3 kHz, whereas with the shortest amplifier time constant the limiting count rate can reach about 25 kHz, but at the cost of a significant loss in the resolution. Moreover, this limiting count rate applies in the EDS to the entire range of x-ray energies excited. With the WDS the diffraction process off the crystal or synthetic multilayer selects a narrow energy band, $\sim$10 eV, which is placed across the peak of interest, excluding all other x-rays. Thus, a low-level peak can be isolated spectroscopically by the WDS, and provided the specimen can withstand high dose rates, the count rate for this relatively weak peak can be increased by increasing the beam current, whereas the proportionally much higher intensities from the matrix peaks are excluded from contributing to the WDS dead time. All analytical aspects of x-ray mapping, for example, the concentration contrast between phases, the lowest concentration that can be detected, and the spatial detail at a compositional boundary, are limited by the number of x-ray counts that can be accumulated in the time available as the beam dwells at each picture element ("pixel"). In conventional x-ray microanalysis at a fixed beam location, the analyst stops the beam at a selected location and records an x-ray spectrum (EDS) or peak and background counts (WDS) with time integration ranging from 1 to 100 s or more, depending on the sensitivity required. When x-ray mapping is implemented, the number of measured points is greatly increased. Usually a minimum digital density of at least 64×64 (total $= 4096$ pixels) is required, with high-quality maps requiring 256×256 (65,536 pixels) or more. The consequence of taking the weak x-ray signal and dividing it into many points to create a map is that only very poor statistics is obtained at any single pixel. This situation can only be improved by operating with long accumulation times for x-ray mapping, ranging from thousands of seconds to many hours (overnight runs). If the specimen can withstand the high beam currents necessary to increase the x-ray production, WDS mapping with multiple spectrometers each measuring a different element has significant advantages.

2. *Parallel collection:* EDS, by its energy-dispersive nature, effectively collects data in parallel for the entire range of photon energies excited by the primary beam. Thus, all constituents can be mapped in a single scan, a huge advantage with a complex specimen. Moreover, this parallel collection is extremely useful when mapping an unknown specimen where unexpected elements may be encountered (see spectrum mapping, in Section 10.6.4.2

below). Because of the diffraction process, WDS is restricted to a narrow energy window, ~10 eV. If more than one elemental species is to be mapped, then after the scan for element A is completed, the WDS must be retuned to the peak position of element B and the scan repeated at a large time penalty. Often the SEM/EPMA is equipped with several WDS units to enable an approach to at least partial parallel data collection.

3. *Resolution and peak-to-background ratio:* The x-ray intensity in any window of width ΔE consists of both characteristic and continuum x-rays. The resolution of the spectrometer is defined in terms of the energy width over which an x-ray peak with a natural width of approximately 1 eV is dispersed. The resolution of the WDS is 2–15 eV, depending on the diffractor, whereas the EDS resolution (129 eV at Mn $K\alpha$) varies between about 70 and 180 eV over the photon energy range from 250 to 10 keV. The chief consequences of the poorer resolution of the EDS are the occurrence of peak interferences and the lower peak-to-background ratio, which increases the influence of the x-ray continuum on the x-ray map, unless a background correction is applied (see compositional mapping, Section 10.6.4 below).

10.6.2. Digital Dot Mapping

In the simplest level of computer-controlled x-ray mapping, the count rate measured in an energy window defined for a particular constituent is used to define a threshold above which a white intensity level is assigned to that pixel location, indicating that the element is present. The map is thus equivalent to the analog dot map recorded on film, with the important exception that the data are stored and can be repeatedly examined and photographed. An example of digital dot maps recorded with EDS on Raney nickel (a Ni–Al alloy) for the Al K and Ni $K\alpha$ windows is shown in Fig. 10.17. A map for Fe $K\alpha$ that was recorded simultaneously was entirely blank because the signal did not exceed the preset threshold. Examination of these digital dot maps reveals the purely qualitative nature of the results.

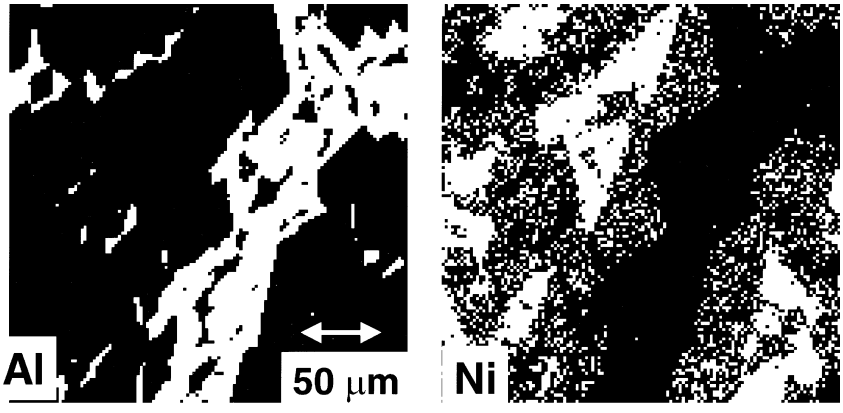


Figure 10.17. Digital dot maps recorded with EDS on Raney nickel (a Ni–Al alloy) for the Al K and Ni $K\alpha$.

There is no indication of the relative concentration at various locations within a single map, only the presence or absence of the constituent.

10.6.3. Gray-Scale Mapping

The most commonly used form of computer-controlled x-ray mapping is gray-scale mapping, in which the x-ray counts at each pixel are recorded

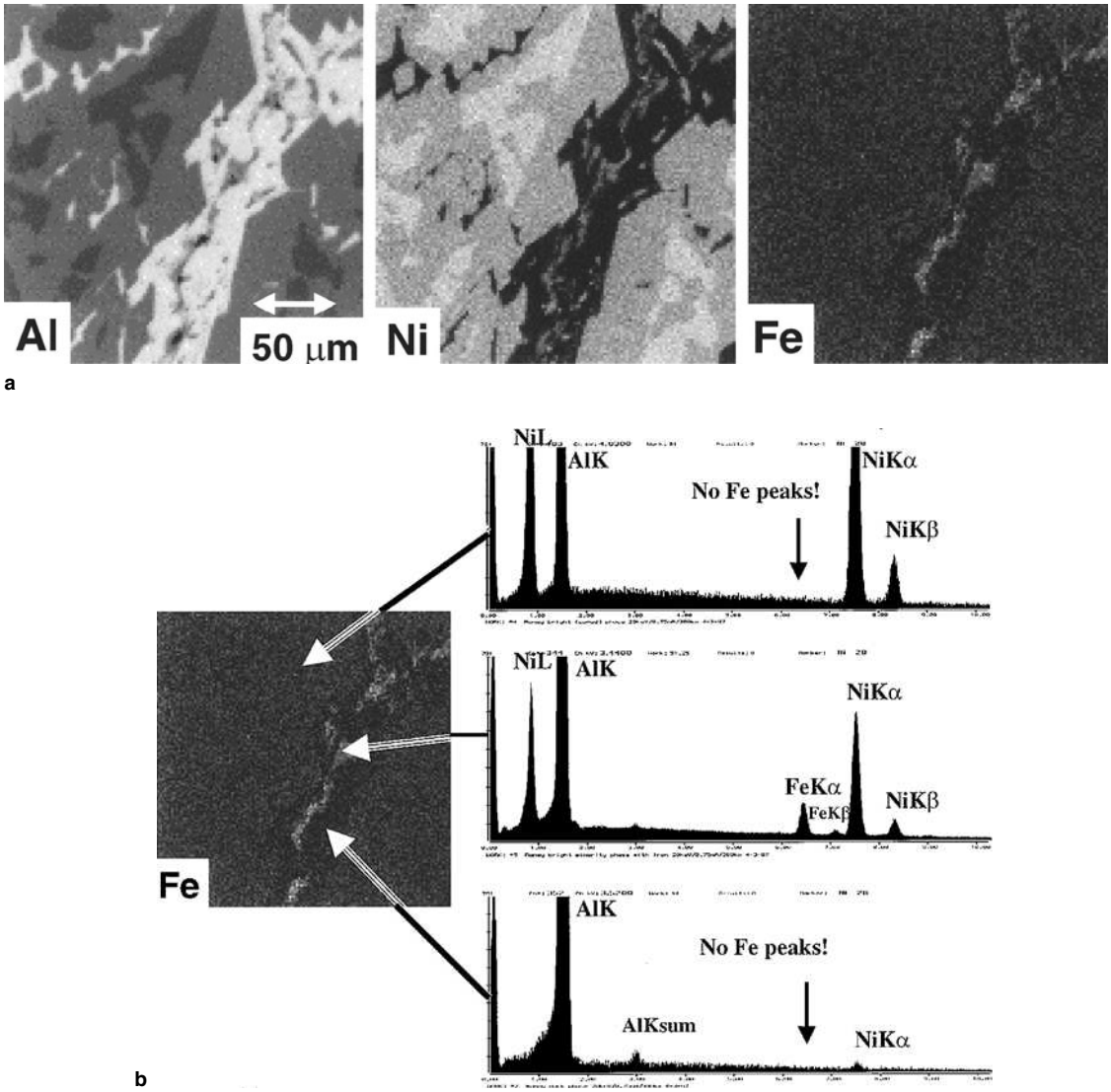


Figure 10.18. (a) Gray-scale (8-bit, 0–255 gray levels) x-ray maps of Raney nickel, same area as shown in Fig. 10.17. Al *K,* Ni *Kα,* and Fe *Kα*. (b) "Ground-truthing" an apparent minor/trace constituent. The localized iron islands are confirmed by the EDS spectra obtained at spot locations, but the apparent continuous iron distribution in the nickel-rich regions is shown by the EDS spectrum to be a continuum artifact.

and displayed as a numerically equivalent gray level. An example of gray-scale maps of the same field of view from Fig. 10.17 is shown in Fig. 10.18 for EDS mapping. In addition to the Al $K\alpha$ and Ni $K\alpha$ maps, a third map for Fe $K\alpha$ can now be obtained because the actual x-ray counts are being utilized and none of the information is being discarded. Comparing the gray-scale maps with the digital dot maps, the increase in information is immediately obvious. Examining the aluminum map, three distinct levels of intensity can be recognized, whereas for the nickel map, four distinct intensity levels can be seen. Because even low count rates are accepted in gray-scale mapping, a map for iron, a minor constituent, is also obtained. Examination of the iron map reveals three apparent levels of iron, but we must consider whether these apparent changes in concentration are real.

10.6.3.1. The Need for Scaling in Gray-Scale Mapping

For display purposes, computers typically operate with an 8-bit-deep intensity scale, which gives 256 discrete levels assigned numerical values from 0 to 255. Depending on the count rate and the dwell time, the intensity range recorded within a map may fall within this range, or the maximum may be greater than 255 counts. When the count exceeds 255, it is common practice with commercial x-ray mapping software systems to automatically scale ("autoscale") the intensity range for that map by applying a multiplicative scaling factor (255/maximum count) to all pixels, so that the scaled data does not exceed full white (255). When the count rate is sufficiently low that no pixel reaches 255 counts, autoscaling is again applied with the multiplicative factor of (255/maximum count), which has the action of raising the highest value pixel to full white (255). Maps scaled in this fashion usually produce good contrast spanning much of the 8-bit display if the concentration values sampled in the mapped area have a wide range.

In some situations, simple scaling based on the highest value pixel in the raw map data will not produce satisfactory display contrast. The situation is often like that illustrated in Fig. 10.19. The information of interest in an x-ray map may consist of an excursion in concentration of the element of interest above a common level throughout the image. In Fig. 10.19a the data range from a minimum of 100 counts (i.e., producing gray level 100) to a maximum of 138 counts (gray level 138). Because only a small portion of the full gray-scale range is used, the visibility of the features is poor. To improve visibility, the intensity produced by the constant composition value, 100, is subtracted from every pixel. The data now range from 0 to 38. This range would produce an equally unsatisfactory image that was very dark. Finally, each pixel is multiplied by 255/38 to increase the data span from 0 to 255, black to white, giving the improved feature visibility shown in Fig. 10.19b. Such an autoscaling operation is normally embedded in the software, and may not even be subject to the user's selection.

Additional processing operations can be applied to gray-scale operations to further improve feature visibility. The statistical fluctuations that

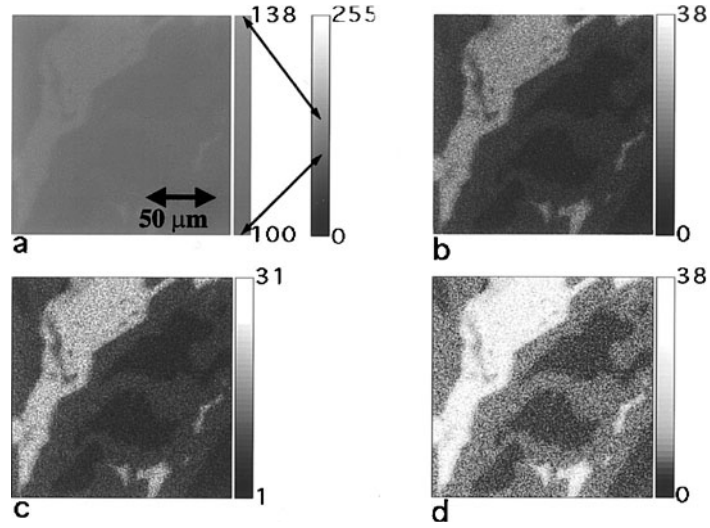


Figure 10.19. (a) Gray-scale display of an x-ray map with levels between 100 and 138; note poor contrast. (b) Improved contrast obtained by subtracting 100 from all pixels and scaling by 255/38. (c) Further contrast improvement by clipped scaling to reduce the influence of statistical noise. (d) Use of histogram equalization to enhance contrast.

naturally occur in x-ray count data will produce outliers that appear as black or white dots when the data set is scaled to span 0–255 gray levels. "Clipped scaling" (Pratt, 1978) consists in selecting a small fraction of the pixels at both ends of the gray scale and assigning minimum (0) or maximum (255) to them, while assigning the balance of the pixels the intensity range 0–255. In Fig. 10.19c, 0.5% of the pixels ($256 \times 256 \times 0.005 = 328$ pixels) at each end on the intensity range were set to black or white. After scaling the remaining pixels to span 0–255, the feature visibility is further improved.

Another popular image transformation is "histogram equalization" (Pratt, 1978; Russ, 1995). Consider interrogating an image one pixel at a time and binning according to the gray level, thus producing a histogram of the number of pixels versus gray level. Histogram equalization is a nonlinear scaling procedure in which the gray levels are redistributed so that the histogram of gray levels is as flat as possible, that is, in the modified histogram, each gray level is given an equal number of pixels. This operation has the effect of increasing the contrast for large features and decreasing the contrast for small features. In Fig. 10.19d the features are all large and the visibility is improved. Histogram normalization must be used with caution, however, because in the transformed image the relative gray levels no longer have the same relationship as in the original data.

Many other image processing functions that are typically used with SEM electron images can be useful with x-ray gray-scale maps. When manipulating images, the raw image data set should always be preserved in an unaltered state so that the microscopist/analyst can return to the beginning data if unsatisfactory transformations are encountered.

The Dark Side of Autoscaling. Autoscaling of the type illustrated in Fig. 10.19 makes the x-ray map information much more accessible to an observer. Unfortunately, in many computer-aided imaging systems the autoscaling operation is applied without retaining a record of exactly what was done to modify the raw data to produce the final x-ray map display. As a result, it becomes impossible to compare, in any quantitative sense, a map for element A to a map for element B obtained from the same area. White (255) may correspond to 77 counts for A and 297 counts for B. Even a series of maps of element A from different locations on the same specimen cannot be sensibly compared without knowing the different autoscaling multipliers applied to each x-ray map. Ideally, to obtain any real quantitative comparison, we must preserve the raw mapping data without any undocumented autoscaling function applied to it. In order to retain the unaltered data from high-count-rate measurements, the data should be stored in arrays that are at least 16 bit deep (maximum 65,536 counts) per pixel (24 bits would be preferable). Such 16-bits data arrays can be scaled to 8-bit arrays for display purposes, but the unaltered values must be retained for rigorous examination using a variety of image processing functions.

False X-Ray Map Contrast Due to the X-Ray Continuum. When data for constructing an x-ray map are collected by defining an energy window across a peak, the measured x-ray count will contain two contributions: characteristic x-rays of the peak and continuum x-rays of the background. Because the continuum intensity scales linearly with the average atomic number of the electron-excited region, the measured x-ray intensity within the window for element A may change as the beam samples different phases, despite there being no change in the concentration of A. In fact, A may even be entirely absent and counts will still be recorded. This false contrast in the A x-ray intensity is due entirely to changes in the x-ray continuum responding to changes in the concentrations of the other elements in the various phases. The susceptibility of EDS mapping to continuum artifacts increases as the analyte concentration decreases. At the major constituent level ($C > 0.1$ mass fraction), the continuum effects are negligible. However, great care must be taken when interpreting x-ray maps of minor ($0.01 \leq C \leq 0.1$ mass fraction) and trace constituents ($C < 0.01$ mass fraction). Any apparent compositional contrast at these levels must be confirmed ("ground-truthed") by examining the spectrum. In Fig. 10.18a, the map for iron contains three intensity levels, which apparently correspond to three different iron concentrations. The highest Fe level is associated with a series of inclusions, but there also appears to be a general increase in iron intensity in the nickel-rich regions compared to the aluminum-rich regions, which apparently contain the lowest level of iron. EDS spectra from these regions are shown in Fig. 10.18b, revealing that a peak for Fe $K\alpha$ is found in the inclusions, but no detectable peak can be seen in the nickel-rich or aluminum-rich regions. In these phases the apparent Fe concentration is purely a continuum-induced artifact. The analyst must thus completely interpret such x-ray maps for a client to prevent

misinterpretation. Because of its superior peak-to-background ratio, WDS is less susceptible to continuum-induced artifacts, but when the analyte concentration decreases far enough into the trace level, even WDS maps are vulnerable to continuum artifacts.

False X-Ray Map Contrast Due to Count Rate Effects. As described in Chapter 7, the EDS is subject to a paralyzable dead time; that is, during the dead time associated with processing a photon, the coincidence inspection circuit rejects other photons that arrive while the detector is busy. Initially, as the input count rate (which is proportional to the beam current) is increased, the output (measured count rate) also increases, but eventually the coincidence rejection leads to more photons being excluded than are being counted. The result of losing these photons is that the output count rate eventually reaches a peak, and any further increases in the input count rate will actually lead to a decrease in the output, as shown in the input–output curve in Fig. 10.20a. The shorter the time constant, the higher is the input count rate for the rollover, but the poorer is the resolution. If there is significant peak overlap with a short time constant in the spectrum of interest, then a long time constant may be necessary, which increases susceptibility to false contrast through the following mechanism: The input count rate axis can be considered both as a beam current axis (input x-ray count scales linearly with beam current) and also as a concentration axis. The input count rate for a constituent that varies from trace to major in the same map field will have a wide range. If the beam current is chosen to produce an output count rate that is just below the maximum on a phase that contains only a low concentration of an element A, then regions of the specimen with high concentrations of A will produce input counts that create a lower output count rate (i.e., the input count rate is beyond the maximum). This effect can actually create "inverted" contrast, where the higher concentrations produce lower intensities in the map. An example of such contrast inversion effects is shown in Fig. 10.20b.

10.6.4. Compositional Mapping

10.6.4.1. Principles of Compositional Mapping

The most rigorous form of x-ray mapping is compositional mapping, in which all of the correction procedures that are performed in a conventional quantitative analysis are applied at each pixel in the x-ray map (Newbury *et al.,* 1991).

Compositional mapping with EDS requires that the following steps be satisfied:

1. All major and minor constituents must be mapped from the area of interest so that the matrix correction procedure accurately reflects the composition at each pixel.
2. X-ray data must be collected on a live time (rather than clock time) basis at each pixel so that the measured count rates are fundamentally quantitative.

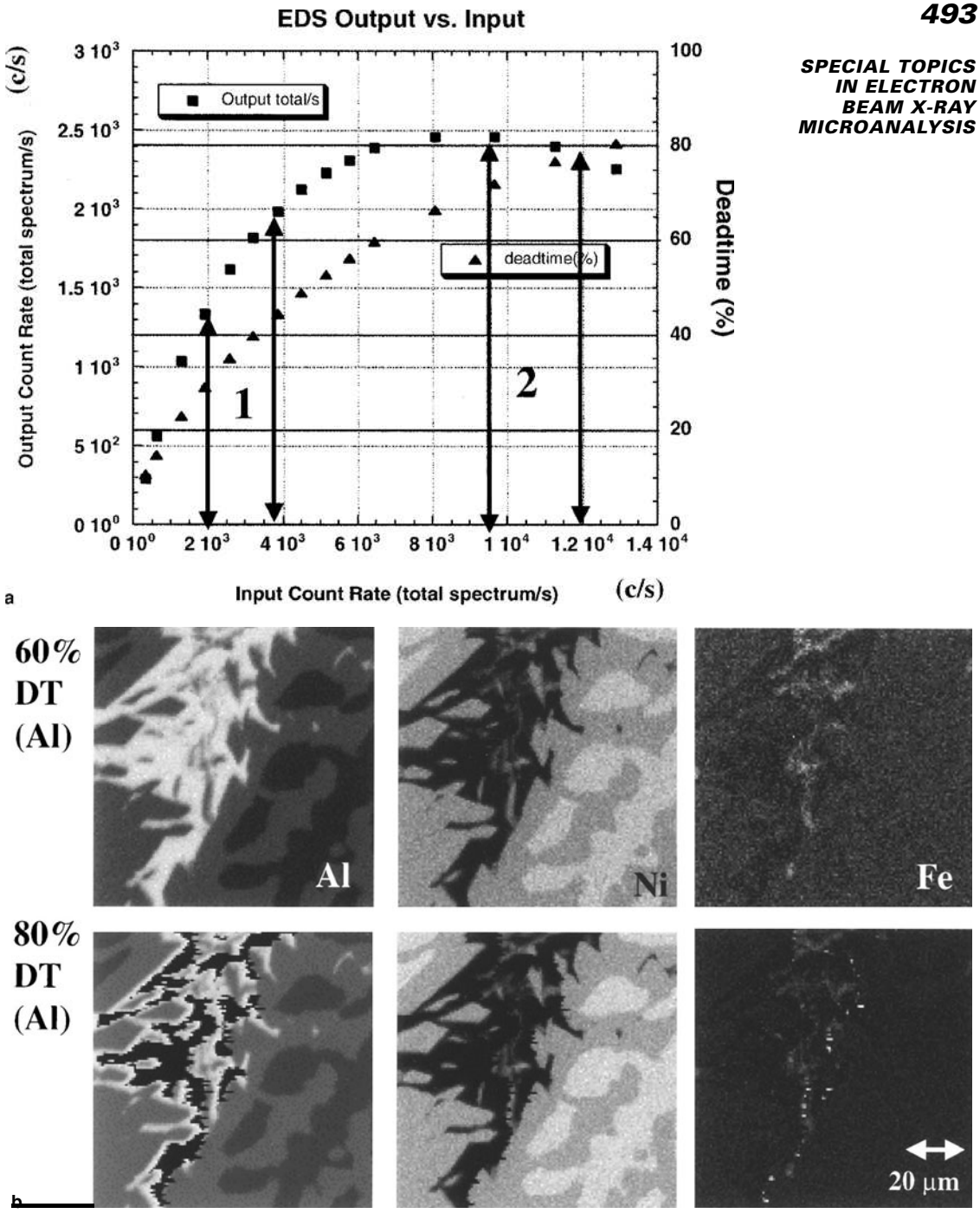


Figure 10.20. False x-ray map contrast due to count rate effects (a) Input–output response measured for an EDS system at optimum resolution. (b) X-ray maps for Al, Ni, and Fe taken with low dead time (upper) and high dead time (lower); note anomalous contrast in the aluminum signal.

3. Sufficient continuum regions must be recorded to permit accurate correction of the background, especially under the minor and trace constituent peaks, where the background has the greatest impact on accuracy.
4. For peaks that suffer interference, overlap corrections must be applied. This is especially important when a peak of a major constituent overlaps a minor or trace constituent.
5. The pixel count rate data must be standardized. By ratioing to the count rate measured on a bulk elemental or compound standard, a set of k-value maps is produced.
6. Matrix corrections are calculated [ZAF, $\phi(\rho z)$, or alpha factors] on a pixel-by-pixel basis by considering the complete set of k values measured at each pixel. Oxygen, if present, can be calculated directly with an appropriate standard or by assumed stoichiometry.
7. The resulting concentration maps should not be normalized, at least not initially, because the analytical total map of raw concentrations contains useful information. If an unexpected constituent appears somewhere in the map, the analytical total will be low for that region(s). Defects such as scratches or pits that introduce geometric effects, especially on absorption, will be detected in the analytical total map. After careful inspection for these effects, normalization can be applied, if desired.
8. Compositional mapping by WDS requires an additional step to compensate for the defocusing of the wavelength spectrometer when the x-ray source moves off the optic axis of the EPMA due to beam scanning. This effect is quite strong at low magnification (less than 1000×) and will dominate the true compositional contrast if not corrected.

An example of a compositional map for zinc in a Cu–Zn alloy is shown in Fig. 10.21. The gray-scale compositional map shows zinc at the grain boundaries of a copper polycrystal undergoing zinc diffusion along the grain boundaries. The superimposed trace shows the zinc concentration (0–4.0 wt%) along the locus *A–B*. The phenomenon observed is diffusion-induced grain boundary migration, and the compositional map clearly shows the oscillatory nature of the concentration of zinc behind the diffusion front. The great advantage of compositional mapping is the preservation of the raw data as well as the k-value and concentration maps. Each pixel can be separately examined, with full measurement statistics available. Advanced software systems for data mining such arrays of compositional maps are available (Bright, 1995).

10.6.4.2. Advanced Spectrum Collection Strategies for Compositional Mapping

Position–Energy Database Mapping. The major problem in x-ray mapping is that of obtaining adequate counts. The x-ray signal is a weak signal. Compared, for example, to backscattered electrons in the same solid angle, the x-ray signal is roughly 100–1000 times weaker. When analysis is performed at selected beam locations, the dwell time to accumulate an

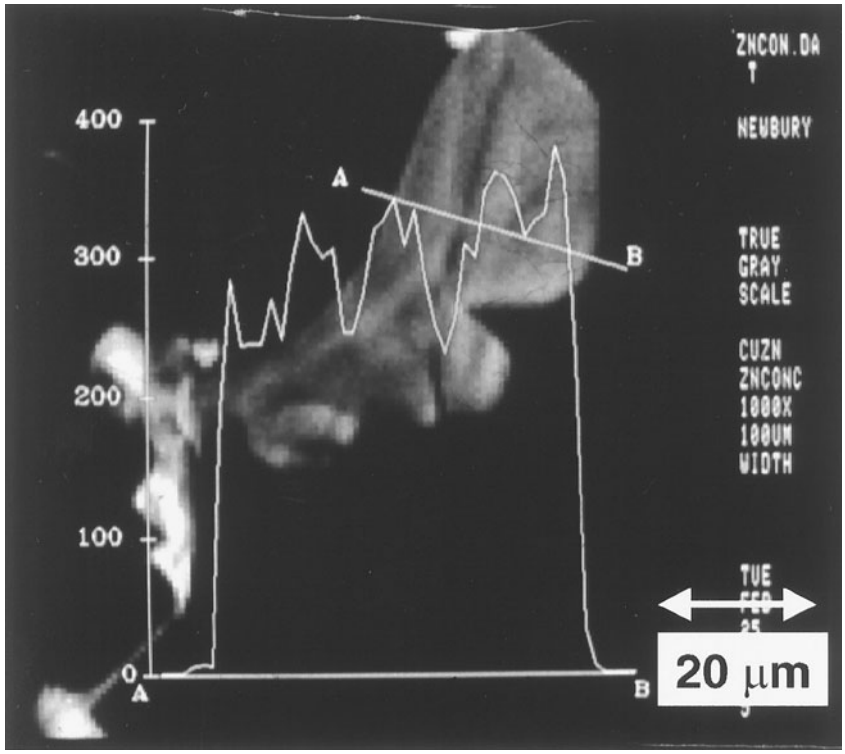


Figure 10.21. Compositional map for zinc in a copper polycrystal undergoing zinc diffusion along the grain boundaries. The superimposed trace shows the zinc concentration (0–4.0 wt%) along the locus A–B.

analytically useful EDS spectrum is typically 10–1000 s, depending on the level of statistical confidence or the limit of detection sought. For mapping in "real time" (patience limit about 1000 s), consider a dwell time of 50 ms per pixel. For a 128 × 128-pixel map, 50 ms per pixel results in a map time of 820 s, not considering dead time (a mapping run over a coffee break!). For a 256 × 256-pixel map, the 50-ms dwell results in a map of 3277 s without dead time (a mapping run over lunch!). An EDS system operating at the shortest time constant (yielding about 150- 180-eV resolution at Mn $K\alpha$) will produce a total spectrum count rate of about 25 kHz. For a composition consisting of O, Mg, Si, Ca, and Fe as major constituents, the spectrum at 20 keV is about 35% continuum. Thus, with 50 ms per pixel dwell, the EDS spectrum will contain about 800 counts total in all the characteristic peaks. Whereas this may be adequate to map major constituents, the mapping limit of detection under such conditions will cause some minor constituents and all trace constituents to be lost *at the single-pixel level*. The only way single-pixel-level statistics can be improved is to operate with much longer pixel dwell times, and therefore map accumulation times must run to several hours or more, such as can be achieved in overnight mapping runs under computer automation control.

To aid real-time analysis, a mapping strategy can be developed that permits the analyst to combine the counts from pixels that are obviously related, such as those that belong to a region having the same composition throughout. The BSE composition image is an excellent tool for making such judgements. Position–energy database mapping is an advanced EDS collection strategy that enables such pixel-based data manipulation as well as provides great flexibility for postcollection data manipulaton (Mott *et al.,* 1995). In this procedure, the beam is scanned repeatedly in a raster, a particular advantage for cold field emission gun systems that are subject to current drift. When an x-ray photon of any energy is detected, the energy and the x and y values of the pixel location of the beam are recorded in a four-dimensional database (x value, y value, E value, and count). When the map accumulation is complete, the analyst can select various software tools to mine the map database. Thus, the analyst could inspect the spectrum at any single pixel by specifying the x and y values, and the spectrum for that pixel would be constructed by finding and binning as an energy spectrum all photons that satisfy the pixel criterion. For a short map accumulation, such a spectrum will be extremely noisy and of use possibly only for major constituents. However, if a related range of contiguous pixels belonging to the same phase can be recognized in a BSE image, then a spectrum can be constructed for a range of pixels ($x \pm \Delta x$ and $y \pm \Delta y$), which is statistically much more robust. Maps corresponding to elemental peaks can be constructed by specifying the energy range $E \pm \Delta E$ that spans the peak. The database will then be inspected for all photons that satisfy that energy range condition, and the pixel locations recovered can be plotted to form a map. The key point here is the retention of all of the x-ray data for postcollection study. Thus, if a question arises about an unanticipated constituent, the analyst can return to the dataset and specify a map for the appropriate $E \pm \Delta E$ of that elemental peak.

Spectrum Imaging. The most complete mapping database that can possibly be created is that of a system that dwells at each pixel and records a complete spectrum, for example, x location and y location, by 2048 energy channels (10 eV wide), each channel of which is capable of recording counts to 16-bit depth (65,536 counts, at least). With the complete spectrum available at each pixel location, any software tool developed for conventional single-location analysis can be applied to the dataset. The remarkable advance of computer speed and data storage capacity makes this approach very promising. Currently, following the arguments presented in the previous section, such an approach is really only warranted at the individual pixel level for long mapping runs. With a sufficiently high dose per pixel, the individual pixel spectra are recorded with sufficient counts to be amenable to mathematical fitting procedures. Such high doses generally require overnight mapping runs with current EDS systems. The on-going, rapid development of the silicon drift detector (SDD) is most promising for performing spectrum imaging in real time. A large-solid-angle SDD system capable of count rates of 500 kHz or even higher means that statistically robust spectra can be obtained at single pixels in maps of 1000 s or

less. It should be possible to analyze this spectrum database qualitatively to automatically determine which elements are present above a defined threshold and categorize them as major–minor–trace by standardless analysis. The analyst will then be able to select elements for rigorous compositional map construction and further analysis. When this rapid spectrum imaging capability becomes widely available, compositional mapping will become a common and extremely powerful microanalysis tool.

10.6.5. The Use of Color in Analyzing and Presenting X-Ray Maps

We are frequently interested in understanding the relative spatial relationships of elemental constituents in a microstructure. X-ray maps provide the basis to do this, but it is difficult to understand the exact relationships from a series of gray-scale images placed side by side such as those in Fig. 10.18a. From the beginning of the development of x-ray mapping, color has been used as a critical tool for combining information from x-ray maps of different elements from the same area.

10.6.5.1. Primary Color Superposition

EDS mapping procedures result in images that are absolutely registered on a pixel basis because all information for a given pixel is simultaneously collected. Any three elemental maps from a set of maps can be simultaneously superimposed by assigning each map to one of the three primary colors (red, green, and blue) of the computer display. We will assume initially that the individual images are digital dot maps, so that the pixels are either black or white (Fig. 10.22a See color figure on the following page.). When such one-bit intensity-range (black or white) images are superimposed, the rules of color addition are simple: where coincidences of any two of the constituents occur, secondary colors will appear: red + green = yellow; red + blue = magenta; and green + blue = cyan; where all three constituents appear, the result is white. In the color overlay of Fig. 10.22a, the pure primary colors are detected, suggesting there is no mixing of the Al, Ni, and Fe.

When this same set of x-ray maps is first converted to continuous gray-scale images and then subjected to color overlay, first as binary combinations (Fig. 10.22b) and then as the ternary combination shown in Fig. 10.22c, more information is obtained. The Al–Ni binary reveals three distinct Al–Ni-containing phases, and the Al–Fe and Fe–Ni binaries and the Al–Fe–Ni ternary reveal that the Fe in the inclusions is associated with both Al and Ni. This theme of enhancing the individual images before color overlay can be exploited with many possible operations.

10.6.5.2. Pseudocolor Scales

High-Contrast Pseudocolor Scales. Pseudocolor scales are arrangements of display colors that are assigned to the input intensity levels in

a regular, functional manner (often referred to as "look-up tables,"LUT). For example, the output color scale may change every 32 units of input creating an 8-band scale for 8-bit (0–255) input levels. "Pseudo" refers to the artificial construction of the scale, which is generally created for a specific purpose. For example, the human visual process is much more sensitive to color changes than to intensity changes displayed as a gray scale. Thus, by creating a scale with alternate color bands selected for harsh color contrast, a scale is produced that is very sensitive to slight intensity changes. Such a pseudocolor scale is illustrated in Fig. 10.23 (See color figure following page 497.), where it is applied to show the form of the defocusing of a WDS (Cr $K\alpha$ scanned across a 250-μm-square field). The regular pattern of the intensity change off the axis of maximum transmission is easily visible in the bands of contrasting colors (augmented by an intensity wedge within each band). Although useful in this case, the high-contrast pseudocolor scale is not as effective when the intensity changes are more complex. Fig. 10.24 (See color figure following page 497.) shows the gray-scale map from Fig. 10.21 with a high-contrast pseudocolor scale applied. The pseudocolor map gives a strong impression of the complexity of the phenomenon, but it is difficult to make out the individual structures.

Thermal Scale. A highly effective pseudocolor scale for displaying compositional information is the thermal color scale, illustrated in Fig. 10.25 (See color figure following page 497.). The thermal scale approximates the sequence of colors given off by a black body as it is heated, starting from black, to deep red, red, cherry red, orange, yellow, to white. The thermal color sequence has a surprisingly strong visual recognition response for most observers, perhaps because of the experience gained in burning our fingers at the campfire for the last 500,000 years! When the thermal color scale is applied to the zinc grain boundary diffusion map shown in gray scale in Fig. 10.21, the compositional structures are easier to recognize in the thermal scale presentation in Fig. 10.25, and the viewer can more readily assign approximate compositional values based upon the colors.

Logarithmic Three-Band Scale. The thermal color scale provides a useful scale with which an observer can understand quantitative relationships within a single map. Often we wish to compare maps of different constituents on a quantitative basis. The logarithmic three-band scale permits such direct intercomparisons among any number of maps (Newbury and Bright, 1999). First, the x-ray mapping data must be corrected to produce normalized k-value maps (background corrected and standardization applied) or rigorous compositional maps (matrix corrections applied to the set of k-values at each pixel). The concentration range for each map is divided into three color bands, each of which has a logarthimic intensity wedge, with the following definitions:

- Trace ($C < 0.01$ mass fraction, or 1%), deep blue to blue pastel
- Minor ($0.01 \leq C \leq 0.1$, 1–10%), deep green to green pastel
- Major ($C > 0.1$, or 10–100%), deep red to pink

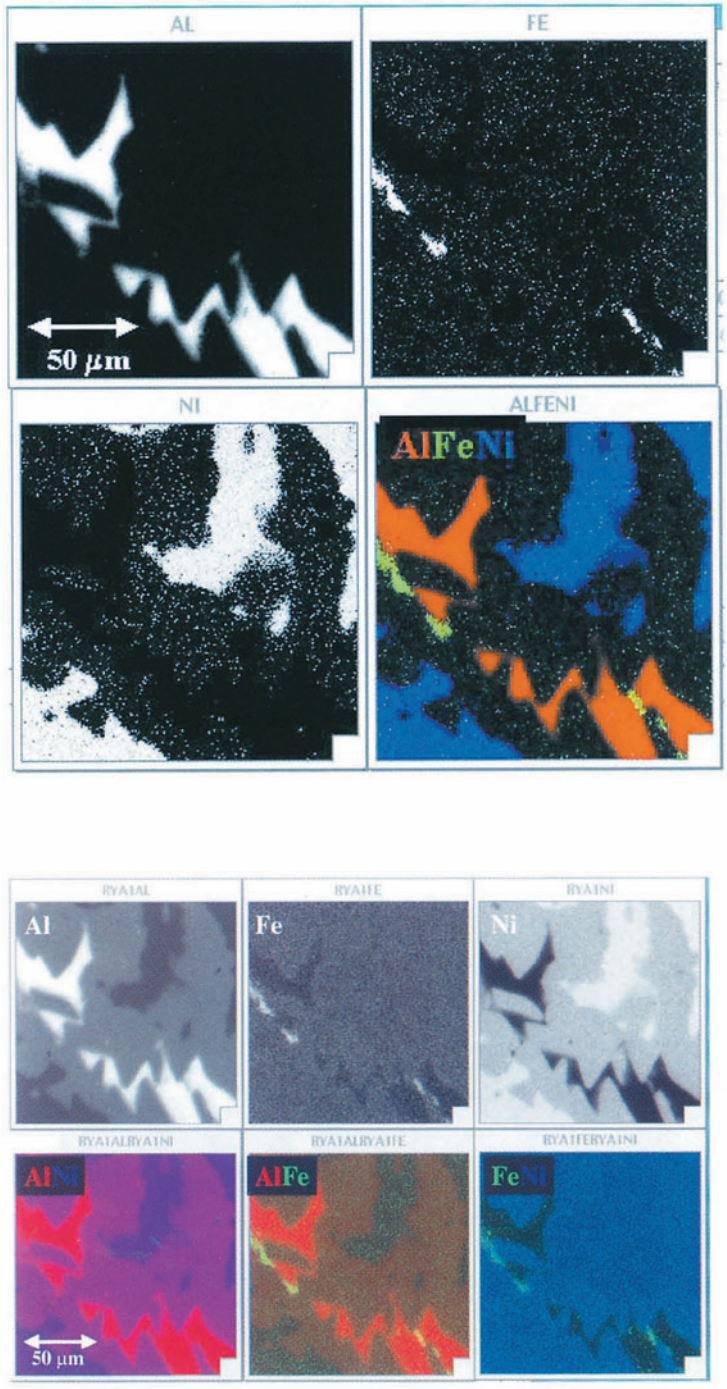


Figure 10.22. (a) Primary color overlay (ternary) of digital dot maps (full white). Al = red; Fe = green; Ni = blue. (b) Primary color overlay (binary combinations) of gray-scale maps Al = red; Fe = green; Ni = blue. (c) Primary color overlay (ternary) of gray-scale maps. Al = red; Fe = green; Ni = blue.

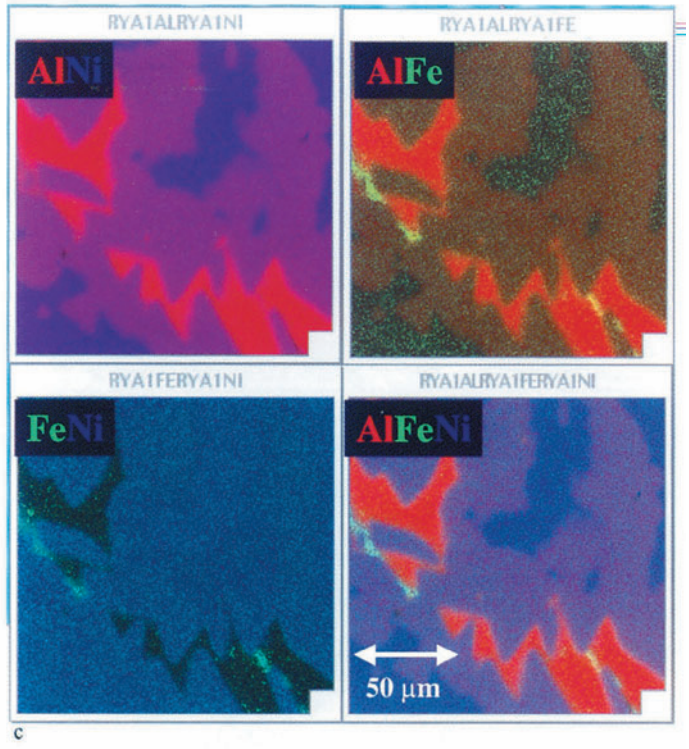


Figure 10.22. *(Continued)*

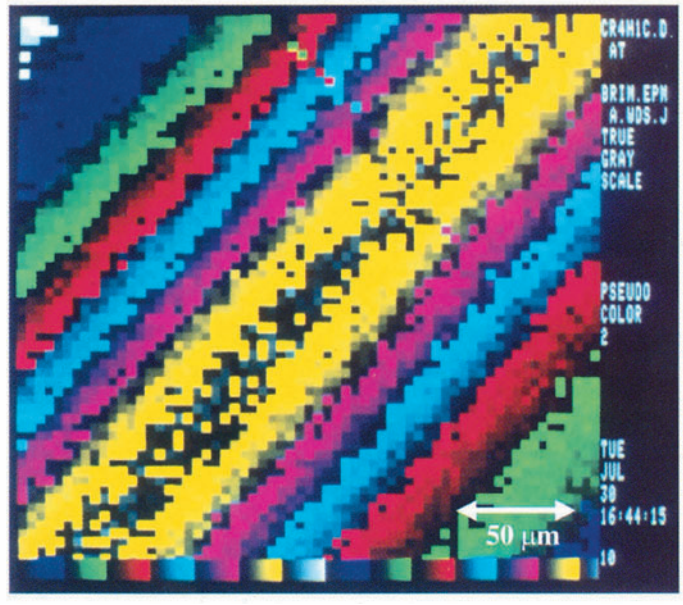


Figure 10.23. High-contrast pseudocolor scale applied to reveal the defocus pattern of a WDS Cr $K\alpha$, 250 × 250-μm scan.

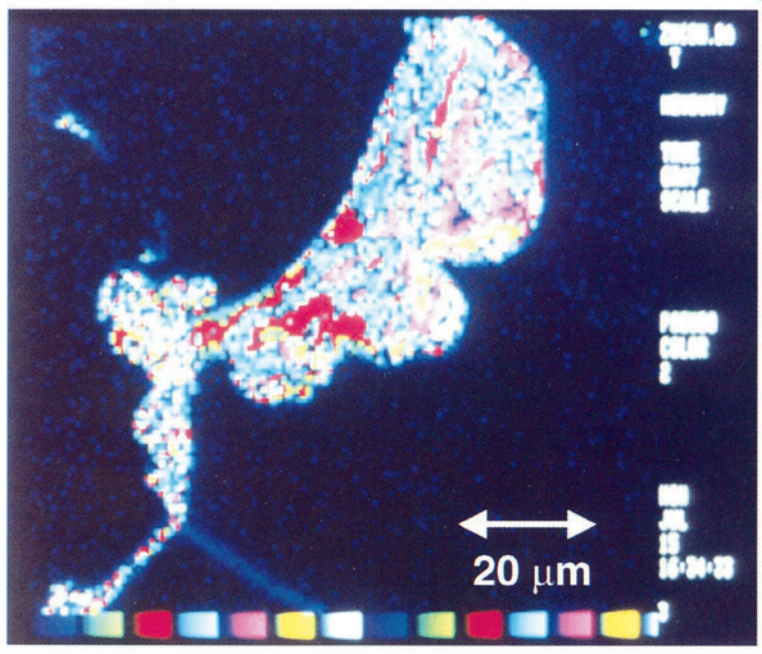


Figure 10.24. High-contrast pseudocolor scale applied to the gray-scale map of Fig. 10.21. Scale limits, 0–10 wt% zinc.

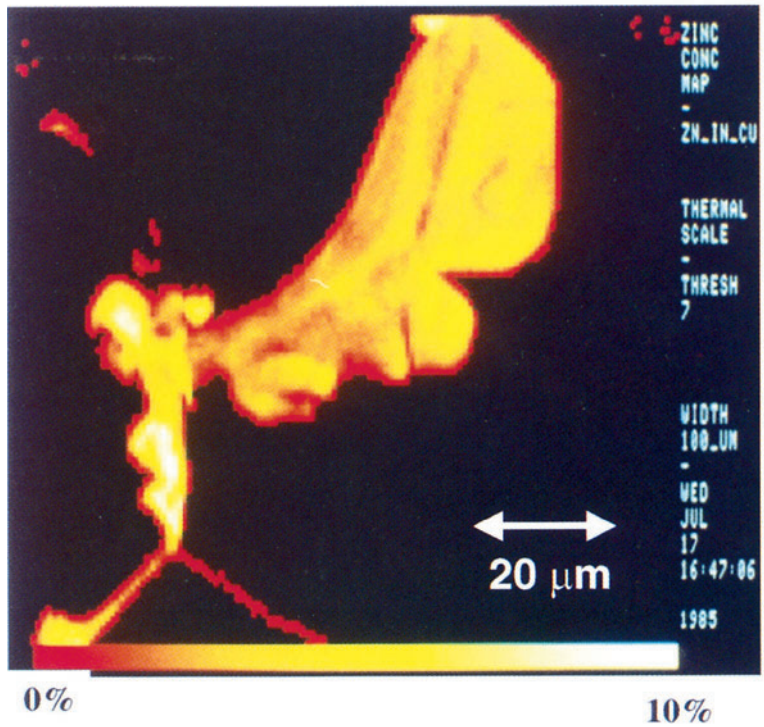


Figure 10.25. Thermal pseudocolor scale applied to the gray-scale map of Fig. 10.21. Scale limits, 0–10 wt% zinc.

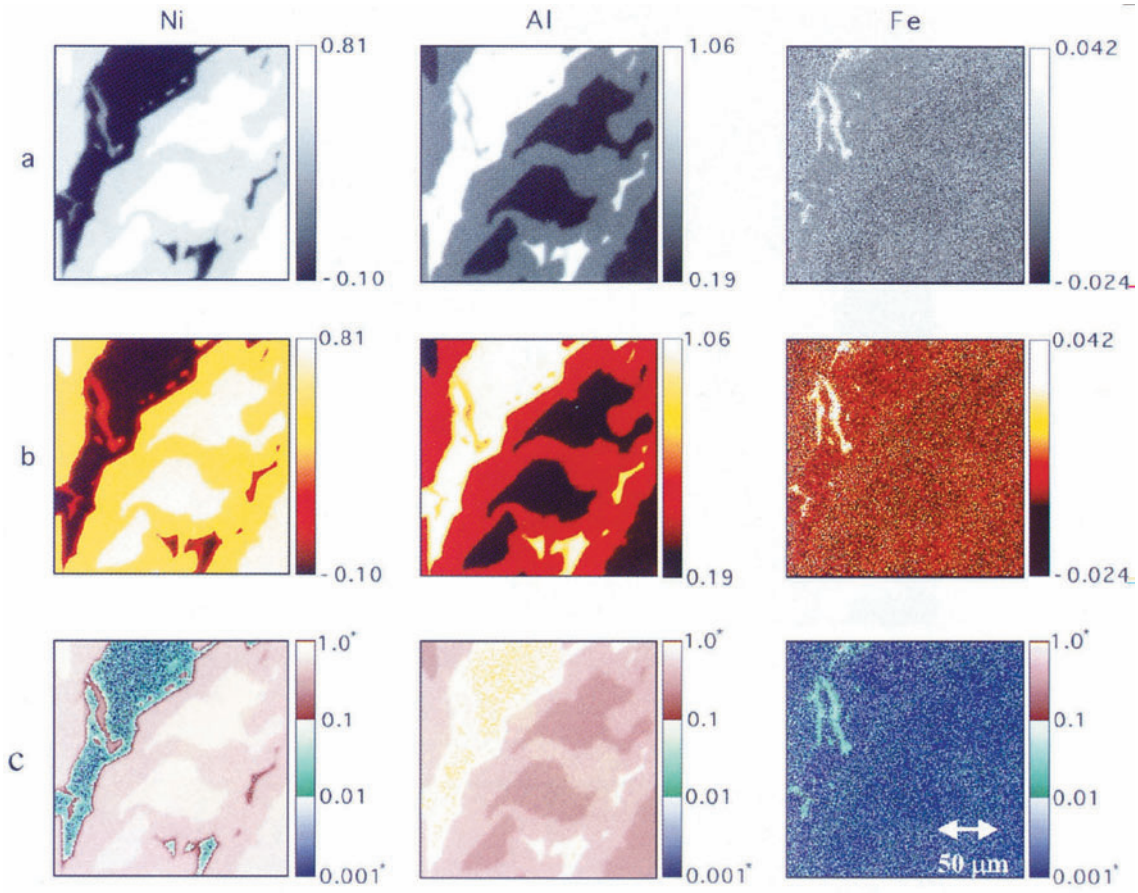


Figure 10.26. Three presentations of x-ray maps for Raney nickel: Al, Fe, Ni; (a) gray scale; (b) thermal scale; (c) logarithmic three-band color scale.

When this scale is applied to maps, as shown in Fig. 10.26 (See color figure following page 497.), where the conventional gray scale, the thermal color scale, and the logarithmic three-band scale are compared, the observer can immediately recognize where each element exists at a major, minor, or trace level within the field of view. This array of images of the same dataset demonstrates the value in having and comparing several display tools for maps.

10.7. Light Element Analysis

Analysis of the light elements, Be (0.109 keV), B (0.185 keV), C (0.282 keV), N (0.392 keV), O (0.523 keV), and F (0.677 keV), is difficult because of their low photon energies. Low photon energy leads to many complications, including high absorption in the specimen and in the components of the spectrometer(s), chemical state effects detectable with WDS and with the microcalorimeter EDS that change the peak position and peak shape, and peak proximity to the fundamental electronic noise of the detection/amplification system (EDS and WDS). Moreover, the analyst must deal with the extreme simplicity of the light element spectra, each element having only one *K*-shell peak. At first this simplicity would seem to be an advantage, but a negative consequence of having only one peak is loss of the flexibility that the analyst has with higher atomic number elements with their rich spectra, which often permit selection of an alternate peak when interference occurs in complex specimens.

10.7.1. Optimization of Light Element X-Ray Generation

The *K*-edge energies of the light elements are so low that an overvoltage of 2–5, which provides efficient excitation, can be obtained with a beam energy as low as 2 keV. For complex specimens, higher beam energies will be needed to also excite the more energetic *K*, *L*, and/or *M* edges of the intermediate and heavy elements. If the light elements are to be measured simultaneously and a useful limit of detection is to be obtained, the lowest possible beam energy must be selected consistent with providing sufficient excitation for all elements. This measurement strategy is required because detection of the light elements is severely compromised by the strong self-absorption these low-energy x-rays undergo in the specimen. Absorption depends on the mass absorption coefficient, which in turn depends upon composition and on the path length out of the specimen to the detector. As the beam energy is increased, the x-ray range increases, proportional to $(E_0^{1.67} - E_c^{1.67})$. Although more light element x-ray intensity is *generated* within the specimen due to higher overvoltage, x-ray absorption increases more rapidly due to the greater depth over which these additional x-rays are produced. Thus, if we measure the x-ray intensity for oxygen in various matrices, shown in Fig. 10.27, starting at low beam energy (2 keV) and increasing in steps to 30 keV, the O *K* intensity is seen to initially rise, reach a peak, and then decrease with further beam energy increases. The beam energy for the peak depends on the matrix through the mass absorption

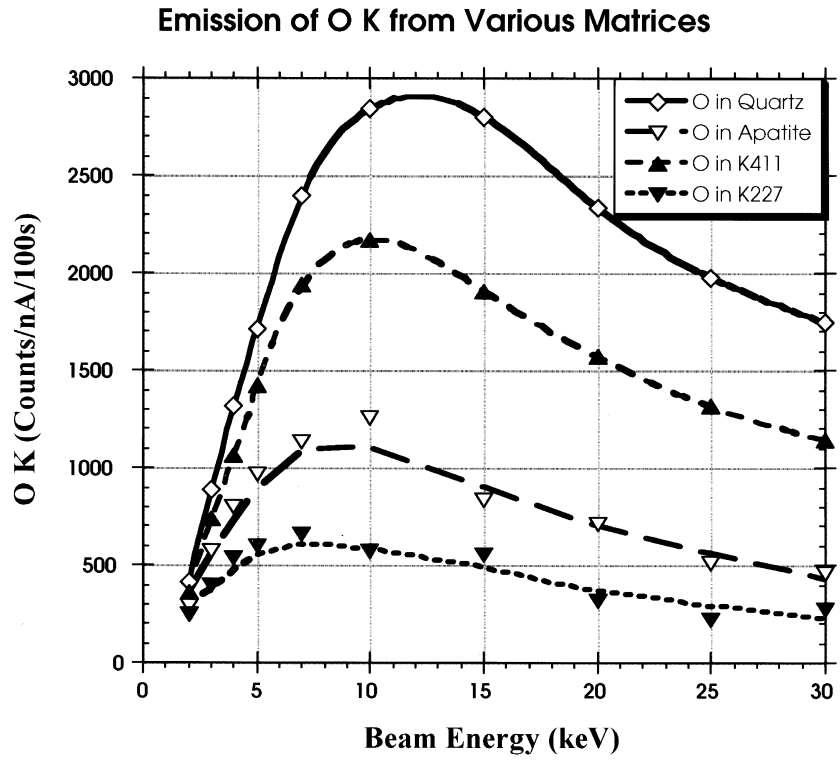


Figure 10.27. Plot of the intensity of O *K* (normalized for beam current) as a function of beam energy for quartz (SiO_2), apatite [$Ca_5F(PO_4)_3$], K-411 mineral glass (Mg 0.0885; Si 0.253; Ca 0.111; Fe 0.112; O 0.424), and K-227 silica–lead glass (Si 0.0935; Pb 0.743; O 0.164). Si EDS (129 eV at Mn $K\alpha$).

coefficient. The lead in glass K-227 much more strongly absorbs O *K* than does silicon in quartz, so the peak occurs at lower energy.

This absorption effect on light element x-rays can be so strong as to completely mislead the analyst if only high beam energy is used to examine the specimen. Fig. 10.28 shows EDS spectra of SiC. At high beam energy, for example, 20 keV shown in Fig. 10.28a, the C *K* peak intensity is greatly reduced by absorption relative to the SiK. With the spectrum expanded vertically so that the Si *K* peak fills the range, it would be very easy to miss the C *K* peak, which represents 30% by weight of the specimen, and report that the only major constituent present is silicon! Even at 10 keV (Fig. 10.28b), the C *K* peak is just visible. Only at low beam energy, for example, 5 and 3 keV in Fig. 10.28c and 10.28d, is the C *K* peak easily detectable.

Clearly, if a beam energy in the "conventional" range of 10 keV and higher (see Section 10.7) is chosen to detect heavier elements, the analyst must be particularly careful to evaluate the possible contribution of the light elements to the spectrum and to be prepared to lower the beam energy to fully confirm their presence or absence.

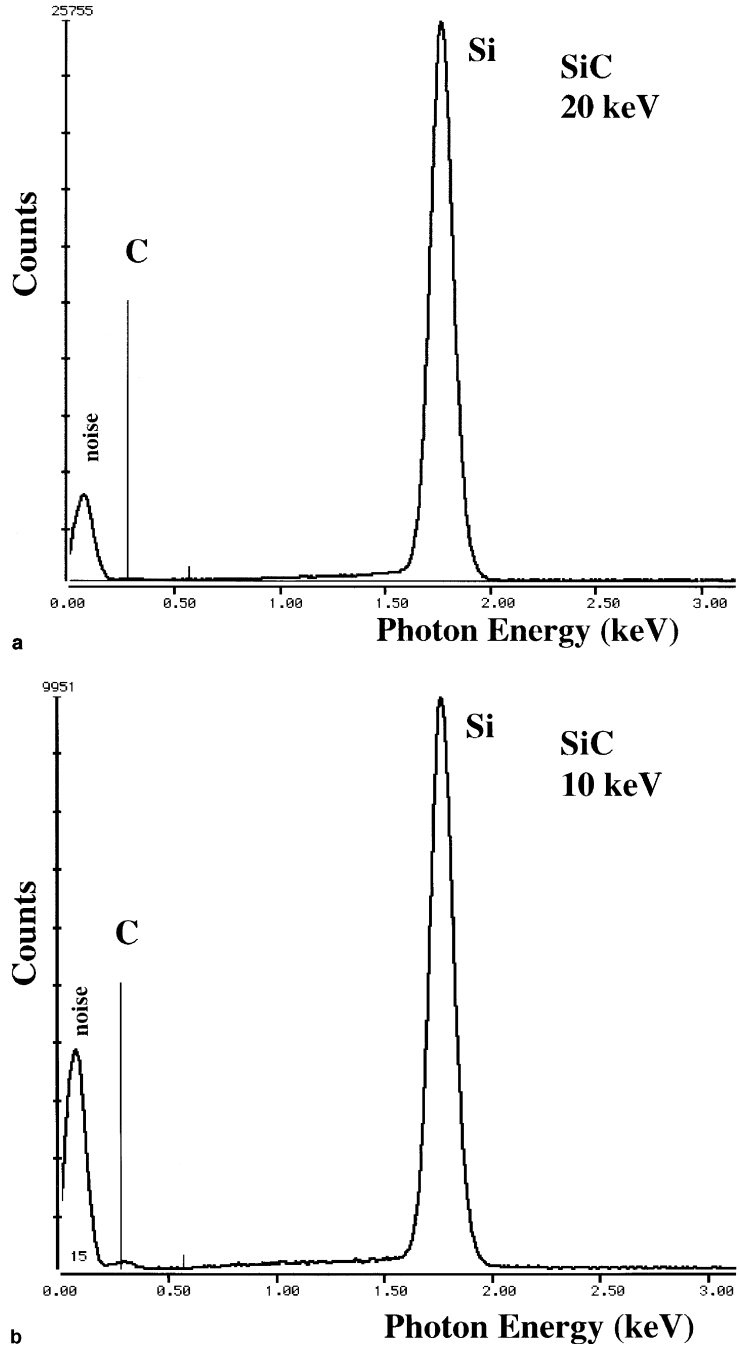
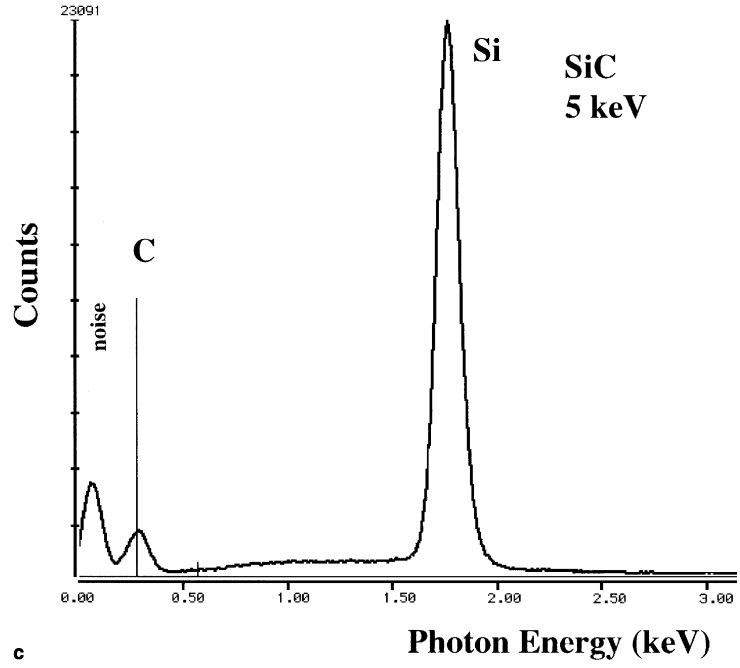
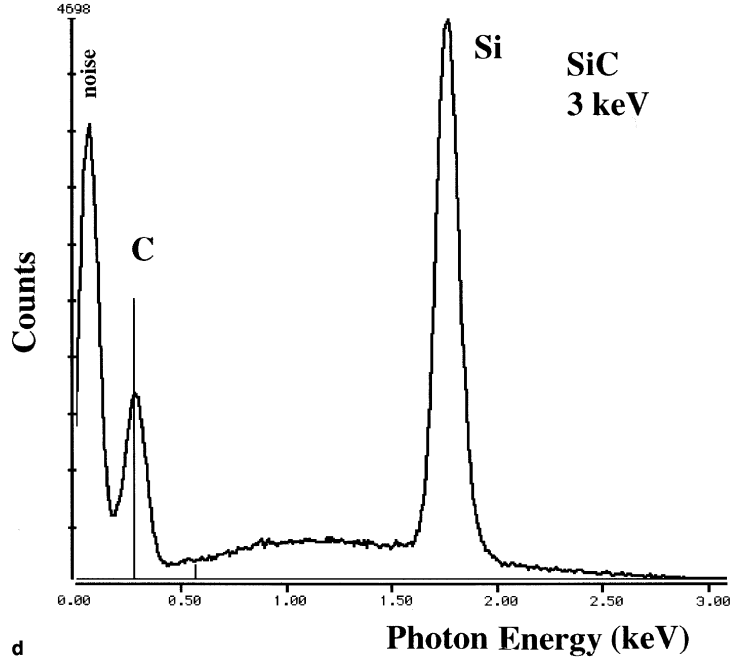


Figure 10.28. Silicon carbide (Si: 0.7 mass fraction; C: 0.3 mass fraction) as measured by Si EDS at various beam energies. (a) 20, (b) 10.



c



d

Figure 10.28. (*Continued*) (c) 5, and (d) 3 keV. Si EDS (129 eV at Mn $K\alpha$).

10.7.2.1. Si EDS

Commercial Si EDS systems are capable of measuring most of the light elements, with some systems extending down as far as Be. An example of a spectrum for high-purity carbon is shown in Fig. 10.29a. At lower photon energy, the peaks for Be and B are barely separated from the fundamental noise peak, as shown for B in Fig. 10.30. Reducing the electronic noise of the detector and amplifier to permit separation of the low-energy peaks for B and Be requires careful design of the detector and optimization of the complete signal collection and amplification system.

EDS Efficiency for Light Elements. A critical aspect of EDS spectrometry of the light elements is the efficiency of the EDS detector in the low-photon-energy range. The efficiency inevitably decreases for photon energies below 1 keV because of absorption in the window and other components of the spectrometer (typical dimensions: 10- to 20-nm gold surface electrode, 30- to 100-nm silicon "dead" layer). Absorption effects generally increase as the photon energy decreases, but there are sharp absorption increases for photon energies just above the critical ionization energy

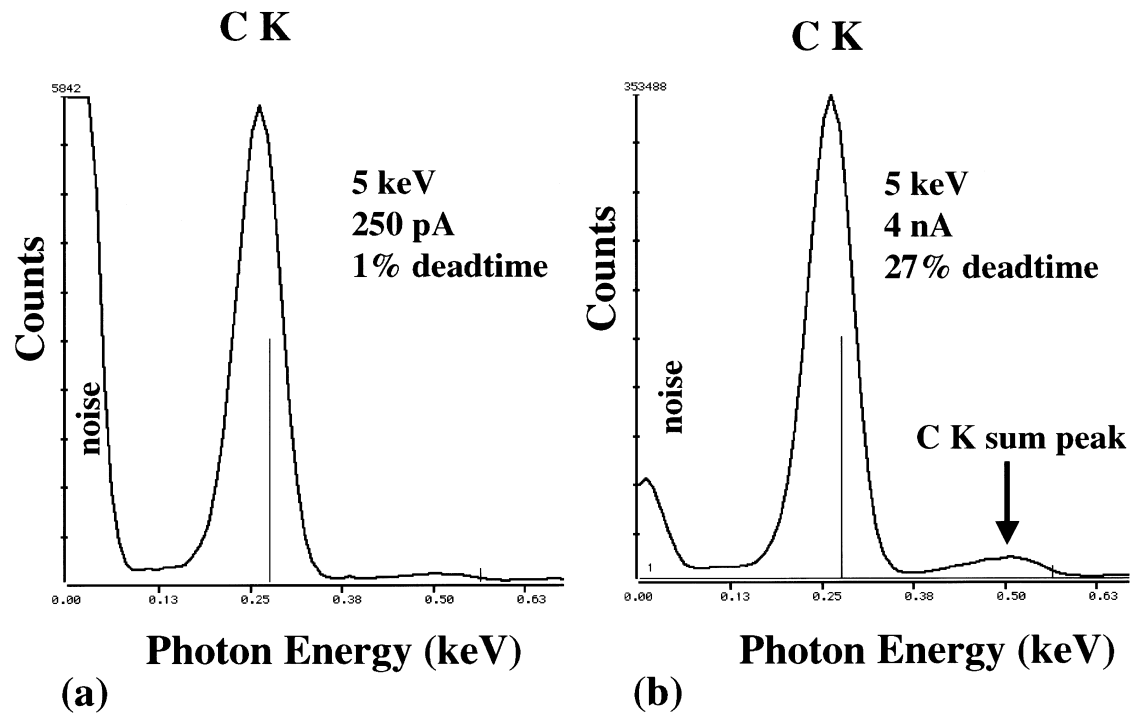


Figure 10.29. EDS spectrum of high-purity carbon: (a) Low count rate, 1% dead time; (b) high count rate, 27% dead time. Note C *K* sum peak. Si EDS (129 eV at Mn $K\alpha$).

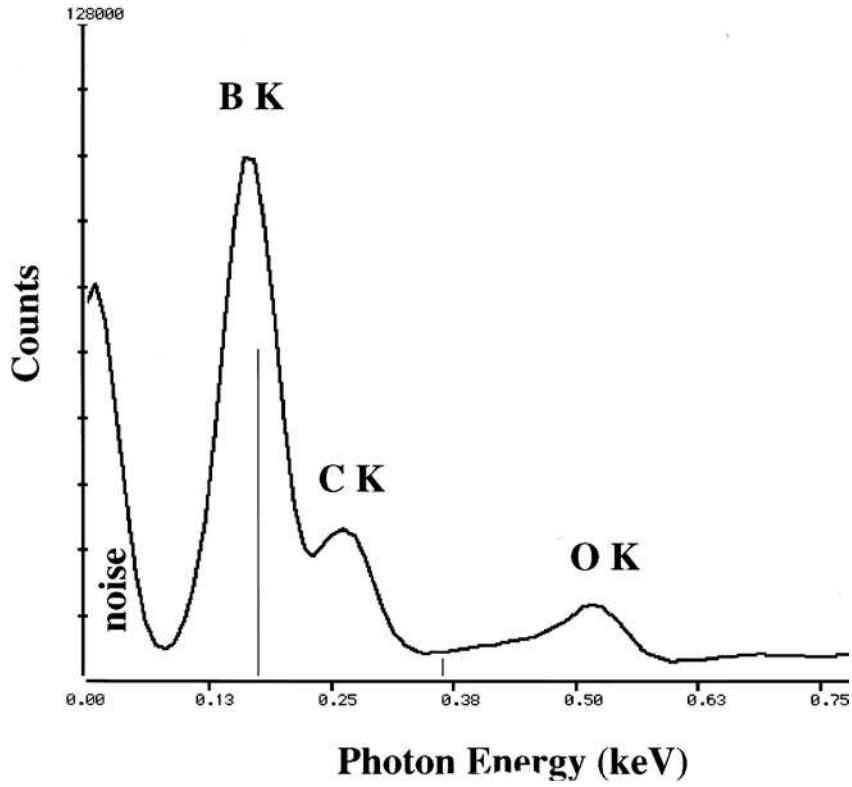
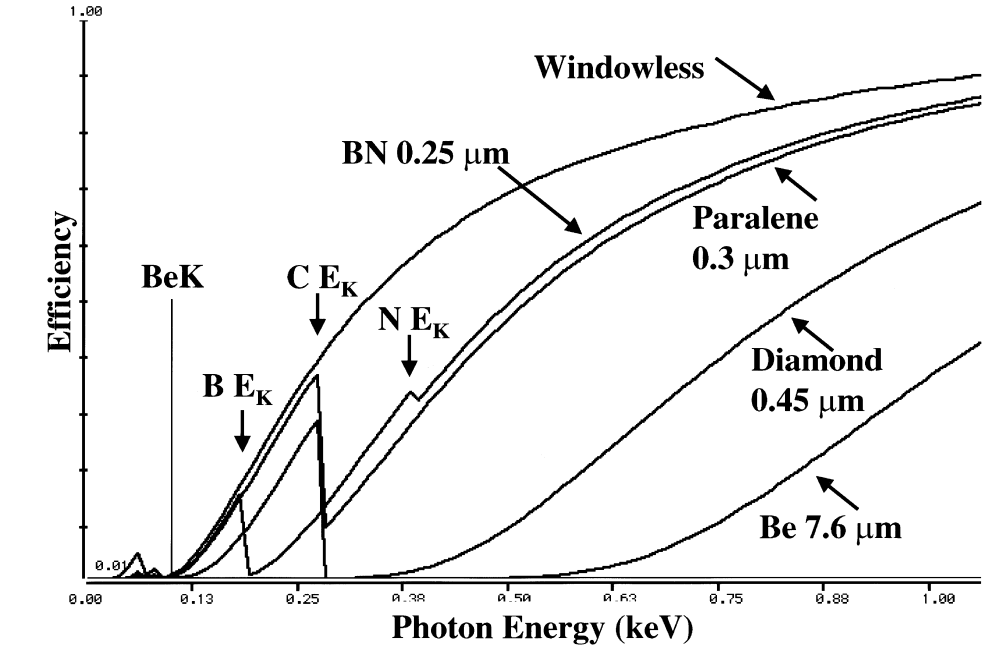


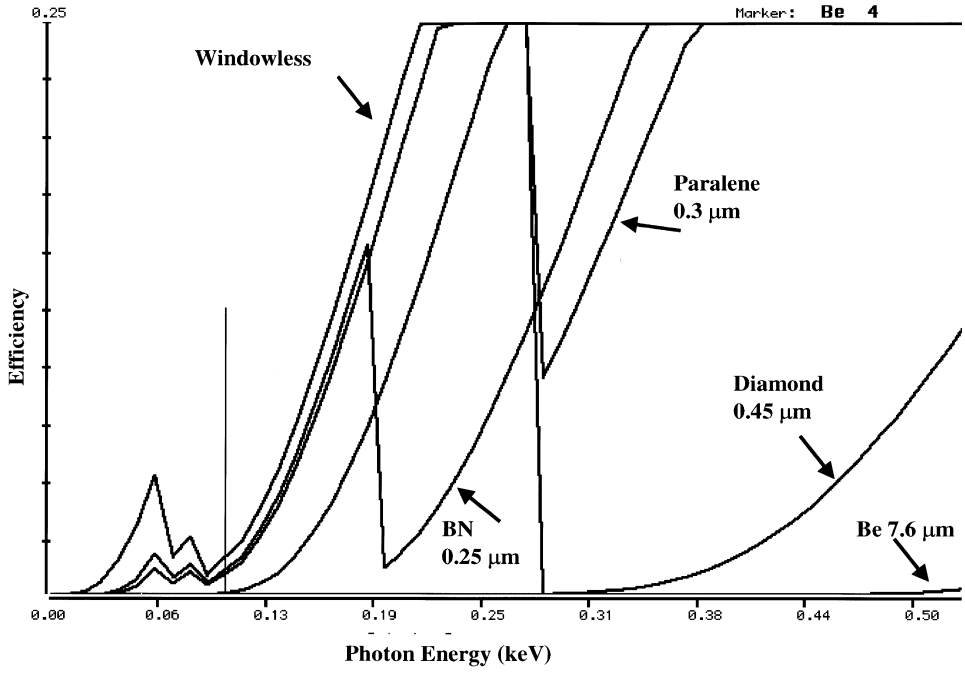
Figure 10.30. EDS spectrum of high-purity boron with a thin (10 nm) carbon coat. Note that oxygen is also observed from the native oxide on the surface of the boron. Si EDS (129 eV at Mn $K\alpha$); $E_0 = 5$ keV.

(energies) of the window materials. Efficiency curves for various windows in the low-photon-energy range are shown in Fig. 10.31a and expanded to show the very low photon energy range in Fig. 10.31b. At the energy of the Be K x-ray (0.109 keV), two window materials (7.6-μm Be and 0.45-μm diamond) have zero efficiency, whereas two have an efficiency of just 1% (0.3-μm paralene and 0.25-μm BN). The maximum efficiency of 2% for Be K is achieved with a windowless detector.

It is possible to operate a Si EDS without any window at all, but such windowless detectors are highly susceptible to contamination from the SEM environment. The Si EDS detector crystal is so cold (~80–100 K) that it makes an excellent vacuum pump for volatile species in the SEM chamber such as water and hydrocarbons. From the efficiency curves shown in Fig. 10.31, a windowless spectrometer would have an efficiency advantage for measuring the low-energy photons of the light elements. When the windowless mode is available, it is usually as one of several positions on a turret (the selection might include atmospheric-pressure-supporting Be, non-atmospheric-pressure supporting ultrathin polymer, and windowless). For windowless operation, the careful analyst will be sure to operate the SEM with attention to maintaining the cleanest possible vacuum, avoiding



a



b

Figure 10.31. EDS efficiency for various window materials: windowless; 7.6 μm Be; 0.45 μm diamond; 0.3 μm paralene; 0.25 μm BN. (a) 0–1 keV, (b) 0–0.25 keV; note the position of the Be *K* peak marker.

all sources of contamination, outgasing the specimen in the airlock before entering the microscope chamber, and using a liquid nitrogen cold shield with the SEM pumping system to minimize high-vapor-pressure components. Even then, the detector should only be exposed through the windowless position to the chamber when x-ray measurements are actually being made. This care will also be rewarded with minimal contamination rates on the specimen during measurements. Even with great care, the EDS detector will inevitably accumulate condensed volatiles such as water on its cold surface, which will act like an additional window absorbing x-rays, especially those of low energy. To detect such change in detector efficiency, a reference set of spectra of the light elements, for example, B, C, and O (as MgO, for example) as well as a Ni K–Ni L spectrum should be recorded when the system is initially installed as part of the laboratory quality assurance plan, and spectrum comparisons should be made on a regular basis. The Ni L/Ni K ratio is especially sensitive to the presence of a water ice layer because of the high absorption coefficient of oxygen for Ni L. The Si EDS detectors often include a "conditioning circuit" to carefully warm the unbiased detector to drive off condensed gases when the recorded spectrum shows a drop in the detector efficiency relative to the reference spectra. (Manufacturer's directions for this procedure should be followed with great care!)

EDS Calibration and Peak Distortion for Light Elements. The peak position of the low-energy photons of the light elements may be significantly shifted, by 20 eV or more from the ideal position, as shown in Fig. 10.29a for carbon, despite being measured at a very low dead time (1%). A similar shift is seen in Fig. 10.30 for boron. This shift has nothing to do with the chemical shifts noted in the section on WDS spectrometry (below), but instead is due to an instrumental deviation from the excellent linearity of the energy response that can be achieved for photons with energies above 1 keV, where peak positions lie within ± 10 eV of the ideal energy after calibration.

The low-energy photons of the light elements are strongly absorbed near the entrance surface of the Si EDS. Consequently, these x-rays are more highly influenced by the Si "dead" layer that exists just below the gold electrode. This layer is not truly "dead"; if it were, it would only influence the spectrum through simple absorption. Because this layer is actually partially active, photons absorbed within the layer are measured as having deposited less charge than a photon of the same energy absorbed within the "normal" portion of the detector, producing the phenomenon known as "incomplete charge collection." These incorrectly measured photons are seen as a shift in the peak position and also as an asymmetric distortion on the low-energy side of the peak. Careful fitting of a Gaussian to low-energy peaks can show this peak distortion in the residuals after subtracting the fitted Gaussian peak from the original spectrum.

Coincidence Peaks. The EDS system, as described in Chapter 7, can only measure one photon at a time. When a second photon enters the detector while the first is being measured, a special pulse inspection circuit

determines whether the first photon being measured must be excluded because of the error in the apparent energy introduced by the charge from the second photon. The low-energy photons of the light elements are so close to the system electronic noise that the pulse inspection circuit is ineffective for these pulses. Thus, sum (coincidence) peaks are much more commonly observed for peaks of the light elements than for high-energy peaks, even at count rates that seem acceptable from higher energy photons. Fig. 10.29b shows the carbon spectrum measured at 27% dead time, where a significant C sum peak is observed. This sum peak could very easily be mistaken for an O *K* peak, mislabeled so that even the qualitative analysis would be wrong, and finally quantified when none of the element actually exists! By lowering the count rate, and thus the probability for coincidence, the sum peak can be virtually eliminated. The sum peak is seen to disappear for the same material measured with a 1% dead time, Fig. 10.29a. To avoid these mistakes, the analyst should study the response with increasing input count rate (beam current) of the detector for pure elements such as Be, B, and C.

10.7.2.2. WDS

Diffractors. The low energies of the light element characteristic x-rays result in long wavelengths. Consequently, the diffraction equation requires that the diffractor have a correspondingly large interplanar spacing d. Of the crystals, thallium acid phthallate (TAP), with a spacing of 2.58 nm, and the other acid phthallates can diffract first-order O *K* (2.37 nm) and F *K* (1.83 nm). The wavelength range from B (7 nm), C (4.4 nm), to N (3.16 nm) was formerly measured with lead stearate (10.4 nm), an artificial diffractor created by the Langmuir–Blodgett technique of building up successive aligned molecular layers of a fatty acid with a heavy metal on one end of the molecule (Chapter 7). The fatty acid molecule serves to space the heavy metal atoms, which act as the scattering layers. These Langmuir–Blodgett structures provided high spectral resolution (~5 eV), but very poor efficiency because of the low density of heavy atom scattering sites.

The lead stearate diffractor has been almost entirely replaced by the new class of artificial diffractors known as layered synthetic materials (LSMs), which are based upon sputter deposition of alternating layers of heavy metal atoms (e.g.,W, Mo) with layers of carbon atoms, sometimes as boron carbide. The carbon (and boron) atoms serve as a low-absorption spacer and the heavy metal atoms serve as the scattering layer. LSMs have much higher scattering efficiencies than the lead stearate structures because the deposition can be controlled at the level of the individual atom layer so that a greater density of heavy metal atoms can be placed in each scattering layer. The LSM efficiency is at least 10 times greater than that of the sterates, an important advance where x-ray counts are in scarce supply due to low fluorescence yield and high self-absorption of low-energy photons. Unfortunately, the higher efficiency comes at the price of the resolution, which is in the range of 10–15 eV, a consequence of the roughness of the interfaces between the sputtered layers. There is the possibility of higher

resolution and efficiency as the deposition process control is improved to yield better-quality interfaces between the light and heavy atoms.

WDS X-Ray Peaks of the Light Elements. The resolution of WDS with the TAP, Pb stearate, and even the LSM diffractors is sufficient to observe "chemical effects" in light element x-ray spectra. Fig. 10.32 shows the B *K* peak using an LSM (LDEB, which is Mo separated by B_4C) in high-purity elemental boron, BN, and a glass containing boron oxide. The peak position in BN and the glass are shifted relative to elemental boron. The peak shapes are also more complex for the glass and BN, and the boron peak in the glass is substantially wider. These effects arise because the electron shells that are involved in the x-ray emission are also involved in forming chemical bonds, which alter the energy levels and density of electron occupation. These chemical effects can be manifest as shifts in the peak position and the shape as the chemical environment of the boron changes. Although these spectral effects carry valuable chemical state information, they cause difficulties when we wish to measure robust *k* values for unknowns and standards in which the chemical environment is different.

Bastin and Heijligers (1986a, b, 1990a, b) and Bastin, *et al.* (1984a, b; 1986) systematically studied these chemical effects for the light elements. To perform quantitative analysis with x-ray peaks that are subject to chemical effects, these authors found that it was necessary to perform a special

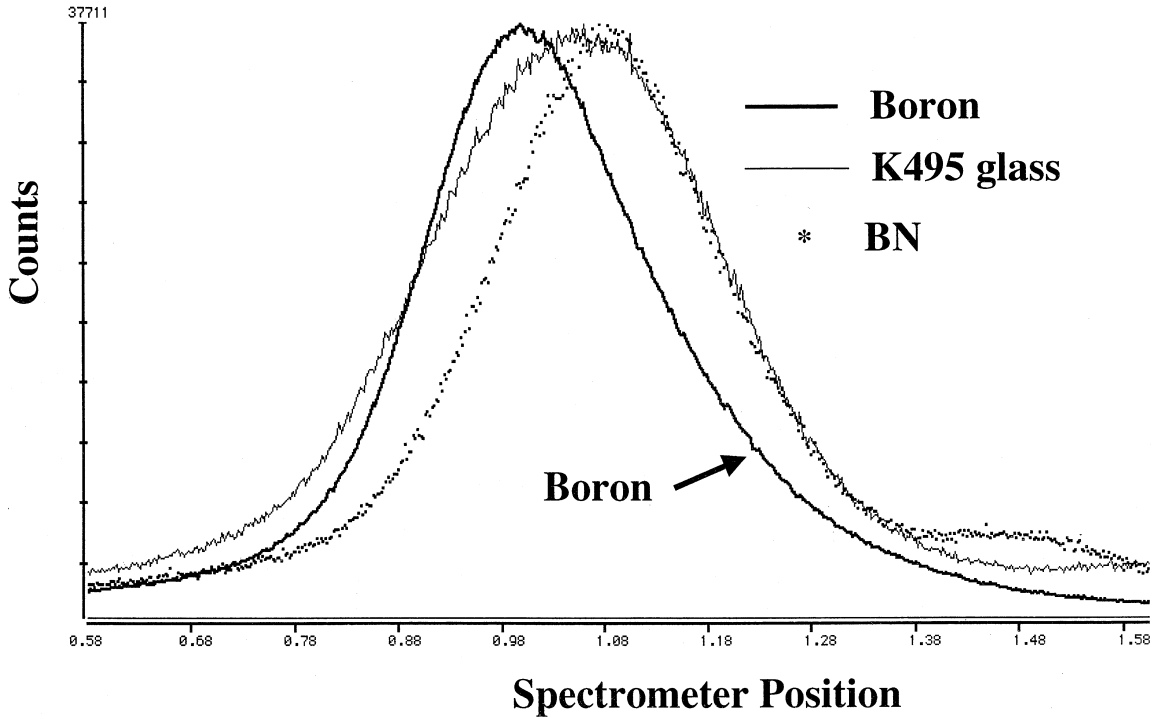
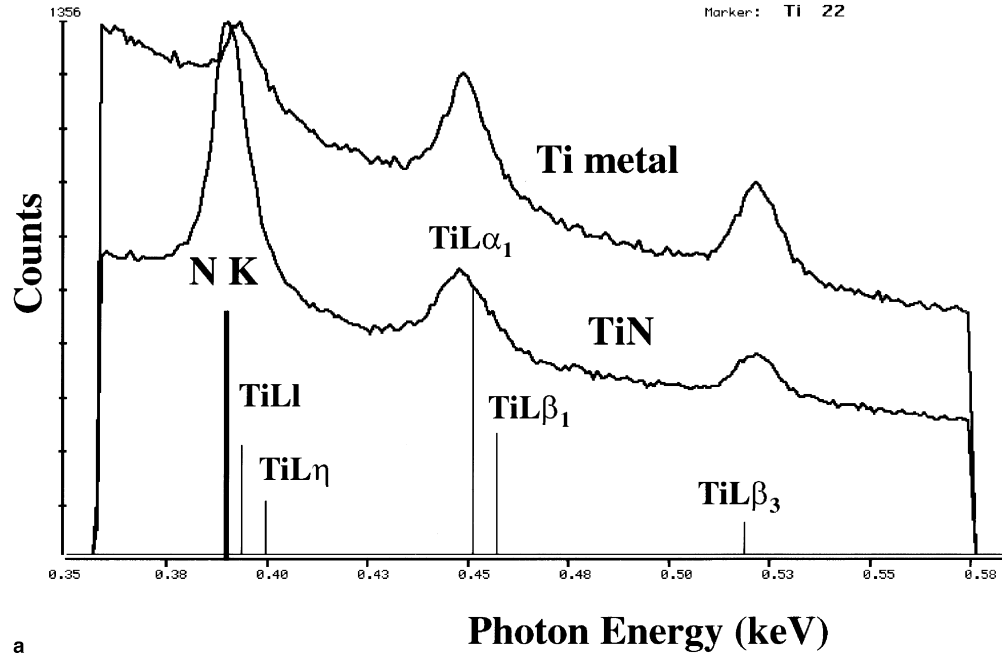


Figure 10.32. WDS (LDEB layered synthetic material, Mo–B_4C) spectra of elemental boron (thick solid line), boron-containing glass (thin solid line; B: 0.233 mass fraction; Al: 0.106; Li: 0.023; O: 0.638); and boron nitride (dots).

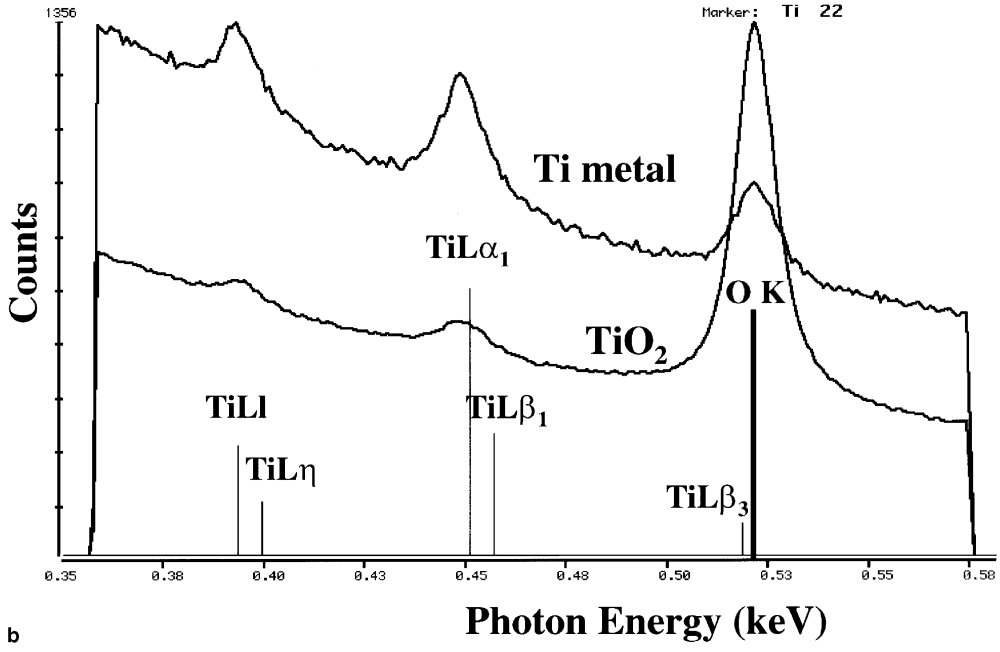
correction because the chemical effects redistributed the photons within the extended peak structure. The ideal peak measurement technique to deal with chemical effects is to scan the full peak width and integrate the intensity. However, performing a peak scan at every specimen location to be analyzed incurs a huge time penalty. To avoid that problem, Bastin and Heijligers developed a peak shape factor that could be used to multiply the intensity measured at the peak maximum and correct for differences between the shape of the peak for an elemental standard and specific chemical forms. Thus, once a scan has been performed to determine the peak shape relative to the standard and the appropriate peak shape factor has been selected, additional measurements can be made with the WDS set on the standard peak position only without the need for scanning the full peak width.

Measurement of the background in the neighborhood of light element K peaks can be complicated by the presence of higher order ($n = 2$, 3, 4, etc.) diffraction peaks from the higher energy K, L, and M peaks of intermediate- and high-atomic-number components of complex specimens. The interfering higher order diffraction peaks can be separated by the use of pulse discrimination. The rogue photons striking the detector due to higher order diffraction actually have the same energy as the $n = 1$ peak, and therefore have higher energy than the light element photons, although they occur at the same diffraction angle. A pulse height analysis of the photon stream would reveal the true low-energy photons and the higher energy, higher order diffraction photons. With a suitable threshold and window in the pulse counting signal chain, the higher order photons can be rejected (Chapter 7). A particular advantage of the LSM diffractors is that they actually suppress the higher order peaks. The lead stearate diffractor previously used very efficiently scattered higher order lines.

However, there may be low-energy L and M peaks of the intermediate and heavy elements that can interfere directly with the light element peaks in LSM scans. Because these are photons of similar energy to the light element photons, pulse discrimination is of no use in separating them. A classic example of this type of interference is shown in Fig. 10.33 for the Ti L family and N K and O K. The Figure shows WDS scans of Ti metal, TiN, and TiO_2 converted into energy space using the conversion function included with the Desktop Spectrum Analyzer. Table 10.8 gives the energies of several of the low-photon-energy K peaks along with interferences from L-shell x-rays of heavier elements. (Note that M-shell interferences have not been considered in compiling this list.) For example, there is strong interference between Ti Ll and Ti $L\eta$ with N K, and between Ti $L\beta_3$ with O K. In fact, for the Ti metal spectrum, a considerable portion of the intensity in the Ti $L\beta_3$ peak is actually due to the O K intensity that arises from the native oxide layer on the surface of the titanium metal. Extraction of accurate intensities from either of these peak interference situations will require peak fitting using multiple linear least squares. Because it is difficult to obtain titanium metal with a totally oxygen-free surface, obtaining a clean reference spectrum for Ti $L\beta_3$ will be difficult. Inspection of Table 10.8 reveals similar interference between V $L\alpha$, $L\beta$ and Cr Ll; $L\eta$ with O K; Sc $L\alpha$, $L\beta$ with N K; and the Fe L family with F K.



a



b

Figure 10.33. (a) WDS scan (LDE1 LSM diffractor) of Ti metal (upper trace) and TiN (lower trace). (b) WDS scan (LDE1 LSM diffractor) of Ti metal (upper trace) and TiO_2 (lower trace). The Ti *L*-family peak positions are marked as well as the peak positions for N *K* and O *K*. Significant interference from Ti *Ll* and *Lη* occurs with N *K*, and from Ti *Lβ*$_3$ with O *K*. Beam energy 5 keV; 100 nA; 1000 channels at 1 s per channel; WDS spectra converted to energy space with Desktop Spectrum Analyzer.

Table 10.8. Selected Examples of Interferences with Low-Photon-Energy *K* peaks[a]

Be *K*	**0.109 keV**	
S *Ll*, *η*	0.149	!Severe interference
B *K*	**0.185 keV**	
S *Ll*	0.149	!Severe interference
Cl *Ll*	0.183	!Severe interference
Cl *Lη*	0.184	!Severe interference
Ar *L*l	0.221	!Severe interference
Ar *Lη*	0.218	!Severe interference
C	**0.282 keV**	
K *Ll*	0.259	!Severe interference
K *Lη*	0.262	!Severe interference
Ca *Lα*	0.340	
Ca *Ll*	0.302	!Severe interference
Ca *Lη*	0.306	!Severe interference
N *K*	**0.392 keV**	
Ti *Ll*	0.395	!Severe interference
Ti *Lη*	0.401	!Severe interference
Ti *Lα*	0.452	
Ti $L\beta_1$	0.458	
O *K*	**0.523 keV**	
Ti $L\beta_3$	0.519	!Severe interference
V *Lα*	0.511	!Severe interference
V $L\beta_1$	0.519	!Severe interference
Cr *Ll*	0.500	!Severe interference
Cr *Lη*	0.510	!Severe interference
Mn *Ll*	0.556	!Severe interference
F *K*	**0.677 keV**	
Mn *Lα*	0.637	!Severe interference
Mn $L\beta_1$	0.649	!Severe interference
Fe *Ll*	0.628	!Severe interference
Fe *Lα*	0.705	!Severe interference
Fe $L\beta_1$	0.718	!Severe interference

[a] Peaks within ±50 eV of the *K*-peak are identified as severe interferences for EDS.
M-shell interferences have not been included.

10.7.3. Special Measurement Problems for the Light Elements

In addition to the cautions raised previously, light element analysis is subject to special conditions and limitations.

10.7.3.1. Contamination

Contamination is a general term used to describe the buildup of foreign material on the specimen surface as a result of electron beam bombardment. The source of this foreign material is usually organic compounds left on the specimen surface as a result of inadequate cleaning procedures or from the vacuum system itself, often as a result of carelessness in handling specimens

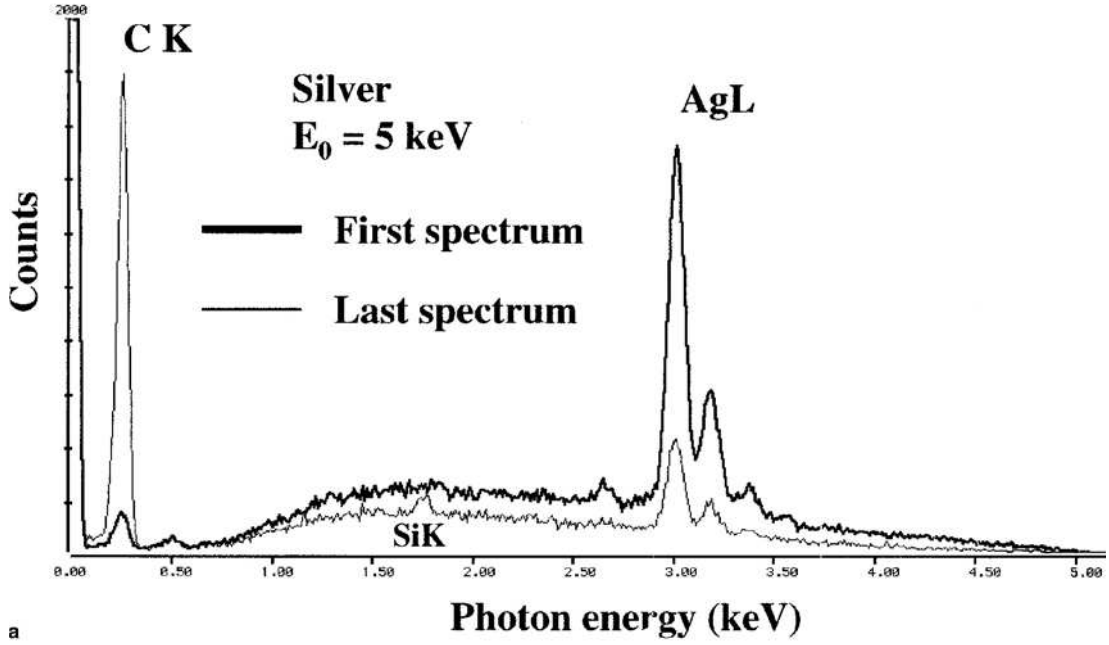
and introducing volatile hydrocarbons into the specimen chamber. The electron beam "cracks" these hydrocarbons, releasing smaller gaseous molecules and building up nonvolatile carbon deposits. If carbon is an element of interest in the analysis, then the additional carbon x-ray emission contributed by the carbon in the contamination layer will obviously distort the result. This effect can be often detected by examining the spectrum as a function of time, as shown in Figs. 10.34a and 10.34b, which depict the results from a time sequence of measurements with the beam scanning a 10×10-μm area on high-purity silver. The initial carbon signal is due to a thin (<10 nm) layer of carbon evaporated on the polished multiple-standard mount containing the silver. The carbon signal shows a steady growth as a function of time. Because the rate of contamination depends on the dose per unit area, the buildup would have occurred much more quickly if a static rather than a scanning beam had been used for this measurement.

Moreover, because the carbon atoms of the contamination are deposited at the surface of the specimen, their emitted x-rays will suffer less absorption than the carbon atoms located in the true specimen below the contamination. This gives the contamination atoms even more weight in the recorded spectrum than an equivalent number of carbon atoms that are part of the specimen.

Carbon contamination can be minimized by a variety of techniques. Hydrocarbons on the specimen surface can be driven off by modest heating during pumpdown in the specimen airlock. The SEM vacuum system usually includes effective anticontamination devices, such as liquid nitrogen cold plates, which act to condense volatile hydrocarbons. The migration of hydrocarbons across the specimen surface to the beam site can be minimized by intentionally irradiating the specimen over a large area with the electron beam so as to lock down a thin carbon layer that is immobile. Finally, a jet of oxygen gas or dry air through a fine, hollow needle directed at the beam impact point has been shown to react with carbon and carry it off as CO and/or CO_2.

10.7.3.2. Overvoltage Effects

A second aspect of the contamination is its impact on the generation and emission of x-rays throughout the spectrum. Beam electrons penetrating though the contamination lose energy. The generation of x-rays depends on $(U - 1)^n$, where U is the overvoltage ratio (E_0/E_c) and the exponent n ranges from 1.3 to 1.6. The generation of x-rays decreases sharply as U decreases for small values of U just above unity, as shown in Fig. 10.35 for an exponent of 1.5. When the analyst chooses a low beam energy, for example, $E_0 = 5$ keV, to improve the sensitivity for light elements by avoiding high absorption, the other elements present with photon energies above 1 keV will be excited with low values of the overvoltage. Beam energy loss in the contamination layer effectively lowers U for these elements. The resulting loss of x-ray generation diminishes the peak-to-background ratio, and the effect on the spectrum can be quite dramatic, as seen for the Ag L-family peaks in the spectra in Fig. 10.34a and in the intensities plotted



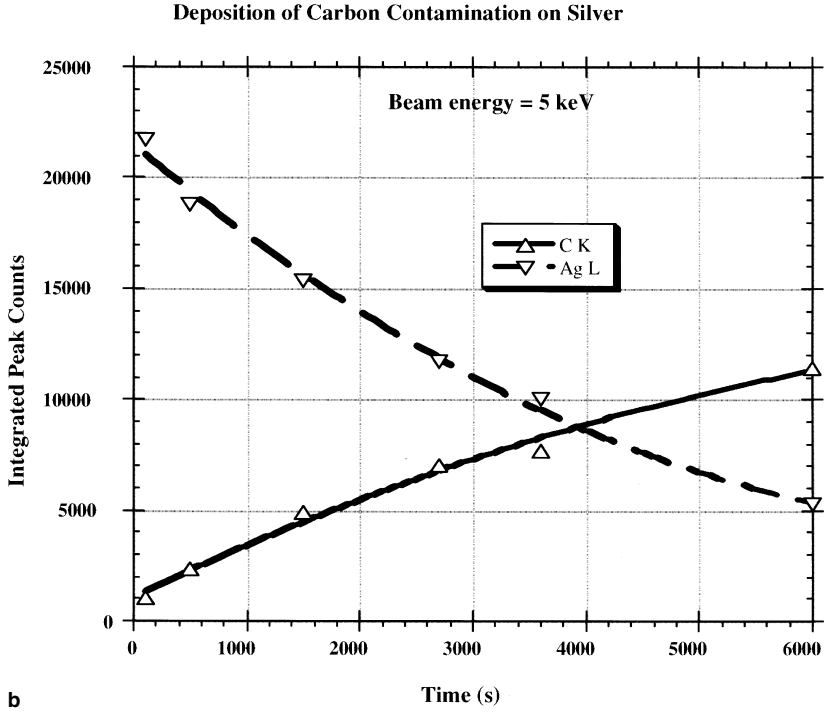


Figure 10.34. (a) Initial and final EDS spectra on silver with a focused probe scanning a 10 mm × 10 mm square. (b) Time sequence of intensities of C *K* and Ag *L* silver, beam energy = 5 keV.

in Fig. 10.34b. The overvoltage for the Ag L-family principal peaks is only $5/3.351 = 1.49$ before any contamination builds up, so that small changes in U have a large effect on Ag L x-ray production. For an exponent of 1.5, a loss of only 100 eV in the initial beam energy (to 4.9 keV effective entry energy) causes a relative intensity drop of 8.4% in the Ag L production, and a 200-eV loss results in a 17% decrease.

10.7.3.3. Absorption Effects

The buildup of contamination can also modify the spectrum through absorption effects. Carbon, the principal component of contamination, can have an especially large absorption effect on those elements whose x-rays have energies just above the carbon K edge at 0.284 keV. Thus, the N K peak at 0.392 keV is particularly vulnerable to anomalous absorption from contamination layers. Figure 10.36 shows a comparison of spectra, measured with a Si EDS operated in a windowless mode, of aluminum oxynitride obtained on a specimen with a 10-nm carbon coating and on a bare aluminum nitride surface viewed through a copper grid to discharge the insulating surface. In this experiment, the 10-nm carbon coating attenuates the N K peak by a factor of roughly two compared to the intensity of the N K peak observed on the "bare" surface. Note that even the "bare" AlN surface spectrum shows a small carbon peak (which causes absorption of N K) due to contamination deposited during the measurement.

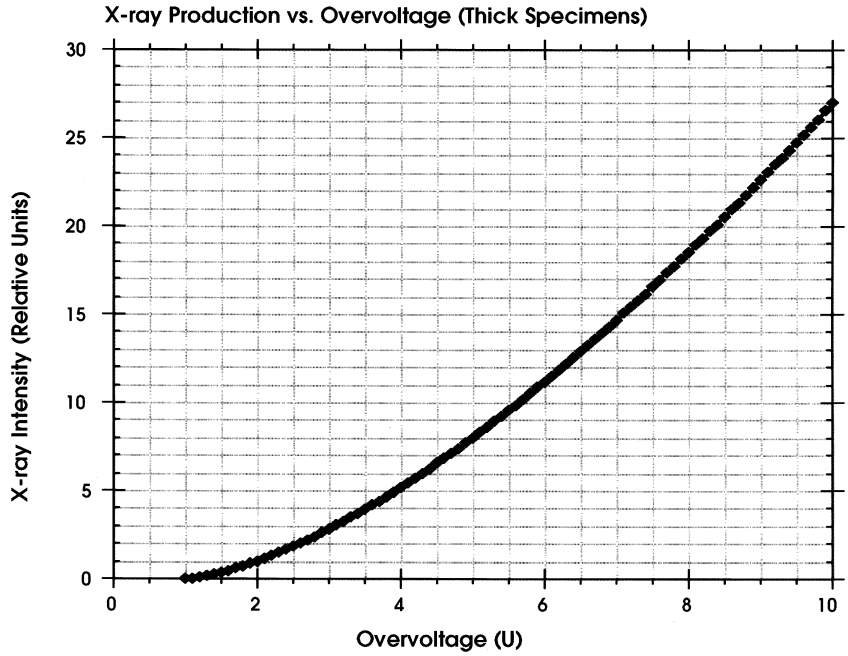


Figure 10.35. Plot of the x-ray generation (relative) versus the overvoltage $U = E_0/E_c$ for an assumed exponent of 1.5 in the equation $I \sim (U - 1)^n$.

The previous sections describe the special difficulties that must be overcome to obtain robust k values for the light elements. Clearly, the measurement of the k value can introduce large errors into the final result and no method of quantification can overcome a poor k value determination to produce an accurate final result. Given that the k-value measurement has been made as carefully as possible to eliminate or at least reduce the error introduced by these effects, the analyst is ready to calculate elemental concentrations through appropriate matrix corrections.

The principal matrix correction challenge for light elements is the absorption correction. Figure 10.37 shows the x-ray depth distribution $\phi(\rho z)$ of C K in Fe_3C as-generated and as emitted for a take-off angle of 40°, calculated by Monte Carlo electron trajectory simulation. At a beam energy of 15 keV, which would provide an overvoltage of 2 for the Fe K shell, the C K is almost entirely absorbed. Only about 3% of the generated C K x-rays are not absorbed. Lowering the beam energy to 10 keV, which gives an overvoltage of only $U = 10/7.11 = 1.40$ for Fe K, but still a useful,

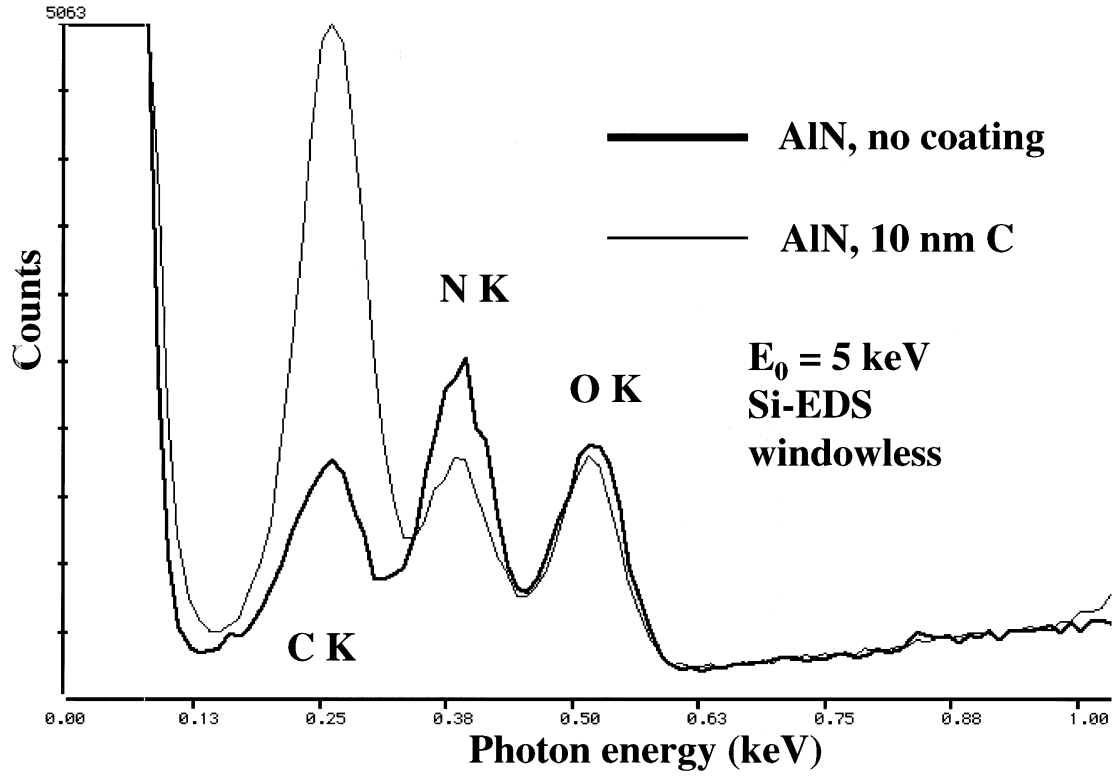


Figure 10.36. Spectra of aluminum oxynitride obtained on a specimen with a thin (~10 nm) carbon coating and on a bare surface viewed through a 200-mesh copper grid to discharge the specimen. Note the attenuation of the N K peak when the carbon layer is present.

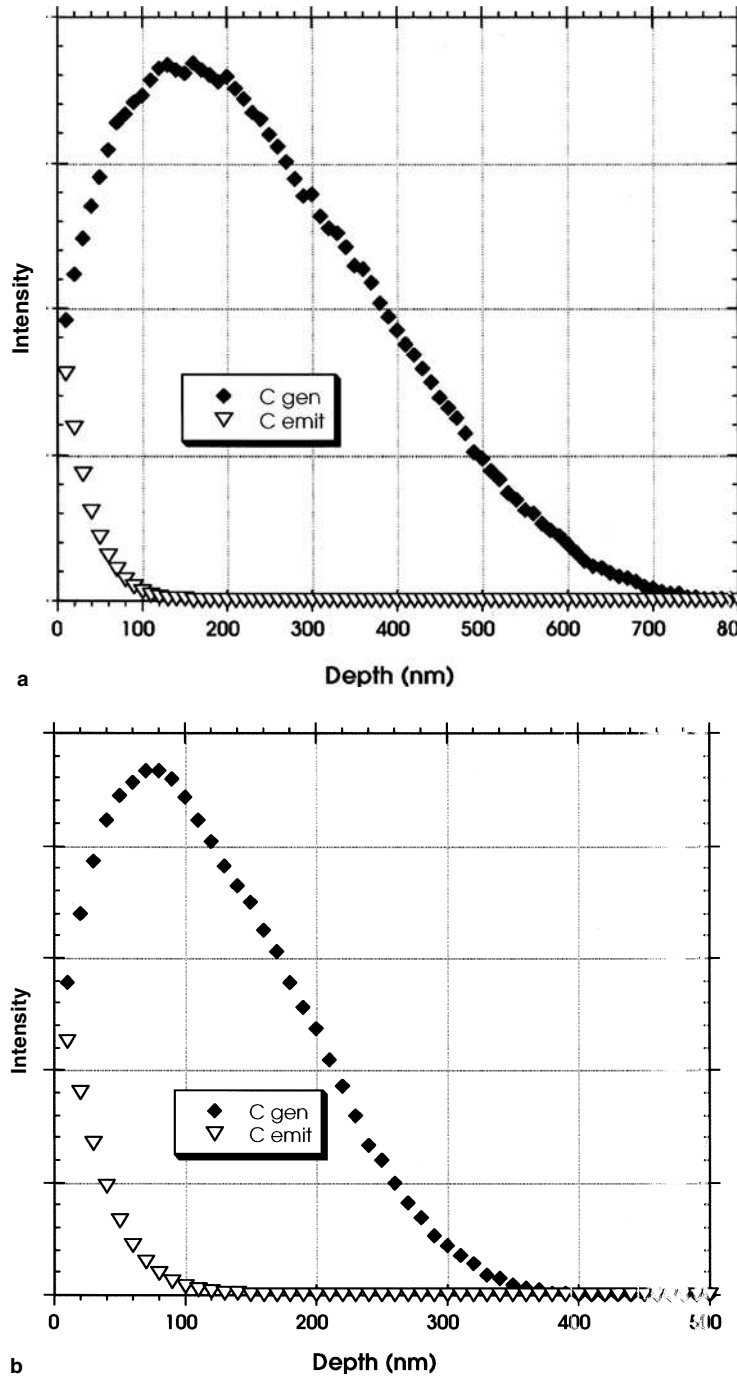


Figure 10.37. X-ray depth distribution $\phi(\rho z)$ of C K in Fe_3C as-generated and as emitted for a take-off angle of 40° calculated by Monte Carlo electron trajectory simulation: (a) $E_0 =$ 15 keV; (b) $E_0 = 10$ keV.

Table 10.9. Mass Absorption Coefficients for Carbon *K* X-Rays According to Various Sources

	Mass absorption coefficient (g/cm^2)				
Absorber	Barkalow *et al.* (1972)	Henke *et al.* (1982)	Heinrich (1986)	Bastin and Heijligers (1990a, b)	Pouchou and Pichoir (1991)
Fe	15,000	13,900	14,100	13,500	13,500
W	18,000	18,800	16,700	18,600	18,000
Mo	19,000	16,400	26,160	20,600	20,500
Cr	10,000	10,600	10,980	10,700	10,700
V	15,000	8,840	9,400	9,700	8,850
C	2,300	2,350	2,150	2,373	2,170

practical value, only increases the C *K* yield to about 9% (Fig. 10.37b). Note that the distribution of the emitted C *K* x-rays is nearly identical in the two plots because it is absorption rather than generation that controls the C *K* signal. Moreover, the C *K* only samples the top 100 nm of the specimen, whereas the much more energetic Fe *K* x-rays escape from essentially their entire generation range. Thus, if we wish to simultaneously quantify C and Fe in Fe_3C using *K*-shell x-rays, we must deal with an extremely strong absorption case for the C. The absorption correction depends on having both an accurate descrption of the depth distribution in the very shallow portion of the generation near the surface and having accurate values of the mass absorption coefficients for the matrix elements (Fe in this case) for the low-energy C *K* x-rays. The problem is illustrated in Table 10.9. The published values from just five sources range from 13,500 to 15,000, an 11% range. This range of values when propagated through the absorption equation results in a similar range of $f(\chi)$. This range is already twice as broad as our error distribution for x-rays above 1 keV, not considering the error introduced by the atomic number factor. How do we proceed to improve the accuracy?

The classic *ZAF* models were mainly developed for higher-energy x-rays excited with high-energy beams in the range 10–25 keV. As such, they are at a disadvantage for light element analysis and tend to produce wider error distributions than the competing $\phi(\rho z)$ methods. The empirical $\phi(\rho z)$ methods have the advantage of being tied to actual experimental data, which include specific measurements of light element x-rays in various matrices. This close empirical relationship to experimental measurements to some extent overcomes our lack of knowledge of the exact physics of electron interaction and low-energy x-ray propagation in the near-surface region. Even so, the $\phi(\rho z)$ method is vulnerable to errors introduced by the fundamental data such as mass absorption coefficients.

The light element absorption situation illustrates one of the great needs of the x-ray microanalysis community: accurate mass absorption coefficients for the low-energy characteristic x-rays, particularly those of

the light elements. With the advent of high-brightness, broad-spectrum synchrotron x-ray sources and with the ready availability of well-characterized thin films, accurate measurements of these fundamental parameters should be possible. In the absence of these improved absorption data, Bastin and Heiligers (1990a, b) and Pouchou and Pichoir (1991) analyzed extensive sets of experimental data obtained on well-characterized light element compounds. Using experimental $\phi(\rho z)$ models, these authors backcalculated values of the mass absorption coefficients that produce the best fit to the experimental data. Thus, based upon this strongly empirical approach, the $\phi(\rho z)$ methods are found to produce the best accuracy for the light elements, although it must be recognized that the results will inevitably show broader error distributions than those obtained for elements measured with more energetic x-rays.

10.8. Low-Voltage Microanalysis

Low-voltage microanalysis has traditionally been applied using WDS to optimize sensitivity for light-element, low-photon-energy analysis, as described in Section 10.6. The development of the field emission gun scanning electron microscope (FEG SEM) has enabled low beam energy analysis to serve a new role, that of high-spatial-resolution analysis in bulk specimens. The FEG SEM in its various configurations (cold, thermal, and Schottky) can produce small, focused probes, 1–10 nm in diameter, with analytically useful levels of beam current, 100 pA to several nanoamperes, at beam energies of 5 keV and below. When combined with a large solid-angle EDS detector and/or a WDS augmented with capillary optics to achieve a large solid angle of collection (see Chapter 7), the FEG SEM can serve as an extraordinary tool for high-spatial-resolution microanalysis. The continuing development of the high-resolution (<5 eV) microcalorimeter EDS will greatly increase low-voltage analysis capabilities.

10.8.1. "Low-Voltage" versus "Conventional" Microanalysis

Although not strictly defined, "conventional" beam energy for microanalysis can be considered to begin at approximately 10 keV. At 10 keV, the overvoltage is sufficiently high to excite *K*-shell peaks from elements with Z as high as 29 (copper: $U = 1.111$, near the practical minimum). This is an important distinction because for even higher atomic numbers, where the *K*-shell cannot be excited by the beam, the L shell must be used. L shells above 1 keV in energy have a sufficiently high fluorescence yield to produce an adequate peak-to-background ratio for reasonably sensitive analysis (low limits of detection). As E_0 is further increased above 10 keV, the analyst has the luxury of choosing *K*- or *L*-shell x-rays for analysis, for example, Ge K or Ge L, for an increasing number of elements.

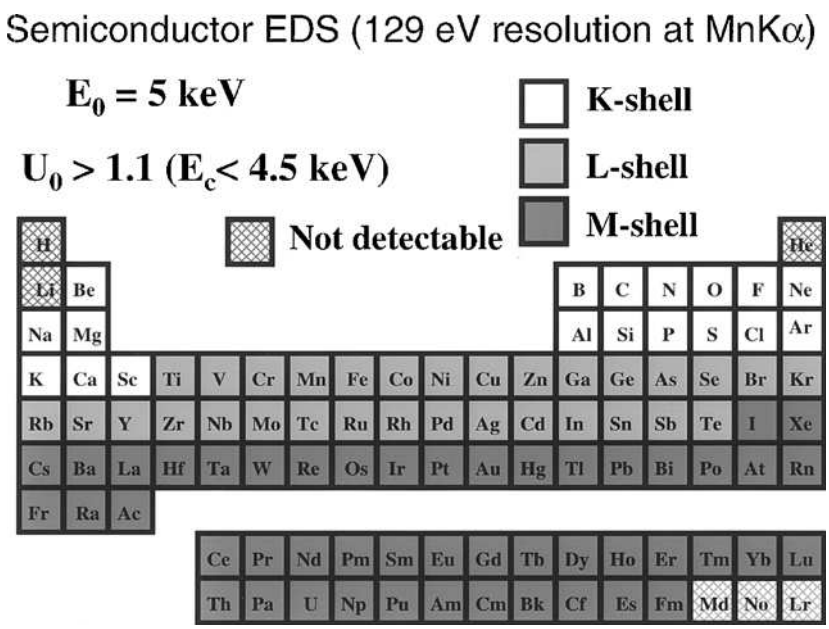


Figure 10.38. Low-beam-energy microanalysis ($E_0 \leq 5$ keV). Analytical peak selection strategy illustrated for $U \leq 1.1$ and $E_c \leq 4.5$ keV.

"Low-voltage-microanalysis" conditions can be considered to start at 5 keV and extend to lower beam energy values (beam energies from 5 to 10 keV will be taken as a transition range). A beam energy of 5 keV is the lowest energy for which a characteristic peak suitable for analysis can be found for virtually all elements of the Periodic Table (except for H, and He, which do not produce an x-ray and Li for which the x-ray energy, 54 eV, is too low for practical measurement), as shown in the analytical selection strategy illustrated in Fig. 10.38. At 5 keV, the first element for which the L line must be chosen because the K line is not excited is titanium. As shown in Fig. 10.39, the Ti L peaks are only excited with a low peak-to-background ratio, which will severely limit sensitivity. The minimum usable beam energy for low-voltage analysis depends strongly on the specific elements to be measured. As will be discussed further, the fluorescence yield of some of these peaks, such as Ba M, is so low that the limit of detection in the low-voltage-microanalysis regime may be of marginal utility for many practical applications, even with high-resolution spectrometers.

10.8.2. X-Ray Production Range

The x-ray range scales approximately as $(E_0^{1.67} - E_c^{1.67})$. The exponential relationship of the range has a strong impact on the analytical spatial resolution. For conventional beam energies, the x-ray generation range is generally in the micrometer range. Examples of the range for the analysis

of iron in various matrices are calculated for conventional analytical conditions (with Fe $K\alpha$) in Table 10.10 and for low-voltage analytical conditions (with Fe L) in Table 10.11. If an overvoltage of at least $U = 2$ is required to obtain a reasonably sensitive analysis, then for conventional analysis conditions E_0 must be 15 keV or above for Fe $K\alpha$ with $E_c = 7.11$, which gives a range of 2.1 μm in a silicon matrix, as an example. Under low-voltage conditions, the use of Fe $L\alpha$ with $E_c = 0.708$ keV and $E_0 = 1.5$ keV is sufficient, giving an x-ray production range of 47 nm in Si. Thus, the analytical resolution in the low-voltage mode improves by a factor of 45 in linear measure and a factor of 90,000 in volume. The interaction volumes for Fe (1%) in silicon as calculated with a Monte Carlo electron trajectory simulation are illustrated in Fig. 10.40 for various beam energies.

10.8.2.1. Contribution of the Beam Size to the X-Ray Analytical Resolution

With conventional energy conditions, a beam with a diameter of a few hundred nanometers can carry hundreds of nanoamps of current, which is sufficient even for most WDS applications. Because the x-ray production range with a conventional energy beam, for example, 15 keV, is of the order of micrometers, a beam diameter of 200 nm makes a relatively small

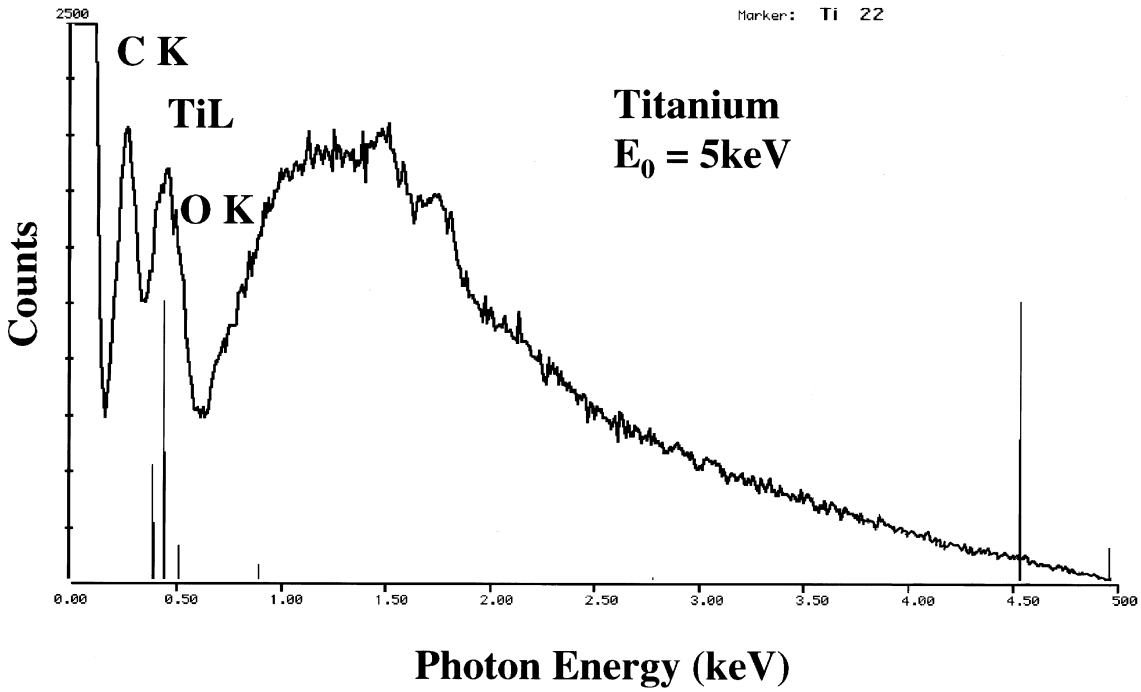


Figure 10.39. Si EDS (129 eV at Mn $K\alpha$) spectrum of titanium excited with $E_0 = 5$ keV. Note Ti L peaks, C K peak, and the peak position for O K.

Table 10.10. Conventional Beam Energy Analysis: X-Ray Production Range for Fe $K\alpha$ in Various Matrices[a]

	X-ray production range			
Matrix	25 keV	20 keV	15 keV	10 keV
Carbon	6.4 μm	4.1 μm	2.2 μm	680 nm
Silicon	6.0 μm	3.9 μm	2.1 μm	645 nm
Iron	2.0 μm	1.3 μm	710 nm	220 nm
Gold	1.1 μm	705 nm	380 nm	117 nm

[a] Fe K edge = 7.11 keV.

contribution to the analytical resolution in terms of the volume of material sampled, which is primarily controlled by the physics of electron scattering and x-ray propagation. Because of the high value of the exponent on the beam and edge energies in the x-ray range equation, the range drops very sharply as E_0 decreases to 5 keV. In the low-voltage microanalysis regime, the x-ray range is well within the nanometer scale. As calculated in Table 10.11 for iron (Fe $L = 0.707$ keV), the range at 5 keV is less than 500 nm, even in carbon. For a 2-keV beam, which still provides an overvoltage of nearly 3 for Fe L, the x-ray range is less than 100 nm in the various matrices. At 2 keV, if an incident beam size of several hundred nanometers must be used to provide the high current for WDS operations with an EPMA, the beam would be the largest contributor defining the spatial resolution in the low-voltage regime. At 2 keV, the FEG SEM can produce beam sizes of a few nanometers, which means that the improvement in resolution due to the reduced x-ray range at low voltage can be fully exploited as a means of achieving high spatial resolution in bulk specimens. The Monte Carlo plots for Si(1% Fe) in Fig. 10.40 show the effect of a 10-nm-diameter incident beam on the lateral distribution of trajectories at the entrance surface. (Note that the range calculated with this Monte Carlo simulation, which uses rigorous models of electron scattering, is a factor of 2–3 less than that given by the simple range equation.) Clearly, improvements in minimizing the beam size in the low-voltage regime impact the spatial resolution of analysis.

Table 10.11. Low-Beam-Energy Analysis: X-Ray Production Range for Fe L in Various Matrices[a]

	X-ray production range (nm)				
Matrix	5 keV	2.5 keV	2 keV	1.5 keV	1 keV
Carbon	475	136	88	52	15
Silicon	449	129	83	47	14
Iron	152	43	28	15	4.7
Gold	82	23	15	8.1	2.5

[a] Fe L_3 edge = 0.707 keV.

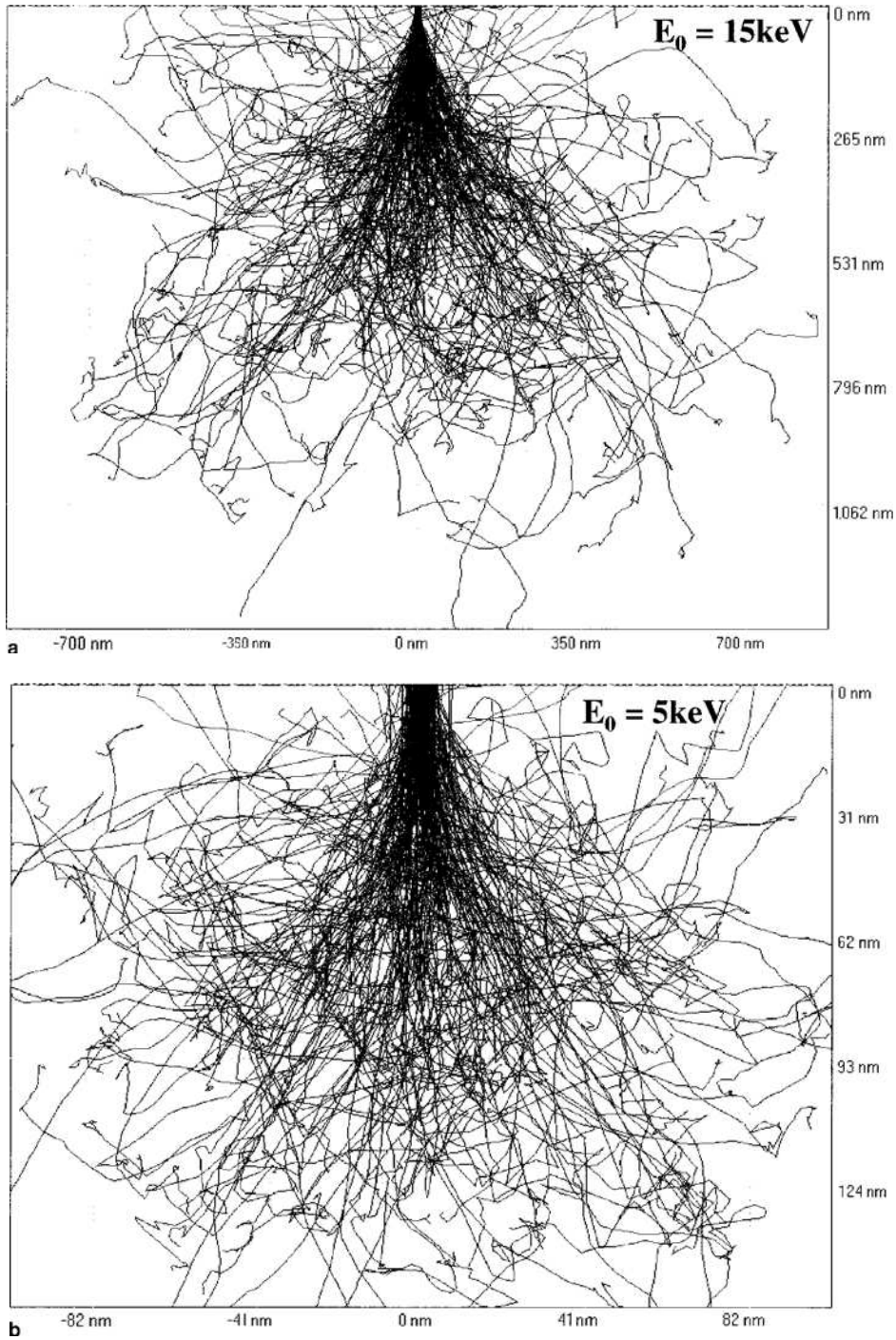


Figure 10.40. Monte Carlo electron trajectory simulations for Fe *L* in a silicon matrix as a function of beam energy: (a) 15, (b) 5.

10.8.2.2. A Consequence of the X-Ray Range under Low-Voltage Conditions

Many of the familiar phenomena of conventional beam energy analysis and the flexible analytical strategies they permit are radically changed in the low-voltage regime. We must constantly remind ourselves that the situation is likely to be different from the "rules" with which we are familiar from conventional analysis experience. As an example, consider the "zeroth" assumption of conventional quantitative analysis of bulk materials, namely that the material must be homogeneous over the entire range of x-ray excitation that we wish to measure. In the conventional analysis regime, the overvoltage is generally sufficiently high for most of the spectrum that the x-rays of interest, although not coming from exactly the same volume due to differing E_c values, at least originate from most of the same interaction volume. A thin surface layer, such as the native oxide layer of approximately 5 nm thickness that is found on most metals, represents only a few percent at most of the excited x-ray range under conventional conditions. For low-voltage microanalysis, the situation can be radically different. Consider the spectrum for CdTe excited with $E_0 = 5$ keV shown in Fig. 10.41a. The L_3 edge for Cd is 3.538 keV, whereas for Te it is 4.342 keV. The range (in nm) for a beam of 5-keV energy with compositionally weighted Z and A parameters for CdTe ($\rho = 6.20$ g/cm^3) is

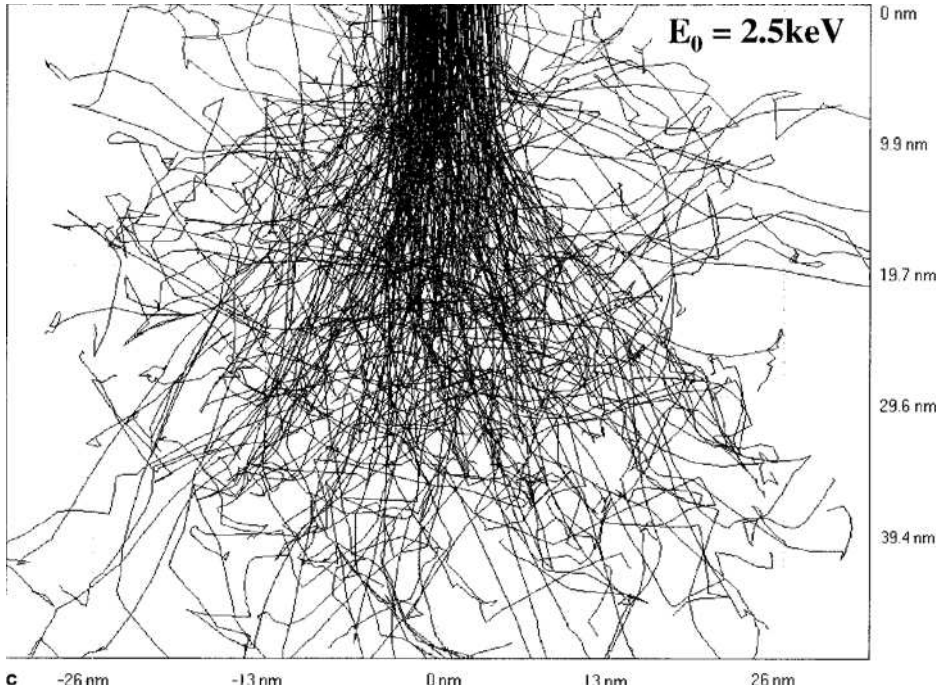
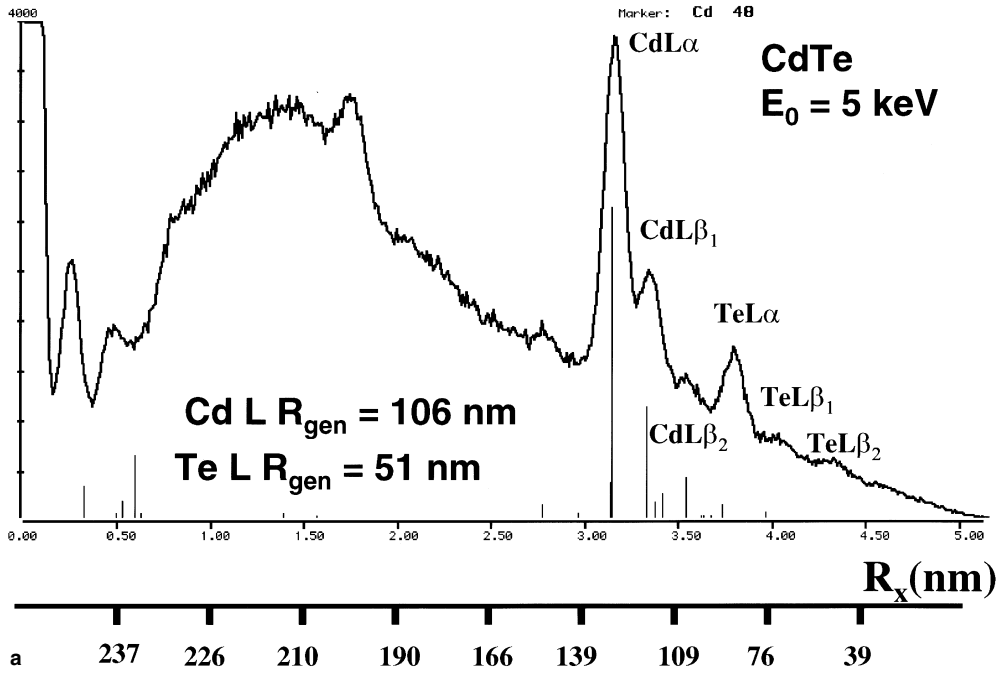


Figure 10.40. (*Continued*) (c) 2.5 keV; beam diameter = 10 nm. [Simulations performed with the CASINO program by Hovington *et al.* (1997).]



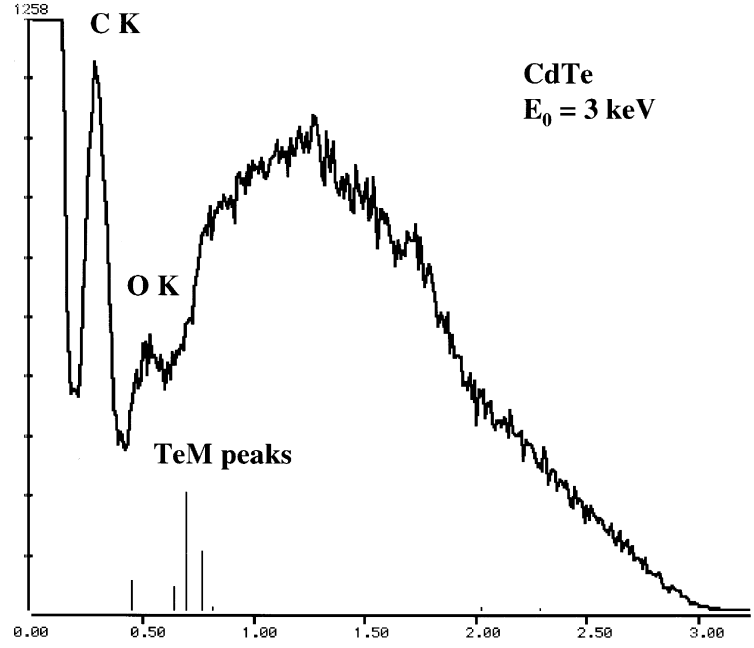


Figure 10.41. (a) Si EDS (129 eV at Mn $K\alpha$) spectrum of CdTe excited at $E_0 = 5$ keV. Note the additional horizontal axis, which denotes the x-ray range for x-rays excited with that edge energy. This range scale is nonlinear because of the exponential dependence on energy. (b) CdTe excited at $E_0 = 3$ keV.

given by:

$$R_x = \frac{27.6\,\text{nm} \cdot 120.5}{6.2 \cdot (50.1)^{0.889}} \left(5^{1.67} - E_c^{1.67}\right). \qquad (10.13)$$

The x-ray range calculated with Eq. (10.13) for Te is 51 nm, whereas for Cd it is 106 nm, a factor of about 2 in linear dimension and 2^3 or 8 in volume. Thus, in low-voltage microanalysis there is a very strong range function across the x-ray spectrum, rendering the measurement particularly susceptible to surface layers or any other heterogeneous features. It is useful to consider the photon energy axis of the spectrum to also be a range axis. An example of this approach is shown in Fig. 10. 41a, where a range axis for this specific material is calculated with Eq. 10.13 for edge energies located every 0.5 keV through the spectrum. Note that in this plot the range scale is nonlinear because of the exponential energy dependence.

10.8.3. X-Ray Spectrometry in Low-Voltage Microanalysis

As described in Chapter 6, excitation of characteristic x-rays by an electron beam requires that the beam energy exceed the ionization edge energy. The overvoltage, the ratio of the beam energy to the edge energy, $U = E_0/E_c$, must exceed unity. X-ray production depends strongly on U, with a functional form $(U - 1)^n$, where n is about 1.5 and varies with the specific element; the functional dependence of I upon U is shown in Fig. 10.35. As the beam energy is reduced from the conventional range to the low-voltage range, the number of edges that can be ionized is greatly reduced. For elements of intermediate and high atomic number, the familiar K-shell or L-shell x-rays that are selected in conventional analysis are no longer accessible, and L-shell or M-shell x-rays must be chosen instead. However, the second factor that must be considered in generating x-rays is the fluorescence yield ω, the fraction of ionizations that result in photon emission. For L and M x-ray peaks below 1 keV in energy, the fluorescence yield drops very rapidly. Unfortunately, these yields are very poorly known, and therefore when considering a low-voltage analysis strategy for a particular selection of elements, it is invaluable to first measure EDS spectra from the elements of interest, either in pure form or as simple compounds, to understand the spectral characteristics, especially the peak-to-background ratio. The CdTe example shown in Fig. 10.41 contains useful L-shell peaks for both elements when the beam energy is 5 keV. When E_0 is reduced to 4 keV and below, there appears to be no possibility with semiconductor EDS of measuring M-shell x-ray peaks as an alternative; see, for example, Fig. 10.41b, where the peak positions are marked for the Te M x-rays. Thus, for Cd and Te, the practical low-voltage range stops near 5 keV.

Losing access to elements because of limited shell selection compounded by low overvoltage and low fluorescence yield is a major hindrance to low-voltage microanalysis by Si EDS. As the beam is progressively lowered below 5 keV, more and more of the periodic table is effectively lost to EDS x-ray spectrometry. Figure 10.42 shows the practical microanalysis

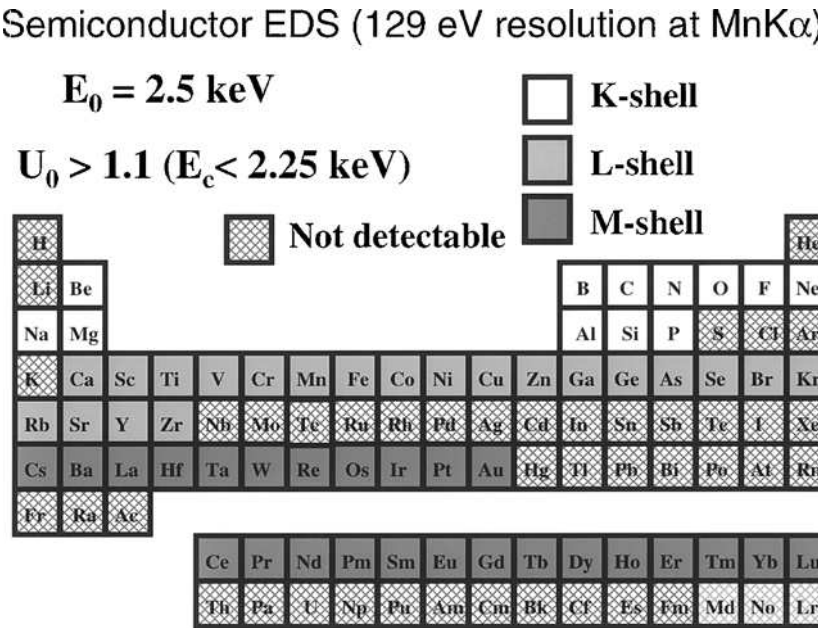


Figure 10.42. Periodic Table showing basic analytical choices for Si EDS (129 eV at Mn $K\alpha$) with $E_0 = 2.5$ keV, assuming a minimum overvoltage $U = 1.1$. Note the large number of elements that are effectively not accessible.

situation at $E_0 = 2.5$ keV, with a measurement criterion requiring a minimum value of $U = 1.1$. Large sections of the Periodic Table cannot be accessed, even for major constituents. As an example of the x-ray spectrometry problems that we face in low-voltage microanalysis, consider the spectrum of apatite, $Ca_5[F(PO_4)_3]$ (mass fractions: Ca = 0.487; F = 0.046; P = 0.226; O = 0.467) shown in Fig. 10.43. The P K peak at 2.01 keV ($U = 2.5/2.145 = 1.17$) representing 22.6% P by weight is visible, but with a peak-to-background ratio of only $1510/1200 = 1.47$. The F K peak, representing only 4.6% by weight is detectable, but just barely. The O K is by far the strongest peak. Calcium represents 48.7% by weight of the excited mass. The beam energy is below the Ca K edge, so the Ca K lines are unavailable. Ca L is the only other choice. When we examine the spectrum in Fig. 10.43, where is Ca L? To understand this portion of the spectrum, we must consider the effect of the presence of carbon and oxygen.

10.8.3.1. The Oxygen and Carbon Problem

The low-voltage microanalysis regime also gives greatly increased sensitivity to the surface regions of a sample. Most materials do not have a clean surface as prepared or after exposure to the "conventional" vacuum of an SEM. Metals, for example, tend to quickly form a thin surface oxide (or "native oxide") upon cleaning and exposure to the atmosphere, even

for a few seconds. For Al this native oxide is about 4 nm thick. Ion beam bombardment in the SEM specimen chamber can eliminate this surface oxide, but even in an intermediate-pressure regime of 10^{-5} Pa ($\sim 10^{-7}$ torr) the arrival rate of oxygen atoms from the residual gas in the chamber is sufficiently fast that the native oxide will quickly reform in a few hundred seconds or less. Carbon, especially in the form of hydrocarbons, is abundant in the environment. With diligent effort in specimen preparation, cleaning, and handling, and with fastidious vacuum practice, carbon due to contamination can be reduced, but rarely entirely eliminated. Thus, carbon and oxygen from surface oxide and contamination are likely to be a regular feature of our spectra, at least under ordinary operation conditions. Figure 10.44 shows the C *K* and O *K* region of the spectrum with the positions indicated for the low-energy *L*- and *M*-shell x-rays that suffer interference from C *K* and O *K*. The low fluorescence yield for x-rays from these *L* and *M* shells relative to *K*-shell peaks of similar energy puts them at a further disadvantage, so that the C *K* (0.282 keV) to O *K* (0.523 keV) region is effectively lost for most practical purposes with Si EDS (129 eV at Mn *Kα*). The Ca *L* peaks in Fig. 10.43 are convolved with the C *K* and O *K* (oxygen forms 46.7% of the compound), which are much more abundant and which overwhelm the Ca *L*. The only evidence of Ca *L* is a distortion on the side of the C *K* peak. Deconvolution of the EDS peaks to separate the Ca *L* by multiple linear least squares fitting might be possible, but great care would be needed to use locally determined peak references for

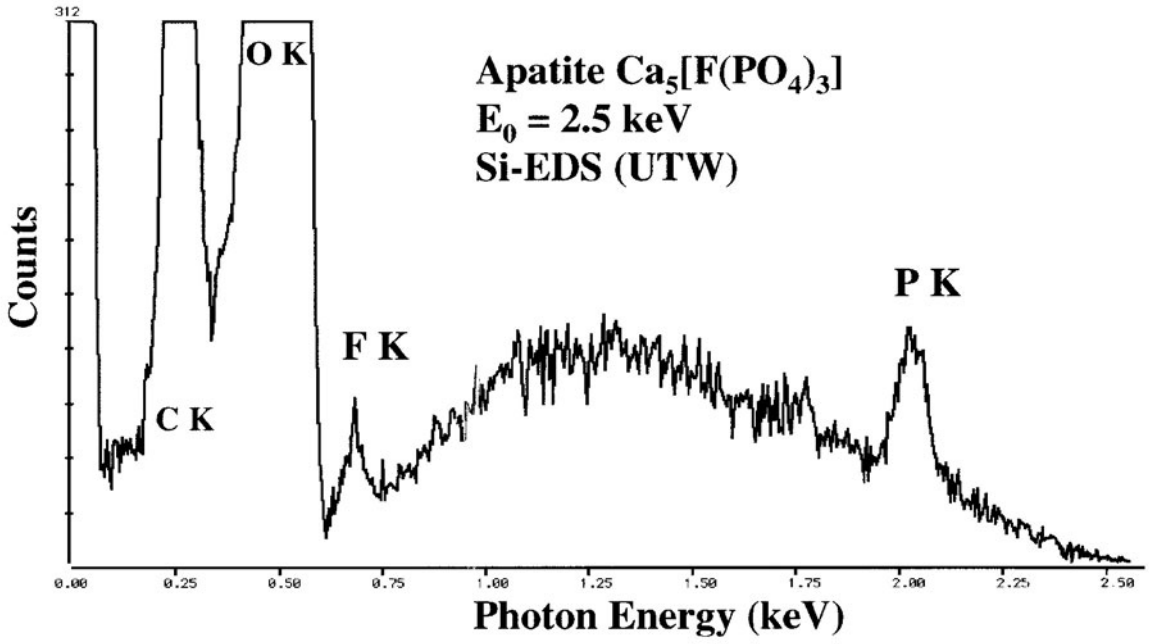


Figure 10.43. Si EDS (129 eV at Mn *Kα*) spectrum of the mineral apatite, $Ca_5[F(PO_4)_3]$ (mass fractions: Ca 0.487; F 0.046; P 0.226; O 0.467); $E_0 = 2.5$ keV.

the carbon and oxygen peaks, which have significant distortion from ideal Gaussians due to incomplete charge. With WDS and the microcalorimeter EDS, these interferences can be separated with the higher resolution. Fig. 10.45 shows a comparison of Si EDS with the microcalorimeter EDS to illustrate the use of high spectral resolution to separate the Ti *L*- and Ba *M*-family peaks in $BaTiO_3$ when conventional Si EDS (129 eV at Mn $K\alpha$) does not permit recognition of these components.

10.8.3.2. Quantitative X-Ray Microanalysis at Low Voltage

As was the case with light element analysis, the preceding discussion indicates that the analyst's chief problem in performing low-voltage microanalysis is obtaining robust *k*-value measurements. Once the various

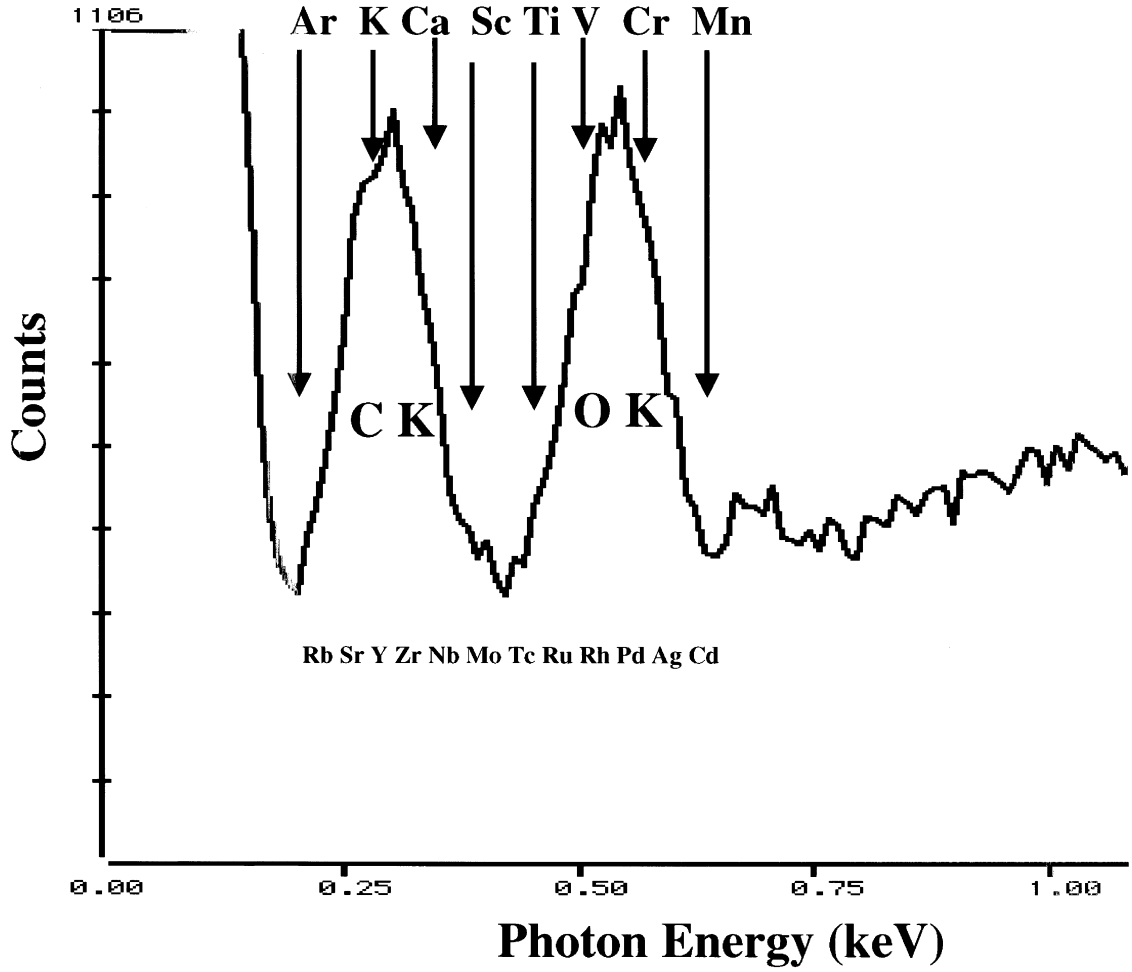
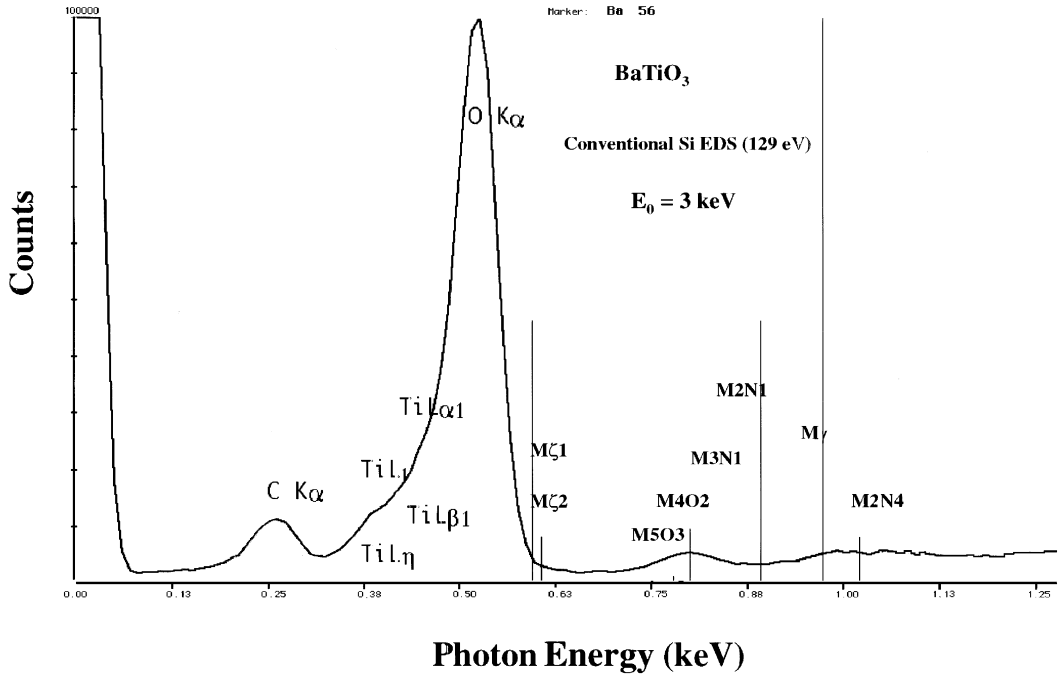
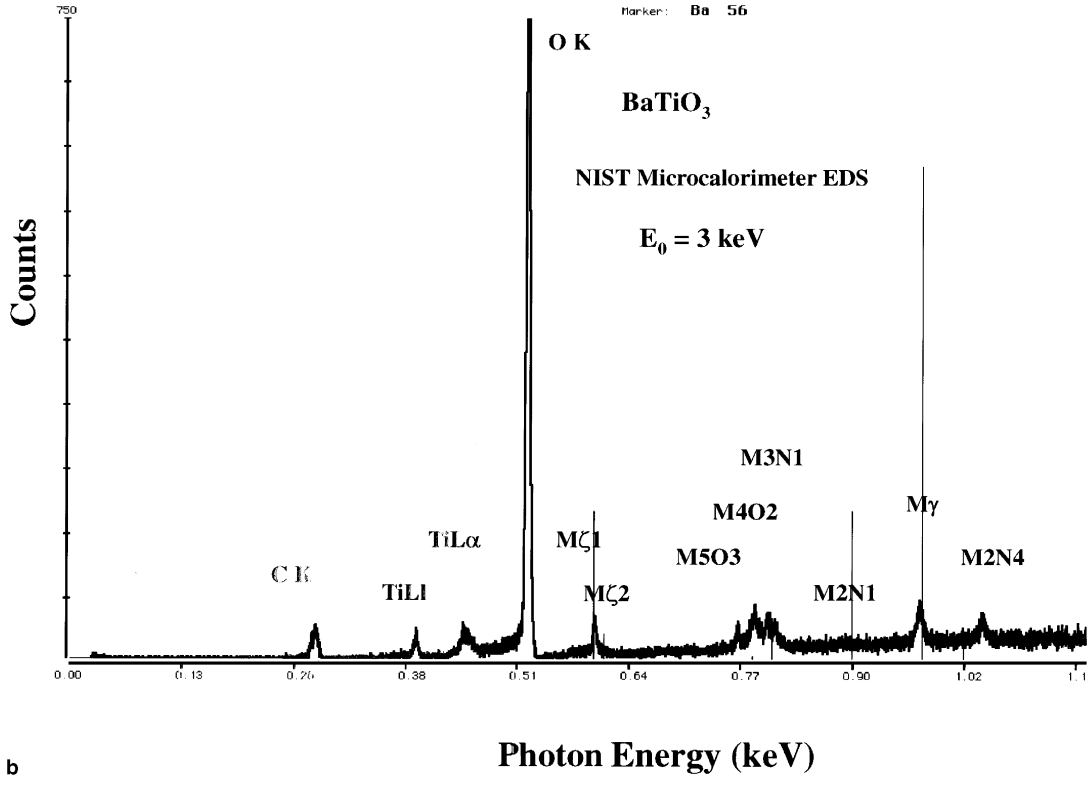


Figure 10.44. Carbon and oxygen *K* peaks with the positions marked for *L*- and *M*-shell peaks that suffer interference in this region; Si EDS (129 eV at Mn $K\alpha$).



a



b

Figure 10.45. $BaTiO_3$ excited with $E_0 = 3$ keV: (a) Si EDS (129 eV at Mn $K\alpha$); (b) Microcalorimeter EDS.

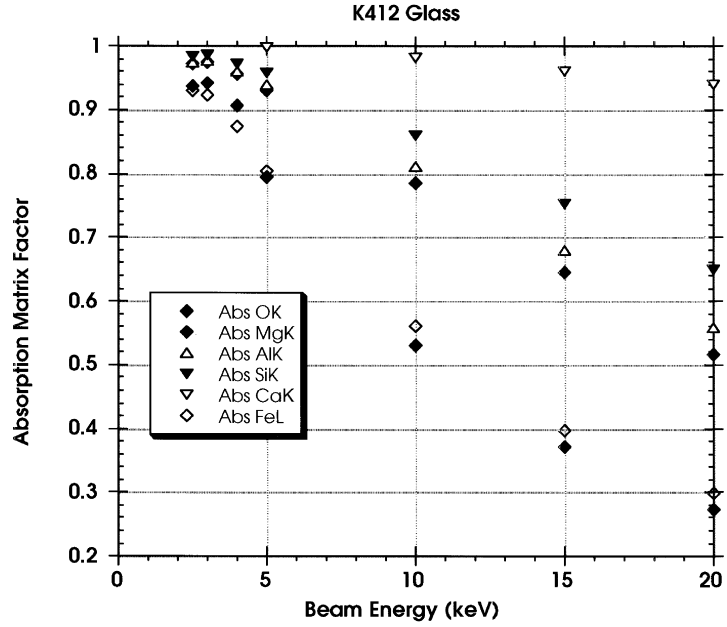


Figure 10.46. Plot of the absorption factor for O *K*, Mg *K*, Al *K*, Si *K*, Ca *K*, and Fe *L* in NIST Glass K-412 as a function of beam energy.

spectral measurement problems have been solved, the analyst will be capable of obtaining useful k values for quantitative analysis.

Because the x-ray range in low-voltage microanalysis is so shallow, there are actually useful consequences for quantitative analysis in the low-voltage range. As discussed in the previous section for the low-energy photon peaks of the light elements, the absorption correction factor tends toward unity as the beam energy is reduced because of the short absorption path resulting from the shallow x-ray range. An example of the change in the absorption factors for various elements in a glass under conventional and low-voltage conditions is plotted in Fig. 10.46. Because the absorption factor is often a major source of analytical error for low-energy x-rays, choosing the lowest practical energy for analysis consistent with the other factors will generally yield the smallest corrections and reduced error. The atomic number matrix factor has a more complex dependence on the beam energy, but it also tends toward unity because the ionization loss due to backscattering is reduced and there is some evidence that backscattering coefficients span a narrower range at low voltage.

The choice of correction methods is affected by similar considerations as encountered in light element analysis. The traditional *ZAF* methods in general were developed for higher beam energy and higher photon energy applications and are thus not expected to perform as well for low-energy photons. The various implementations of the $\phi(\rho z)$ procedure generally incorporate in the database actual experimental measurements of low-photon-energy peaks in various matrices measured with low beam

Table 10.12. Analysis of ZnS with Low-Voltage Microanalysis[a]

a. NIST *ZAF* (DTSA) matrix corrections

El/line	*k* value	NIST *ZAF*	Error (%)	Normalized *ZAF*	Error (%)
Zn *L*	0.683	0.730	+8.8	0.689	+2.8
S *K*	0.373	0.329	0	0.311	−5.6
Sum		1.059			

b. CITZAF matrix corrections

El/line	*k* value	CITZAF	Error (%)	Normalized	Error (%)
Zn *L*	0.683	0.721	+7.5	0.679	+1.2
S *K*	0.373	0.341	+3.5	0.311	−2.5
Sum		1.062			

c. Bastin PROZA matrix corrections

El/line	*k* value	PROZA	Error (%)	Normalized	Error (%)
Zn *L*	0.683	0.731	+8.9	0.684	+1.9
S *K*	0.373	0.339	+3.0	0.317	−3.8
Sum		1.070			

[a] $E_0 = 4$ keV. Standards: Zn metal and FeS (troilite). Zns ideal stoichiometry: Zn 0.671, S 0.329.

energy. By including these data in the fitting procedures that give rise to the $\phi(\rho z)$ equations, better accuracy is expected.

An example of a quantitative analysis of ZnS with a beam energy of 4 keV is given in Table 10.12. The results are based upon the intensities extracted by peak fitting with multiple linear least squares and locally measured peak references (measured on the standards), with the quality of the peak fit shown in Fig. 10.47, where the background can be seen to be devoid of any remnant peak structures. The *k* values extracted (with Zn metal and FeS as standards) were corrected with two *ZAF* formulations and one $\phi(\rho z)$ procedure. The results in Table 10.12 reveal that there is not much difference among the three methods in this particular case. It is suggested that the analyst performing quantitative analysis in the low-voltage regime pay close attention to the accuracy of spectral fitting to obtain robust *k* values, and to examine several matrix correction schemes to compare the spread of the calculated results. Additionally, if multielement standards with the elements of interest to the analysis are available, the reduction in matrix factors obtained with such standards will further improve the accuracy.

10.9. Report of Analysis

The Report of Analysis (ROA) is the formal record of the results of an analytical project provided to a client. The analyst may never have direct verbal contact with the client to explain the results, so the ROA must present a complete story that can stand on its own in conveying to the client

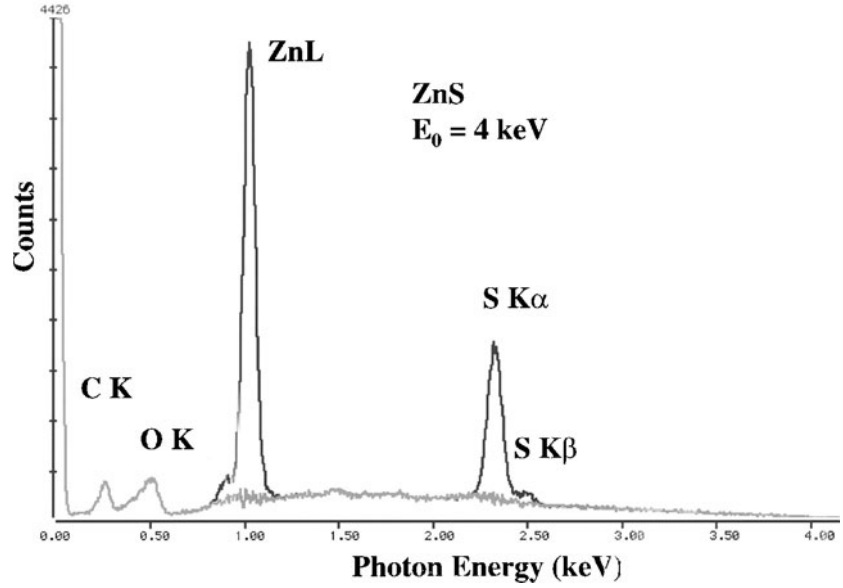


Figure 10.47. Spectrum of ZnS excited with a beam energy of 4 keV. The original spectrum and the background after peak fitting for Zn L, S Kα and S Kβ are shown.

the analytical results and their proper interpretation. In preparing a proper ROA, the analyst should consider the following general points:

1. The ROA is the permanent, official record of the analytical project, and, as such, it constitutes discoverable evidence in later legal cases such as patent disputes, product safety actions, etc. Therefore the analyst should only deal with the aspects upon which he or she is expert: the microscopy, the x-ray microanalytical procedures used, and the interpretation of the measurements within the context of analysis. Qualitative analysis, quantitative analysis, systematic and random errors and their appropriate measures, limits of detection, and the relation of analytical results spatially to structures recognized in accompanying SEM images are major topics that are likely to form elements of the ROA. The analyst should never speculate or make interpretations that are outside of the field of his or her expertise.

2. The ROA must be tied closely to the laboratory quality assurance system so that results are supported by the archival records. Thus, the numbering system for the report should transparently connect all data connected with the project, including written laboratory notebook entries, computer-generated and stored spectral and image data, and calculations that support interpretation of that data. SEM/x-ray microanalysis is strongly dependent on computer-aided imaging and analysis systems. In this regard it must be noted that computer storage can be an ephemeral thing. As a computer-based society, we are constantly losing the recent past as data are accidently left behind on obsolete storage media. Equally deadly to the recovery of data is the constant evolution of instrumentation operating software and

the computer operating systems that support that software. Even if the data are recorded on a medium that is still readable in the future, the software system that generated the images and spectral files may no longer be viable, rendering the stored data useless if it is in a format that is rigidly locked to a specific commercial product. Part of a robust quality assurance plan is to determine the critical time period for which computer files must be recoverable. Critical data should be stored, if possible, in both the file structure of the software system and separately as file structures that can be expected to be universally interpretable. Thus, x-ray spectra can be stored as simple text files consisting of the intensity value for each channel, with a text header that gives the eV/channel, which is the critical information to reconstruct the spectrum using any plotting software, the beam energy, beam current, accumulation time, and spectrometer solid angle and window information, as well as information that identifies the specimen. Images should be stored in formats that are now widely exchanged, such as Bitmap, TIFF, or jpeg, and should include in a text header the magnification, beam energy, detector information, and specimen description.

The following list is an example of a possible structure for a Report of Analysis. This format should only be considered as a guideline, with the specific form to be adjusted to the analyst's particular circumstances. Although this proposed ROA structure form may seem long, the information it provides should permit understanding of the analyst's approach and reconstruction of the results in the future even if the analyst is not available. If the data are properly archived, it should be possible to reexamine the basic SEM image and x-ray spectral measurements at any time in the future with general imaging and graphical analysis software.

Report of Analysis Format

1. Title of the ROA.
2. ROA number related to the laboratory's quality assurance system. A typical format is: year–organizational unit identification–report sequence number.
3. Name(s) of staff involved and responsibilities. Staff education and experience records may be appropriate.
4. Customer request: What was requested by the customer? If possible, the customer's request document should be included.
5. Specimen handling record: date of receipt, tracking, and specimen security arrangements. Note that specimen "chain of custody" issues are likely to be of great importance when legal issues are anticipated.
6. Specimen description, as received. The specimen should be documented with photographs, possibly including low-magnification optical stereomicroscope images in color. If the client provides images that to locate the area(s) of interest, these should be included.
7. Description of methods. A short description of the methods of SEM imaging and x-ray microanalysis should be provided so that a reader with little or no background in the techniques can obtain a first level of understanding, supported by appropriate references to texts for more detail and to describe specific procedures used.

8. Analytical strategy: What analytical route was chosen? For example, EDS only or EDS/WDS? Low-voltage microanalysis or conventional analysis? Single-point analysis, line scans, and/or x-ray maps, etc.
9. SEM images showing the locations of analysis. If the client provided images to aid in locating the area(s) of interest, the relation between the analyst's images and the client's images should be established.
10. Qualitative analysis: What elements were identified? For minor ($0.01 \leq C \leq 0.1$ mass fraction) and trace ($C < 0.01$) constituents, how was the identification confirmed?
11. Quantitative analysis: What standards were used? How was the standard composition determined (e.g., independent chemical analysis for which the certificate of analysis was archived). Describe the specific matrix correction procedure utilized, including known strengths and weaknesses.
12. Error estimates: (a) Random error: The statistical analysis presented in Chapter 9 enables the analyst to calculate from the measured peak and background counts a robust estimate of the standard deviation of the concentration, which can be expressed in terms of a confidence interval within which the true value is expected to reside. (b) Systematic error: The matrix correction procedures introduce errors that are related to limitations in the underlying physical models or in the empirical fits to experimental measurements that are used to calculate the correction factors. When these procedures have been tested against suites of homogeneous multielement standards whose compositions are known by independent chemical analysis (e.g., wet chemistry and gravimetry), the resulting error distributions, examples of which have been presented in Chapter 9, can be used to estimate the systematic error of the analysis. Thus, analysis of flat, polished unknowns with *ZAF* matrix corrections and pure-element or simple compound standards has been shown to produce an error distribution such that 95% of all analyses fall with $\pm 5\%$ relative error, that is, within $\pm 5\%$ of the amount present. Note that with the investment of a high dose to produce spectra with large numbers of counts, it is almost always possible to obtain a random error which is much smaller than the systematic error. Although achieving a small random error is useful when specimens of very similar composition are to be compared, the systematic error is the factor that controls the analyst's ability to report absolute concentration values.
13. A statement of the limit of detection for each constituent of interest under the chosen measurement conditions may be required. For those constituents that were indicated to be of interest by the client, but were not actually detected, the limit of detection establishes the threshold value below which that constituent could be present, but not detected with the measurement conditions chosen by the analyst.
14. For results presented graphically such as line traces, show the relationship of the plotted values to the specimen features in the micrograph.
15. For data presented as x-ray maps, carefully explain the nature of the image that is presented. Can the gray level or color scale of the image be

interpreted in any quantitative fashion? What constitutes a significant result/feature in the map? What, if any, are the artifacts in the image? Reporting this last topic is critical because the analyst may automatically dismiss known mapping artifacts because of prior knowledge and experience, whereas the client who does not have such experience may find the image artifacts to be interesting features and may proceed to interpret artifacts as useful results.

16. Summary of the most important results.

References

Armstrong, J. T. (1991) In Electron Probe Quantitation (K. F. J. Heinrich and D. E. Newbury, eds.) (Plenum Press, NY), p. 261.

Armstrong, J. T. and P. R. Buseck (1975) Anal. Chem. *47,* 2178.

Barkalow, R. H., R. W. Kraft, and J. I. Goldstein (1972). *Met. Trans.* **3,** 919.

Bastin, G. F., and H. J. M. Heijligers (1984). In *Microbeam Analysis—1984,* (A. D. Romig, Jr., and J. I. Goldstein, eds.), San Francisco Press, San Francisco, p. 291.

Bastin, G. F., and H. J. M. Heijligers (1986a). *X-ray Spectrom.* **15,** 143.

Bastin, G. F., and H. J. M. Heijligers (1986b). Quantitative Electron Probe Microanalysis of Boron in Binary Borides, Internal Report, Eindhoven University of Technology, Eindhoven, The Netherlands.

Bastin, G. F., and H. J. M. Heijligers (1990a). *Scanning* **12,** 225.

Bastin, G. F., and H. J. M. Heijligers (1990b). In *12th International Congress on Electron Microscopy* (L. Peachey and D. B. Williams, eds.), San Francisco Press, San Francisco, Vol. 2, 216.

Bastin, G. F., F. J. J. van Loo, and H. J. M. Heijligers (1984a). *X-ray Spectrom.* **13,** 91.

Bastin, G. F., H. J. M. Heijligers, and F. J. J. van Loo (1984b). *Scanning* **6,** 58.

Bastin, G. F., H. J. M. Heijligers, and F. J. J. van Loo (1986). *Scanning* **8,** 45.

Bright, D. S. (1995). *Microbeam Anal.* **4,** 151.

Cosslett, V. E., and P. Duncumb (1956). *Nature* **177,** 1172.

Duncumb, P. (1957). In *X-ray Microscopy and Microradiography* (V. E. Cosslett, A. Engstrom, and H. H. Pattee, eds.), Academic Press, New York, p. 617.

Echlin, P. (1999). *Microsc. Microanal.* **4,** 577.

Hall, T. A. (1968) "Some Aspects of the Microprobe Analysis of Biological Specimens," in Quantitative Electron Probe Microanalysis, (K. F. J. Heinrich, ed.) (NBS Special Publication 298, Washington) p. 269.

Hayashi, S. R., and R. B. Bolon (1979). *Microbeam Analysis,* San Francisco Press, San Francisco, p. 310.

Heinrich, K. F. J. (1986). In *Proceedings 11th International Conference on X-ray Optics and Microanalysis* (J. D. Brown and R. H. Packwood, eds.), University of Western Ontario, London, Ontario, Canada, p. 67.

Henke, B. L., P. Lee, T. J. Tanaka, R. L. Shimabukuro, and B. K. Fujikawa (1982). *Atomic Data Nuclear Data Tables* **27,** 1.

Hovington, P., D. Drouin, and R. Gauvin (1997). *Scanning* **19,** 1.

Ingram, P., D. A. Kopf, and A. LeFurgey (1998). *Scanning* **20,** 190.

Kitazawa, T., H. Shuman, and A. P. Somlyo (1983) Ultramicroscopy, 11, 251.

Kyser, D., and Murata, K. (1974). IBM J. Res. Dev., *18,* 352.

Marshall, D. J. and T. A. Hall (1966) in X-ray Optics and Microanalysis (R. Castaing, J. Deschamps, and J. Philibort, eds.) (Hermann, Paris) p. 374.

Mott, R. B., R. Batcheler, and J. J. Friel (1995). *Microscopy Society of America Proceedings* (A. Garrat-Reed, ed.), Jones and Begell, New York, p. 592.

Newbury, D. E., and D. S. Bright (1999). *Microsc. Microanal*. **5,** 333.

Newbury, D. E., and R. L. Myklebust (1991). In *Microbeam Analysis—1991* (D. G. Howitt, ed.), San Francisco Press, San Francisco, p. 561.

Newbury, D. E., R. L. Myklebust, K. F. J. Heinrich, and J. A. Small (1980). "Monte Carlo Electron Trajectory Simulation—An Aid for Particle Analysis" in Characterization of Particles (K. F. J. Heinrich, ed.) (Washington, NBS) 39–60.

Newbury, D. E., C. E. Fiori, R. B. Marinenko, R. L. Myklebust, C. R. Swyt, and D. S. Bright (1990a). *Anal. Chem.* **62,** 1159A.

Newbury, D. E., C. E. Fiori, R. B. Marinenko, R. L. Myklebust, C. R. Swyt, and D. S. Bright (1990b). *Anal. Chem*. **62,** 1245A.

Newbury, D. E., R. B. Marinenko, R. L. Myklebust, and D. S. Bright (1991). In *Electron Probe Quantitation* (K. F. J. Heinrich and D. E. Newbury, eds.), Plenum Press, New York, p. 335.

Pouchou, J. L. and F. Pichoir (1991). In *Electron Probe Quantitation* (K. F. J. Heinrich and D. E. Newbury, eds.), Plenum Press, New York, p. 31.

Pratt, W. K. (1978). *Digital Image Processing* Wiley, New York.

Roomans, G. M. (1981). In *SEM/1981/II,* SEM, Inc., AMF O'Hare, Illinois, p. 345.

Roomans, G. M. (1988). *J. Electron Microsc. Technique* **9,** 19.

Russ, J. C. (1995). *The Image Processing Handbook,* CRC Press, Boca Raton, Florida.

Small, J. A., K. F. J. Heinrich, C. E. Fiori, R. L. Myklebust, D. E. Newbury, and M. F. Dilmore (1978). In *SEM/1978/I*, SEM, Inc., AMF O'Hare, Illinois, p. 445.

Small, J. A., K. F. J. Heinrich, D. E. Newbury, and R. L. Myklebust (1979). In *SEM/1979/II,* SEM, Inc., AMF O'Hare, Illinois, p. 807.

Statham, P. J. (1979). *Mikrochem. Acta* **8**(Suppl.), 229.

Statham, P. J. and J. B. Pawley (1978). In *SEM/(1978)/I,* SEM, Inc., AMF O'Hare, Illinois, p. 469.

11

Specimen Preparation of Hard Materials: Metals, Ceramics, Rocks, Minerals, Microelectronic and Packaged Devices, Particles, and Fibers

This chapter outlines a variety of sample preparation procedures for imaging and x-ray analysis of hard materials in the SEM. Several special and relatively new techniques, such as the use of focused ion beams for preparation of cross sections of various materials, are also discussed.

11.1. Metals

The proper and successful preparation of metallic samples will ultimately depend on the type of information that is to be derived from the sample using the SEM. Preparation of fracture surfaces for topographic imaging in the SEM is quite different than preparation of samples for microchemical analysis in the SEM. In this section we will briefly discuss the specimen preparation techniques for metals that are applicable to topographic imaging of rough samples (e.g., fracture surfaces), microstructural analysis using secondary or backscattered electrons, and microchemical analysis using x-rays.

11.1.1. Specimen Preparation for Surface Topography

One of the most useful features of the SEM is that many samples can be imaged directly with little or no specimen preparation. The first step is to get the sample to a size that will fit in the sample stage of the SEM. This process may require careful cutting of the sample using techniques that will not contaminate the sample or damage it. The sample may be cleaned in an appropriate solvent, such as acetone, with the use of an ultrasonic cleaner (ASTM, 2000). A final wash with methanol will help remove any

remaining surface films. It may be appropriate to further clean the sample in a low-energy plasma as discussed in Section 11.6.1. It is important to note that the solvent chosen should not damage or modify the surface. If the surface contains fine particles that are easily separated from the surface, it is best to image the sample as-is, with no initial sample preparation. The cleaned sample must be mounted onto an appropriately sized specimen stub. Generally, a conductive paint is used to secure the sample to the sample stub. If a nonconductive glue must be used, it is necessary to apply a stripe of conductive paint to the sample to provide an electrically conductive path to ground. In some cases, for example, ceramics or metal particles in a ceramic matrix, it is necessary to coat the sample with a thin, electrically conductive layer. The techniques for this are discussed further in Chapter 15. It is best to image most rough samples without a coating first, because the coatings are relatively easy to apply, but impossible to remove if they interfere with imaging or microanalysis of the sample.

The use of a partial-vacuum SEM (Chapter 5) may obviate the need for much initial specimen preparation for topographic imaging. The ability to use a relatively poor vacuum, from as low as 10 Pa (0.076 torr) in a partial-vacuum SEM to as high as 2600 Pa (20 torr) in an environmental SEM, allows samples that are oily or wet to be imaged without any cleaning. The relatively poor vacuum in the SEM prevents the immediate drying out of water or volatilization of oily films while allowing the sample to be imaged. Also, nonconductive samples may be imaged successfully without conductive coating at pressures starting in the range 10–50 Pa (0.076–0.38 torr).

11.1.2. Specimen Preparation for Microstructural and Microchemical Analysis

The primary goal of microstructural or microchemical analysis is to reveal the constituents and the structure of metals and alloys through the use of imaging and elemental analysis in the SEM. The preparation steps for producing a sample are identical to those used to produce a polished sample for light microscopy. The proper preparation steps are extremely important because the true microstructure or microchemstry of the sample may be obscured or altered by the use of poor technique. Because of the diversity in equipment and techniques and the wide variety of materials that may be encountered, the discussion will be limited to a general overview of the techniques that may be employed to produce a suitable sample. The reader is encouraged to seek more detailed information from Van der Voort (1984), ASM (1985, 1986), ASTM (2000), and Bousfield (1992).

11.1.2.1. Initial Sample Selection and Specimen Preparation Steps

One of the most important steps is to select an appropriate area of the sample to prepare. This is important because the polishing steps required can be relatively time-consuming and therefore expensive.

A specimen of suitable size must be removed from the sample. In most metallurgical problems, polishing is limited to an area of a few square

centimeters. There are many ways to section the samples, including blade sawing, wire sawing, abrasive cutting, fracturing (best for more brittle materials), shearing, spark erosion, and microtomy. Usually, the sample is cut with some sort of saw. It is important not to damage the specimen during the sawing operation by allowing it to get too hot (use a water-cooled saw) or to be heavily deformed by a saw blade or cut-off wheel that is too coarse. Special care should be taken if the cut surface is ultimately to be polished. If this is the case, a slow-speed diamond wheel cut-off saw or a slurry saw should be utilized to minimize the zone of deformation around the cut.

The next step will depend on the size of the sample and the desired information to be obtained from the sample. When working with larger bulk samples, it may not be necessary to embed the sample in a mounting material. However, if the sample is small or difficult to hold during polishing, mounting becomes necessary. If sample edges are to be examined, mounting along with other techniques used to preserve the sample edge may also be employed. If required, the sample should be mounted using either a Bakelite or epoxy mounting medium. There are two important considerations to keep in mind when selecting a mounting material. The first is that although the metal sample may be conductive, the mounting medium generally is not. There are some conductive mounting media, but these are more expensive and more difficult to work with than the standard mounting media. Also, the materials in these compounds may smear across the surface during polishing and interfere with microanalysis of the sample. The other important criterion is the height of the mounted sample. Most SEM instruments have a limited maximum sample height that can fit in the specimen stage. This dimension should be kept in mind during sample preparation. Generally, a sample thickness of 1–2 cm should fit in most instruments.

11.1.2.2. Final Polishing Steps

The surface to be prepared for microstructural analysis is polished using a graded sequence of abrasive materials. The first step involves coarse grinding using an abrasive of 60–180 grit. When the damage from the cutting operation is removed, a sequence of finer abrasives is used. A common sequence uses 120-, 240-, 320-, 400-, and 600-grit abrasive papers. This sequence may be modified depending on the initial surface roughness and the hardness of the material. These grinding steps are usually carried out wet to prevent heating of the specimen and to promote a better surface finish.

After the specimen has been ground to a 600-mesh finish, the sample must be polished to produce a flat, scratch-free surface. Polishing or lapping the sample with abrasives having particle sizes starting with 30 μm for coarse polishing and proceeding to 1 μm or smaller for final polishing accomplishes this objective. For some samples and for softer metals it may be necessary to electropolish the sample as a final step to remove any deformation induced by the grinding and polishing operation. The final result of careful grinding and polishing will be a flat, scratch-free surface with a high reflectivity.

Microstructural features, in most materials, must be revealed by etching the sample. There are many ways to produce an etched condition that allows the microstructure of the specimen to be studied. Etching may be accomplished through the use of chemical attack solutions, heat tinting, or tint etching. Generally for SEM applications, it is best to avoid techniques like heat tinting and tint etching, which result in the formation of oxides on the sample surface. Chemical etchants are developed by trial and error and extensive lists of useful etchants can be found in Van der Voort (1984) and ASM (1985, 1986). The etchant reveals the microstructure of the specimen through a selective dissolution of the microstructural constituents. It is best to start with short etch times to avoid overetching of the sample and obscuring important details of the microstructure. Etching can modify the composition of a specimen in the critical near-surface region that is sampled by x-ray microanalysis. As an alternative to etching, often it is possible to use backscatter electron imaging of the specimen with no etch applied to observe the microstructure.

The final step before the specimen is examined in the SEM may be to coat the sample with a thin conductive layer. It is possible to examine mounted specimens with no conductive coating as long as a conductive path to ground has been established using conductive paint or other means. However, it is nearly impossible to avoid scattered electrons impacting the mounting medium and causing a charge buildup that will eventually distort the image or cause instability. Charging of the mounting medium may be avoided by coating the entire surface of the mount with a conductive layer as discussed in Chapter 15.

11.1.2.3. Preparation for Microanalysis

In most cases, the preparation of samples for qualitative microanalysis is identical to that used for microstructural analysis. However, one must be careful to avoid the introduction of artifact x-ray peaks from the polishing materials used or the specimen mounting material within the microscope. Preparation-induced artifacts may be simply the incomplete removal of polishing compound residues from the sample. Polishing compounds may contain many different elements that can result in additional peaks in an x-ray spectrum. The techniques used for mounting the sample in the SEM may also introduce additional x-ray peaks due to the sample holder. A record of the materials used in specimen preparation should be retained in case a question arises subsequently on possible contamination.

Quantitative x-ray analysis requires that the specimen be perfectly flat (see Chapter 9). Generally, to obtain a flat surface, the same procedures used to polish samples as discussed for microstructural imaging are employed. For the best quantitative analysis results the samples should be not only perfectly flat, but as scratch-free as possible. Also, there should be as little surface topography as possible. The first criterion for a flat and scratch-free surface may be met through the use of very careful polishing of the specimen. The second criterion for little surface topography requires

that metals and alloys be analyzed unetched (Heinrich, 1981; Scott and Love, 1983). These criteria can also be applied to the preparation of other materials for quantitative analysis.

In many cases the specimen to be studied will consist of many phases of different hardnesses. This will often result in steps at the interphase interfaces. This effect can also be observed across grain boundaries in single-phase materials due to the dependence of polishing rate on grain orientation and can be minimized by using low-nap polishing cloths. Steps in the surface of the sample can result in errors in quantitative analysis due to x-ray absorption effects. In many cases, it is necessary to perform light microscopy of the sample in an etched condition prior to performing quantitative analysis in the SEM. This requires the sample to be polished and etched for light microscopy and then repolished to remove the etched surface. Polishing, etching, and repolishing may result in a sample that is no longer flat, but may consist of surface steps at interphase interfaces and also a hilly surface that can result in errors in quantitative analysis. In some cases it is absolutely necessary to etch the sample to visualize the region of interest in the specimen. In these cases the best course of action is to polish and etch the sample and then use microhardness indentations to outline the area of interest. The specimen is then carefully repolished to remove the surface topography, but not the microhardness indentations.

Quantitative analysis in the SEM is best carried out using standards of known composition (see Chapter 9). The specimen to be analyzed should be treated in the same manner as the standards. Therefore, if the standards are coated for conductivity, then the specimen should be coated using the exact same conditions to reproduce the coating thickness. Best practice for quantitative analysis requires that the standards and the unknown be coated at the same time to ensure that there are no differences in coating thickness (Heinrich, 1981; Scott and Love, 1983). This is of particular importance when low-energy x-ray lines are used for quantitative analysis.

11.2. Ceramics and Geological Samples

The preparation of ceramics and geological samples is in many ways similar to the preparation for metals and alloys described in the previous section. Many geological samples, however, such as soils and sediments, contain unwanted components and are wet in their normal state. These materials can be prepared for SEM examination by techniques discussed for polymer or biological and hydrated materials (Chapters 12–14). Other ceramic and geological samples are in the form of particles and must be embedded for specimen preparation prior to examination in the light microscope. Sample preparation for individual particle examination is discussed in Section 11.7. Detailed description of preparation procedures for ceramic and geological samples are given in Lee and Rainforth (1994), Buehler (1973), Bousfield (1992), Miller (1988), Taylor and Radtke (1965), and Humphries (1992).

11.2.1. Initial Specimen Preparation: Topography and Microstructure

The main difference between metals and ceramic or geological samples is that the later samples are generally insulators and will require conductive coating for good imaging. If coating for conductivity is not possible, imaging with low voltages and beam currents or examination in an environmental or partial-vacuum SEM may provide adequate results.

11.2.2. Mounting and Polishing for Microstructural and Microchemical Analysis

Mounting procedures for ceramics are similar to those used for metals. Hand specimens of rocks will require cutting to obtain pieces of suitable size and shape for mounting and polishing. It may be helpful to use a conductive mounting medium to prevent charging of the sample during imaging in the SEM. In many applications, particularly for geological specimens, thin sections are required so that viewing by transmitted light is possible. Prior examination in a transmitted-light optical microscope is necessary to locate areas of interest. In many dedicated SEM instruments for x-ray microanalysis, light-optical microscopes are available in the instruments as an aid in the analysis. The thin-section preparation procedure is the same as for ordinary thin sections in the first stages using a strong adhesive to attach the rock slice to the glass slide. Care must be taken so that the glue is unaffected by subsequent etching. The slice is ground to a thickness somewhat greater than the 30-μm final thickness required. Care should be taken in the selection of the polishing compound because many mineral phases are quite hard. Diamond pastes are commonly used for polishing of ceramics and geological samples due to the extreme hardness of the diamond. Progressively finer grades of abrasive are employed. As in preparation of metal samples, the specimens should be thoroughly cleaned after each stage to avoid having abrasive materials embedded in pores and cracks of the specimens.

Procedures for microchemical analysis are similar to those used for metals and alloys.

11.2.3. Final Specimen Preparation for Microstructural and Microchemical Analysis

Ceramic and geological samples, like metals, may require etching to permit microstructural features to be imaged or analyzed in the SEM. Generally, etches for ceramic samples and some geological samples contain HF. Polished thin sections with a thickness of about 30 μm can also be etched, but one must be very careful not to degrade the glue that attaches the sample slice to the glass slide. It is possible to use thermal etching to produce surfaces suitable for microstructural imaging. However, the use of backscattered electron imaging due to atomic number contrast is commonly used to discriminate among phases in the specimen. For microanalysis any

specimen etching must be removed. Coating of the ceramic or geological specimen is required due to the insulating nature of most ceramics and oxides (see Chapter 15). Coating with carbon or other metal films should allow imaging and microanalysis to be performed. The preferred coating element for x-ray analysis is carbon because it has a minimal effect on the x-ray spectrum. In addition, carbon coating is the best choice for cathodoluminescence work.

11.3. Microelectronics and Packages

The preparation of microelectronic devices, integrated circuits, and packaged devices utilizes many of the same techniques that have been discussed previously for bulk ceramics, geological samples, and metals. The examination of microelectronic parts can be performed at a variety of levels. Classical SEM techniques can be used to examine integrated circuits and their packages. Important classes of problems that may be solved by SEM imaging are the condition of the integrated circuit (metallization and passivation integrity, etc.), the electrical attachments to the integrated circuits (condition of the wire bonds, etc.), the attachment of the integrated circuit within the package, and the attachment of the packaged integrated circuit to a printed wiring board. Because these devices and packages are constructed from metals, ceramics, or polymers, most of the techniques discussed previously are of use (ASM, 1996; Bousfield, 1992).

The examination of specific parts of the integrated circuit are more difficult and may involve voltage contrast, electron-beam induced current (EBIC), or even cross-sectional transmission electron microscopy. Recent advances in the use of focused ion beams (FIB) have allowed site-specific cross sections of integrated circuits to be prepared. FIB will be discussed in more detail in Section 11.6.2.

11.3.1. Initial Specimen Preparation

Initial specimen preparation depends on the type of information desired from the sample. A packaged integrated circuit is treated differently than an unpackaged die or patterned wafer. Generally, failure analysis will be performed on a packaged part. The first step is to expose the integrated circuit. For hermetic packages with soldered lids it is fairly simple to use a sharp tool to cut through the solder on one corner of the cover and then gently remove the cover. This should expose the integrated circuit, wire bonds, and lead frame for study. When plastic packages are encountered, it is possible to find a seam between the two halves of the package (usually where the lead wires exit the package). Gentle prying with a sharp tool at this interface can sometimes open the package for further analysis. In either case, care should be taken to avoid damaging the delicate lead wires or integrated circuit inside the package.

Once the package has been opened, it is possible to examine the integrated circuit in the SEM. It is often best to avoid removing the integrated

circuit entirely from the package because the package is a convenient way to mount the sample in the SEM. It is best and easiest to connect all of the pins or leads of the package together and then connect these to ground. This procedure will eliminate some of the charging that will occur due to the glass passivation and other insulating materials present on most integrated circuits. Once the integrated circuit is mounted in the SEM, various features such as passivation integrity and step coverage, bond pad integrity, or wire bond quality may be studied. Some charging artifacts may result during the imaging of the passivated integrated circuit. It is possible to reduce the charging by either operating at low voltage (where a balance between positive and negative charging can be found) or imaging with backscattered electrons. Backscattered electron imaging at higher accelerating voltages may be used to image details of the metallization under the passivation. This technique has been successfully used to determine the degree of stress voiding in Al metallizations. If it is not possible to image the sample with the passivation intact, it can be removed by careful reactive ion or chemical etching.

If more detailed information is needed, like various metal layer thicknesses, passivation thickness, or other structural details, the integrated circuit must be cross-sectioned. If the integrated circuit is bonded into a package, it is best to section through the entire integrated circuit and the package. Often a typical packaged sample can be cross-sectioned with a slow-speed saw. In some cases, it may be better to perform the cross sectioning without the use of a lubricant to prevent contamination of the integrated circuit surface. After the cross section has been cut with a diamond saw, standard mechanical polishing techniques are used. Some chemical etching or ion beam etching may be useful to bring out structural details. In modern laboratories, it is much more common now to use focused ion beam (FIB) cross sectioning to observe subsurface structural details. Figure 11.1 shows an FIB-prepared cross section through an integrated circuit.

Care should be taken during the mounting of packaged integrated circuits. Once the cover or lid is removed from the package, it is helpful to carefully fill the opening containing the integrated circuit with a liquid epoxy. If it is desired to section through a specific part of the circuit, a clear epoxy should be used so that the progress of the grinding and polishing can be monitored. It may be necessary to embed the samples under vacuum to remove or minimize the amount of trapped air. Another useful approach is to use an epoxy that is filled with glass spheres. These fillers reduce the amount of expansion and contraction that occurs during the curing of the epoxy and may prevent wire leads or wire bonds from being damaged.

11.3.2. Polishing

Polishing of the mounted samples is accomplished using the same techniques as have been described previously for metals, geological samples, and ceramics. Packaged integrated circuits may be difficult to polish due to the wide range of hardness of the materials involved. Solders are

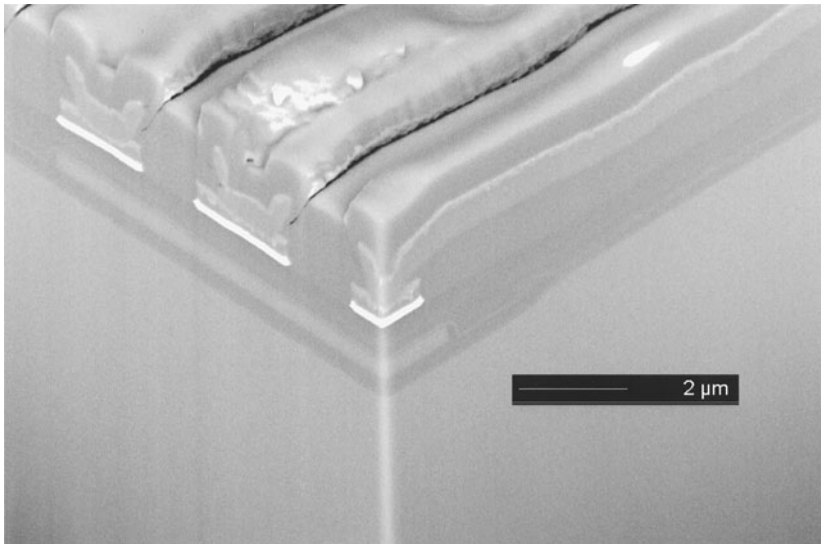


Figure 11.1. SEM image of a cross section through an integrated circuit made using a focused ion beam system. Two cross sections at 90° to each other have been prepared to allow the contact between the upper- and the lower-level aluminum interconnects to be examined.

extremely soft compared to the ceramic of the package. Careful polishing using light loads and low-nap clothes usually will produce suitable results.

11.3.3. Final Preparation

For most microelectronic devices there will be sufficient contrast from atomic number differences to allow the various features of interest to be visualized. In some cases it may be necessary to etch the sample to improve the ability to image the details of the structure. Wet chemical etches will adequately delineate the metallic portions of the sample, whereas plasma etching may allow the structure of the ceramic and oxide components of the structure to be visualized. There are also stains or etches that delineate doped regions of the semiconductor for SEM imaging.

11.4. Imaging of Semiconductors

Not only is it sometimes desirable to understand the structure of an integrated circuit or a semiconductor device, but it may be equally important to understand and characterize the electrical performance of the device. Voltage-contrast and charge collection microscopy may require special surface preparation to be performed. For a discussion of the use of SEM to characterize semiconductors see Chapter 5 as well as Holt (1989), Davidson (1989), and Joy (1989). Specific specimen preparation procedures are discussed in the following sections.

11.4.1. Voltage Contrast

Voltage contrast is a useful technique for the study of active microelectronic circuits. Voltage differences in the metallizations on the integrated circuit result in a change in contrast observed as a function of the voltage level. Accurate voltage contrast imaging is generally performed on devices that have not been glass-passivated or have had the passivation removed. It is very difficult to observe voltage contrast on passivated samples due to the buildup of charge in the passivation layer. Voltage contrast can sometimes be observed on passivated devices for 10–100 scans before the contrast quickly fades away (Davidson, 1989).

The best approach to obtaining samples for voltage contrast is to obtain devices prior to the deposition of passivation. If this is not possible, the passivation must be removed for best results. It is possible to selectively remove the passivation that covers the area of interest. Semiconductor devices are made up of many layers, which may include doped silicon, silicon dioxide, silicon nitride, polycrystalline silicon, and metal. Polycrystalline silicon and metal are the main materials used for interconnections. The upper-level metallizations are usually insulated from other metallization layers with an oxide. Also, the complete device may be protected by an overall passivation layer, usually an oxide or possibly a nitride layer.

Removal of the passivation and other insulating layers can be achieved through the use of chemical or plasma techniques. A buffered HF solution (10–20% HF by volume in saturated ammonium fluoride) is usually used for passivation and oxide insulator removal. Careful control of the etching time is required to prevent attack of the underlying Al metallizations. Plasma methods using CH_4 may also work well in removing the oxides and nitrides (Davidson, 1989). Organic layers may be removed with oxygen-assisted plasma etching. FIB techniques may also be used to selectively remove passivation layers.

11.4.2. Charge Collection

The signals used in the charge collection mode are the currents or voltages induced in the specimen by the electron beam. The most important type of signal in charge collection mode is electron-beam induced current (EBIC). EBIC is used to study Schottky barriers as well as diffused and ion-implanted *p–n* junctions. The examination of *p–n* junctions in devices requires that the device be free of passivation. EBIC imaging can be achieved with the beam perpendicular or parallel to the *p–n* junction. When *p–n* junctions are studied using EBIC it is important to make good ohmic contacts to the sample.

When EBIC is used to study bulk semiconductor materials without *p–n* junctions, a Schottky barrier contact must be applied to the surface of the sample. The deposition of high-quality Schottky barrier contacts requires some care. The sample must be first carefully cleaned and the barrier metal must be deposited in a clean metal deposition system. For *p*-type material,

evaporated Al or Ti layers about 100 nm thick work best. The best metals for *n*-type material are AuPd and Au of about the same thickness. More detailed description can be found in Reimer (1998) and Holt (1989).

11.5. Preparation for Electron Diffraction in the SEM

Two techniques that are commonly used to study crystallography in the SEM are electron channeling patterns (ECP), obtained by rocking the electron beam through the Bragg diffraction angle while recording the backscattered electron signal (Reimer, 1998; Joy, 1989; Joy *et al.,* 1982), and electron backscatter diffraction (EBSD), which requires that the electron beam be held stationary on the sample surface while a position-sensitive electron detector (usually a phosphor imaged with a video-rate or a slow-scan CCD camera) is used to collect the pattern (Randle and Engler, 2000; Michael, 2000). For more detail on ECP and EBSD, see Chapter 5. Specimen preparation for diffraction studies is discussed in the following sections.

11.5.1. Channeling Patterns and Channeling Contrast

Samples for electron channeling patterns and contrast must be flat, clean, and strain-free because the contrast comes from a surface layer only 50–100 nm thick. Conventional metallographic (mechanical) polishing techniques are often unsuitable due to surface contamination that is difficult to remove and surface damage that can be introduced during polishing. The best approach is to use either chemical or electrochemical polishing. Final polishing may also be accomplished using vibratory polishing. Channeling contrast may be obtained from unprepared samples as shown in Fig. 11.2, which is a channeling contrast image of the terminus of a weld in Ta. The sample was cleaned in ethanol before imaging in the SEM. Metals and alloys will produce the best patterns after electropolishing, whereas it may be necessary to polish and thermally etch ceramic materials. Silicon and germanium may be chemically polished using a solution of 90% HNO_3 and 10% HF. Silicon wafers will give adequate patterns without much effort. Processed wafers will need oxide layers removed with an etch (2% HF solution). Samples for electron channeling should be washed in electronic-grade ethanol to remove any trace residues from the surface (Reimer, 1998; Joy, 1989; Joy *et al.,* 1982).

11.5.2. Electron Backscatter Diffraction

Proper specimen preparation is most often cited as the main difficulty of applying the EBSD technique to a wide range of materials. The criteria for a good sample for EBSD are very similar to those for electron channeling patterns. The samples must be clean and free of surface deformation resulting from specimen preparation. Specimen preparation may be

as simple as ensuring the sample surface is clean and free from contamination. This is true for crystals grown in clean growth apparatus as well as many naturally occurring single crystals. Thin films grown on substrates may also be studied directly with EBSD.

EBSD is more difficult with insulating samples. If the sample is allowed to charge under the electron beam, it can begin to act as an electron mirror, which can cause artifacts in or completely obscure the EBSD patterns. In these cases it is necessary to lightly coat the sample for conductivity. Ion sputtering of thin chromium coatings is one of the better coating techniques for EBSD. Very light carbon or aluminum applied by evaporation can also work well. An example of the effect of specimen coating is shown in Fig. 11.3. Figures 11.3a–11.3f are EBSD patterns collected at 10, 20, and 30 kV from bare Si and Si coated with 20 nm of Al. The EBSD patterns from the bare Si are quite clear at all accelerating voltages. The pattern collected at 10 kV from the Si coated with 20 nm of Al shows only very slight contrast. As the accelerating voltage is increased, the EBSD patterns become clearer as a result of decreased scattering in the coating of the higher energy electrons. Figure 11.3 demonstrates that coated specimens may require higher accelerating voltages to obtain good-quality patterns.

The preparation of bulk samples requires the sample to be sectioned and polished. Metallographic polishing followed by a light attack etch can

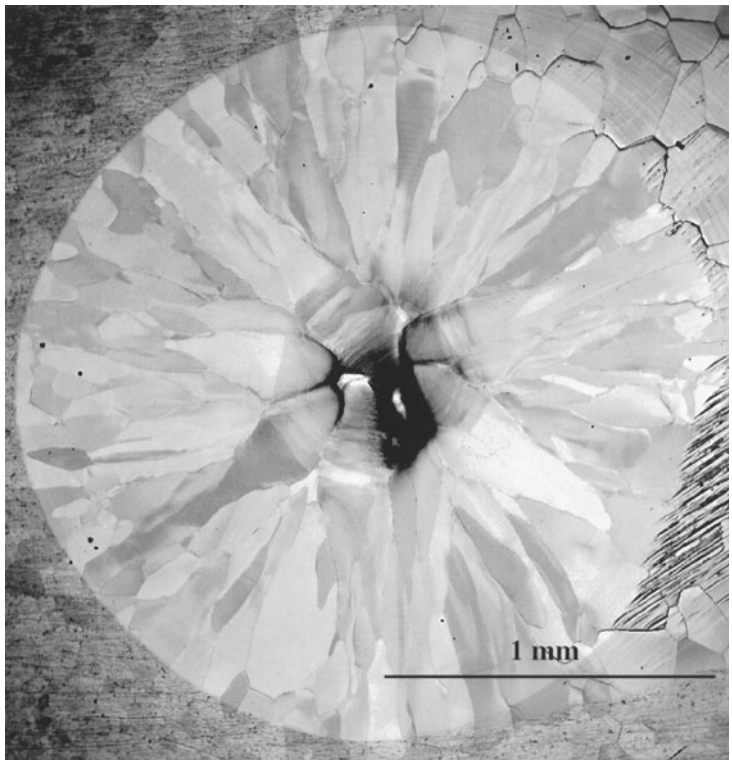


Figure 11.2. Channeling contrast image of the terminus of a laser weld in a Ta alloy. The grain-to-grain differences in gray level are due to channeling of the electrons. This image was obtained from an as-welded sample that had been cleaned in ethanol.

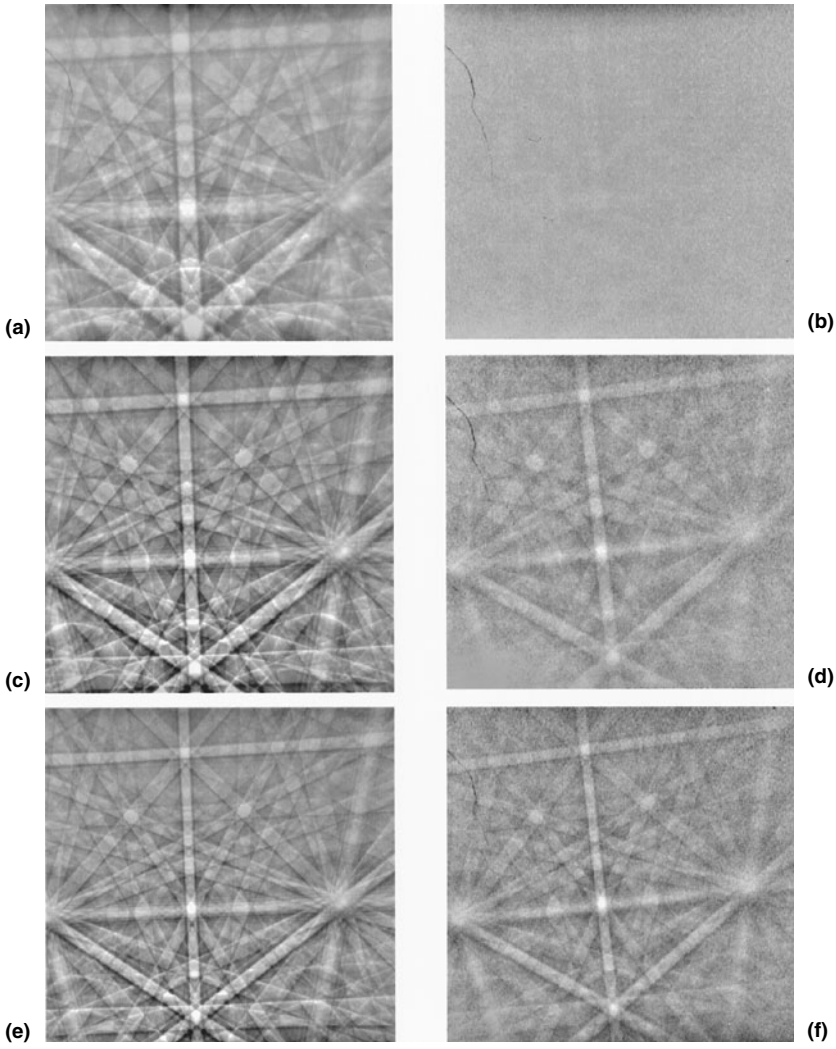


Figure 11.3. EBSD patterns from bare silicon and silicon coated with 20 nm of aluminum. (a) 10 kV, bare silicon. (b) 10 kV, 20 nm of aluminum on silicon. (c) 20 kV, bare silicon. (d) 20 kV, 20 nm of aluminum on silicon. (e) 30 kV, bare silicon. (f) 30 kV, 20 nm of aluminum on silicon.

be used to prepare samples for EBSD examination. Harder metals and alloys may be examined in the as-polished condition with no etching. Softer metals, like lead, aluminum, and tin, may require extra care during polishing and attack etching to prepare good samples for EBSD. Electropolishing is the best and most widely applicable technique for the preparation of metals and alloys.

Ceramics, geological samples, and oxides may be prepared using standard metallographic polishing techniques followed by the application of a conductive coating. Figure 11.4 shows an EBSD pattern obtained from alumina.

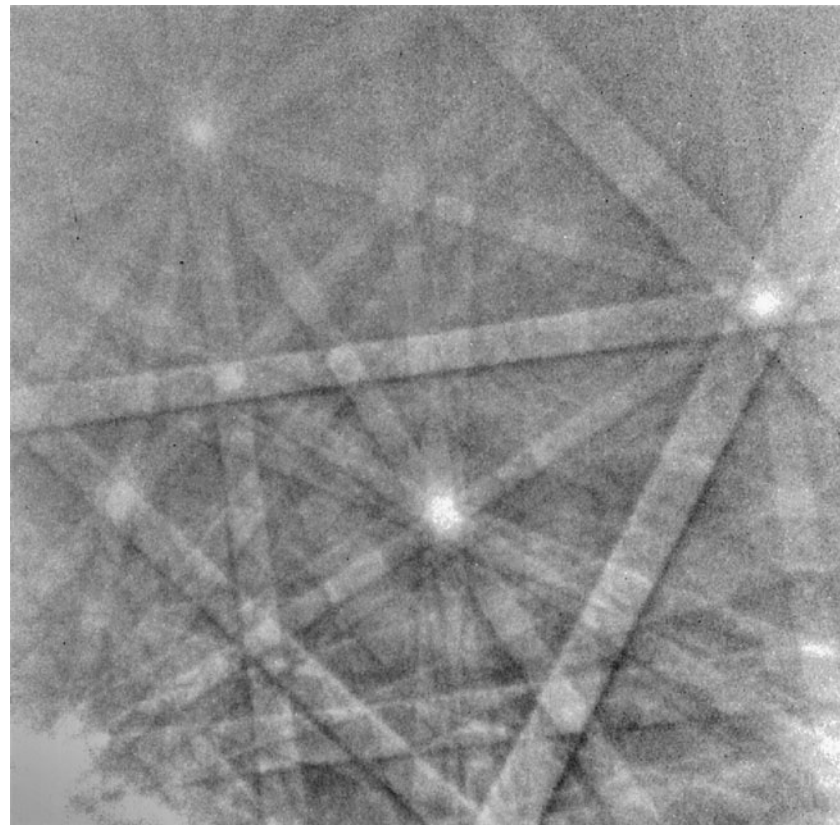


Figure 11.4. EBSD pattern of alumina (Al_2O_3) collected at 20 kV.

Examples of preparation techniques for routine materials analysis using EBSD are (Randle and Engler, 2000) as follows:

- Commercially pure aluminum and titanium alloys: Electropolish in 5% perchloric acid in ethanol at −25°C.
- Mild steels: Swab with 2% nital for several seconds.
- Rocks and minerals: Diamond polish followed by colloidal silica vibratory polish for several hours.
- Alumina: Etch in phosphoric acid at 250°C for 40 s. Thermal anneal at 1200°C for 24 min or ion polish at 3 kV for 20 min (Mulvihill *et al.,* 1998; Farrer *et al.,* 2000).
- Polysilicon: Wash in a detergent solution followed by immersion in 10% hydrofluoric acid in a plastic beaker.

Sample preparation for EBSD can be fairly routine and in many cases uses standard metallographic preparation techniques. The main point to remember is to use techniques that introduce a minimum amount of surface deformation into the material and to avoid any etchants or treatments that result in oxide or other surface film formation.

EBSD is also useful for the identification of small crystalline particles. High-quality EBSD patterns from submicrometer-sized particles mounted on a bulk substrate are difficult to obtain. This is most probably due to the contributions to the overall pattern background from electrons that passed through the particle and interacted with the sample substrate. However, EBSD patterns of high quality can be obtained from particles mounted on thin TEM support films (Small and Michael, 2000). The techniques used to prepare particles on thin support films are discussed by Brown and Teetsov (1980) and in Section 11.7.

11.6. Special Techniques

11.6.1. Plasma Cleaning

Clean samples free of contamination are very important, particularly for low-accelerating-voltage examination. Plasma cleaning is an excellent technique for the removal of hydrocarbon contamination from both specimens and components for the SEM (Zaluzec *et al.,* 1997; Isabell *et al.,* 1999). During scanning of the sample in the SEM there are two main causes of contamination buildup in the scanned area. The first source of contamination is the migration of existing surface contamination to the electron beam. Hydrocarbon contamination of the sample may result from specimen preparation due to the touching of the sample or specimen holder, backstreaming of oil in a vacuum coating system, or inadequately removed solvents or chemicals. It can be extremely difficult to prepare a perfectly clean sample. The other source of contamination is hydrocarbons that arise from the vacuum system of the SEM.

Plasma cleaning relies on the application of a high-frequency, low-energy reactive gas plasma for the removal of contamination. Various process gasses may be used during plasma cleaning to induce chemical reactions involving various components of the sample. Oxygen plasmas are highly effective at removing hydrocarbon contamination from the sample. Disassociated oxygen atoms created by the plasma chemically react with hydrocarbon materials present on the sample by converting them to CO, CO_2, and H_2O. These gases are then removed by the vacuum system of the plasma cleaner. Other gasses may be used to remove other types of contamination or to selectively etch microstructural constituents (Isabell *et al.,* 1999).

The residence time within the plasma varies depending upon the amount of contamination that needs to be removed. Freshly prepared samples may only require a few seconds to a few minutes to clean. Longer times are required to clean heavily contaminated samples. Sample holders and other microscope parts may need to be cleaned overnight. Figure 11.5 shows an example of plasma cleaning of a contaminated magnification standard. About 30 min of plasma cleaning was required to remove the contamination. More or less time may be required, depending on the plasma energy and the extent of contamination.

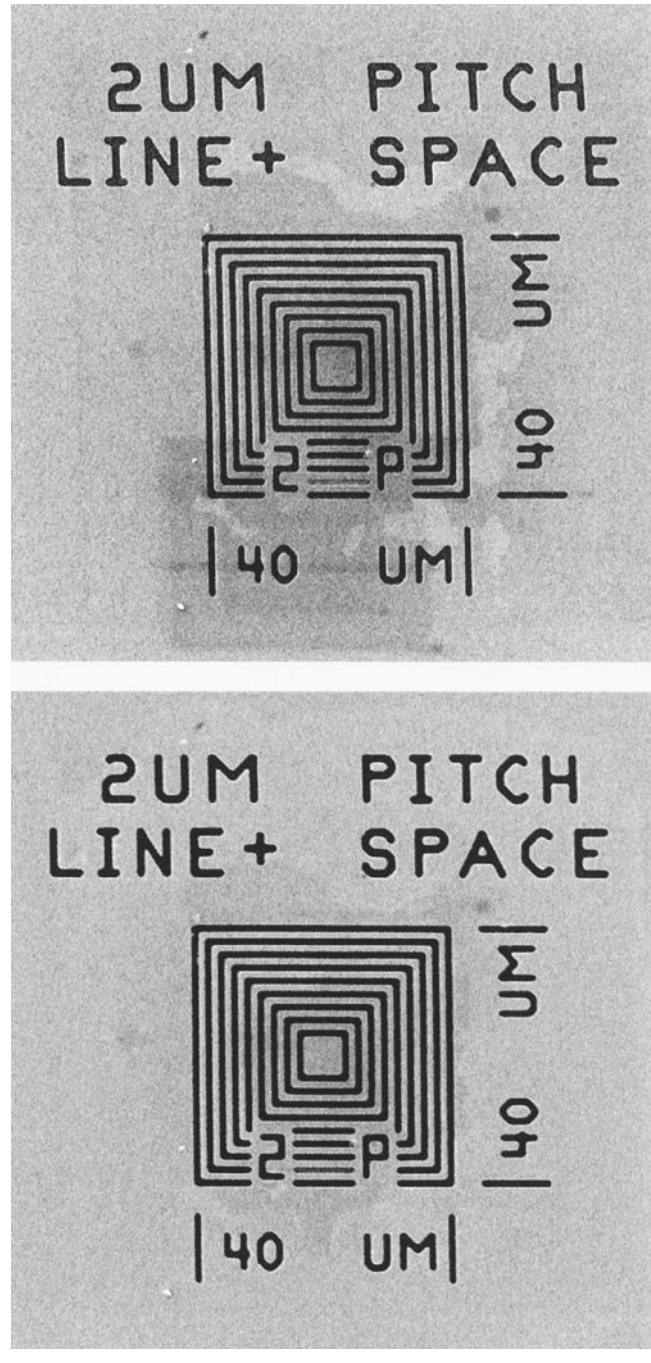


Figure 11.5. Plasma cleaning of samples is an effective technique for the removal of contamination. (a) Magnification standard after exposure to the electron beam in the SEM. (b) Same sample after 30 min in a low-energy oxygen plasma. The contamination due to the SEM has been removed with no damage to the sample.

The focused ion beam (FIB) microscope works by rastering a beam of charged particles over a sample surface with the resulting signal at each raster position plotted to form an image. This is analogous to image formation in the SEM. The major difference is that instead of electrons as in the SEM, FIB uses a beam of positively charged ions. Recent improvements in ion sources have enabled the resolution of the FIB to approach 5 nm or better (Prewett and Mair, 1991; Phaneuf, 1999). The main reason for improved resolution has been the advances made in the design of liquid-metal ion sources (LMIS). The most common LMIS ion source uses gallium ions as the primary beam (although ion beams of other species can be produced from LMIS). The typical FIB instrument employs a Ga LMIS high-brightness ion source operating at 25–50 kV. These FIB systems are often equipped with gas injectors for delivering a range of chemistries to the rastered area for the purpose of ion-beam-assisted deposition or etching of the sample (Prewett and Mair, 1991; Young, 1993; Krueger, 1999). More versatile systems include a high-resolution SEM column along with the FIB column. In this way, sample preparation and imaging can be accomplished in the same instrument. FIB instruments have initially been used for integrated circuit manufacturing and failure analysis. Now the FIB is attracting interest from other areas in the field of materials science.

FIB instruments are extremely useful for both material science and semiconductor specimen preparation because of their unique capabilities. These include:

1. The ability to make a precise cut or cross section with the ion beam. The cut or cross section may be less than 1 μm to as large as 100–200 μm in width and 10–30 μm in depth. The cut or cross section may be located with an accuracy of better than 20 nm.
2. Immediate imaging, using the ion beam with a resolution of 5 nm, with either secondary electrons or secondary ions. Excellent channeling contrast of crystalline materials can be obtained using either secondary electrons or ions generated by the focused ion beam.
3. The possibility of injecting several different gases at the ion beam impact point to deposit both conductive and insulating materials, or to selectively etch the cross section or the sample surface (Prewett, 1993).

The traditional single-beam FIB (ion imaging only) is limited when compared to a dual-beam FIB system (electron and ion beam columns on one instrument) due to the destructive nature of ion beam imaging. The imaging of small features is difficult due to the damage caused by the ion beam during the imaging process. A properly designed dual-column or dual-beam instrument integrates the ion beam milling capability with the electron column's nondestructive, high-resolution imaging. Energy-dispersive x-ray microanalysis is usually available on the instrument as well. By combining these functions into one instrument, significant time

can be saved because there is no need to transfer the sample from one instrument to another. The time saved is not only in the transfer time, but also in relocating the area or region of interest.

Another important advantage of the dual-beam system is the possibility of metal deposition. Quite often it is important to prepare cross sections of thin surface films. It is important to protect the surface film from damage by the ion beam and this is usually accomplished by the deposition of metal on the area of interest before cross sectioning. In a dual-beam instrument, a protective metal layer can be deposited using the electron beam. This technique ensures that only the area of interest is coated and that there is no damage to the sample.

It is also possible with a dual-beam instrument to employ slice-and-view techniques. In many instances a three-dimensional visualization of a sample region is required. It is possible in a dual-beam instrument to remove small slices of material with the ion beam followed by imaging with the electron beam. In this manner it is possible to slowly slice through a feature of interest and obtain high-quality electron beam images of each slice. These images can be subsequently animated for ease of visualization.

11.6.2.1. Application of FIB for Semiconductors

Applications in the semiconductor industry were the driving force for the initial development of FIB (Krueger, 1999). FIB cross sections are extremely useful for the characterization of microelectronic devices because specific areas of an integrated circuit can be rapidly sectioned and immediately examined. This capability is very useful for the understanding of problems associated with integrated circuit fabrication. FIB is indispensable in failure analysis of integrated circuits. Typically, the exact or suspected failure site is located using electrical or electron beam testing of the circuit. FIB may then either be used to affect a repair of the circuit by depositing an alternative electrical path (in the case of an open metal line) or the suspected area can be cross-sectioned and the structure of the device and the failure site investigated. Figure 11.6 shows an FIB-prepared cross section through a via between two Al interconnects. The circuit had failed to operate correctly due to a suspected opening between the upper metal layer and the lower metal layer. Figure 11.6 shows that a large stress void had formed at the location of the via causing the failure. A typical cross section can be prepared in about 20 min. Cross sections can also be used to evaluate the general structure and features in specific areas of an integrated circuit as shown in Fig. 11.7. The cross section was prepared using normal FIB processing followed by gas-assisted etching of the cross section to delineate the structural features. The combination of the FIB with an FEG SEM permits high-resolution images, which show structural details of the integrated circuit, to be acquired. This technique can also be used to analyze particulates. The particle is first held in place with a thick metal deposit. The metal deposit containing the particle is cross-sectioned for analysis.

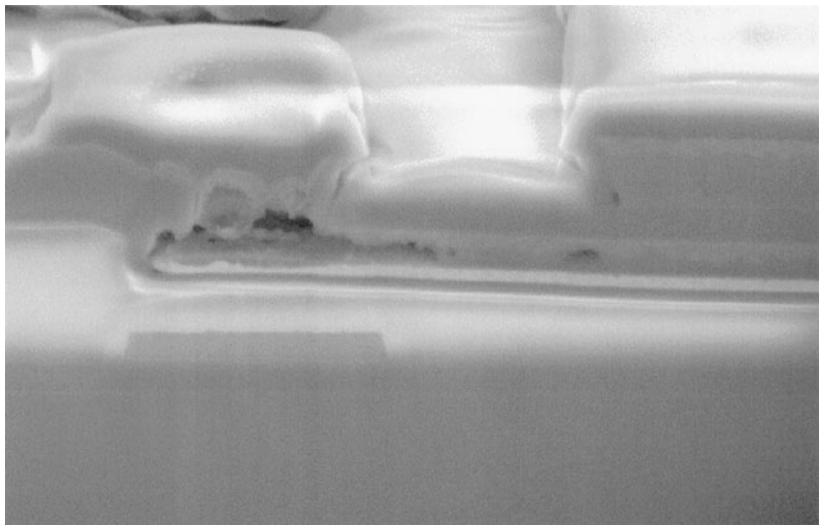


Figure 11.6. SEM image of FIB cross section through a via between the upper- and lower-level metallizations on an integrated circuit. The void that is clearly visible caused the integrated circuit to fail.

FIB is also an important tool for deprocessing of integrated circuits, for example, EBIC imaging of metal lines and conductors well below the surface passivation. The ion beam is used with a etch gas that selectively removes the passivation from a small area of the integrated circuit allowing the circuit to continue to function for EBIC. An example of gas-assisted FIB etching is shown in Fig. 11.8.

11.6.2.2. Applications of FIB in Materials Science

All of the advantages of FIB in semiconductor applications are also advantages in materials science studies (Phaneuf, 1999). FIB application in materials science has mainly been hampered by the high cost and therefore limited availability of these instruments to the materials science community. Recent developments of single- and dual-beam instruments (without many of the sample handling equipment and large stages required for semiconductor applications) have produced instruments that are less expensive and are ideal for materials science applications.

Cross sections of many difficult-to-prepare samples have been produced. Examples of these samples include ceramics, metal-to-ceramic interfaces, Zn coatings on steels, Au ball bonds on Al bond pads, and many other materials (Phaneuf, 1999; Kato *et al.,* 1998). Often, these different materials can be successfully prepared with little or no variations in technique or ion beam conditions. The most noticeable difference from one material to the next are variations in the rate of material removal.

The main disadvantage of FIB techniques for materials science applications is the limited size of the specimen that can be produced using the

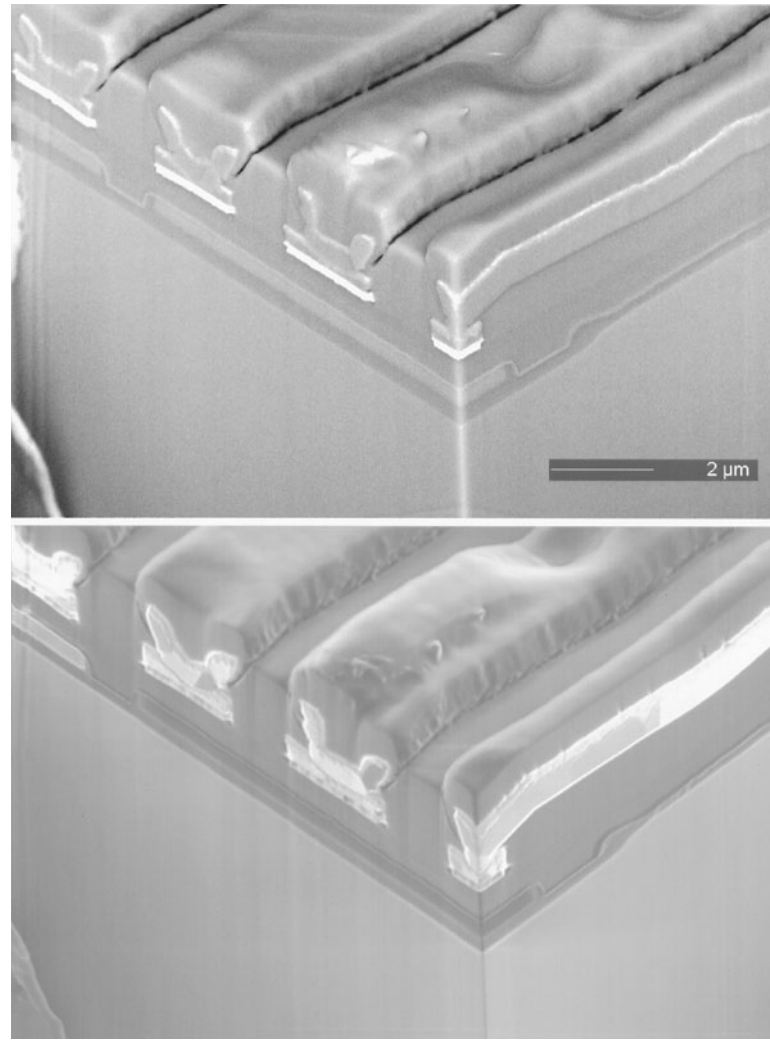


Figure 11.7. Gas-assisted ion-beam-etched FIB cross sections through a row of vias in an integrated circuit. (a) Secondary electron SEM image. Compare this image with that shown in Fig. 11.1 to demonstrate the sample detail resulting from the etch. (b) Ion-beam-excited secondary electron image of the same sample as in (a). Note the increased contrast between the metal layers and the glass passivation.

FIB. A large FIB-produced cross section would measure up to 100–200 μm in length by 10–20 μm in depth. Thus, FIB in materials science is best used when a specific area of interest (grain boundary, corrosion pit, crack tips, films grown on surfaces, etc.) can be defined. FIB is not the ideal tool for the preparation of specimens for general microstructural examination. This is most certainly true for large-grained materials, where in any cross section only two or three grains may be imaged, which may or may not give a true representation of the general microstructure of the sample.

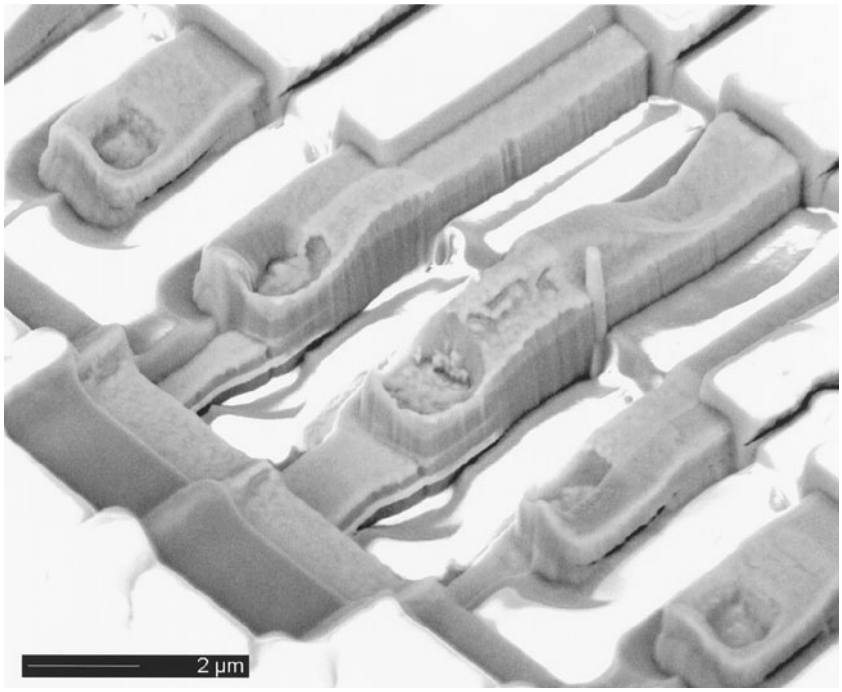


Figure 11.8. Gas-assisted ion etching can be used to remove glass passivation films from around a via between aluminum interconnects. This technique is useful for simply exposing the via for examination or removing the glass passivation so that EBIC or voltage contrast images can be obtained.

One advantage of cross sections prepared in the FIB is the ability to reveal small voids or other fragile features without the potential damage involved in conventional sectioning techniques using mechanical means. The lack of crystalline damage to the sample is important because one of the most useful ion imaging techniques in the FIB is channeling or orientation contrast. Ion imaging produces stronger channeling contrast than electron beam imaging. Figure 11.9 shows a comparison of electron and ion images of an electrodeposited Ni sample. The increased crystallographic channeling contrast of the ion image is clearly visible.

11.7. Particles and Fibers

The proper preparation of particles and fibers requires that the particles be well-dispersed, secure, electrically conductive, and attached to a substrate that does not interfere with x-ray microanalysis. Well-dispersed particles make it easier to differentiate individual particles from agglomerates and measure particle size and shape, and allow the chemistry of individual particles to be better determined. The substrate is very important for

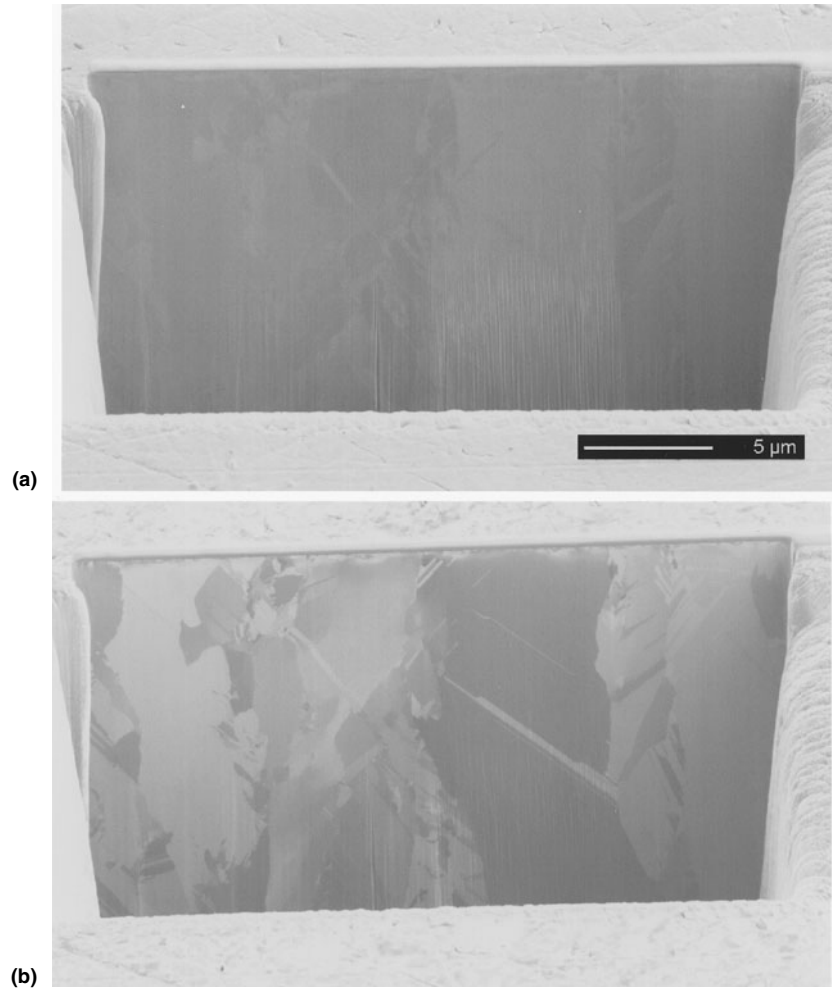


Figure 11.9. FIB cross sections prepared from an electrodeposited Ni sample. (a) Secondary electron SEM image of the cross section. Surface topography is visible as well as slight channeling contrast. (b) Ion-beam-excited secondary electron image of the same sample. The increased channeling contrast is immediately apparent.

particle analysis. First, the substrate must be smooth enough so that the individual particles are not hidden by the roughness of the substrate and second, the substrate must not interfere with x-ray analysis of the particles.

The techniques used for particle collection are quite diverse and each has specific advantages and disadvantages. Several techniques for SEM analysis are described in the following sections. For a full description of particle collection see Murphy (1984) Brown and Teetsov (1980). The main goal of particle collection is to produce a sample that is representative. In some cases, such as filtration or collection on adhesive surfaces, the samples are satisfactory for SEM imaging and analysis as collected.

11.7.1. Particle Substrates and Supports

Bulk materials as well as supported thin films are commonly used as substrates for particle analysis. Substrates for particles should provide enough mechanical support and stability to permit the particle to be imaged and the support or substrate should not contribute to the electron or x-ray signals and should be conductive. Also, the support or substrate should not have topographic features that obscure the particles to be examined.

11.7.1.1. Bulk Particle Substrates

Bulk materials are best used as substrates for large particles with thicknesses greater than the x-ray analysis volume for the accelerating voltage used. This will minimize the effect of the substrate on microanalysis of the particle using EDS in the SEM. One only needs to look in a catalogue for suppliers of consumables for the SEM laboratory to realize there are many different choices of substrates for particle analysis. Commonly used bulk substrates are are polished low-atomic-number materials such as graphite, silicon, boron, and beryllium.

The lowest atomic number material that is used for bulk substrates is beryllium. Beryllium can be highly polished to minimize surface features that would interfere with imaging of small particles. Users of beryllium substrates should be aware of the toxic nature of beryllium oxide debris, which may form during the polishing of beryllium substrates. Commercially "pure" beryllium will often contain minor amounts of impurities, such as iron or calcium, which may be detected during x-ray microanalysis. Even though beryllium is not easily detected with EDS, beryllium and all other substrates will contribute a continuum background to the x-ray spectrum. Commercially produced beryllium planchets are available from a number of suppliers of consumables for SEM laboratories.

Carbon, often in the form of graphite, is an excellent substrate for particle imaging and analysis because of its low atomic number, availability, mechanical stability, and high purity. Very smooth graphite or carbon sample mounts are available; it is also relatively simple to polish the mounts to a fine finish. Graphite tends to be relatively soft, so polished surfaces will contain scratches that may tend to obscure smaller particles. Another form of carbon, pyrolitic graphite, is also available and is a highly purified form of carbon that takes a high polish and is an excellent substrate material. Boron can be obtained in high-purity bulk form and polished. Unlike Be, it does not produce a highly poisonous oxide and can therefore be handled much more easily.

Other acceptable substrates for larger particles are silicon and aluminum. Silicon is readily available as highly polished and extremely pure wafers from semiconductor manufacturing suppliers. The polished surface of a silicon wafer is extremely smooth and featureless even at the highest of magnifications. This substrate is most useful for larger particles of heavier metals. There is the virtual inevitability of observing a silicon peak in the EDS x-ray spectrum from the substrate due to scattering from the sides of the particle, and from direct penetration in the case of particle diameters

less than the electron range. Aluminum is also a suitable substrate for some particle analysis. Sample mounts are readily available and can be polished for mounting particles. The drawback to aluminum is similar to that for silicon, in that an aluminum x-ray peak may be observed in the EDS spectrum.

11.7.1.2. Thin Particle Supports

Thin supports for particles, similar to those used for TEM sample mounting, can be used in the SEM for mounting small particles. The thin film is usually supported on a metal grid. Carbon is typically used for the thin film, but many other types of films are available and these include silicon nitride, silicon oxide, and diamond. Some metal films are available, but these tend to be quite fragile. Metal grids are also available in a wide range of materials, the most common is copper, but grids of beryllium, nickel, gold, and titanium are commonly available. The choice of metal for the grid is important because electrons will be scattered by the particle and some of the scattered electrons will strike the metal grid. This will result in extra peaks in the x-ray analysis results, which may confuse the correct identification of the elements within the particle. Beryllium grids will greatly reduce this problem. Figure 11.10 is a comparison of EDS x-ray spectra obtained at 20 kV from 200-nm Fe–Co particles mounted on bulk and thin-film substrates. The decreased background intensity for the particle mounted on the thin-film substrate is apparent. This lowered background intensity may permit smaller concentrations of elements in the particles to be detected. The main disadvantage to the thin-film substrate is that the elements present in the grid that supports the thin film may be detected in the x-ray spectrum. The copper peak in the spectrum from the Fe–Co particle mounted on the thin substrate is a result of scattering of electrons into the support grid.

The use of thin-film supports optimizes the contrast observed from small particles while minimizing contributions to the x-ray spectrum from the particle support. For best imaging and microanalysis, it is important to eliminate scattering of x-rays generated by electrons that are transmitted through the sample. This is accomplished by mounting the grid over a hole in the sample stub. If the stub is aluminum, covering the inside of the hole with carbon paint will reduce the numbers of backscattered electrons and detectable x-rays generated by the interaction of the transmitted electrons with the stub material.

11.7.2. Particle Mounting Techniques

Once a suitable substrate has been selected it is important to decide how the particle sample will be affixed to it. The goal of a proper mounting technique should be to provide a mounted sample of the particulate material that is representative of the original sample without selective loss of large or small particle fractions (fractionation), without aggregation, and with adequate mechanical stability and conductivity to prevent loss of particles during sample handling and good imaging and analysis in the SEM. The

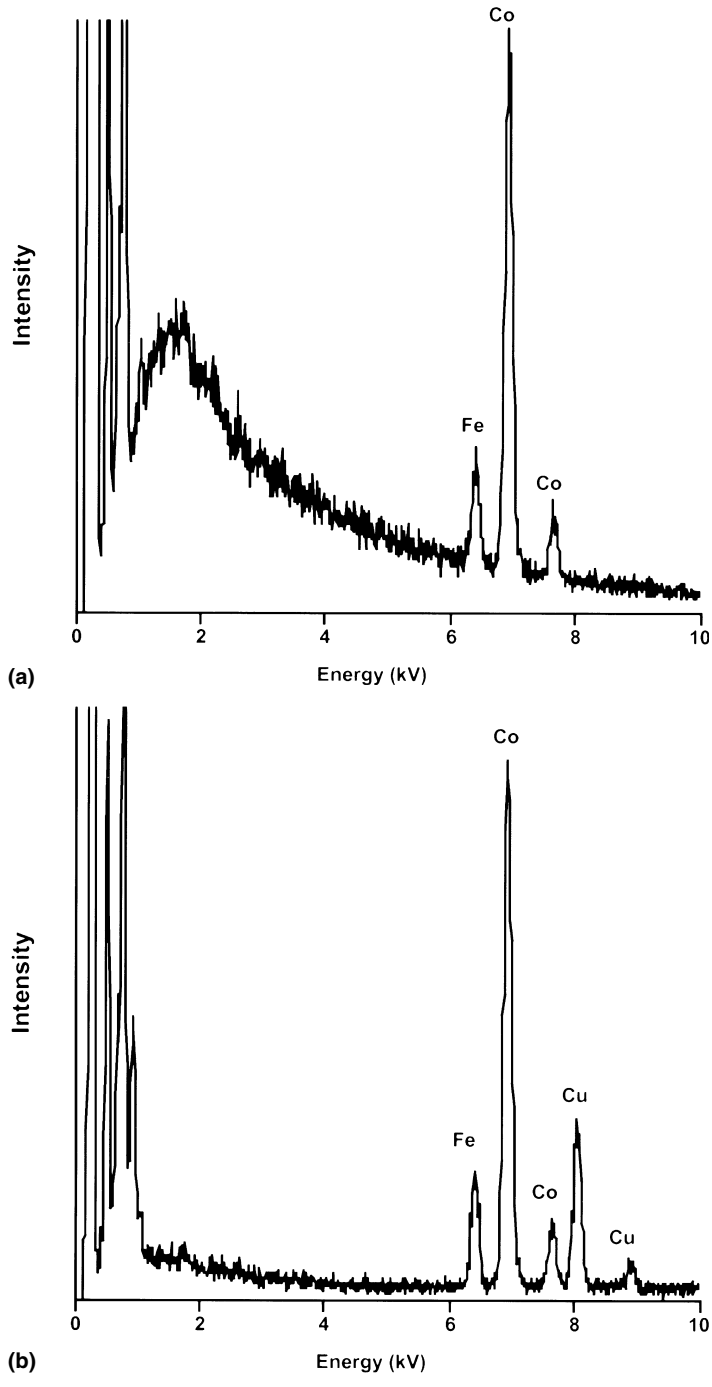


Figure 11.10. EDS of 0.2-μm particles at 20 kV, showing the much lower background that can be obtained from particles on a thin support film. (a) Particle mounted on a bulk, polished graphite substrate. (b) Particle mounted on a thin carbon film on a copper grid over a blind hole in a graphite substrate.

selection of mounting technique will depend on the nature of the particle sample (Brown and Teetsov, 1980).

Adhesive Mounting. The most often used technique for mounting powders for SEM examinations is to mount dry particles on double-sided adhesive tape. This technique is most useful when the specimen is received in the form of abundant, loose particles. Care must be taken when mounting the tape on the sample stub in order not to trap air under the tape, which may cause mechanical instability during SEM examination. It is important to use tape that has an adhesive with high viscosity so that it does not engulf the particles. Double-sided carbon tape is now widely available and is suitable for mounting larger particles. The tape has the advantage that it is conductive, but these tapes tend to be relatively rough, which may obscure smaller particles, and the adhesive on the tape may not have sufficient viscosity to prevent smaller particles from being engulfed by the adhesive. Small portions of the particle sample may be transferred to the double-sided tape by spatula or simply poured from the bottle. After the particles are on the tape, the stub is tapped several times to dislodge any loose particles, and the particles are spread across the surface of the tape. After the particles are deposited, the edges of the tape may be covered with conductive paint followed by conductive coating if needed.

Particles may also be attached to the substrate material by coating the substrate with carbon paint or a thin layer of parlodian or collodian. It is important to wait until the paint or polymer film has begun to dry. This prevents the particles from being engulfed or sinking into the film. It may be necessary to coat the mounted particles for conductivity.

It is sometimes necessary to disperse particles in a liquid with the use of ultrasonic vibration to break up aggregates. It is generally simplest to then place a drop of the solution containing the particles directly on a suitable substrate and allow the liquid to evaporate. It is important to select the liquid medium so that the particles are not modified and the liquid fully evaporates leaving little or no residue. A final thin conductive coat can be applied to insulating particles.

11.7.3. Particles Collected on Filters

Examination of particles on filters is a common particle analysis problem. Many particle samples, such as airborne particles, are collected as a deposit on a filter. Many different types of filters are available and appropriate preparation techniques will depend on the type of filter.

Some filters act by the impact of particles onto a smooth filter surface. These filters are generally smooth and featureless enough to permit direct examination of the particles in the SEM without removal of the particles from the filter material, but a conductive coating may be needed. If x-ray microanalysis is to be used to identify small particles, it is important to understand the background peaks resulting from x-rays excited from the filter material.

Tortuous-path filters collect particles in the interior of the filter medium. These filters require that the particles be exposed or removed from the filter. Particle exposure is accomplished by etching the filter or partially dissolving it in a suitable solvent. After coating, the filter may be directly imaged in the SEM. Removal of the particles from the filter may be accomplished by first etching the surface of the filter to expose the particles. A relatively thick carbon coating is applied and then the remaining filter material is removed with a solvent. The exposed particles are embedded in the carbon layer and can be transferred to a suitable substrate.

Low-temperature ashing in an oxygen plasma is used to remove filter material. This is only an option if the particles are not degraded by exposure to the oxygen plasma. The ashed residue may be mounted on a substrate and coated for conductivity for SEM examination.

11.7.4. Particles in a Solid Matrix

Inclusions or second-phase particles in a solid matrix may be difficult to analyze in polished cross sections due to excitation of the surrounding matrix. It may be necessary to determine if there are common constituents in the particle and the matrix. In this case it is necessary to remove the particle from the matrix. This is best accomplished through the use of extraction replicas. The specimen is first metallographically polished to produce a flat surface. The sample is then lightly etched with a suitable etchant that attacks only the matrix and not the particles of interest. In mild steels, for example, a good etchant is 2% nital (a mixture of nitric acid in methanol). The etched sample is then coated with carbon. Other coating materials may be used if carbon is not suitable. The coated specimen is then etched further. The etchant attacks the matrix and allows the carbon film to detach from the surface with the particles. The sample is transferred to water and the carbon film containing the particles is allowed to float from the surface and is then collected and mounted on a SEM substrate.

11.7.5. Transfer of Individual Particles

A very difficult preparation problem is the transfer and mounting of specific individual particles to the SEM for imaging or microanalysis. Individual particles should be mounted on substrates that allow the particles to be identified. The substrate should be scribed with easily identified areas for particle mounting. Small particles, larger than 3 μm, may be picked up with a sharp needle and positioned in a drop of amyl acetate. A small drop of collodian is placed next to the drop of solvent containing the particle. The drop containing the particle is moved across the collodian to transfer some of the collodian to the drop containing the particle. The solvent is allowed to evaporate, resulting in the particle being held by a thin layer of collodian. This is a difficult procedure and should be practiced with less important particles. For more details see Brown and Teetsov (1980).

References

ASM (1985). *ASM Handbook,* Vol. 9, *Metallography and Microstructures,* ASM International, Metals Park, Ohio.

ASM (1986). ASM Handbook, Vol. 10, *Materials Characterization,* ASM International, Metals Park, Ohio.

ASM (1996). *Microstructural Science,* ASM International, Metals Park, Ohio.

ASTM (2000). Metals Test Methods and Analytical Procedures, in *Annual Book of Standards 2000,* Vol. 03.01, American Society for Testing of Materials, Philadelphia.

Bousfield, B. (1992). *Surface Preparation and Microscopy of Materials,* Wiley, New York.

Brown, J. A., and A.Teetsov (1980). In *The Particle Atlas* (W. C. McCrone, ed.), Ann Arbor Science Publishers, Ann Arbor, Michigan, p. 1472.

Buehler Ltd. (1973). *Petrographic Sample Preparation,* Vol. 12/13.

Davidson, S. M. (1989). In *SEM Microcharacterization of Semiconductors* (D. B. Holt and D. C. Joy, eds.), Academic Press, New York, p. 153.

Farrer, J. K., J. R. Michael, and C. B. Carter (2000). In *Electron Backscatter Diffraction in Materials Science* (A. J. Schwartz, M. Kumar, and B. L. Adams, eds.), Kluwer/Academic Press, New York, p. 299.

Heinrich, K. F. J. (1981). *Electron Beam X-ray Microanalysis,* Van Nostrand Reinhold, New York.

Holt, D. B. (1989). In *SEM Microcharacterization of Semiconductors* (D. B. Holt and D. C. Joy, eds.), Academic Press, New York, p. 242.

Humphries, D. W. (1992). *The Preparation of Thin Sections of Rocks, Minerals, and Ceramics,* Oxford University Press, Oxford.

Isabell, T. C., P. E. Fischione, C. O'Keefe, M. U. Guruz, and V. P. Dravid (1999). *Microsc. Microanal.* **5,** 126.

Joy, D. C. (1989). In *SEM Microcharacterization of Semiconductors* (D. B. Holt and D. C. Joy, eds.), Academic Press, New York, p. 69.

Joy, D. C., D. E. Newbury, and D. L. Davidson (1982). *J. Appl. Phys.* **53,** R81.

Kato, T., M. H. Hong, K. Nunome, K. Sasaki, K. Kuroda, and H. Saka (1998). *Thin Solid Films* **319,** 132.

Krueger, R. (1999). *Micron* **30,** 221.

Lee, W. E., and W. M. Rainforth (1994). *Ceramic Microstructures: Property Control by Processing,* Chapman and Hall, New York.

Michael, J. R. (2000). In *Electron Backscatter Diffraction in Materials Science* (A. J. Schwartz, M. Kumar, and B. L. Adams, eds.), Kluwer/Academic Press, New York, p. 75.

Miller, J. (1988). In *Techniques in Sedimentology* (M. Tucker, ed.), Blackwells, Oxford, p. 86.

Mulvihill, M. L., M. A. Gulgun, E. Bischoff, and M. Ruhle (1998). *Z. Metallk.* **89,** 546.

Murphy, C. H. (1984). *Handbook of Particle Sampling and Analysis Methods,* Verlag Chemie International, Weinheim, Germany.

Phaneuf, M. W. (1999). *Micron* **30,** 277.

Prewett, P. D. (1993). *Vacuum* **44,** 345.

Prewett, P. D., and G. L. R. Mair (1991). *Focused Ion Beams from Liquid Metal Ion Sources,* Research Studies Press, Somerset, England.

Randle, V., and O. Engler (2000) *Introduction to Texture Analysis: Macrotexture, Microtexture and Orientation Mapping,* Gordon and Breach, New York.

Reimer, L.(1998). *Scanning Electron Microscopy: Physics of Image Formation and Microanalysis,* Springer, New York.

Scott, V. D., and G. Love (1983). *Quantitative Electron-Probe Microanalysis,* Wiley, New York.

Small, J. A., and J. R. Michael (2000). *J. Microsc.* **201,** 59.

Taylor, C. M., and A. S. Radtke (1965). *Econ. Geol.* **60,** 1306.

Van der Voort, G. F. (1984). *Metallography: Principles and Practice,* McGraw Hill, New York.

Young, R. J. (1993). *Vacuum* **44,** 353.

Zaluzec, N. J., B. J. Kestel, and D. Henriks (1997). *Microsc. Microanal.* **3**(2), 983.

12

Specimen Preparation of Polymer Materials

12.1. Introduction

This chapter outlines a variety of sample preparation procedures for performing scanning electron microscopy and x-ray microanalysis of polymer materials. Typical specimen preparation methods will be described and shown in a range of applications that are encountered every day in academic and industrial laboratories. The reader new to the study of polymer materials is directed to Chapter 12 of the Enhancements section on the accompanying CD, which provides background material for this chapter. The topics found on the CD are an introduction to polymer materials, E12.1, and polymer morphology, E12.2, and a description of some typical processes used to produce materials from polymers, E12.3. The book by Sawyer and Grubb (1996) is comprehensive, providing details of specimen preparation methods, dozens of practical examples, and references that describe both the structure and the microscopy of polymers.

The interested reader is also directed to texts and papers on microscopy and materials science of polymers. Young (1981) is a general polymer text, Hemsley (1984, 1989) give a short description on light microscopy, and White and Thomas (1984) is directed to SEM. General texts and papers of interest include Grubb (1993), Bassett (1988), Keith and Padden (1963), Ciferri and Ward (1979), and Ward (1975, 1983) on microstructure. Representative materials science research work is shown in Baer and Moet (1991), Paul *et al*. (1988), Paul and Sperling (1986), Vezie *et al*. (1992), and Hobbs *et al*. (1988).

12.2. Microscopy of Polymers

Microscopy is a key technique for problem solving and understanding structure–property–process relationships of polymer materials. There is no one technique that provides the "solution" to all problems and thus there may be a need for complementary techniques. The focus here on

SEM and x-ray microanalysis covers a wide range of structures, but for polymers, complementary correlative techniques are light microscopy, transmission electron microscopy (TEM), and atomic force microscopy (AFM). These techniques provide information on structures ranging from greater than 1 μm to those less than 10 nm in size. Analytical techniques such as thermal and X-ray analysis and infrared and mass spectroscopy are also complementary to microscopy and are often required for problem solving.

Most specimen preparation methods for polymers involve damage or fracture processes that must be understood for image interpretation. Because the preparations often damage the material, multiple preparation techniques may be helpful for interpretation.

12.2.1. Radiation Effects

Electron micrographs of polymers cannot be interpreted usefully without considering the possibility that radiation or beam damage has changed the polymer during imaging. Absence of easily visible changes does not prove that there is no damage because the damage may occur quickly, prior to focusing and saving an image. The processes involved in radiation damage are extremely complicated and well beyond the scope of this book (see Everhart *et al.,* 1972, Sawyer and Grubb, 1996). However, the nature of the potential changes must be considered for proper image interpretation. Some of these changes include the breaking of chemical bonds, mass loss, and reduction of crystallinity as a result of ionizing radiation exposure and thermal effects. For example, polyethers and polyoxymethylenes will degrade with mass loss and formation of volatiles in the vacuum of the SEM. Beam damage caused by radiation and thermal effects is irreversible, whereas specimen charging (see Chapter 15.2) is a reversible process.

The relative radiation stability of different chemical groups is known. Aromatic compounds are much less sensitive to radiation than aliphatic ones. The presence of oxygen can have a detrimental effect due to formation of peroxides. The tendency for radiation damage rises with temperature and is lower if a solid material is crystalline than if it is amorphous. Several types of morphologies can be formed during sample irradiation in the SEM. Mass loss results from bond-breaking processes, forming cracks and bubbles in thick specimens, although the unirradiated bulk of the sample acts to stabilize the form of the irradiated surface. The image may appear to have a uniform depression in the sample surface or holes and cracks may appear. Loss of light elements also occurs during irradiation. This factor makes quantitative analysis for light elements quite difficult because the sample changes during the high-energy irradiation required for microanalysis. Mass loss may also be regarded as an etching process and this may be used to develop differential contrast between two polymers in a polymer blend. Loss of crystallinity results in changes to the spherulitic textures and lamellae, with cracks often appearing in the amorphous region of these structures. Finally, large dimensional changes can be induced by radiation

that distorts the object and changes the image. This is quite serious because it might be overlooked and misinterpreted.

Examples of radiation or beam damage are shown in Figs. 12.1 and 12.2. Poly(oxymethylene) (POM) is very beam-sensitive, as shown in a SEM image (Fig. 12.1) of the uncoated surface of a molded article which has actual holes in the specimen. The holes were formed when focusing on the sample at higher magnification and were imaged after reducing the magnification. Even in an FEG SEM, beam damage can occur in a short time. Application of a thin and uniform metal coating can give a little more time for imaging, but the damage occurs rapidly. Surface cracking and formation of depressions is another result of radiation damage (Fig. 12.2), in this case to the surface of a polypropylene microporous membrane.

12.2.2. Imaging Compromises

The most important factor in imaging radiation-sensitive materials is to be aware of the potential formation of image defects, as described in the last section. Such defects, termed "artifacts," may lead to incorrect image interpretation. A standard procedure for minimizing radiation damage is to reduce the number of incident electrons into a specific volume of material (minimize the dose).

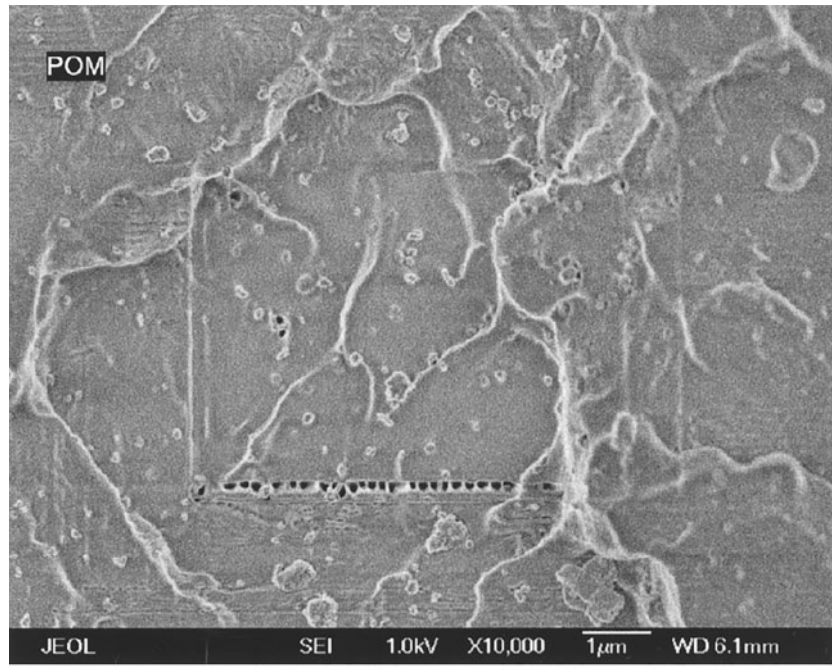
These same conditions generally result in poor signal-to-noise ratios and thus poor image quality. Improvements in image quality can be made by using a SEM with a lanthanum hexaboride gun or a field emission SEM (FEG SEM). Low beam energies may also be used in the FEG SEM to obtain better sample surface detail. Variable-pressure SEMs permit imaging of uncoated samples, although radiation damage still occurs, especially with oxygen present in the chamber.

For the highest resolution of beam-sensitive polymers, FEG SEM has been used on samples that have been metal-coated with less than 5 nm of Pt or W by ion-beam sputter coating. Beam energies in the range 0.8–1.5 keV are used for uncoated specimens, whereas the best setting for lightly coated specimens is typically about 5 keV.

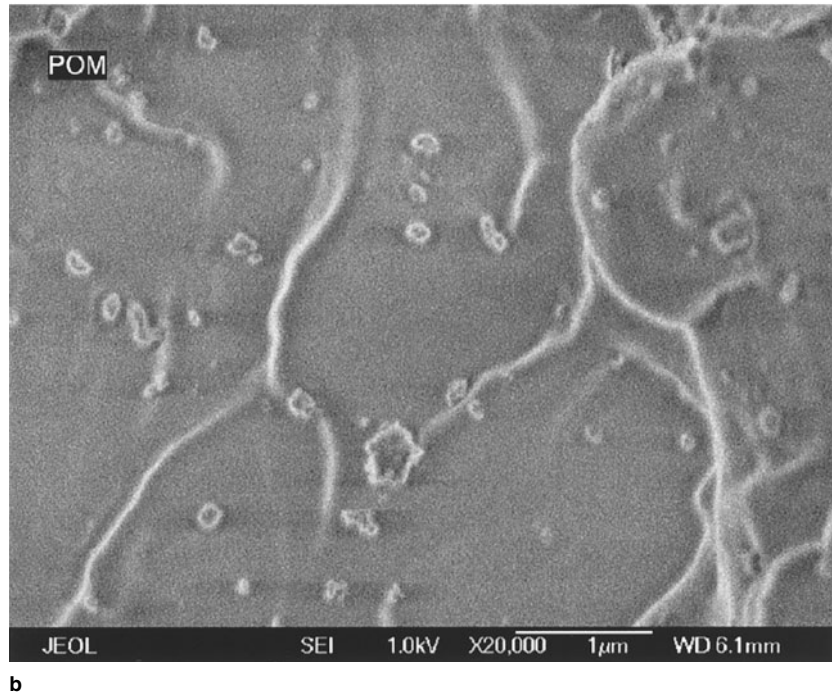
12.2.3. Metal Coating Polymers for Imaging

Chapter 15 provides details of the procedures for eliminating charging in nonconductive specimens for SEM imaging and x-ray microanalysis. It should be remembered that application of metal coatings is a specimen preparation method that can result in damage to the sample and misinterpretation of image details.

In some cases, very thick coatings of gold are sputter-coated onto samples. These coatings appear as shiny mirrors and obliterate all details less than 50 nm in size. The coated surface often exhibits a series of granular particles of metal, which at high magnification can be resolved and misinterpreted as a granular feature of the specimen. SEM images (Fig. 12.3) of a gold-coated polypropylene microporous membrane reveal a granular



a



b

Figure 12.1. SEM images of a molded surface of POM with holes and beam damage (a) resulting from focusing at higher magnification (b). (Courtesy of Vern Robertson, JEOL.)

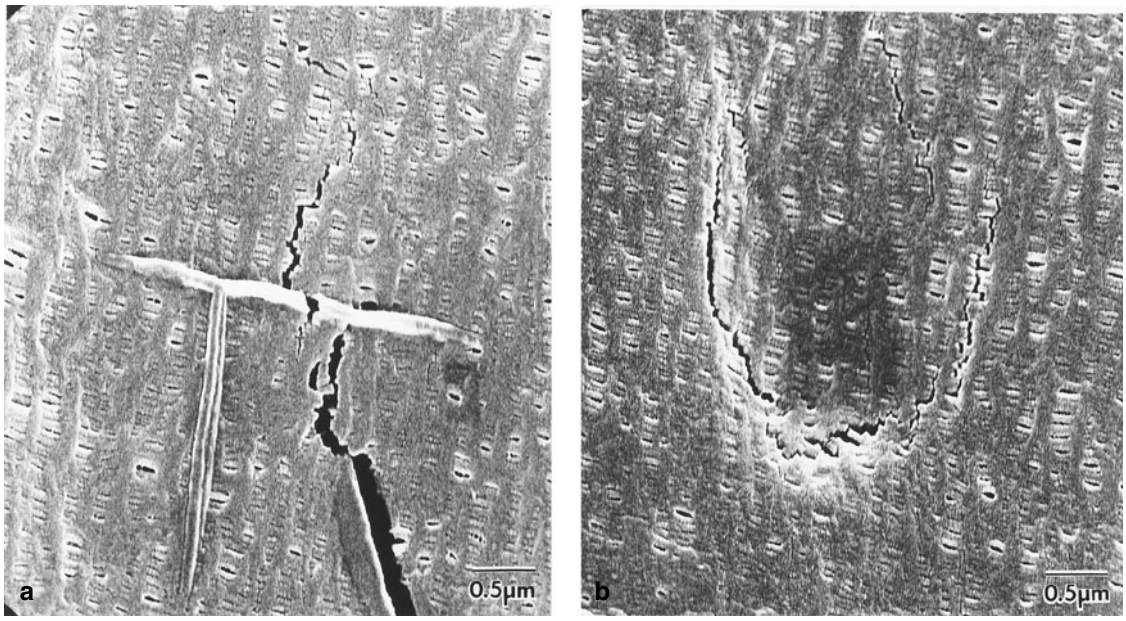


Figure 12.2. SEM of polypropylene microporous membrane with cracks and depressions due to beam damage (Sawyer and Grubb, 1996).

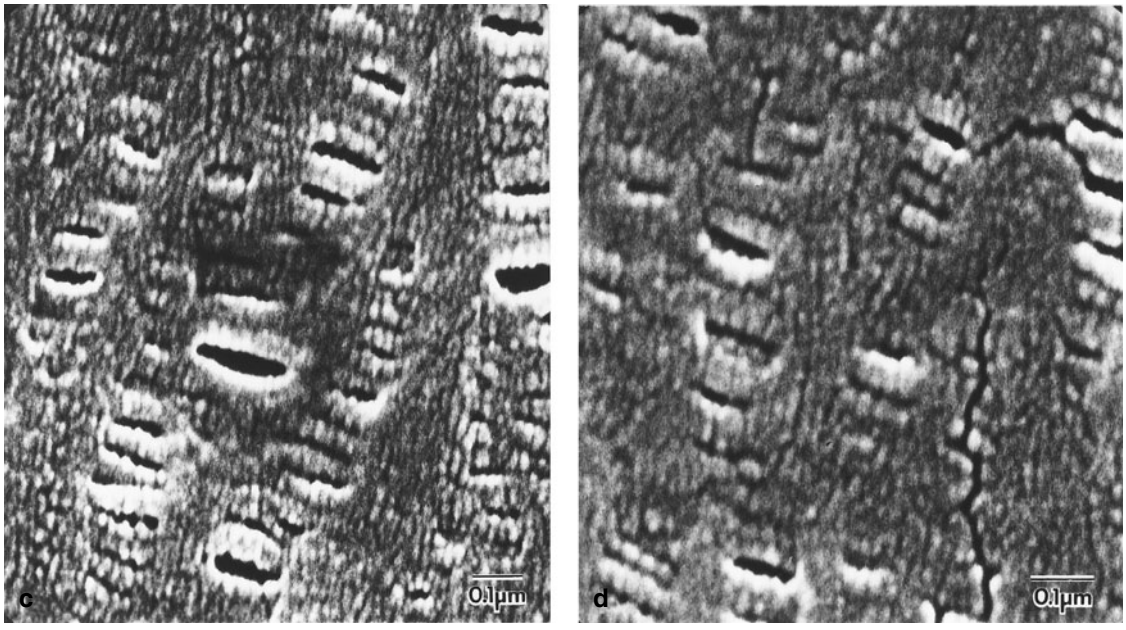


Figure 12.3. SEM images of polypropylene microporous membrane showing a thick granular texture and cracks due to Au coating (Sawyer and Grubb, 1996).

texture forming a discontinuous coating with surface cracks and poor image quality. Polymer damage is also encountered due to heating effects and chemical reactions.

12.2.4. X-Ray Microanalysis of Polymers

Polymer specimens are particularly difficult to analyze using energy-dispersive spectrometers (EDS) or wavelength-dispersive spectrometers (WDS) for the same reasons they are difficult to image in the SEM. In the case where there is a small amount of a high-atomic-number element well-dispersed in the polymer there is significantly more difficulty in the analysis than the case in which that element is found localized, for instance, at the locus of failure in a molded part. The higher beam currents required for x-ray microanalysis can greatly affect beam-sensitive specimens. Quantitative analysis is quite difficult because the rapid mass loss of some polymers and differential mass loss of various elements seriously hamper interpretation of results. Qualitative analysis can be routinely conducted, but it is important to monitor the x-ray count rate as a function of time in order to determine if the element of interest or the polymer in which it is embedded is exhibiting mass loss over time. Preparation of biological and polymer specimens for x-ray analysis is described in Chapter 13.

12.3. Specimen Preparation Methods for Polymers

There are nearly as many sample preparation techniques as there are types of polymers to be prepared. Issues such as low atomic number, beam sensitivity, and the presence of hard inorganic materials all make for difficulties in preparation. Many of the methods used for biological materials described in Chapters 13 and 14 may be important for polymer specimen preparation. Goodhew (1984) described many relavant materials science methods.

There are many characterization problems for microscopy where quite simple preparation methods are applied, at least in the initial stages of morphological study. However, most polymers must be prepared with three criteria in mind:

1. Isolate the surface or bulk
2. Enhance contrast
3. Minimize radiation damage

For surface study, simply cutting out a specimen from the bulk sample or more tedious replication procedures supply the specimen of interest. Imaging of the internal structure of materials requires cutting, fracturing, polishing, or sectioning, all processes which could damage the specimen. Often contrast must be enhanced and methods employed include staining, etching, replication, shadowing, and metal decoration; again, these processes may change the original material. Other methods primarily developed by biologists and adapted for microscopy preparation of soft, deformable, and

perhaps wet materials are freeze-drying, critical-point drying, and freeze-fracture etching.

12.3.1. Simple Preparation Methods

A major advantage of the SEM for surface observations is that sample preparation is generally simple (also see Chapter 13.2). In some cases, the material is chosen carefully from a larger sample—by eye or using a stereo binocular microscope—and then is placed on sticky tape on a specimen stub. Small amounts of conductive paints may be used to connect the specimen to the stub or to draw lines across the tape and a bulk sample. Many beautiful images have been taken of silver paint that wicked up onto even larger specimens. It is very wise to examine all the materials used in SEM preparations, to enable identification when exploring unknown specimens.

Special specimen holders can be made or purchased to hold specific objects such as fibers and fabrics. In the case of solutions or suspensions, dispersal may be done by atomizing or spraying, or larger drops of the fluid may be allowed to spread over a support film or even a glass microscope cover slip adhered to the SEM stub. Additional substrates for scanning electron microscopy include silicon, highly oriented pyrolitic graphite (HOPG), and mica, all of which are flat and easy to prepare. These substrates are also useful for metal coating experiments in which atomic force or scanning probe microscopy can be used to determine the metal grain size and the root mean square surface roughness. Fibrillated liquid-crystalline polymer fragments or other polymers can be made using cell disrupters or high-powered ultrasonic generators. Samples are suspended in water or a water–ethanol mixture and dropped onto silicon or HOPG for SEM study. Care must be taken, however, because imaging of thin films or fine fragments on HOPG surfaces may reflect the underlying cracks and steps in the graphite surface. In a worst-case example some graphite features resemble molecules and can be misinterpreted. The recurrent theme of conducting complementary imaging continues to be of importance for SEM.

12.3.2. Polishing of Polymers

Mechanical polishing of surfaces, often conducted for metals, ceramics, and rocks, also finds use especially for polymer composites containing hard materials. This is one of the best methods for producing a flat surface for x-ray microanalysis. Samples to be etched with solvents or plasmas also can be polished first, examined by SEM, generally using BSE imaging, and then etched and reexamined. Hard-to-section materials such as glass or carbon-fiber-filled composites, tire cords, and large samples have been polished for SEM and x-ray microanalysis. Polishing specimens can produce artifacts at each stage of preparation. Initial cutting can cause formation of microcracks, easily interpreted as voids or cracks in the specimen. Water-cooled diamond saws are useful when cutting a small specimen or when near the surface of interest. Grinding should be done wet and with gentle pressure. Cleaning the specimen is important between stages

so as not to contaminate the polishing cloth. Undercutting the specimen, preferential polishing of the embedding resin compared to the specimen, should be minimized when using this method.

Small samples are mounted to hold their orientation, embedded in epoxy, and then cut with a diamond saw. Wet grinding is done with abrasive papers ranging from 120 to 600 grit just long enough to remove the scratches from the previous grinding. Polishing is conducted using graded alumina suspensions in water, chromium oxide slurries, or diamond paste suspended in oils. Polishing is best done on a rotating wheel with a layer of fluid on the surface of the polishing cloth using manual or automatic wheels. Generally, bulk samples are polished, but thin sections can also be produced by similar polishing techniques and then examined by optical or scanning electron microscopy.

12.3.3. Microtomy of Polymers

The three forms of microtomy useful for polymers are simple peel-back methods, thin sectioning, and ultrathin sectioning. These methods are useful for direct examination of the bulk of fibers and films and for the formation of flat surfaces for x-ray microanalysis of embedded specimens or sections of larger samples. Both thin-sectioning methods are adapted from biological preparations.

Synthetic fibers were among the earliest polymers produced for commercial applications and thus the first to be characterized in order to understand structure–property relations. During the 1950s the SEM was not yet available, but the Scott (1959) peel-back technique was used to make thin slivers of fibers for TEM. This split section or cleavage technique is still used to open up a fiber or film with minimal disruption for SEM imaging and comparisons with fractured surfaces and thin sections. The sample is held onto a glass slide or SEM stub with an adhesive or tape, and a scalpel or razor used to nick the surface. Fine forceps are used to peel back the cut either near the central or surface part of the specimen. An example of a polyester fiber prepared by this method is shown in Fig. 12.4.

Specimens for microtomy are cut to fit in the chuck of the microtome if they are self-supporting, or they may be embedded in a supporting medium (Chapter 13.3.8). Pieces of molded or extruded plastics are trimmed to fit the chuck using small saws and razor blades. Films and fibers may be cut using special SEM holders or they may be embedded in a resin for support and orientation. Small cardboard frames can be made easily that fit into capsules or small vials for embedding once the samples are mounted. Standard embedding media (e.g., epoxies) are used. Most polymer materials are dry, or can be air-dried without structural change, whereas others are wet and require drying prior to embedding. Air drying may destroy or distort structures due to the deleterious effects of surface tension and these materials may require special drying prior to embedding (see Section 12.3.8). Glass knifes are best used to prepare flat block faces for SEM. The ESEM may permit direct examination of wet polymers with water held in equilibrium with water vapor.

Typically cross or longitudinal sections and sections at 35° angles are cut for optical microscopy or TEM, although the sections or the cut block faces may be used for SEM. Block faces of specimens are flat surfaces that may be etched, stained, or examined directly by SEM, generally using backscattered electron imaging and x-ray analysis. Samples are generally metal-coated or observed without coating for imaging, whereas carbon coating is used for x-ray analysis (see Chapter 15).

12.3.4. Fracture of Polymer Materials

The fracture of polymer materials is one of the most common specimen preparation methods to observe bulk microstructure in the SEM. The superior depth of focus of the SEM makes it the ideal imaging tool. It is typical that materials are fractured in use or simply shattered with a hammer or some other method that is not reproducible.

The bulk fracture surface may come from one of three sources:

1. Deformation of the sample in a standard mechanical testing device, such as a tensile tester (e.g., Instron), an impact testing machine (e.g., Charpy or Izod), or a tear tester
2. Deformation by fracture using special *in situ* testing devices generally during observation in the SEM
3. Yielding or fracture of a real article or product during service or testing

Extensive discussions of polymer fractography, including crazing, are found in basic texts (e.g., Kausch, 1978, 1983), and it is quite obvious that the preparation method involves a damage process. In the first case, mechanical deformation is part of the sample preparation and the morphology

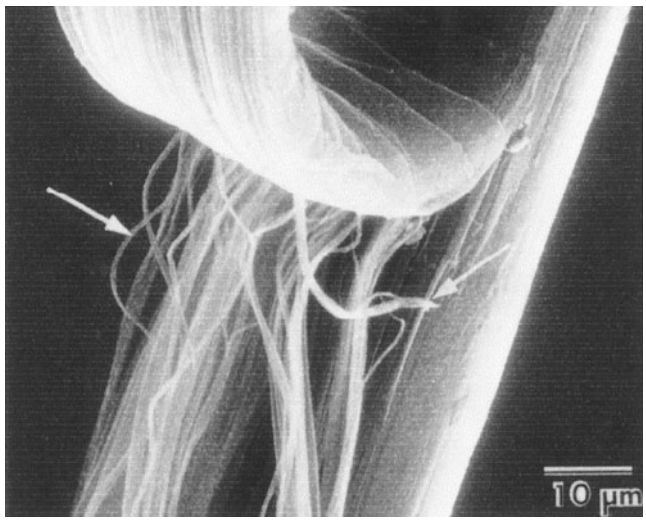


Figure 12.4. Secondary electron image of a polymer fiber prepared by the peelback method, revealing the internal fibrillar texture (arrow) (Sawyer and Grubb, 1987).

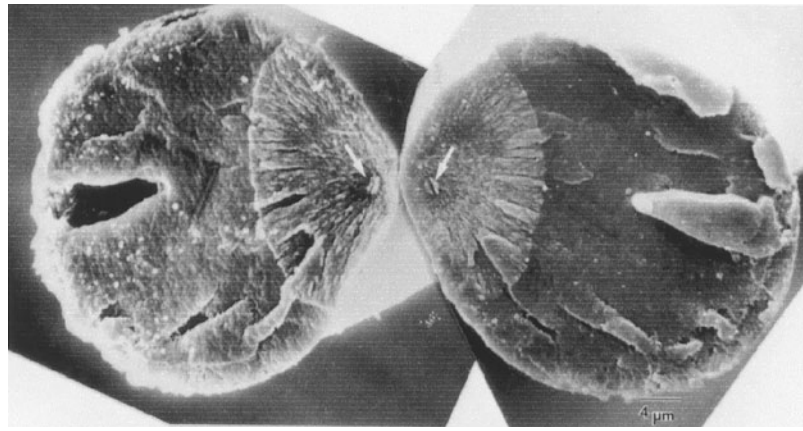


Figure 12.5. SEM images of matched fiber end broken in an Instron (Sawyer and Grubb, 1996).

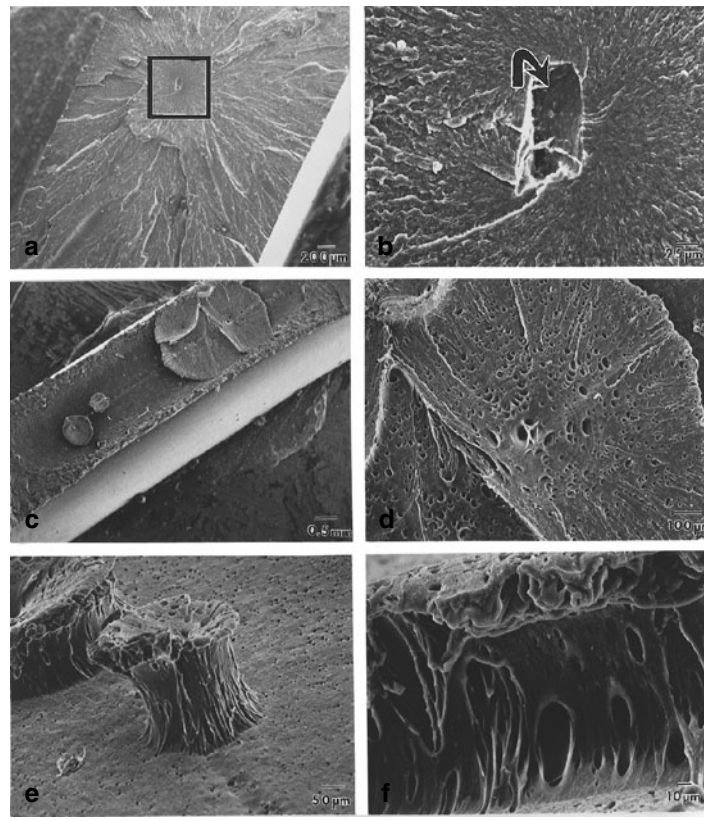


Figure 12.6. SEM images of brittle and ductile fractures. The SEM was used to compare the fracture morphology of two semicrystalline specimens that were processed similarly and yet produced different elongation properties. The nylon specimen with smaller elongation properties (a, b) has a brittle fracture morphology with a rectangular particle at the locus of failure. The specimen with higher elongation properties (c–f) exhibited ductile failure. (Sawyer and Grubb, 1987.)

observed can readily be compared to actual mechanical property data, such as correlation of dispersed particles in a molded resin with the impact strength. This method is highly preferred and if enough samples are examined it might be reproducible, in contrast to the use of a hammer or even the commonly used procedure of cooling in liquid nitrogen followed by percusive fracture on a lab bench.

The second case requires extensive testing and time to collect enough information to have statistical significance. Such studies of fibers, however, can reveal the mode of deformation for comparison with fibers fractured in a tensile tester. SEM images of the matched ends of a polyester fiber from testing in an Instron tensile tester (Fig. 12.5) reveal detail that will be discussed below (see Section 12.4.1).

The third case is the most prevalent: the fracture of a real product during service. The fracture of molded plastic articles is often conducted for problem solving or during product development. The difference between ductile and brittle failure in molded nylon is clearly seen, with the locus of failure in the brittle fracture surface (Fig. 12.6) aiding identification that can provide insights into improving the product. Fracture surfaces of composite samples are commonly analyzed to understand the adhesion of the fibers or minerals to the matrix for correlation with mechanical properties and for failure analysis.

A set of SEM images of a fractured glass filled composite is given in Fig. 12.7. Figure 12.7a shows a sample which exhibits poor adhesion. Figure 12.7b shows a sample with excellent adhesion of the matrix to the fiber surface. BSE imaging is helpful when examining mineral-filled and fractured composites because the atomic number contrast shows the presence of particles that may not be observed by the Everhart–Thornley detector (Fig. 12.8). Finally, x-ray analysis is useful for identifying the elemental composition of defects causing failure. The SEM image and EDS x-ray map

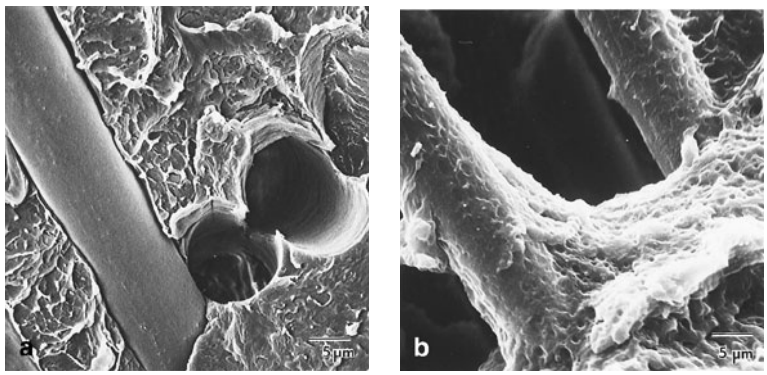


Figure 12.7. Secondary electron images of a glass-filled composite. Images of two glass-filled composites with different adhesive properties are shown. (a) A composite with poor adhesion. Holes are seen where fibers with poor adhesion pulled out of the matrix. (b) Failure in the matrix with resin on the fiber surface suggests good adhesion. (Sawyer and Grubb, 1987.)

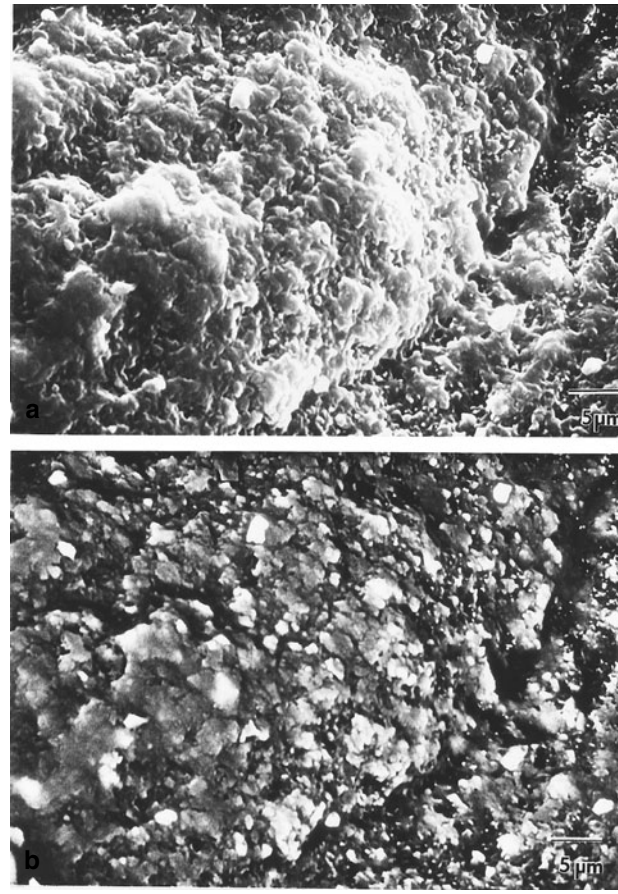


Figure 12.8. (a) Secondary electron image of a mineral-filled polymer composite. This image does not reveal the nature of the dispersed filler particles. (b) Backscattered electron image shows the higher atomic number mineral filler. (Sawyer and Grubb, 1987.)

show a particle at the locus of failure in a fractured molding (Figs. 12.9a and 12.9b). The EDS x-ray spectrum suggests a TiO_2 particle caused the failure in the nylon molded part. $TiO_2$is often used as a pigment, but larger particles may result in premature failure.

12.3.5. Staining of Polymers

Most polymers, in common with biological materials, are composed of low-atomic-number elements and thus exhibit little variation in electron density. Optical microscopy can be used to enhance the contrast of thin specimens or sections using polarized light or phase contrast techniques. However, for electron microscopy the lack of contrast results in little microstructural information. Staining involves the incorporation of high-atomic-number atoms into the polymer to increase the density and thus enhance contrast. Thus staining will refer here to either the chemical

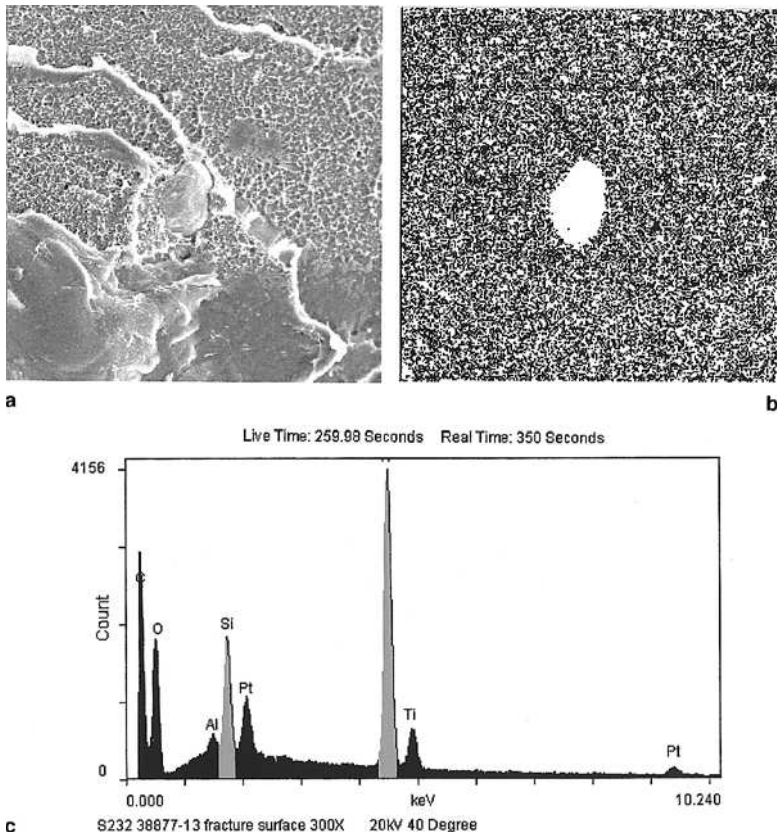


Figure 12.9. TiO_2 particle in a molding fracture surface. (a) Secondary electron image, (b) Ti x-ray map using EDS, (c) EDS spectrum of the molding fracture surface. (Courtesy of Thomas Bruno, Ticona.)

or physical incorporation of heavy atoms. Staining methods were first applied to natural polymer fibers such as cellulosic textiles. Many of the stains used for polymers have been adapted from biological methods (Chapter 13). Biological materials often have sites that react with various stains, whereas there is little differential contrast or sites for such stains in most polymers.

Staining of polymers has been reviewed in detail (Sawyer and Grubb, 1996) and will only be summarized here (see Enhancements, Tables 12.1 and 12.2). Most of the stains applied to polymers are positive stains. In positive staining, the region of interest is stained dark by either a chemical interaction or selective physical absorption. Stains may be applied directly to the specimen, to block faces produced by microtomy, or to the sections themselves. In negative staining the shape of small particles mounted on smooth substrates is shown by staining the regions surrounding the particles rather than the particles themselves. Such staining methods are often applied to latex or emulsion materials. Staining can be used to enhance contrast for SEM, generally with BSE imaging, to reveal details in multiphase polymers and in composites with inorganic fillers.

12.3.5.1. Osmium Tetroxide and Ruthenium Tetroxide

The most commonly used stains are osmium tetroxide and ruthenium tetroxide. Osmium tetroxide, either as a vapor or in aqueous solution, reacts with carbon–carbon double bonds and finds general use for many polymers. Multiphase polymers containing an unsaturated rubber or elastomer phase form the largest single group of polymers studied by microscopy. Osmium tetroxide not only acts as a stain, but also tends to harden the rubber phase, making sectioning of stained specimens easier. Unsaturated latex particles are also readily stained providing contrast enhancement and hardening important for accurate particle-size measurements. Preferential absorption of OsO_4 into the amorphous regions of spherulites has revealed their structures in semicrystalline polymers. In some cases two-step reactions are needed with pretreatments such as alkaline saponification at boiling temperature. Polymers with no double bonds for reaction can be treated, for instance, with unsaturated surfactants, which can fill the voids in a membrane, prior to staining to enhance contrast, as shown in the micrographs of a microporous membrane (Section 12.4.2).

Ruthenium tetroxide is a stronger oxidizing agent than osmium tetroxide and is thought by some to be superior for staining rubber. This chemical oxidizes aromatic rings and can stain saturated polymers and others not readily stained by osmium. Ruthenium tetroxide can be used to stain latex and resin materials by crosslinking ester groups. Aqueous solutions (1%) or vapor staining can be conducted for ether, alcohol, aromatic, or amine moieties. Combined use of OsO_4 and RuO_4 reveals interesting details in some multiphase polymers. RuO_4 is much less stable chemically and thus is used less frequently.

12.3.5.2. Ebonite

Thin sectioning of composites containing polymers is often quite difficult because the soft phases tend to smear during preparation. The ebonite method is useful for hardening rubbers and making them easier to cut in materials such as multiphase blends and tire cords. The method involves treatment of a small sample block in molten sulfur followed by sectioning. Comparison with cryomicrotomy has been conducted and the ebonite method is thought to be easier.

12.3.5.3. Chlorosulfonic Acid and Phosphotungstic Acid

Chlorosulfonic acid crosslinks, stabilizes, and stains the amorphous material in semicrystalline polymers such as polyolefins by incorporation of chlorine and sulfur on the lamellar surfaces. Poststaining with uranyl acetate is typical because it intensifies and stabilizes the contrast. This method is most often used to image the lamellar textures in polyolefins. Phosphotungstic acid reacts with surface functional groups, such as hydroxyl, carboxyl, and amine and is used to stain nylon and the lamellar regions of polypropylene. Its reaction with epoxies precludes its use prior to embedding.

Etching is a preparation method that potentially enhances the information available in a polymer, although the methods are not reproducible and the resulting images may be difficult to interpret. The three general categories of etching are dissolution, chemical attack, and bombardment with charged particles as in plasma and ion beam etching.

Dissolution implies the removal of whole molecules of a material as it is dissolved. Solvent extraction tends to be incomplete and material is often left on the surface. In chemical etching there is chemical attack on molecules and removal of fragments. For instance, acid treatments can selectively oxidize one phase in a multiphase material, aiding contrast. The most serious dilemma is that etching removes one phase, but that material may have another polymer included or disrupted from the bulk. These chemical methods require careful experimentation for the best reactant for each polymer specimen. Time, temperature, and handling are other variables that tend to render these methods nonreproducible. An example of SEM imaging of a chemically etched molded polymer surface is shown in Fig. 12.10, in which amorphous regions of the polymer are preferentially etched.

Charged species activated by high voltages (ion etching) or in radio-frequency plasma are also used to etch away material differentially on the surface of specimens. Plasma and ion beams are employed to bombard or sputter and remove surface atoms and molecules. A major difficulty in plasma and ion etching is that textures such as steps, cones, and holes can be produced, which are artifacts having little to do with the underlying microstructure. Complementary methods should always be used when etching of any type is used for sample preparation.

There is no one general method applicable for either ion or plasma etching polymers. Etch times are generally short, on the order of 5–30 min, because it is important to remove all surface residues. Rotation of the specimen and the target and cooling of the ion guns can minimize artifact

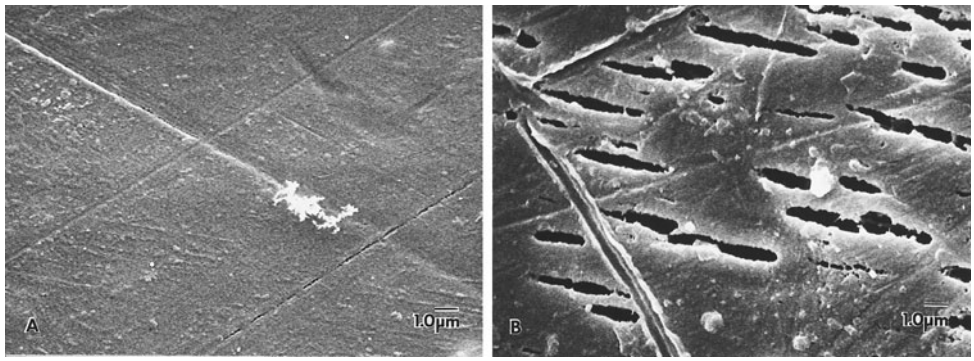


Figure 12.10. Secondary electron images of unetched (a) and chemically etched (b) polyacetal. (a) Molded polyacetal surface showing a smooth texture with little surface detail. (b) Elongated pits, oriented in the direction of polymer flow after etching for short times. (Sawyer and Grubb, 1987.)

formation. Plasma etching is less complicated than ion etching. Oxygen and argon are generally used in these methods.

12.3.7. Replication of Polymers

Replication is one of the oldest methods used for the production of thin TEM specimens (see Chapter 13.3.9.4). Replica methods have been used for SEM samples too large (e.g., orthopedic implants) or too volatile (e.g., light oils on sample surface) for direct imaging. Solvents in the replicating material can form artifacts and good release of the replica may be difficult if the replicating material adheres to the polymer. Single-stage replicas are negative impressions of the specimen surface and this must be borne in mind when interpreting these images. Two-stage replicas are easier to interpret. A general method for making two-stage replicas is to apply a mixture of a silicone and a cure agent and after curing to peel off the replica. Epoxies can be used to fill the molded silicone replica to form a positive replica for the SEM. Rough surface textures can be readily imaged in the SEM at magnifications as high as 5000×.

12.3.8. Rapid Cooling and Drying Methods for Polymers

Generally the methods decribed here involve special drying of the specimen or some kind of rapid cooling method (see Chapter 14). Many polymer materials are already dry enough to examine directly, although most benefit from the vacuum applied in the metal coating devices. Materials as diverse as emulsions, latexes, polymer blends, wet membranes, and ductile polymers, however, may require special methods of preparation prior to SEM imaging. These materials may be amenable to ESEM observation.

In the case of emulsions, latexes, some adhesives, and wet membranes the specimen is wet, invariably with water, and must be dried without the deleterious effects that result from the stress of surface tension forces. Some latexes can be "fixed" in their original shapes by chemical or physical treatment to make the particles rigid.

12.3.8.1. Simple Cooling Methods

Colloidal suspensions, microemulsions, and liquids may be cooled, albeit slowly, by placing thin layers directly onto a microscope cold stage. Polymer materials that are soft at room temperatures may become rigid when cooled to low temperatures. Polymers can be immersed in liquid nitrogen, preferably for 20–30 min, depending on their thickness, mechanically shattered, and then allowed to reach room temperature prior to mounting and coating for observation. Multiphase polymers and blends benefit from being notched prior to cooling and fracturing in an impact-testing machine for production of a more reproducible surface and comparison with impact properties. Details of these and related procedures are considered in Chapter 14.

12.3.8.2. Freeze-Drying

The simple method of drying by replacing the water with an organic solvent of lower surface tension is not generally used due to the potential detrimental effect of the solvent on the polymer. Freeze-drying has been used successfully to prepare film-forming latex and emulsions that have a glass transition slightly below room temperature. Details of freeze-drying processes are given in Chapter 14.

12.3.8.3. Critical-Point Drying

Critical-point drying is described in detail in Chapter 14. Care must be taken in the choice of both the dehydration and transition fluids in order to limit sample damage. Carbon dioxide dissolves in epoxies and polystyrene. Long-term exposure to high pressures and sudden release can turn polymers into a popcorn-like structure. As an example, a wet polymer membrane used for desalination, immersed in liquid nitrogen, and then hand-fractured revealed a distorted and ductile fracture (Fig. 12.11a) (Sawyer and Jones, 1984). Critical-point drying from carbon dioxide following dehydration into ethanol yielded a membrane that fractured with no evidence of ductility (Fig. 12.11b) and showed robust macrovoids supported by granular walls.

12.4. Choosing Specimen Preparation Methods

The purpose of this section is to discuss the preparation techniques that can be applied to a range of different polymer materials. Methods will be described for fibers, films, membranes, engineering resins and plastics, including polymer blends and composites, and adhesives. More details may be found in Sawyer and Grubb (1996).

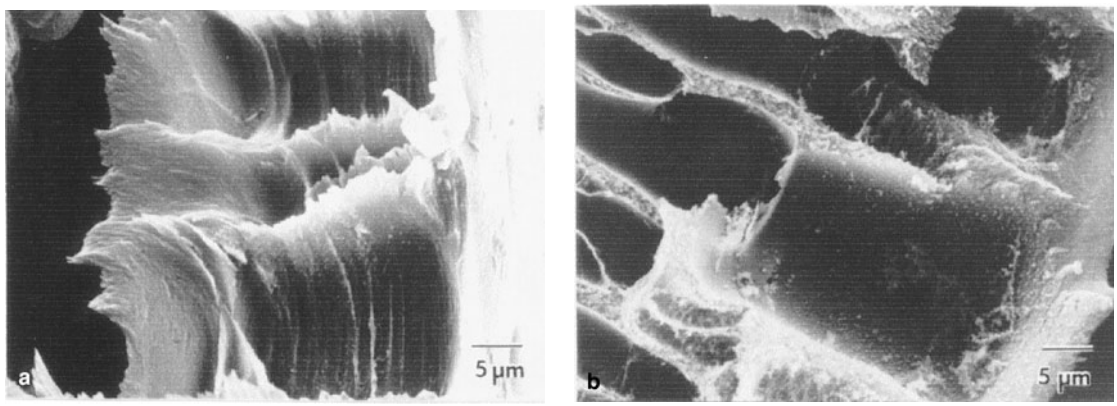


Figure 12.11. SEM images of a reverse osmosis membrane, fractured in liquid nitrogen (a) and fractured after critical-point drying (b) (Sawyer and Grubb, 1996).

12.4.1. Fibers

Polymer fibers are used in many applications and microscopy is used to observe features such as fiber shape, diameter, structure (crystal size, voids, etc.), molecular orientation, size and distribution of additives, structure of yarn and fabric assemblages, and failure mechanisms. These features are directly related to specific mechanical and thermal properties.

The three-dimensional images of fibers in the SEM clearly show surface modifications, finish applications, and the nature and cause of fiber failure. A combination of secondary electron (SE) and BSE imaging is used to distinguish surface debris and contamination from surface morphology. The bulk morphology of many high-performance fibers, such as polyester fibers used in tire cord, is highly fibrillar in nature due to the elongated molecular structure. The peel-back method is useful for providing details on the size and nature of these fibrils (Fig. 12.4). Complementary ultrathin sectioning is used to confirm the size and organization of these fibrils.

Fibers fail by a variety of methods, such as brittle fracture, ductile cracking, axial splitting, and failure due to kink band formation resulting from compression. The mechanism of fiber failure can be determined by fractography studies in the SEM. Although many studies have been conducted in which the fibers are actually stressed within the SEM and videotapes made of the failure, these are very time-consuming. It is more useful to examine fibers after tensile testing in a tensile testing device. The matched fiber ends can be mounted on carbon stubs for x-ray analysis. For example, scanning electron microscopy of matched polyester tire cord fiber ends in Fig. 12.5 reveals a classical fracture surface with a flat mirror region and hackle or ridge marks that literally point to the flaw causing the failure. X-ray EDS analysis of the flaw revealed elements consistent with the catalyst used in the polymerization process.

12.4.2. Films and Membranes

Films find applications in coatings and as packaging materials such as food wraps and in membranes used for separations, controlled release, coating, and packaging barriers. Very thin and/or flat films may not reveal much detail in the SEM and thus optical, TEM, and AFM evaluations are often more informative.

As with fibers, films that are semicrystalline may have spherulitic textures that can be observed directly or in cross sections by polarized light microscopy. Features such as surface morphology, lamellar structure, defects, and molecular-scale ordering can be imaged with the SEM. A standard reverse osmosis membrane, composed of poly(benzimidazole), is shown in Fig. 12.11 after critical-point drying and fracture.

The regular surface features of microporous membranes make them ideal materials for SEM. The surface features are generally very small and difficult to image. Long times are often needed for focusing, resulting in beam damage (Fig. 12.2). Microporous membranes are useful for evaluation of metal coating devices because the granular textures associated

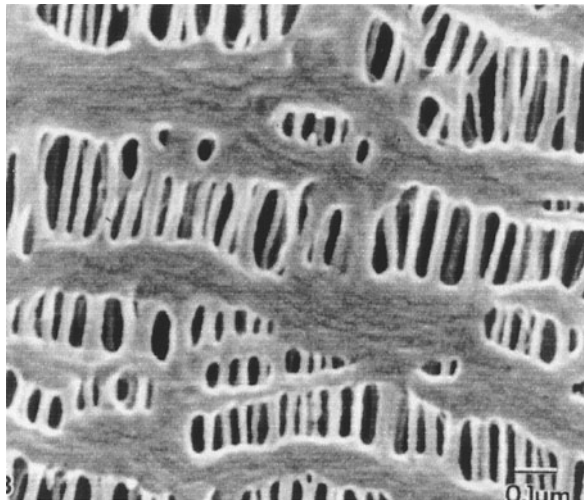


Figure 12.12. FESEM of Pt ion-beam-coated polypropylene microporous membrane (Celgard) (Sawyer and Grubb, 1996).

with thick coatings of gold by sputter or evaporative coating are readily observed (Fig. 12.3). A FEG SEM image of the same membrane ion-beam sputter-coated with 5 nm of Pt is shown in Fig. 12.12. This image shows no granular texture, but the oriented structure with parallel arrays of pores aligned parallel to the machine direction and with drawn fibrils separating regions of undrawn crystalline lamellae define the pore volume (Sarada *et al.,* 1983). Interestingly, this surface image alone would suggest there is a more regular morphology than the tortuous morphology revealed by the TEM sections cut along, across, and in the plane of the machine direction (Fig. 12.13).

12.4.3. Engineering Resins and Plastics

The microstructure of engineering resins and plastics is determined by the chemical composition and manufacturing processes (i.e., molding, extrusion). The SEM is widely used for understanding structure–property relations because microstructure often plays a major role in mechanical properties of the article. Single- and multiphase polymers fit into this section, as do blends of polymers with elastomers, although the multiphase polymers are more complex and thus microscopy is more informative for these materials e.g. (Paul and Sperling, 1986; Paul *et al.,* 1988).

Crystalline polymers are imaged by polarized light to reveal the size and distribution of spherulites. Surface details, such as wear and abrasion, are imaged by SEM, as are the bulk textures of fractured or sectioned specimens. Fractured moldings or extrudates imaged in the SEM can reveal the skin-core oriented texture, the presence of voids, or the dispersion of mineral particles or fibers. Phase-contrast light microscopy and SEM provide images of dispersed-phase particle sizes, shapes, and distributions

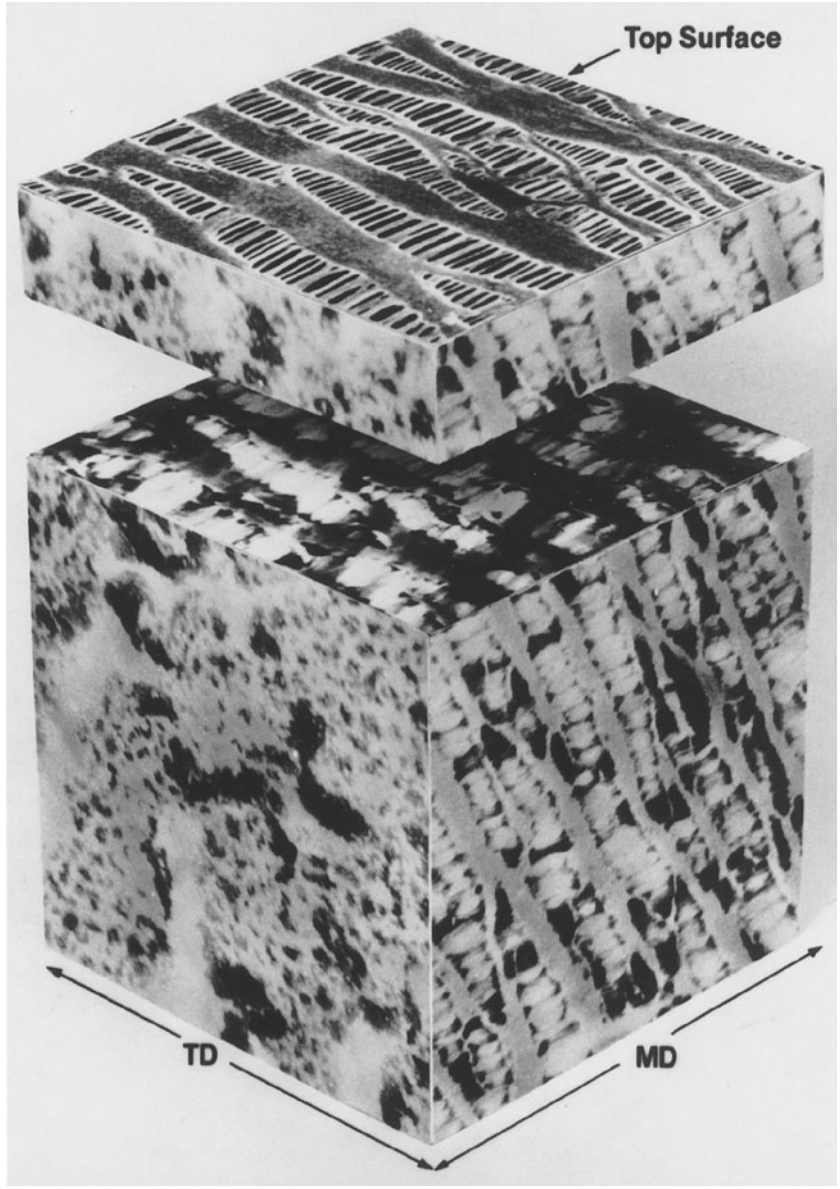


Figure 12.13. Three-dimensional model of a Celgard membrane developed using TEM stained sections for the three dimensions and FESEM for the surface view (Sawyer and Grubb, 1996).

with enhanced contrast offered by staining or BSE imaging. Differential radiation-induced contrast can also aid imaging of multicomponent polymers if they react differently to the electron beam.

Spherulitic textures (see the accompanying CD, Enhancements for Figs. 12.1 and 12.2) are typical of engineering resins. The surface of a

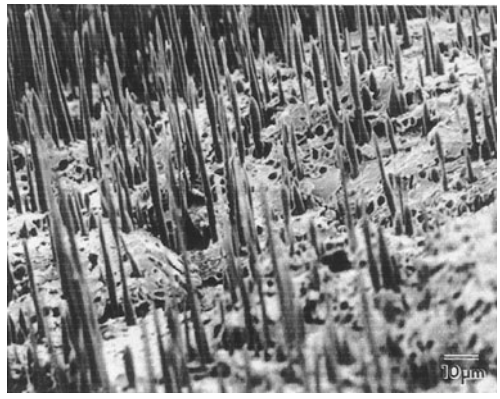


Figure 12.14. FESEM image of a two-phase polymer blend with the sample tilted to reveal the long fibrils of the dispersed phase (Sawyer and Grubb, 1996).

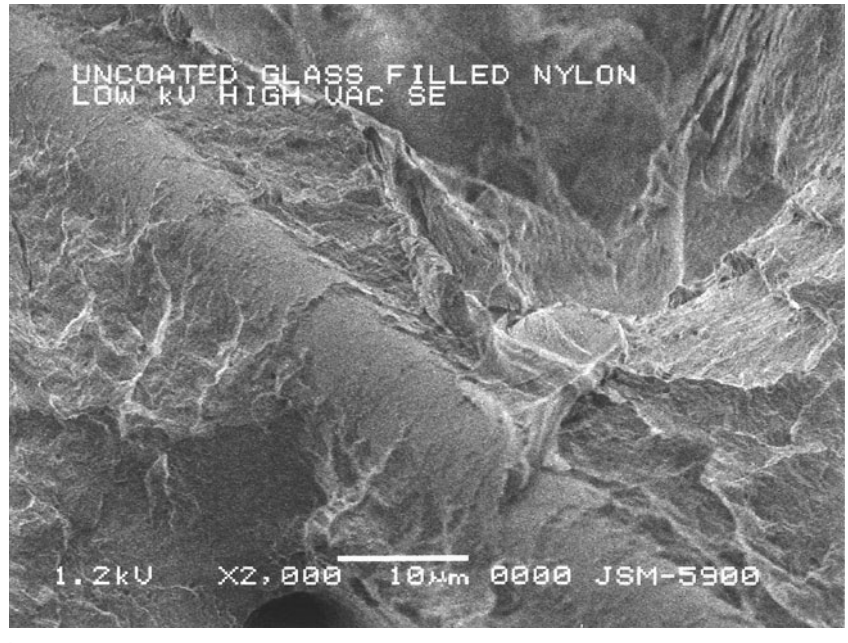
molded polyacetal shown with radiation damage is typical of a molded resin surface (Fig. 12.1). Fractured tensile bars of nylon reveal typical brittle and ductile fracture behavior (Fig. 12.6), useful for comparison with actual resin properties. Brittle fracture is generally due to a contaminant within the molded part, as was shown in the example in Fig. 12.9.

Multicomponent polymers and blends are commonly assessed using SEM of fracture surfaces. Such images can provide information about the interfacial region and adhesion of the various phases in the blend. Phase-contrast and polarized light microscopy and TEM of thin sections are used in combination with SEM to thoroughly describe complex morphologies of such multicomponent polymers and blends. A major value of SEM is in rapid imaging of three-dimensional fields of view, such as in the fracture surface of a polymer blend in Fig. 12.14. The very long, fibrillar phase has poor adhesion to the matrix polymer, although the fine fibril diameter results in improved stiffness.

Scanning electron microscopy of composites, engineering resins filled or reinforced with minerals, glass, carbon, ceramic, or polymer fibers can provide morphological details to compare with mechanical properties. Assessment of the dispersion of mineral particles (Fig. 12.8), adhesion of the matrix polymer to the fiber, and their protruding length (Fig. 12.7) can provide some insights. A comparative study using different types of SEM instrumentation is worthwhile when imaging fiber-reinforced composites. A tungsten-based conventional SEM was used to image an Au–Pd sputter-coated fractured glass-reinforced composite in Fig. 12.7. An uncoated glass fiber-filled composite is shown in a BSE image taken at low vacuum in a variable-pressure SEM in Fig. 12.15a and in a SE image at high vacuum in Fig. 12.15b. The atomic number contrast in the BSE image shows the fiber and the polymer coating more clearly than in the SE image.

Liquid crystalline polymers (LCP) are high-performance engineering resins (Baer and Moet, 1991; Sawyer and Jaffe, 1986). Their highly oriented structures show the typical skin–core morphologies observed for most molded and extruded polymers by polarized light and by SEM of

a



b

Figure 12.15. An uncoated glass-fiber-filled composite examined in a variable-pressure SEM. (a) BSE image taken at low vacuum and 5 keV, (b) secondary electron image taken at high vacuum and 1.2 keV. (Courtesy of Vern Robertson, JEOL.)

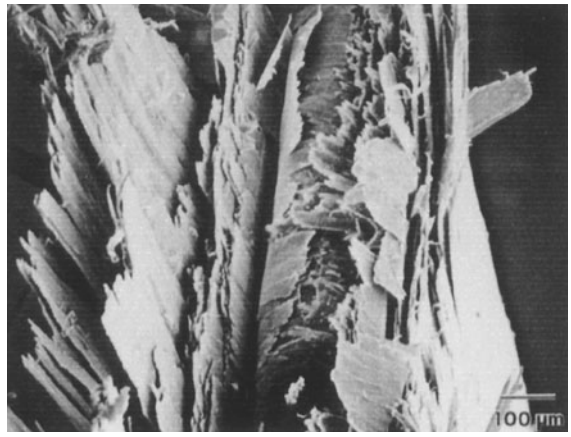


Figure 12.16. SEM of a fracture surface of a liquid crystalline polymer molded part (Sawyer and Grubb, 1996).

a fractured molding in (Fig. 12.16). Complementary microscopy methods have been used profitably to reveal the three-dimensional structure of liquid crystalline polymers and fibers (Sawyer and Grubb, 1996).

12.4.4. Emulsions and Adhesives

Emulsions, dispersions of one liquid in another liquid, find broad application in the fields of food, medicine, pesticides, paints and coatings, paper, printing, and cosmetics. Broadly these materials can be classified as macroemulsions, latexes, colloids, and microemulsions. Ranging in size from 10 nm to about 50 μm, the major structural unit of interest in these materials is the "particle," whose shape, size, and distribution control the properties and end-use applications. Polymer liquids in the form of emulsions and adhesives are wet and sticky, making preparation for microscopy very difficult. Because particle size and shape are so important, the preparation must focus on not changing the original particles. Preparation methods include cryotechniques, staining–fixation (Shaffer, 1983), microtomy, and simple methods such as dropping the solution on a specimen holder for air-drying or freeze-drying.

Adhesives are polymers that are used in many applications. A particularly demanding application is the cementing of metal joints in military aircraft with polymers such as epoxy resins. The interfaces in such materials must be characterized to determine the strength of the adhesive bond and the relation of such properties as peel strength with morphology. This morphology is generally investigated in the SEM. Adhesion science and adhesion and absorption of polymers have been described in volumes edited by Lee (1980). Sticky tapes and labels are obvious uses of adhesives that cover a broad range of applications. Generally rounded domains or particles are desired to control the adhesive properties; an example is shown in

Fig. 12.17 of the adhesive layer on a Post-it product. The regions between the domains clearly do not stick to surfaces and provide the unique ability to adhere and yet be peelable.

12.5. Problem-Solving Protocol

The preceding sections have provided the concepts to describe polymer materials and the specimen preparation methods required for SEM imaging and analysis. A problem-solving protocol is useful for microscopy characterization to follow an easy and short path to a solution. Generally more than one microscopy technique is required for problem solving using the SEM with EDS/WDS analysis and imaging to add critical information about the morphology and chemistry of the material. As has been shown, light and transmission electron microscopy also play a major role in morphological analysis.

The difficulty in taking that first step is that it is not always obvious what to do first and there may not be much information or actual specimen available. Failure of molded or extruded products in use due to customer testing or as part of product development protocols is common and often little sample and even less sample history are available. Some simple steps are worth noting as a problem-solving protocol:

- Collect all the facts and data about the problem.
- Clearly define the problem, objectives, and timetable.

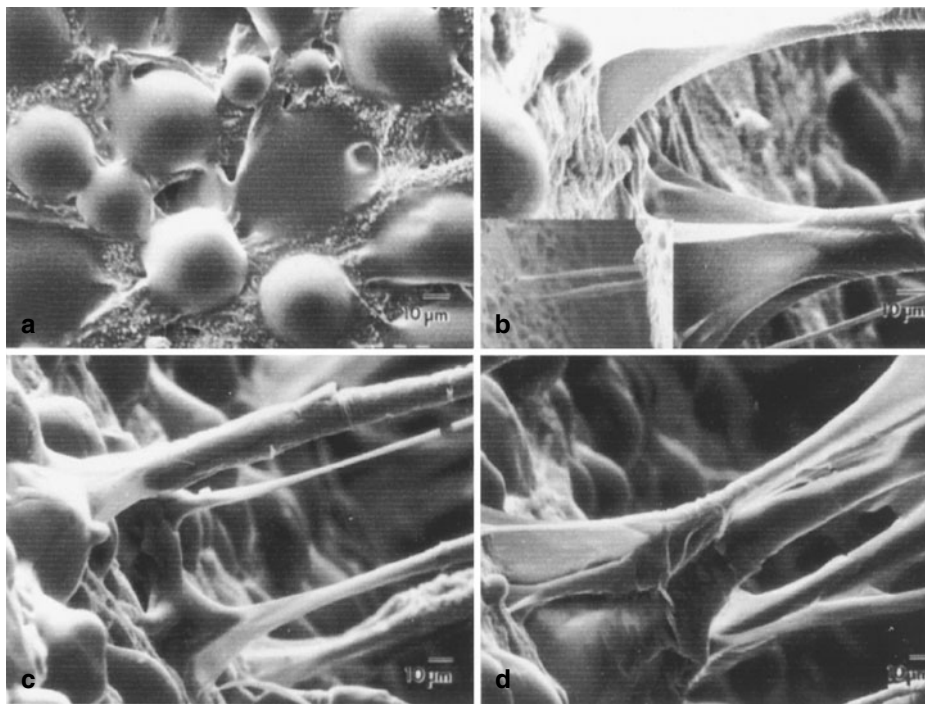


Figure 12.17. SEM images of an adhesive surface (Sawyer and Grubb, 1996).

- Select the specimen for solving the problem.
- Define scientific controls and materials to aid interpretation.
- Examine the specimens by eye and using a stereo-binocular microscopy to select the region for assessment.
- Determine the size of the morphology that needs to be imaged.
- Define the microscopy technique and best specimen preparation methods for imaging and microanalysis.
- Take representative images and label those not representative, but important with regard to the problem being solved.
- Assess potential artifacts as part of the image interpretation process.

12.6. Image Interpretation and Artifacts

The key to success is image interpretation, which can only come from an understanding of the topics discussed in this chapter. Critical to image formation is a good understanding of the polymer material, the history of the specimen in the case of problem solving, and the polymer process used to form the sample, for example, into a molded part or fiber. It is of equal importance to understand how the specimen can be changed during specimen preparation. Often artifacts are formed on or in the specimen during preparation and these artifacts may be interpreted as being related to the specimen.

A major consideration in the selection of preparation methods for microscopy study is the nature of the potential artifacts formed. SEM preparations are generally direct and rapid, yet there are quite a number of potential artifacts. Paints, glues, and tapes used to attach the specimen to the SEM sample holder can wick up or contaminate the specimen surfaces, adding false structures to the image. Preparation of specimens from the bulk generally involves deformation of the specimen, for example, peeling, fracturing, sectioning, and etching. Peeling results in fibrillation, which is a function of the material and the force used. Fracturing at room temperature often causes deformation of soft or rubbery phases, whereas fracture in liquid nitrogen causes less deformation, but tends to be less reproducible. Fracturing by standardized testing, for example, Izod fracture, is more reproducible and permits comparison of structure and properties. However, even in this case, several samples might need to be assessed to ensure the fracture surface is representative of the material. Sections and sectioned block faces exhibit knife marks and other textures associated with the preparation that must be noted and not attributed to the material. Etching is one of the methods with the most opportunity for imaging confusing details that are not associated with the material. Chemical etching removes polymer material, often redepositing it on the surface and creating new structures unrelated to the material. Ion etching can result in heating and melting of the specimen, and formation of cone-like structures is often observed, especially if water-cooled guns are not used with low glancing angles. Plasma etching is less likely to form spurious structures, although if there is oxygen present, there can be surface oxidation. Replication often results in loss of details depending on the method and material used. The

application of metal or carbon coatings to polymers can cause formation of grain-like textures or melt or otherwise damage the specimen surface. Finally, irreversible radiation or beam damage and reversible charging effects also can cause formation of image details not related to the material and problem being solved.

References

Baer, E., and A. Moet, eds. (1991). *High Performance Polymers,* Hanser, New York.

Bassett, D. C., ed. (1988). *Developments in Crystalline Polymers—2,* Elsevier Applied Science, London.

Ciferri, A., and I. M. Ward, eds. (1979). *Ultrahigh Modulus Polymers,* Applied Science, London.

Everhart, T. E., R. F. Herzog, M. S. Chang, and W. J. DeVore (1972). In *Proceedings 6th International Conference on X-ray Optics and Microanalysis* (G. Shinoda, K. Kohra, and T. Ichinokawa, eds.), University of Tokyo, Tokyo.

Grubb, D. T. (1993). In *Materials Science and Technology,* Vol. 12 (E. L. Thomas, ed.), VCH Publishers, Weinheim, Germany.

Goodhew, P. J. (1984). *Specimen Preparation for Transmission Electron Microscopy,* Oxford University Press, Oxford.

Hemsley, D. A. (1984). *The Light Microscopy of Synthetic Polymers,* Oxford University Press, London.

Hemsley, D. A., ed. (1989). *Applied Polymer Light Microscopy,* Elsevier-Applied Science, London.

Hobbs, S. Y., M. E. J. Dekkers, and V. H. Watkins (1988). *J. Mater. Sci.* **23,** 1219.

Kausch, H. H. (1978). *Polymer Fracture,* Springer-Verlag, Berlin.

Kausch, H. H., ed. (1983). *Crazing in Polymers,* Springer-Verlag, Berlin.

Keith, H. D., and F. J. Padden (1963). *J. Appl. Phys.* **34,** 2409.

Lee, L. H., ed. (1980). *Adhesion and Adsorption of Polymers,* Plenum Press, New York.

Paul, D. R., and L. H. Sperling, eds. (1986). *Multicomponent Polymer Materials,* American Chemical Society, Washington, D.C.

Paul, D. R., J. W. Barlow, and H. Keskkula (1988). In *Encyclopedia of Polymer Science,* 2nd ed., Vol. 12 (H. F. Mark, N. M. Bikales, C. G. Overberger, and G. Menges, eds.), Wiley, New York, p. 399.

Sarada, T., L. C. Sawyer, and M. Ostler (1983). *J. Membrane Sci.* **15,** 97.

Sawyer, L. C., and R. S. Jones (1984). *J. Membrane Sci*. **20,** 147.

Sawyer, L. C., and D. T. Grubb (1987). *Polymer Microscopy,* Chapman and Hall, London.

Sawyer, L. C., and D. T. Grubb (1996). *Polymer Microscopy,* 2nd ed., Chapman and Hall, London.

Sawyer, L. C., and M. Jaffe (1986). *J. Mater. Sci.* **21,** 1897.

Scott, R. G. (1959). *ASTM Spec. Tech. Publ.* **257,** 121.

Shaffer, O. L., M. S. El-Aasser, and J. W. Vanderhoff (1983). In *Proceedings of the 41st Annual Meeting of the EMSA,* San Francisco Press, San Francisco, p. 30.

Vezie, D. L., E. L. Thomas, S. D. Hudson, and W. W. Adams (1992). *Polym. Prepr. ACS* **33,** 735.

Ward, I. M., ed. (1975). *Structure and Properties of Oriented Polymers,* Applied Science, London.

Ward, I. M. (1983). *Mechanical Properties of Solid Polymers,* Wiley, New York.

White, J. R., and E. L. Thomas (1984). *Rubber Chem. Technol.* **57**(3), 457.

Young, R. J. (1981). *Introduction to Polymers,* Chapman and Hall, London.

13

Ambient-Temperature Specimen Preparation of Biological Material

13.1. Introduction

We need only to consider our own bodies as an example to realize the complexity of biological material. Our body is three-dimensional and composed primarily of light elements, most of which are organized into a mixture of periodic and aperiodic structures. These structures range from simple molecules and macromolecules to complex heteropolymers, all bathed in an aqueous solution of ions and electrolytes. We are thermodynamically unstable, live at ambient temperatures and pressures, and are sensitive to ionizing radiation.

A simple summary will help to remind us about the problems we face when examining biological material in an electron beam instrument. The pressure inside a modern electron beam instrument is less than 1 ten millionth of an atmosphere. It is very, very dry inside the microscope and the interaction between the electron beam and the sample may result in a considerable rise in local temperature. There is a high flux of ionizing radiation and, finally, the highest signal-to-noise ratio necessary for image formation depends critically on scattering from high-atomic-number elements. The inside of an electron microscope is undoubtedly an alien environment for living material.

There are really only two approaches to biological sample preparation. We can either modify the environment inside the electron microscope or modify the specimen to make it sufficiently robust to withstand the high vacuum and high-energy beam of a standard SEM. There has been some success with the first approach with the arrival of environmental SEMs, variable-pressure SEMs, and low-voltage imaging. These instrumental modifications have been discussed in Chapters 5 and 10.

The second approach is to modify the specimen by transforming it into a form which retains its integrity and remains stable during imaging

and analysis in a high-vacuum instrument. This chapter reviews procedures which have proved successful for preparing biological samples for the SEM. Because specimen preparation is dictated primarily by the sample, it is impossible to provide a methodology for all specimens, although a few generic procedures will be suggested. The methods are invariably the same procedures used so successfully for transmission electron microscopy. Where appropriate, procedures developed specifically for scanning electron microscopy and *in situ* chemical analysis will be given emphasis together with a few key references. For more detailed information on sample preparation, the recent publications by Bozzola and Russell (1999), Dashek (2000), Glauert and Lewis (1998), Gupta (2000), Hayat (2000), and Maunsbach and Afzelius (1999) all contain a wealth of well-tested procedures. Other good sources of up-to-date information are the News and Commentary section of the Microscope Society of America (MSA) publication *Microscopy and Microanalysis* and the MSA ListServer.

The next section of this chapter discusses procedures for structural microscopy where the methods are generally invasive and are less concerned with preserving the precise chemical constitution of the specimen. The subsequent section gives the principles and procedures that must be adopted if the *in situ* chemical constitution of the sample is of prime importance and where, by necessity, the procedures have to be less invasive.

Before embarking on the details, we must offer two cautionary warnings. First, before carrying out any specimen preparation, it is imperative to assemble as much information as possible about the natural properties of the specimen such as pH, tonicity, minimal size, and physiological state. Second, it is essential to have a clear idea about what information is to be obtained about the specimen. Does the interest center on structural details, chemical information (or both)? Are the outside and/or the inside of the specimen to be studied? What information is already available, if not about the actual sample, then about closely related samples?

13.2. Preparative Procedures for the Structural SEM of Single Cells, Biological Particles, and Fibers

13.2.1. Particulate, Cellular, and Fibrous Organic Material

Such samples include seeds, pollen grains, fungal spores, plant, animal and microbial cells, organic debris, and macromolecules. The material may have been collected on a filter either from an air or liquid suspension or as a centrifuged precipitate from a liquid suspension. Single-cell cultures may be grown as monolayers on bio-compatible glass, plastic, or noble metal supports placed in the culture medium. The collected material can be examined directly as a powder, as individual components, or, in the case of single cells, after fixation and dehydration. All samples must be finally deposited on an appropriate smooth support such as polished pyrolitic graphite, pieces

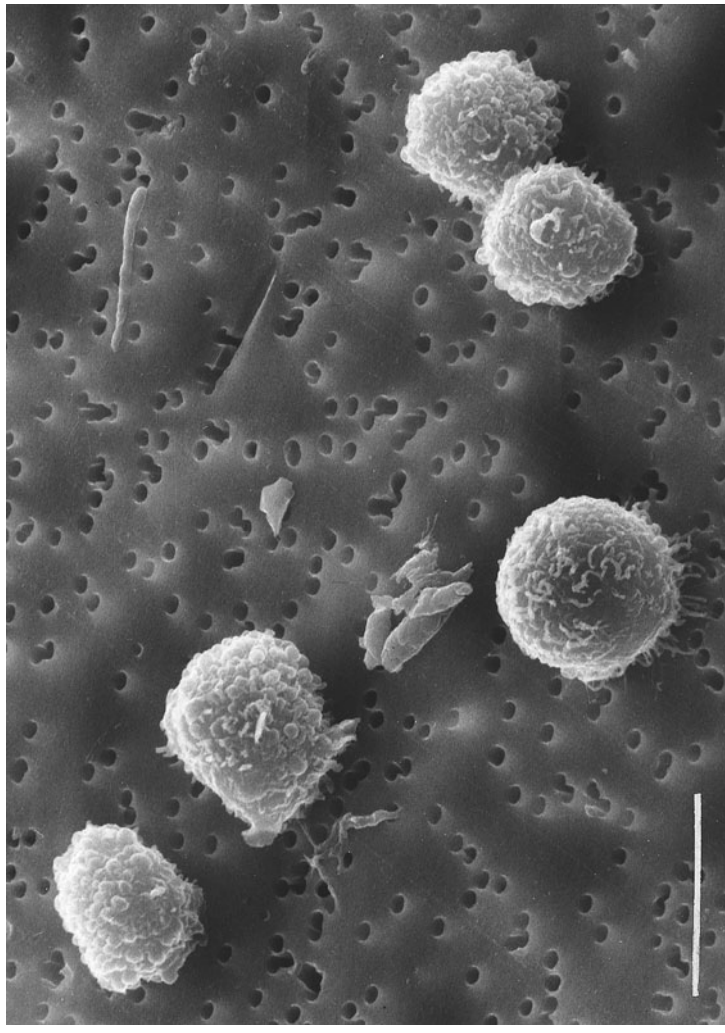


Figure 13.1. SEM image of Chinese hamster ovary (CHO) tissue culture cells on a membrane filter support. Magnification marker = 10 μm. (Wetzel *et al*. 1974.)

of silicon wafer, glass cover slips, or highly polished metal. Figure 13.1 is an image of some tissue culture cells collected on a Millipore filter.

13.2.2. Dry Organic Particles and Fibers

The dry material can be either sprinkled or, using a binocular microscope and a pair of fine tungsten wire needles, more carefully placed at specific sites on a smooth support (specimen stub). If the material has a tendency to move around on the surface, the underlying support should be painted with a very thin layer of glue made from 5 cm of standard sticky tape stripped of adhesive in 10 ml of chloroform. The material can

also be placed on a small piece of double-sided sticky conductive tape attached to the surface of a clean specimen stub. Carbon paint or silver dag is used to make sure there is a thin conductive line from the surface of the tape to the metal stub. The stub is dried overnight, if necessary, in a 40°C oven to remove all the sticky tape volatiles. Finally, the specimen is coated with a thin layer of conductive material (Chapter 15). The paper by Spurney (1994) contains details of the more specialized procedures to be used when collecting very fine dust particles and aerosols.

13.2.2.1. Organic Particles and Fibers on a Filter

The filter may be viewed directly in the SEM. A filter of appropriate pore diameter to retain the sample of interest is chosen and dried. A small section of filter is attached at two or three points to the stub with small drops of conductive paint. It is important that the conductive paint does not wick into the filter. The stub is dried in a 40°C (313 K) oven and coated with a thin layer of conductive material.

13.2.2.2. Organic Particles and Fibers Entrained within a Filter

Provided the particles and fibers are sufficiently robust and/or contain a substantial amount of inorganic material, the organic filter material may be dissolved completely in solvents, entirely removed by cold-plasma etching, or partially removed by solvent etching. The procedure must not affect the sample. If the filter is to be dissolved in solvent, the released material should be treated as detailed in either Sections 13.2.2.3 and 13.2.2.4 below. If the filter is to be removed by cold-plasma etching, a small portion of the filter should be placed on a polished inorganic support and the etching allowed to proceed. The inorganic support and released material should be carefully attached to the specimen stub using a few small drops of conductive paint, dried in an oven, and coated with a thin layer of conductive material.

If the filter is to be partially solvent-etched to release the entrained particles, a thin layer of carbon should be evaporated onto the exposed surface after etching. The filter is then turned over and etching repeated to remove the remainder of the filter. The particles remain trapped within the electron-transparent evaporated carbon layer.

13.2.2.3. Organic Particulate Matter Suspended in a Liquid

Particulate material may be sedimented and washed by centrifugation. Soft particulate matter such as cells and macromolecules will need stabilizing and dehydrating using the procedures given in Section 13.3. Robust material which does not require stabilization should first be dehydrated and then resuspended in a low-surface-tension organic liquid such as diethyl ether and air-brush-sprayed directly onto a clean, polished specimen stub. Alternatively, a small drop of the concentrated particulate suspension may be placed onto a clean, hydrophilic, 10-mm-diameter glass cover slip and allowed to spread as evenly as possible to enable individual particles to separate. This may require careful dilution of the suspension. The glass

support is attached to the metal specimen stub, making sure there is a conductive pathway from the support to the stub. The stub is dried in an oven and coated with a conductive film.

If the particles are suspended in water, the addition of a very small amount of an appropriate surfactant or dispersing agent will help the droplet to spread evenly. The particles should be air-dried in a dust-free environment. Very small particles ($<1\ \mu$m) such as fungal spores and bacteria, may be dispersed in ethanol or methanol followed by ultrasonication.

13.2.2.4. Manipulating Individual Organic Particles

Individual 1- to 100-μm particles may be viewed under a stereo light microscope and picked up and moved to specific locations using finely pointed tungsten needles. Considerable manipulative skills are needed and it may be appropriate to use a micromanipulator. A small drop of collodian solution is placed on an etched finder slide and allowed to dry. A very small drop of amyl acetate is placed on a portion of the dried collodian film and, using a tungsten needle, the particle is transferred to the droplet. The amyl acetate dissolves the film and the particle is bound to the substrate by a 20-nm-thick collodian film. Figure 13.2 shows how this procedure may be

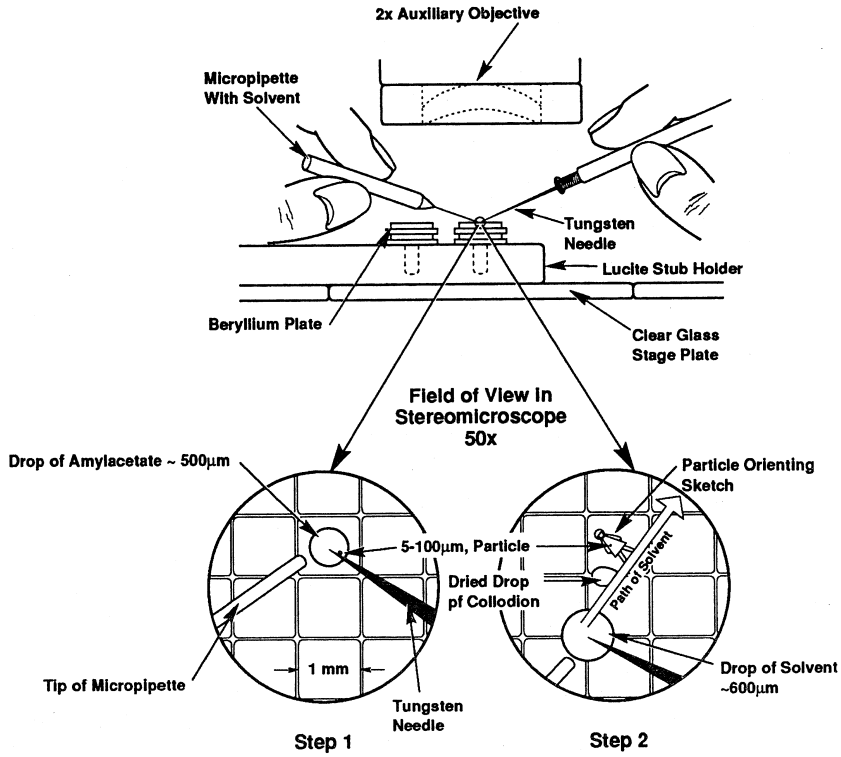


Figure 13.2. Stages in the manipulation of a selected microscopic particle to a prepared substrate (Brown and Teetsov, 1976.)

performed; the papers by Brown and Teetsov (1976) and Teetsov (2000) gives further details of these procedures.

13.3. Preparative Procedures for the Structural Observation of Large Soft Biological Specimens

13.3.1. Introduction

The procedures described in this section are based on using solutions of chemicals to convert the naturally hydrated organic material to dry material composed of crosslinked biopolymers, which have in turn been soaked in solutions of heavy metal salts. The process is akin to fossilization.

13.3.2. Sample Handling before Fixation

If the sample is to be taken from a large specimen, the sample should be as small as is compatible with any physiological or structural features that are to be studied. It is important to remember that the diffusion rate of many of the chemicals used in specimen preparation is about 1–5 μm sec^{-1}. The sample should be uncontaminated and in as near natural a state as possible. If necessary, the surface of the sample should be rinsed or flushed in an appropriate buffer using a soft artist's paintbrush to remove persistent debris. One should only resort to enzymatic digestion and surfactants if this is really necessary. The surface of dry samples may be cleaned with a dry, soft artist's brush or with gentle jets of clean air. Although plasma cleaning is a very effective way of removing organic material from inorganic samples (Isabell *et al.,* 1999), it is not advisable to use this procedure with biological and organic material.

13.3.3. Fixation

This is primarily a chemically invasive process of stabilization, which involves crosslinking and preserving macromolecules, usually at the expense of small molecules and electrolytes. Most of the methods have been derived from TEM and light-optical microscopy (LM) and there is no one method for all samples. Every fixation protocol must be tailor-made for a given sample because of differences in sample size, water content, diffusion rates, penetrability, physiological state, osmolarity, pH, and chemistry. Fixation is the most critical part of sample preparation and it is important to give this process careful attention. A good first step is to consult the literature and discover what procedures have been used in the past to fix samples. The books by Glauert and Lewis (1998) and Hayat (2000) are full of useful information. Two papers by Hockley and Jackson (2000) and Hockley *et al.* (2001) describe a simple-to-construct apparatus for the continuous-flow processing of SEM specimens. More details, including experimental protocols are given in the accompanying CD, Enhancement 13.1.

13.3.4. Microwave Fixation

There is good evidence developed over the past 30 years that the use of microwave radiation during the fixation process will improve the preservation of some samples. Microwaves are a form of nonionizing radiation and increase the diffusion rate of the fixative chemicals by agitating the molecules. The method is good for impervious samples or where rapid processing is needed. Much emphasis has been placed on creating even microwave energy fields and eliminating the heating effects induced by the radiation. The book by Login and Dvorak (1994) is an excellent source of information on this approach to microwave fixation.

More recent studies involving the use of new equipment and different experimental protocols place less reliance on microwave heating and more on the direct effects of the energy of the microwaves. Such an approach appears much less invasive and is being used sucessfully to process thermally sensitive molecules, which are the target of immunocytochemical studies (see Section 13.4.4). The book by Giberson and Demeree (2001) gives details of this new approach to microwave fixation.

13.3.5. Conductive Infiltration

This is a procedure unique to the preparation of specimens for scanning electron microscopy and can be used to increase the bulk conductivity of the sample and thus avoid the need to apply a thin conducting film on the sample surface (see Chapter 15.). Conductive infiltration is achieved by soaking the sample in reactive heavy metal salt solutions together with bifunctional mordents to increase the loading of the heavy metal salts onto the specimen.

The improved sample conductivity allows the specimen to be examined uncoated in the microscope at a much lower beam current and accelerating voltage. It is a useful procedure for backscattered imaging and for porous and fibrous samples. The metal-impregnated sample may be repeatedly dissected to reveal internal detail and highly sculptured samples may be rendered more conductive than by using thin-film coating alone. However, the infiltration procedure may cause severe tissue damage and precludes any subsequent x-ray microanalysis and the use of specific heavy metal stains to localize chemical groups in the specimen. On balance, this procedure should be used as a means of last resort because the spatial resolution of information which can be obtained is not as good as can be obtained from thinly coated samples imaged at low kV. Figure 13.3 shows the type of result which may be obtained; the paper by Murakami *et al*. (1983) and Chapter 6 of the book by Hayat (2000) should be consulted for the operational details of these procedures. A generic procedure for conductive staining is given in Enhancement 13.2 on the accompanying CD.

13.3.6. Dehydration

Unless the sample is going to examined in an environmental or variable-pressure SEM or in a low-temperature SEM, the water in the

specimen must be either immobilized or removed. If the water is to be removed, this must be achieved with minimal extraction (difficult) and minimal distortion (possible). There are several ways to remove water from specimens.

Physicochemical dehydration is the standard procedure used in all forms of microscopy and involves passing the fixed and washed specimen through increasing concentrations of a large volume of ethanol, methanol, or acetone. A gradual progression is better than a stepwise progression. All three dehydrating agents cause varying amounts of extraction of lipid-soluble material, although this effect is less marked with ethanol.

Chemical dehydration takes advantage of the reaction between 2-2 dimethoxypropane (DMP) and water to form a mixture of acetone and methanol. The reaction is exothermic and it is advisable to carry out the process at 4°C (277 K). The methanol–acetone mixture should be allowed to reach room temperature and then substituted with bone-dry 100% ethanol. The sole advantage of the process is speed and it should only be used when time is at a premium. More details, including an experimental procedure, are given in Enhancement 13.3 on the accompanying CD.

Although physicochemical and chemical dehydration procedures remove all the water from the sample, the specimen still ends up in a liquid. Critical-point drying is another procedure unique to the preparation of samples for scanning electron microscopy and allows the sample to be taken from the liquid to the gas phase without any deleterious surface tension effects.

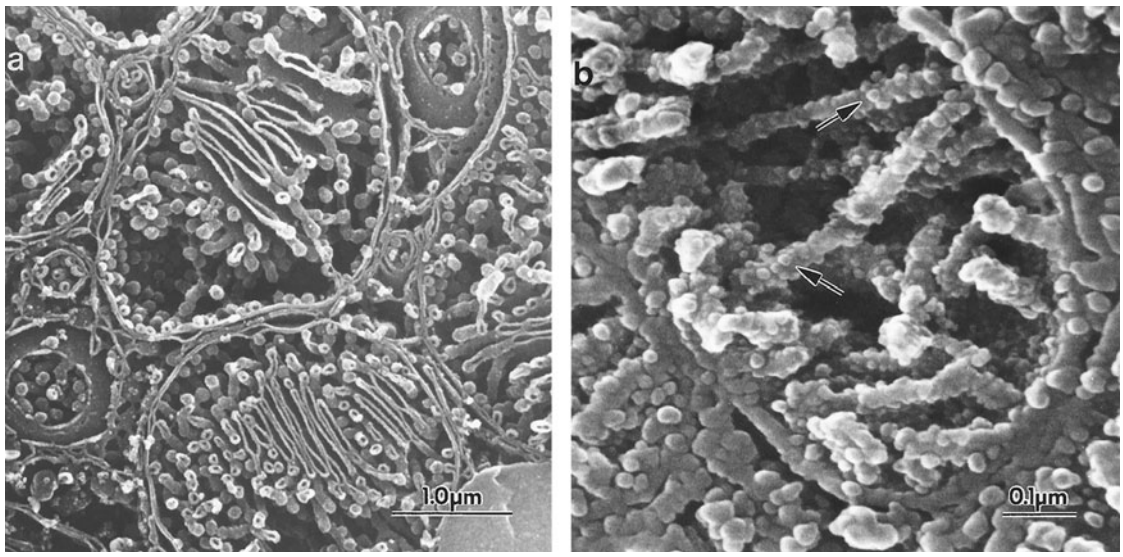


Figure 13.3. Conductive staining methods. (a) Mitochondria of a rat suprarenal gland; lamellar, tubular, and vesicular christae are present (Hanaki et al., 1985). (b) Inner membrane particles on mitochondrial tubules (arrows) from a rat liver cells (Osatake *et al*., 1985).

The equipment is complicated and the procedures are potentially hazardous because high liquid and gas pressures are involved. The 100% dehydrating fluids (ethanol or acetone) are substituted by liquid carbon dioxide, which is then gently heated under pressure to its critical point, where the density of the liquid and gas phases are the same, thus avoiding any damaging surface tension effects. Critical-point drying is a very effective drying procedure, but can cause shrinkage and solvent extraction. Figures 13.4 and 13.5 show respectively section diagram of the apparatus and with a picture of the baskets and troughs needed to hold the specimens. Figure 13.6 is a high-resolution image of the surface of a cell that had been critical-point-dried. More details, including a generic procedure for critical point drying, are given in Enhancement 13.4 on the accompanying CD.

After dehydration in either ethanol or acetone, the sample may be substituted into a highly volatile liquid such as diethyl ether, dimethyl ether, or 1-2 epoxypropane and then air-dried. This procedure is only satisfactory for robust specimens because the surface tension effect at the liquid–air interface can cause damage. Alternatively, the dehydrated specimen may be substituted into tertiary butyl alcohol, which is then removed under low vacuum. The same procedure works using tetramethylsilane. Both methods are rapid, but some shrinkage must be expected.

It is possible to use liquids with a very low vapor pressure, such as glycerol or triethylene glycol, to infiltrate biological specimens so they can be observed in the SEM without drying. The advantages to this rather messy procedure are that it retains lipids and waxes, is inexpensive, and is

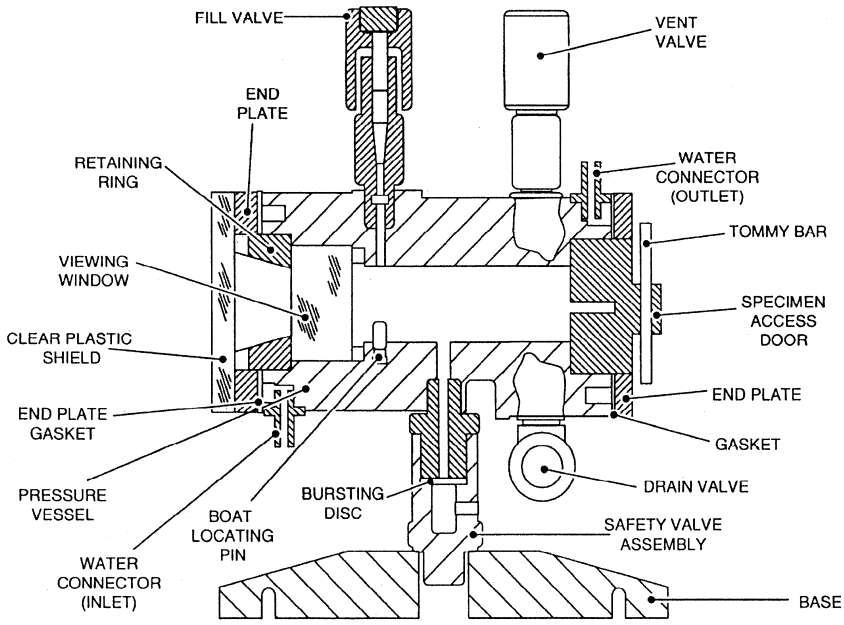


Figure 13.4. Sectional diagram of a critical-point dryer (Quorum Technology UK).

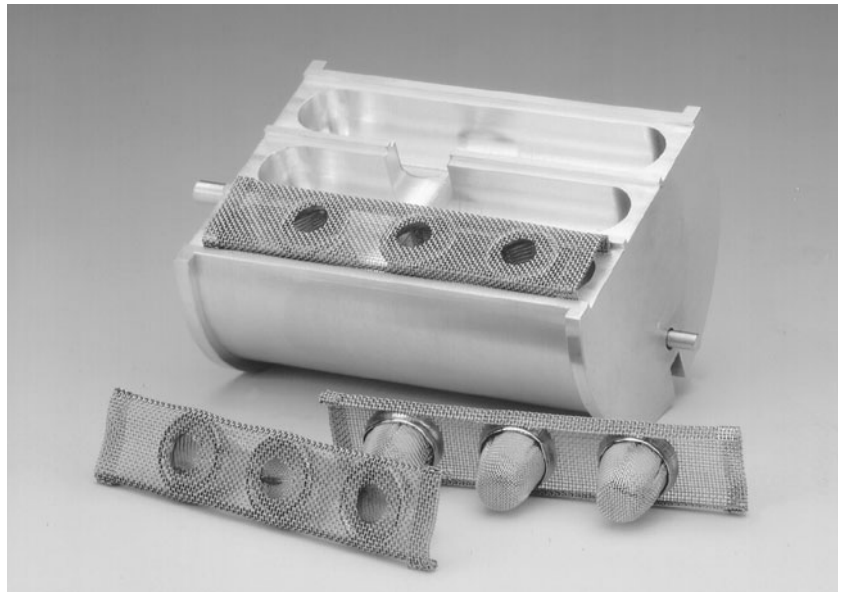


Figure 13.5. Wire baskets and holding troughs for use in a critical-point dryer; this picture shows how the samples are kept below the surface of the dehydrating fluids prior to loading into the critical-point dryer (Quorum Technology, UK).

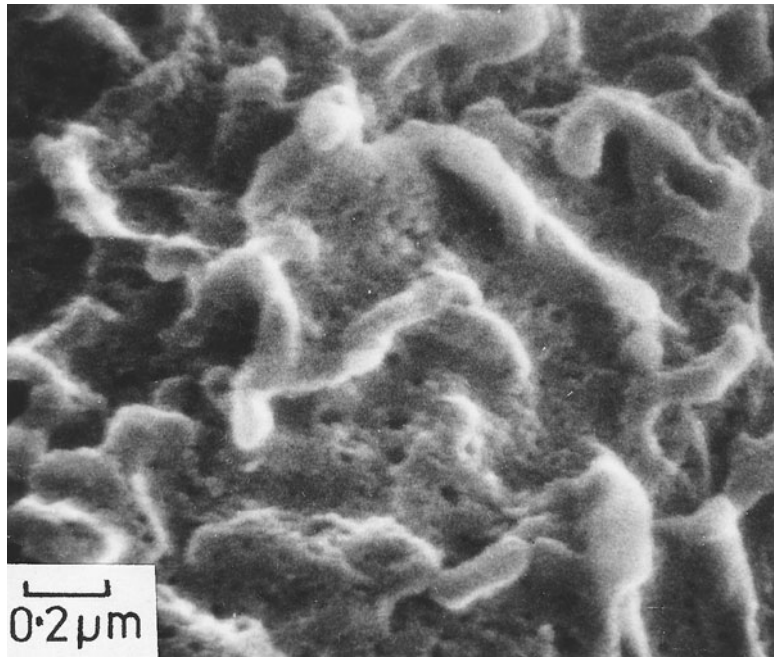


Figure 13.6. An image of the surface of a mastocytoma cell prepared by critical-point drying and imaged in a FEG SEM; note the porosity of the cell surface, which is absent in cells prepared by direct drying methods. (Liepens and de Harven, 1978).

easy to carry out. The paper by Ensikat and Barthlott (1993) gives details of the procedures. Far better results may be obtained by examining wet specimens in an environmental microscope.

Frozen water can be sublimed at low temperature under vacuum. Small samples are quench-cooled and kept below the ice recrystallization point. The sublimation is started at -140°C (133 K) and 10^{-6} torr and the temperature is gradually ramped up over a couple of days to allow secondary drying to take place at 20°C (293 K). Once dried, the sample must be kept in a very dry environment. This procedure avoids the use of chemical dehydrating agents and although there may be problems with ice crystal damage, there is much less shrinkage and extraction. More details can be found in Chapter 14.

As the four images in Fig. 13.7 show, air-drying normally results in severe surface tension effects at the water–air interface, which cause shrinkage and collapse in all but the most robust specimens such as wood, bone,

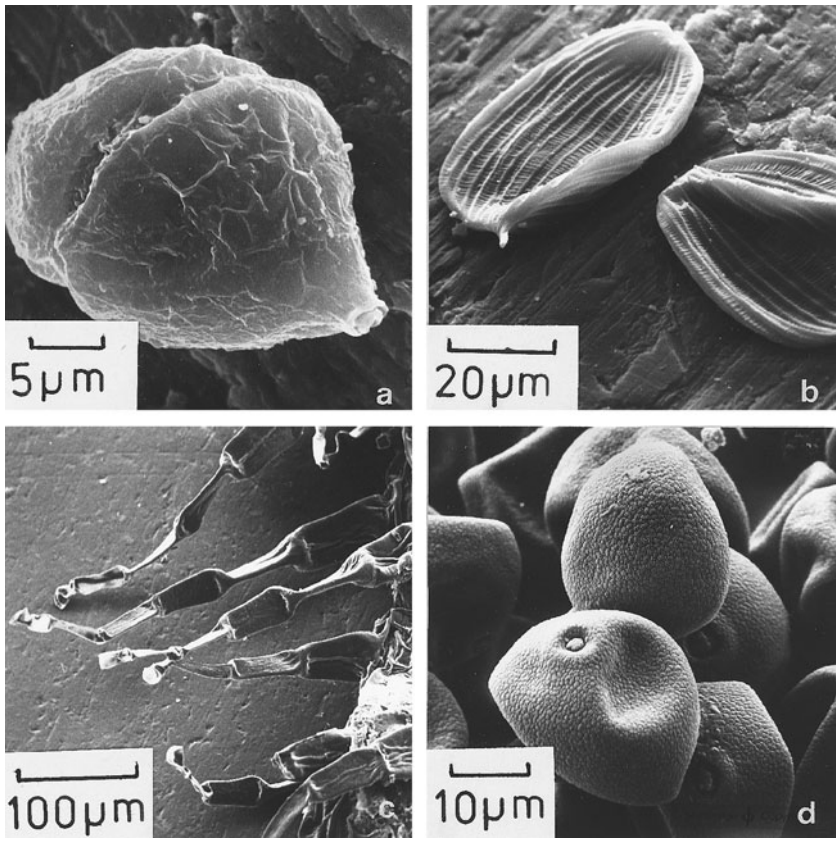


Figure 13.7. Examples of distortion due to air-drying. (a) Dinoflagellate alga *Prorocentrum micrans*. (b) Egulenoid flagellate alga *Phacus* sp. (c) Leaf hairs of *Solanum sarrachoides*. (d) Pollen grains of the grass *Poa pratense*. Note that all four specimens have a severely distorted surface.

and teeth. Even tissue toughening using the conductive infiltration procedures described earlier rarely gives acceptable results.

13.3.7. Embedding

If the internal contents of soft samples are to be examined, it is necessary to increase their structural integrity and mechanical strength so that they can be sectioned, fractured, or etched. Hydrophobic epoxy resins can be used for embedding as is done in transmission electron microscopy. Reference should be made to the standard texts on the subject. More details, including a generic procedure for rapid epoxy resin embedding, are given in Enhancement 13.5 in the accompanying CD.

Although hydrophobic epoxy resin embedding gives excellent results for structural studies in the SEM, it can cause a lot of tissue extraction and may thus compromise any chemical analysis involving cytochemistry and immunocytochemistry. There are a number of hydrophilic materials that can be used as embedding substances. Polyethylene glycol, polyvinyl alcohol, polyester waxes, and some acrylic resins are hydrophilic and/or water-soluble, so there is less extraction.

13.3.8. Exposing the Internal Contents of Bulk Specimens

There is a wide range of methods and techniques for exposing the internal contents of soft samples either directly or after they have been given sufficient mechanical strength. It is important to remove any debris remaining after exposing the inside of samples. This can be done by gentle washing in buffer, using mild detergents or appropriate enzymes, or air brushing to remove dry material. Inorganic material can be removed from organic material with EDTA. Organic matter can be removed from inorganic material with KOH, NaOH, and hypochlorite. Care must be taken to ensure that the specimen is unchanged.

13.3.8.1. Mechanical Dissection

The inside of the sample can be revealed by simple mechanical dissection using tools viewed under the binocular microscope. Fixed, dehydrated, and critical-point-dried material can be dry-fractured by gently pressing a piece of sticky tape on the surface and then pulling it away to reveal the underlying structures. Standard microtomy and ultramicrotomy can be used to give sections that can be examined using either the transmitted or reflective imaging mode. Alternatively, the sample can be ultraplaned, the section discarded, and the smooth block face examined in the SEM.

13.3.8.2. High-Energy-Beam Surface Erosion

An oxygen plasma can be used to surface-etch thick sections of resin-embedded material to reveal ultrastructural details of subcellular organelles. Recent advances allow focused ion beams as small as 5 nm to etch precisely and cut samples or remove parts of samples (see Chapter 11).

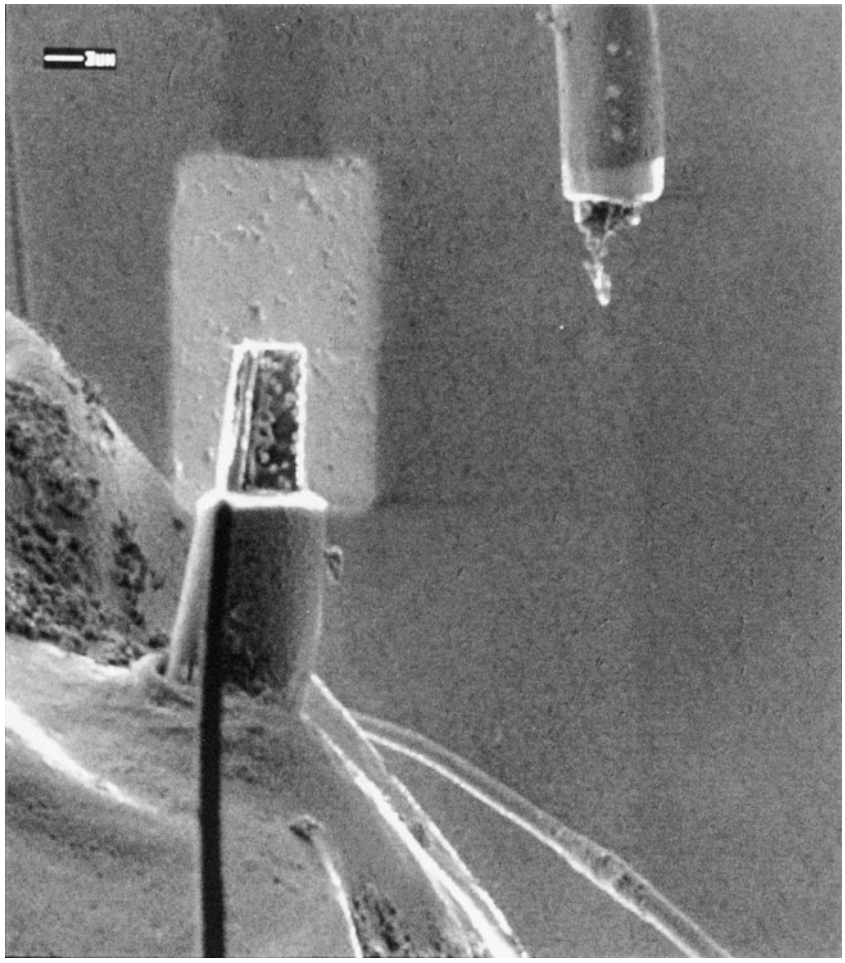


Figure 13.8. An image showing how a focused ion beam can be used to etch and cut soft biological material. The beam has made a precise cut in the base of the antenna, which has moved to the top right-hand corner of the image. (Image supplied by K. Robinson and P. Pugh, formerly of the British Antarctic Survey Laboratories, Cambridge.)

Figure 13.8 shows how a focused ion beam can be used to dissect an antenna from a marine organism. In other samples, nonfocused ion beam sputtering can be used to thin away the sample surface in a semicontrolled manner. Other surface erosion techniques can be used. For example, glow discharge etching can be used to remove layers of resin from specimens that have been fixed, dehydrated, and embedded. Cold-plasma etching and ion beam etching in a sputter coater can be used in the same way.

13.3.8.3. Chemical Dissection

Fixed and washed material can be incubated in potassium methoxide, which progressively dissolves extracellular and cellular material. It is necessary to examine the sample over time and once dissection is complete,

the sample should be washed, dehydrated, and critical-point-dried. Chemical dissection can also be carried out using selective enzymes. Epoxy-resin-embedded material can be etched progressively with sodium methoxide. Figure 13.9 shows details of the type of high-resolution information which may be obtained using this technique. Resin-embedded material, while still liquid, may be hardened by cooling in liquid nitrogen and then fractured at low temperatures. The fractured material is allowed to warm up in a resin solvent, which gradually reveals the embedded sample. The same procedure can be used for 100% ethanol-infiltrated material. Finally, samples can be infiltrated with an antifreeze agent such as dimethylsulfoxide, quench-cooled in liquid nitrogen, and fractured at low temperatures. The fractured pieces are thawed in fixative, which preserves the large structural molecules and removes small molecular species.

13.3.8.4. Surface Replicas and Corrosion Casts

Surface replicas are a useful for very delicate samples or very large surfaces. Casts of fracture faces or exposed surfaces can be made using dental wax, silicone rubber, and various plastic films, which can then be coated and examined in the microscope.

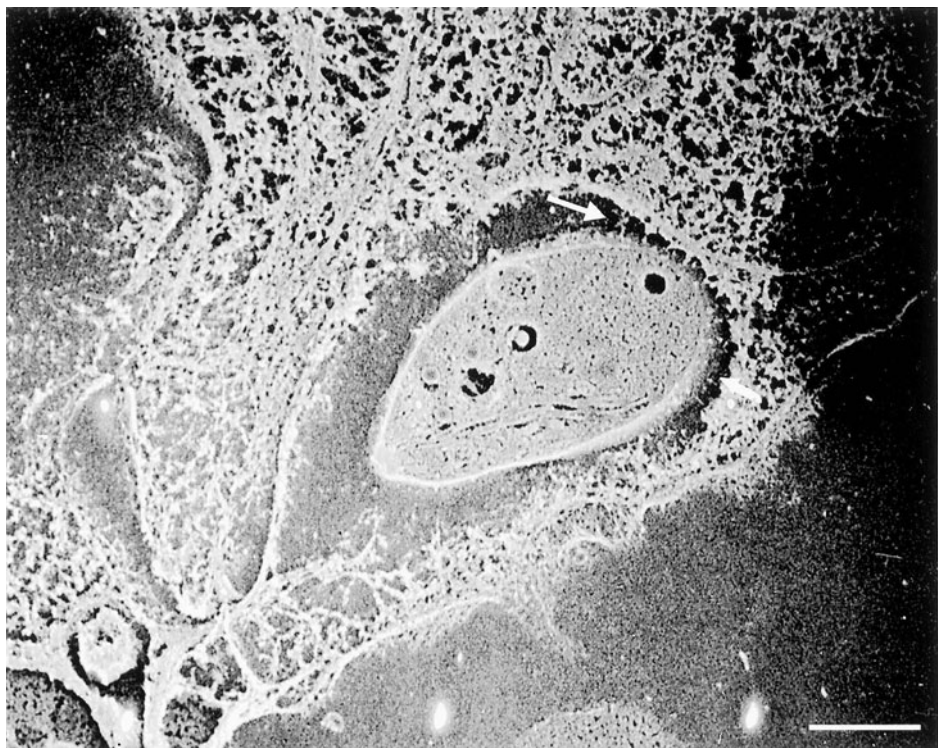


Figure 13.9. High-resolution FEG SEM image of an Epon-deembedded section of the protozoan parasite *Toxoplasma gondii*. The arrow shows the presence of short filaments on the surface of the parasite. Magnification marker = 200 nm. (Schatten *et al.,* 2000.)

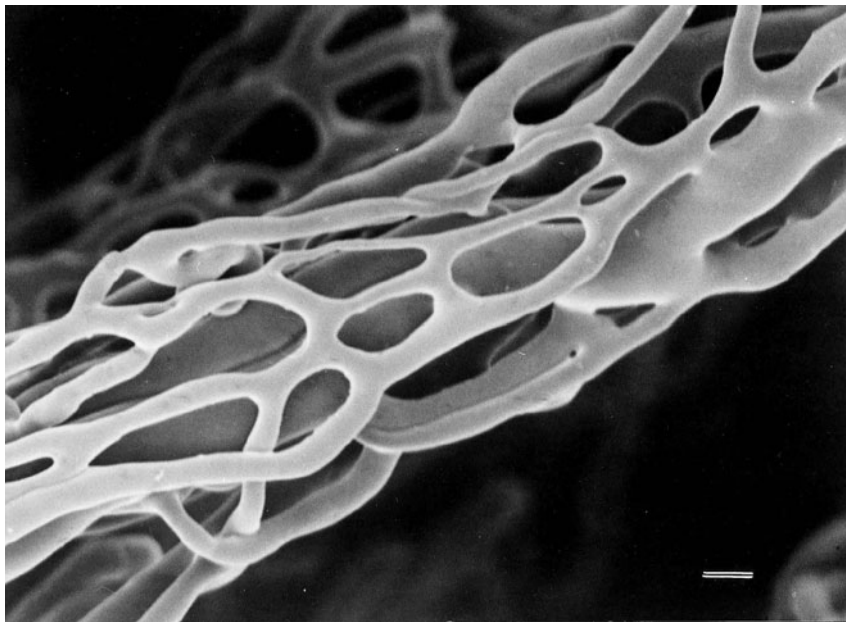


Figure 13.10. Microcorrosion cast of the bulbar papilla in bovine claw. The image clearly shows the central arteriole and venule of the papilla surrounded by a dense capillary network. Magnification marker = 16 μm. (Hirschberg *et al.,* 1999.)

Corrosion casts work on the principle of making a cast of a space, for example, a blood vessel, using latex, methacrylate, or polyester resins and discarding the surrounding solid material. Figure 13.10 shows the type of information that can be obtained. The space is first rinsed clean by diffusion with a suitable buffer and while the specimen is still living, the space is filled with a liquid resin such as Mercox and left for 1 h for *in situ* polymerization. Mechanical dissection is followed by overnight aldehyde fixation, after which the tissue is removed by maceration in 25% NaOH. The papers by Hirschberg *et al*. (1999) and Sims and Albrecht (1999) review the procedures for microvascular corrosion casting. The paper by Hosler and Douglas (2001) shows how these procedures can be used for quantitative image analysis.

13.3.9. Specimen Supports and Methods of Sample Attachment

The methods described so far are designed to transform the specimen to a form suitable for examination in the scanning electron microscope. Two additional processes must be considered to enable this examination to take place. The sample must be secured to a supporting structure which allows it to be transferred to and examined in the microscope. With most biological material, it is usually necessary to devise procedures to eliminate charging during subsequent examination and chemical analysis. These latter procedures are considered separately in Chapter 15.

Specimen supports must either be conductive or made to be conductive. There is a wide range of choices available depending on the type of specimen being studied. There is no standard support or stub, although many microscopes will take the so-called Cambridge stub, made of an aluminum alloy. The stub has a 12-mm-diameter surface supported on a 7.5 by 2-mm pin, which fits into a hole on the microscope stage. Because the stubs are made of metal, they can, as Fig. 13.11 shows, be easily modified. Secondary supports can be made from metal or plastic foils, plastic, metal, or cellulose filters, glass or plastic cover slips, silicon chips, freshly cleaved mica, or polished pyrolitic graphite. These secondary supports, which should be scrupulously clean, must be readily secured to the underlying metal stub and a conductive pathway established between the two. Finally, it is important to mark the underside of the stub with a code that identifies the specimen.

Robust wet samples can be placed directly onto metal stubs or secondary supports and simply allowed to dry. Wet samples collected on filters that have been allowed to dry can be similarly attached. Various compatible glues can attach dried samples, although it is important that all glue volatiles be removed before the samples are coated and/or go into the microscope. Suitable adhesives include L-polylysine or 3-aminopropyl triethoxy silane for particulates, cells, and microorganisms. Water-soluble and organic glues

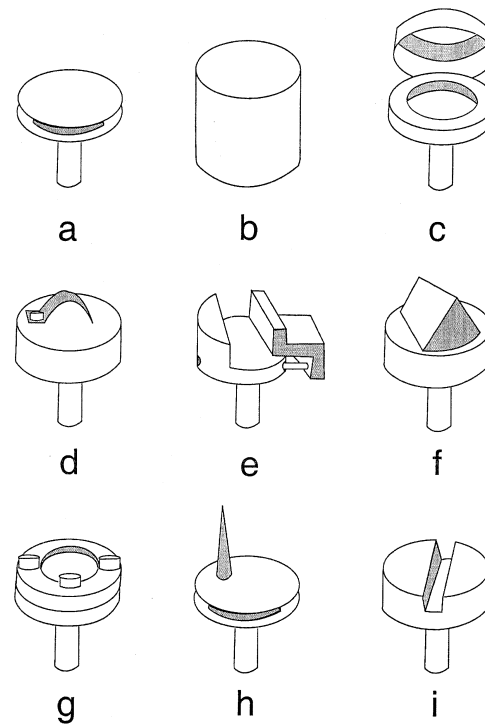


Figure 13.11. The types of specimen stubs used for mounting samples for imaging in the SEM. The standard types are shown in (a) and (b), whereas the others are modifications which may be easily constructed for different types of specimens. (Judy Murphy, in Bozzola and Russell, 1999.)

may be used for larger specimens. Double-sided adhesive carbon tape is particularly useful and a thin strip can be used to provide the conducting pathway to the underlying stub. There are also a number of conducting adhesives and paints such as silver and carbon dag. Care must be taken when attaching samples to the underlying support. The sample must not sink too far into the glue and porous samples must not be allowed to wick up the glue. Ideally, the sample should be examined in the SEM as soon as possible after it has been prepared. If immediate examination is impractical, it is important to store the sample properly in a dry environment.

13.3.10. Artifacts

It is important to distinguish between artifacts (controlled and expected changes to the sample) and damage (uncontrolled and unexpected changes to the sample). Artifacts, damage, and sample contamination can occur at any point during the preparation process. The books by Crang and Klomparens (1988), Hayat (2000), Bozzola and Russell (2000), and Maunsbach and Afzelius (1999) are useful guides to identifying and understanding the problem. It is useful to maintain a compendium of images that show identifiable artifacts, damage, and contamination. In addition it is important wherever possible to carry out correlative microscopy and view the sample by both light optics and/or transmission electron microscopy.

13.4. Preparative Procedures for the in Situ Chemical Analysis of Biological Specimens in the SEM

13.4.1. Introduction

The scanning electron microscope can be used to give chemical information about the sample in addition to structural details. There are a number of different ways this may be achieved. Although structure and composition are joined, the guiding principles underlying preparing specimens for chemical analysis, as distinct from structural investigation, are summarized in the following guidelines

1. The preparative methods must retain elements and molecules at their original concentration and original location in the sample.
2. The preparative techniques and their associated chemicals should not remove or add to the chemical identity of the material being analyzed.
3. The preparative technique, although biased toward retaining chemical information, should also retain recognizable structural components.
4. The best methods are usually the least invasive.
5. The best-prepared sample is the one that is closest to its natural state.

Adequate sample preparation by wet chemical procedures is often difficult to achieve in practice because techniques designed to preserve ultrastructure tend to retain macromolecules at the expense of soluble constituents. In addition, techniques designed to retain the chemical content

of samples usually do not adequately preserve the macromolecular architecture.

There are three main approaches to preparing specimens for *in situ* chemical analysis:

1. Low-temperature methods, in which only the phase of the liquid environment is changed (see Chapter 14)
2. Environmental SEM, in which the liquid environment remains unchanged (see Chapter 5)
3. Ambient-temperature wet chemical methods, in which the natural chemical and liquid environment of the sample is changed to varying degrees.

These procedures will be discussed in the following sections under three different headings: *elemental analysis* by x-ray spectroscopy (Section 13.4.2); *molecular analysis* by histochemistry, where specific heavy metal stains can be used to localize chemical ligands within organic macromolecules and such locations can be visualized by backscattered imaging and characterized by x-ray microanalysis (Section 13.4.3); and *macromolecular analysis* by immunocytochemistry, where antibodies conjugated to colloidal gold are used to localize specific active sites within biological material, which can then be visualized by backscattered imaging and characterized by x-ray microanalysis (Section 13.4.4).

13.4.2. Preparative Procedures for Elemental Analysis Using X-Ray Microanalysis

Most of the techniques are derived from methods developed for use with light and transmission electron microscopy. This creates problems, for the emphasis is primarily on preserving macromolecular structural features generally at the expense of differential solubilization, relocation, and loss of elements. The ambient-temperature wet chemical methods discussed earlier in this chapter can be used, provided one is aware of what they may do to the sample. In a study on the distribution of aluminum in living tea leaves, Echlin (1996) showed that ambient-temperature wet chemical procedures caused a 3- to 10-fold reduction in the amount of aluminum measured in different cell compartments compared to the amounts measured in samples prepared using low-temperature preparation procedures.

The low-temperature methods, which are discussed in Chapter 14, are generally acknowledged to be the best way to prepare samples for elemental analysis. Each specimen will require a different approach and it is important to have a clear understanding of the chemical identity of the substance being analyzed before devising a preparation protocol. A general discussion of the principles involved will be presented here. The book by Warley (1997) is a useful reference and contains a number of proven preparation techniques.

13.4.2.1. The Nature and Extent of the Problem

X-ray microanalysis makes little distinction between the binding states of elements, whereas the preparative methods are designed around the

binding state of the element. For example, consider the binding state of the elements in the three following samples: sulfur covalently bound in a protein, calcium in a crystal of apatite, and potassium ions in solution. The same element may have different binding states and hence different solubilities. The phosphorus in an enzyme and in a membrane is amorphous and an integral part of the chemical structure, whereas phosphorus is part of the crystal in bone apatite.

All cells and solutions have diffusion movements and some cells have cytoplasmic streaming. These movements are quite fast. Electrolytes in solution move 1 μm in 0.5 ms, although it is estimated the rate slows to 0.1 μm in 0.5 ms in cell cytoplasm. Physiological events occur at a rate of 1–5 ms and many chemical events are virtually instantaneous. In contrast, the diffusion rate of many of the chemicals used for fixation are of the order of about 1 μm/s and even cryofixation at 1 μm in 20 ms may be too slow to arrest some processes.

13.4.2.2. Types of Sample that May Be Analyzed

Different types of samples exhibit different x-ray spatial resolution, depending on their form and how much of the sample needs to be analyzed. These two factors will, to some extent, dictate how the sample should be prepared and the type of instrumentation that will be needed.

1. Bulk samples, which do not transmit the electron beam, are limited to an analysis of natural surfaces, polished surfaces, smooth(ish) fractured surfaces, and sectioned surfaces. The x-ray spatial resolution is between 50 nm and 10 μm.
2. Isolated cells and organelles, although only a few micrometers thick, are usually treated as bulk samples. X-ray spatial resolution is between 50 and 500 nm.
3. For particles and fibers that may have been eluted, collected, or analyzed *in situ,* x-ray spatial resolution is between 50 and 100 nm.
4. For thick sections up to about 2 μm thick, in which reasonable morphological detail can be resolved using the transmitted electron signal, x-ray spatial resolution is between 50 and 100 nm.
5. For thin sections up to 200 nm thick in which high-resolution morphological detail can be resolved, x-ray spatial resolution can be as good as 2–20 nm.

13.4.2.3. The General Strategy for Sample Preparation

Bearing in mind the various types of biological samples that are of interest, the analyst is advised to first make a list of the parameters that will affect the analysis before preparing the sample for the SEM. The following questions should be answered before designing the experimental protocol for the particular specimen being analyzed.

1. What type of analytical instrumentation and methods of imaging are available?
2. What information is available about the structure of the sample?

3. What information is there about the location of the elements being analyzed?
4. What information is available about the binding state of the elements?
5. What are the normal hydration properties of the sample?
6. What structural and analytical resolution is required?
7. What degree of analytical sophistication is required?
8. What information is being sought in the sample?

13.4.2.4. Criteria for Judging Satisfactory Sample Preparation

The experimenter must set up various criteria for judging whether a preparative procedure has worked. Satisfactory morphological information usually is not enough. A useful protocol is as follows:

1. Preserve structural detail at a spatial resolution 10 times better than the expected analytical resolution.
2. Identify the location and measure the amount of elemental loss from the sample and translocation within the sample. This is usually very difficult to measure accurately.
3. Use preparative chemicals that will not mask or contribute to the x-ray spectrum.
4. Measure any changes which may have occurred in the organic matrix and the water content during sample preparation because these will affect the accuracy of the analytical results.
5. Assess any and all information available from correlative studies that would help evaluate the effectiveness of the preparative procedure.

The caveats and precautions suggested in Sections 13.1–13.3 are even more important when it comes to specimen preparation for chemical studies. It is suggested that a list should be made of the chemicals that are going to be used in all the stages of sample preparation to make sure that they do not contain the elements being sought in the specimen. An uncontaminated sample should be removed from the unstressed living specimen and immersed or perfused with a suitable fixative as quickly as possible.

13.4.2.5. Fixation and Stabilization

All chemical fixatives cause changes in cell permeability and our concern must be how to staunch the flow of elements from and within the sample. Heavy metal fixatives are much worse than organic fixatives and in addition may mask the elements being analyzed. Wherever possible, organic fixatives such as glutaraldehyde, paraformaldehyde, or a mixture of the two should be used. Organic buffers such as piperazine-N-N'-bis(2-ethane sulfonic acid) (PIPES) or N-(2-hydroxyethylpiperazine-N'-2-ethane sulfonic acid) (HEPES), tertiary amine-heterocyclic buffers lessen (but do not prevent) loss of material from the sample, should be used. Chemical fixation is too slow for use with diffusible elements and one must either try to precipitate the element concerned as an insoluble salt or use the cryopreservation

methods discussed in Chapter 14. Mild fixation at low temperatures may be possible for some elements, and aldehydes and diimido-esters which crosslink proteins, may be useful for fairly well bound elements. The fixative and buffer vehicle must be tailor-made for the particular specimen.

13.4.2.6. Precipitation Techniques

Heavy metal salts can be used to immobilize specific diffusible substances by the formation of insoluble precipitates that, it is hoped, stay in place in the sample. The insoluble deposits can be imaged by BSE or identified by their x-ray spectrum. For example, the chloride ion is precipitated by silver nitrate and phosphate is precipitated with lead salts, and potassium pyroantimonate will precipitate a number of mono- and divalent cations. However, recent studies suggest that the precipitating agent acts as a focus for electrolytes and that the spatial resolution of the method is not very good. In addition, it is difficult to ensure that there is no movement of precipitates during subsequent preparative techniques.

13.4.2.7. Procedures for Sample Dehydration, Embedding, and Staining

Dehydration using organic solvents will cause loss of elements, although this effect is ameliorated if there is a progressive lowering of temperature as the dehydrating agent becomes more concentrated. Freeze-drying is the most satisfactory method, and freeze substitution, in which the ice is dissolved in organic solvents at low temperatures, can also be used. See Chapter 14 for more details.

Embedding is necessary if sections are to be cut to reveal the inside of the specimen. Frozen-dried tissue embedded in epoxy or acrylate resins may result in loss of elemental material from the sample and the resins may contribute elements to the sample, e.g. S and Cl. Hydrophilic Lowicryl and LR White acrylate resins show more promise because they can be polymerized at subzero temperatures.

Nonspecific staining is only helpful if low-contrast sectioned material is to be examined by transmitted electrons. The biggest problem is that the x-ray lines of the heavy metal stains will overlap the light element lines of the biological material and the stain solutions will cause additional extraction.

13.4.2.8. Specimen Supports

The support, in addition to being a good electrical and thermal conductor, should not contribute to the emitted x-ray spectrum. For bulk materials, single cells, microdroplets, and thick sections that are to be viewed by the secondary electron detector in the SEM, highly polished disks made of ultrapure pyrolitic graphite or beryllium should be used. For thin sections, beryllium or graphite grids should be used because they produce low continuum x-ray background.

13.4.3. Preparative Procedures for Localizing Molecules Using Histochemistry

Molecular analysis involves using added chemicals, particularly heavy metal salts, to localize sites of a given molecule. The procedure is only satisfactory if the reactions are known to be quite specific with no cross-reactivity and the reaction products remains localized and unaffected by other preparative procedures. Many of the methods, which provide an important link between structure and function, have their origins in TEM and light microscopy, but have been adapted to work in the SEM. All the methods are chemically invasive and involve modifying the sample. Care, of course, must be taken in interpreting the results.

13.4.3.1. Staining and Histochemical Methods

Unlike light microscopy, which depends on differences in wavelength contrast to identify different regions in the sample, the SEM depends on differences in atomic number as one way to obtain chemical differentiation. This method involves introducing heavy metal ligands to specific chemical sites in an otherwise light element matrix to increase the electron scattering in these regions. The introduction of heavy metal ligands can be achieved in two different ways.

1. *Selective staining,* which relies on the specificity of a chemical ligand within the sample for a heavy metal salt, can be used for the *in situ* localization of a specific chemical group or a specific metabolic or chemical process and for following chemical changes at a specific location. For example, ruthenium ($Z = 44$) tetroxide can be used to localize aromatic rings and ether alcohols; lead ($Z = 82$) salts can be used for phosphatases

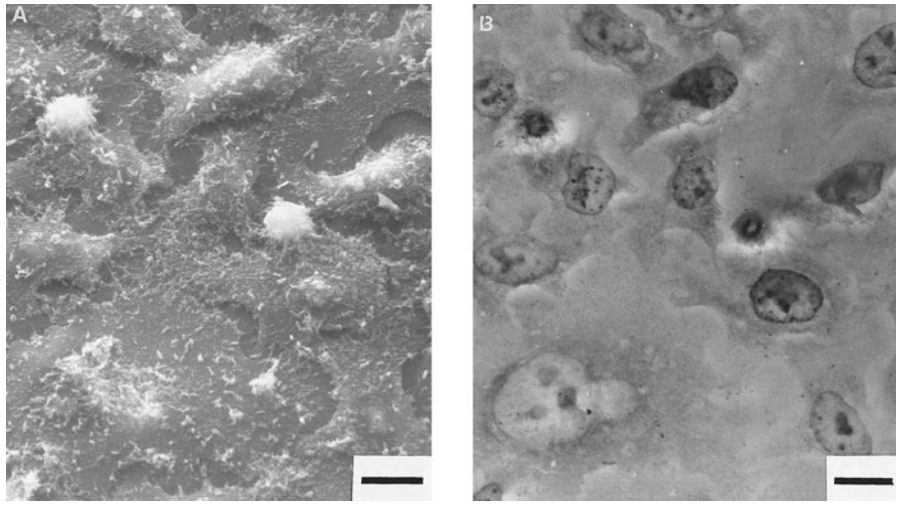


Figure 13.12. (a) Secondary electron image and (b) reverse-contrast, backscattered compositional image of HeLa tissue culture cells stained with Wilder's silver stain. The silver selectively stains the chromatin and perinuclear regions of the cells. Magnification marker = 10 μm. (Abraham, 1977.)

and basic groups; and tungsten ($Z = 74$) as phosphotungstic acid can be used for hydroxyl, carboxyl, and NH groups. The selective staining is best done on thick sections before and after a particular chemical treatment. Chapter 6 in the book by Hayat (2000) contains a large number of recipes that may easily be adapted for use in the scanning electron microscope. Figure 13.12 shows how a stain containing silver is used to localize chromatin and perinuclear material in tissue culture cells.

2. *Enzyme histochemistry,* which links the end product of an enzymatic reaction to a heavy metal salt, is an effective way of localizing areas of specific metabolic activity. Care is needed during specimen preparation to retain enzymatic activity, and low-temperature preparation techniques are the best approach. Examples of this procedure include diaminobenzidine (DAB)-osmium for oxidases, insoluble lead salts for acid phosphatases and esterases, and copper ferricyanide for dehydrogenases. Figure 13.13 shows how cobalt salts and ammonium sulfide can be used to localize alkaline phosphate activity. The books by Lewis and Knight (1992) and Van Noordan and Frederiks (1992) should be consulted for details of the specific procedures, which can be used to identify and localize a wide range of chemical ligands. The heavy metal reaction product in both procedures can be visualized in sections as regions of high contrast and, if necessary, identified by their x-ray spectrum. In bulk samples, use should be made of the atomic number contrast obtained from backscattered electrons.

13.4.3.2. Atomic Number Contrast with Backscattered Electrons

The nature of backscattered electrons and their detection are discussed in Chapters 3 and 4. The differences in backscattering coefficient from various areas of interest form the basis of the atomic number procedure. Examples of the use of the technique are a heavy metal dust particle inhaled into pulmonary tissue, a heavy metal ligand attached to the end product of

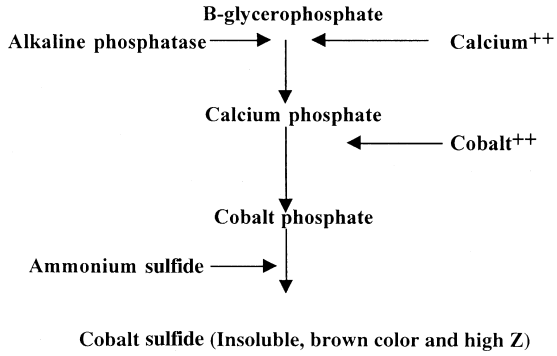


Figure 13.13. An example of how enzyme histochemistry can be used in the SEM to localize regions of metabolic activity. The enzyme alkaline phosphatase with the substrate β-glycerophosphatase and in the presence of calcium ions forms calcium phosphate. The calcium phosphate in the presence of cobalt ions and ammonium sulfide forms a brown insoluble precipitate containing cobalt. The enzyme can be localized either by x-ray spectroscopy or compositional imaging using backscattered electrons.

an enzymatic reaction, or a heavy metal stain with a specificity for a given chemical grouping within a cell.

For biological samples, material is stained with a specific heavy metal salt and then taken through a fixation–dehydration–embedding schedule. The smooth surface of the bulk specimen is imaged and it is possible to get information from different depths of the sample by varying the kV. It is important to emphasize that the sample must have a reasonably flat surface to avoid topography contrast. Fixation should avoid the use of heavy metals and the sample should be coated with light elements such as C or possibly Al.

The backscattered signal usually appears bright against a dark background. Biologists have frequently found it convenient to reverse this contrast so that regions of strong backscattering appear dark against a light background. The procedure can be used to study either natural systems such as the distribution of calcium in gall stones or bone, or where a particular heavy metal chemical group is linked to a specific region of the sample. The chemical identity of the signal may, if necessary, be checked by its x-ray spectrum.

13.4.4. Preparative Procedures for Localizing Macromolecues Using Immunocytochemistry

13.4.4.1. Introduction

Antibodies, lectins, and immunoglobulins have a high affinity and specificity for organic chemical groupings in biological macromolecules and smaller molecules. Antibodies are proteins and can only be seen directly in their native state at very high magnifications. They are readily located in cells and tissues at lower magnifications when attached to a suitable label such as a fluorescent dye, which may be seen in the light microscope, or a heavy metal such as gold, which may be imaged in an electron beam instrument. The book by Polak and Van Noorden (1997) and the review by Skepper (2000) describe in detail the major immunocytochemical strategies in current use in electron microscopy. Immunocytochemical procedures are important in the analysis of macromolecules in the SEM.

13.4.4.2. The Antibody–Antigen Reaction

The antibody–antigen interaction is well understood and characterized. A good antibody has a high affinity for its specific antigen. Monoclonal antibodies are homogeneous and usually react with only one molecule and a single antigenic determinant (epitope). Monoclonals can be prepared in large amounts, but because of their high specificity, the epitope may be crosslinked during sample preparation. The monoclonal antibodies are best used with samples prepared with weak fixatives or by cryomethods. Polyclonal antibodies are heterogeneous and are usually directed against a number of different epitopes on the specimen and thus increase the chance of crosslinking, but at the expense of specificity.

The chemicals and preparations used in the immunolabeling should be tested for the specificity of the antibody by standard procedures such as enzyme-linked immunosorbant assay (ELISA) and radioimmunoassay. The book by Polak and Van Noorden (1997) gives details of the general methods used in immunocytochemistry for electron microscopy.

The labeling procedures fall into two general classes: direct methods, where the primary antibody is directly conjugated to the marker molecule, and indirect methods, where the primary antibody is in turn conjugated to a labeled second antibody raised against an antibody from the species that gave the primary antibody. Most SEM immunocytochemistry is based on using colloidal gold particles as the markers.

The colloidal gold probes are gold spheres, which come in a range of sizes from 2–40 nm and when coated with a selected protein can be attached to a specific antibody. The gold probes are electron-dense, have no known cytotoxicity, and are stable cytochemically. They can be used directly or indirectly to localize specific antigenic sites with a high sensitivity. Colloidal gold immunocytochemistry is very useful for examining natural or exposed surface antigens and for the three-dimensional examination of large areas of tissue. The book by Verkleij and Leunissen (1989) should be consulted for specific details of the procedures.

There are many procedures in the literature for preparing the gold probes *de novo,* but most probes are now readily available commercially and usually already attached to the required antibody via a range of different proteins. Large (20–40 nm) gold particles should be used for low-resolution survey images and small (2–20 nm) particles used for the high-resolution imaging of receptor sites.

13.4.4.3. General Features of Specimen Preparation for Immunocytochemistry

All the methods are based on using either a mild fixation and dehydration routine or low-temperature techniques discussed in Chapter 14. The immunocytochemical procedures fall into two general classes:

1. *Preembedding techniques using the indirect procedure*. The sample is first incubated with the tissue-specific primary antibody followed by incubation with a secondary antibodies conjugated to a gold probe. The specimens are then fixed, dehydrated, and embedded in resin and, if necessary, the inside of the sample exposed by ultraplaning or smooth fractures. The procedure ensures good structural preservation and accurate immunolocalization of the epitope. The procedure is good for external antigens.

2. *Postembedding techniques using the indirect method.* The ultraplaned surface or smooth fracture faces are prepared first and then incubated with the specific primary antibodies followed by incubation by secondary antibodies conjugated to the gold probe. The procedure is suitable for both internal and external antigens. More details, including a generic wet chemical fixation routine for immunocytochemical studies, are given in Enhancement 13.6 in the accompanying CD.

13.4.4.4. Imaging Procedures in the SEM

The distribution of the colloidal gold particles is best achieved by combining the atomic number contrast image obtained from backscattered electrons with the topographic image from secondary electrons. A dedicated solid state detector placed around the lower face of the final lens is used to collect the backscattered electrons and an ET detector is used to collect secondary electrons along with a highly directional backscattered electron signal (Chapter 4). The secondary electron signal provides good morphological information, along with a weak signal for a small number of gold particles. The backscattered electron signal from the dedicated BSE detector provides a signal from the gold particles. The BSE signal can be enhanced by coating the surface of the gold probes with a deposit of silver. If the backscattered signal polarity is reversed, the signal from the gold probes appears dark against the lighter secondary electron signal and makes it much easier to see the precise location of the antigenic sites. Figures 13.14

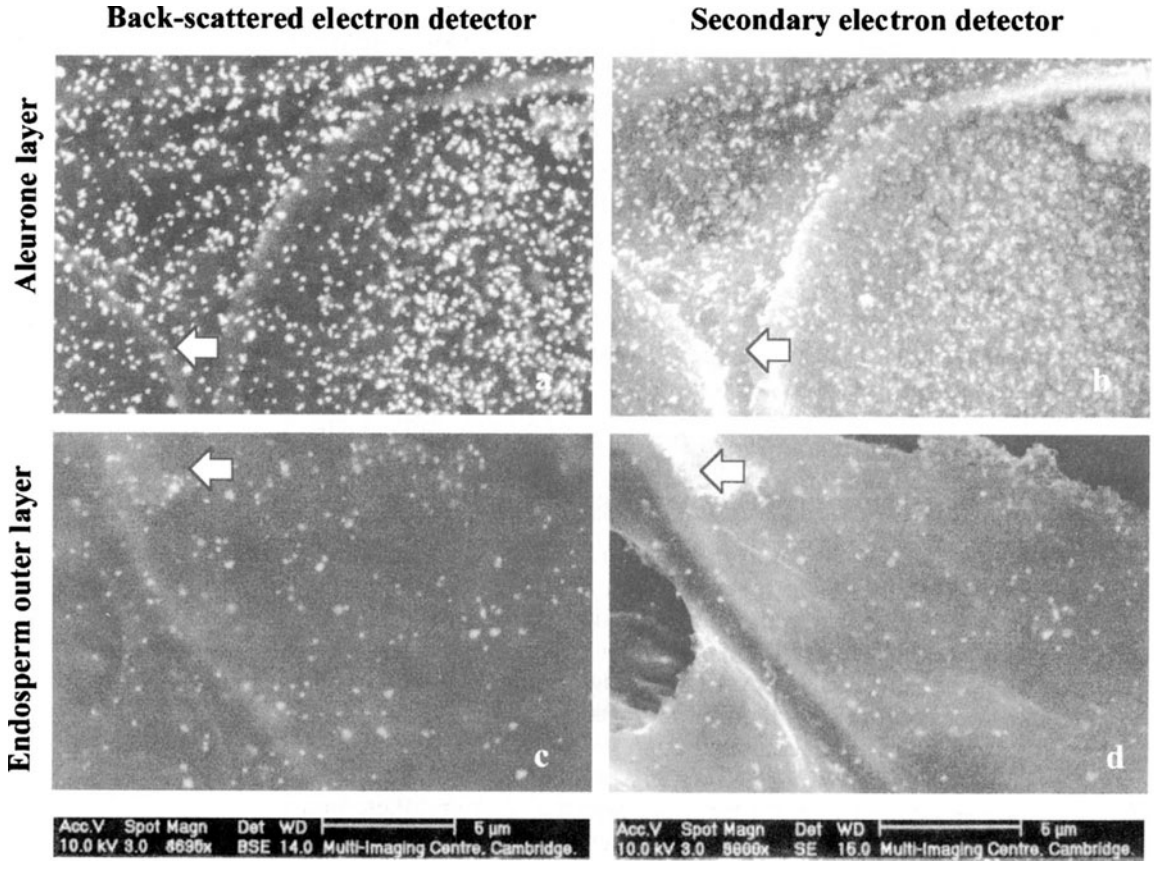


Figure 13.14. Immunolabeling of wheat arabinoxylans using colloidal gold-labelled antibody enhanced with silver. Two different regions of the wheat caryopsis have been studied and a comparison made using the images from (a, c) a solid state detector and (b, d) an Everhart–Thornley detector. The finely granular structures shown in the topographic contrast images (b, d) are not seen in the BSE compositional contrast images (a, c). The set of four images shows that there are considerably more arabinxylans associated with the cell walls of the aleurone layer than with the endosperm cell walls. (Xia, 2001.)

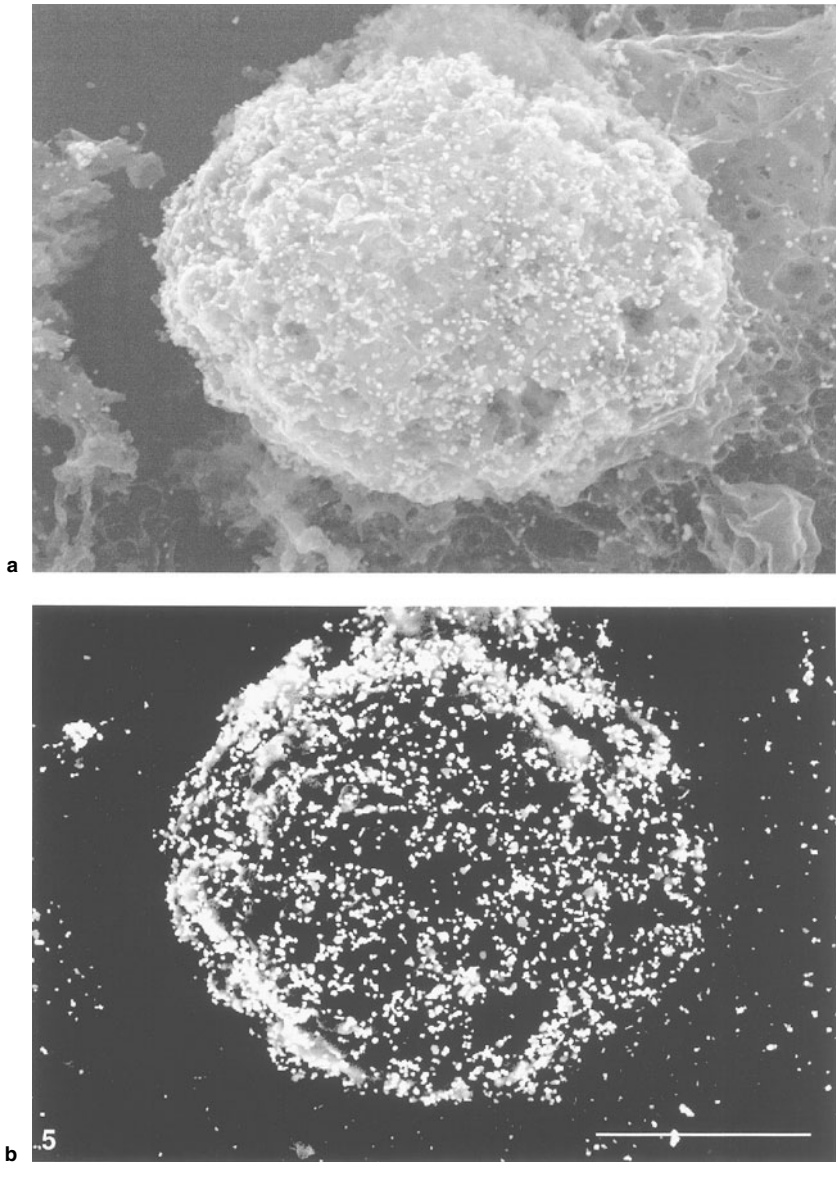


Figure 13.15. Images from ET detector (upper) and dedicated BSE detector (lower) of a monocyte cell surface labeled with a colloidal gold–antibody complex against the cell surface receptor P2X7. The ET image provides good topographic contrast, whereas the BSE image gives good compositional contrast, which can be used to localize the cell surface receptor. Magnification marker = 5 μm. (Emma Kidd, in Skepper, 2000.)

and 13.15 show how the two images can be either combined or contrasted to give valuable *in situ* information about reactive groups in cells and tissues.

A wide range of secondary antibodies with conjugated gold probes are readily available commercially at a relatively low cost. The gold probes come in different sizes, enabling two different antigens to be labeled on the

same specimen. When using colloidal gold immunocytochemistry, care must be taken to recognize gold aggregates, distinguish between topographic and compositional contrast, and not confuse natural particles and gold particles on the specimen surface. Because the procedure is best carried out on smooth surfaces, there may be some difficulty in unambiguously recognizing morphological features in the image from the Everhart–Thornley detector.

References

Abraham, J. L. (1977). In *SEM/1977/II,* IIT Research Institute, Chicago, p. 119.

Bozzola, J. J., and L. D. Russell (1999). *Electron Microscopy. Principles and Techniques for Biologists,* Jones and Bartlett, Sudbury, Massachusetts.

Brown, J. A., and Teetsov, A. (1976). In *SEM/1976/I,* p. 385.

Crang, R. F. E., and K. L. Klomparens (1988). *Artefacts in Biological Microscopy,* Plenum Press, New York.

Dashek, W. V., ed. (2000). *Plant Electron Microscopy and Cytochemistry,* Humana Press, Totowa, New Jersey.

Ensikat, H. J., and Barthlott, W. (1993). *J. Microsc*. **172,** 195.

Giberson, R. T., and R. S. Demaree eds. (2001). *Microwave Techniques and Protocols,* Humana Press, Totowa, New Jersey.

Glauert, A. M., and P. R. Lewis (1998). In *Practical Methods in Electron Microscopy,* Vol. 17 (A. M. Glauert, ed.), Portland Press, London.

Gupta, P. D., ed. (2000). *Electron Microscopy in Biology and Medicine,* Whitaker Press, Hanaki, D., *et al.* (1985).

Hanaki, M. K., K. Tanaka, and Y. Kashima (1985). *J. Electron Microsc*. **34,** 373.

Hayat, M. A. (2000). *Principles and Techniques of Electron Microscopy. Biological Application,* 4th ed., Cambridge University Press, Cambridge.

Hirschberg, R. M., C. K. W. Mulling and H. Bragulla (1999). *Microsc. Res. Techniques* **45,** 184.

Hockley, D. J., and M. H. Jackson (2000). *Proc. R. Microsc. Soc*. **35,** 187.

Hockley, D. J., M. H. Jackson, and R. A. Fleck (2001). *Proc. R. Microsc. Soc*. **36,** 110.

Hosler, F. E., and J. E. Douglas (2001). *Microsc. Microanal.* **7,** 253.

Isabell, T. C., P. E. Fischione, C. O'Keefe, M. U. Guruz, and V. P. Dravid (1999). *Microsc. Microanal.* **5,** 126.

Liepins, A., and E. de Harven (1978). In *SEM/1978/II,* SEM, Inc., AMF O'Hare, Illinois, p. 37.

Lewis, P. R., and D. P. Knight (1992). In *Practical Methods in Electron Microscopy,* Vol. **14** (A. M. Glauert, (ed.), Elsevier, Amsterdam.

Login, G. R., and A. M. Dvorak (1994). *The Microwave Toolbook,* Beth Israel Corporation, Boston.

Maunsbach, A. B., and B. A. Afzelius (1999). *Biomedical Electron Microscopy,* Academic Press, New York.

Murakami, T. N., T. Iida, O. Taguchi, A. Ohtani, A. Kikuta, A. Ohtsuka, and T. Itoshima (1983). *Scanning Electron Microsc.* **1,** 235.

Osatake, H., K. Tanaka, and T. Inoue (1985). *J. Electron Microsc. Tech.* **2,** 201.

Polak, J. M., and S. Van Noorden (1997). *An Introduction to Immunocytochemistry,* Bios Scientific Publishers, London.

Schatten, H., D. Sibley, and H. Ris (2000). *Microsc. Microanal.* **6**(Supp. 2), 648.

Sims, P. A., and R. N. Albrecht (1999). *Microsc. Microanal.* **5,** 99.

Skepper, J. N. (2000). *J. Microsc*. **199,** 1.

Spurney, K. B. (1994). *Analyst* **119,** 41.

Teetsov, A. (2000). *Microsc.* Today **2000,** 8.

Verkleij, A. J., and J. L. M. Leunissen (1989). *Immuno-Gold Labelling in Cell Biology,* CRC Press, Boca Raton, Florida.

Van Noorden, C. J. F., and W. M. Frederiks (1992). *Enzyme Histochemistry,* Oxford University Press, Oxford.

Warley, A. (1997). *X-Ray Microanalysis for Biologists,* Portland Press, London.

Wetzel, B. G. H., E. L. Cannon, B. W. Alexander, and G. H. Westbrook (1974). In *SEM/1974,* p. 1.

Xia, J. (2001). Ph.D. Thesis, Faculty of Biology, University of Cambridge, Cambridge.

14

Low-Temperature Specimen Preparation

14.1. Introduction

Low-temperature specimen preparation is not to be considered solely in the context of hydrated biological systems, although much of what will be discussed here will be directed toward these types of samples. Low temperatures are an essential prerequiste for studying naturally frozen materials such as snow and ice and frozen foods such as ice cream (Fig. 14.1). The technology is central to the preparation of aqueous systems such as paints, suspensions, emulsions, solutions, soils, clays, muds, and cements; of any nonaqueous liquid systems such as oils and organic liquids; and even of gases, vapors, and volatile materials. In addition, as has been shown in Chapter 12, low temperatures may play an important part in the preparation of plastics, polymers, and elastomers. The common physical parameter of these diverse samples is that they all can be solidified, provided the temperature is below their melting point. Once solidified they can then be further manipulated for subsequent examination and analysis in electron beam instruments.

Converting a liquid or a soft or flexible material to a solid state provides a firm matrix suitable for manipulation, microscopy, and analysis. It maintains the triphasic state that exists naturally in some samples, and dissolved elements, compounds, and molecules remain *in situ* within the solidified matrix. Sufficiently low temperatures effectively immobilize dynamic and physiological processes. This process is, to a first approximation, a chemically noninvasive preparative procedure and at very low temperatures there is a diminution in high-energy-beam-induced radiation damage.

The disadvantage of the low-temperature approach to specimen preparation is the formation of crystalline materials during the liquid-to-solid phase transition. Many organic materials, when cooled below their melting point, form amorphous glasses in which any suspended or dissolved materials remain randomly distributed. In contrast, as hydrated samples are cooled, they generally form highly ordered eutectic mixtures of pure

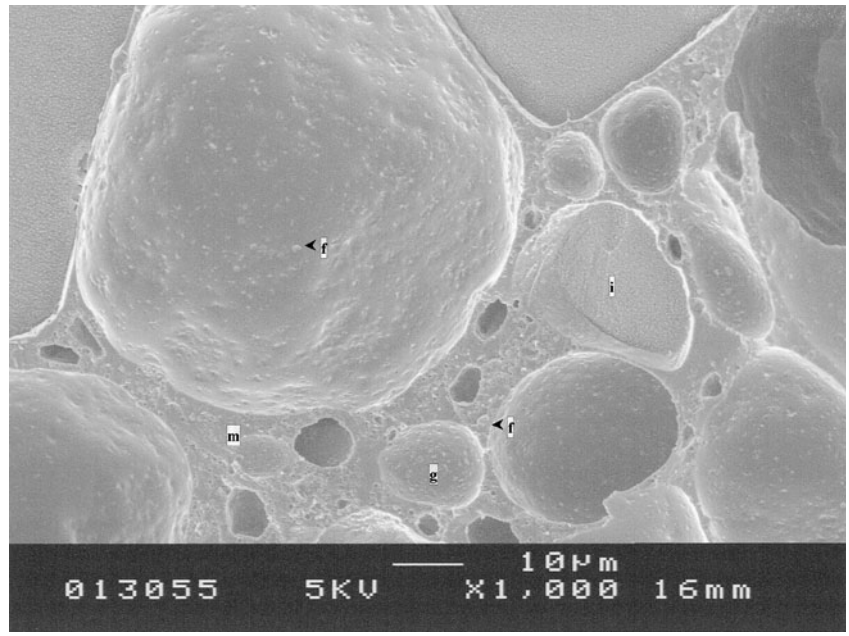


Figure 14.1. Fracture face of a sample of fully frozen-hydrated ice cream imaged ay 130 K and 5 keV. i, Ice crystal; f, fat globule; g, gas bubble; m, protein–carbohydrate matrix. Magnification marker = 10 μm. (Courtesy of Mark Kirkland, Unilever Research Centre, Bedford, UK.)

ice crystals and partitioned solids which bear little structural relationship to their original random organization. However, if the samples are cooled fast enough, this structural disorganization can either be significantly diminished or entirely avoided. A substantial portion of this chapter will center on the practices and procedures developed to preserve the natural structural and chemical constitution of biological specimens, which contain varying amounts of water. In addition, a brief discussion of additional procedures that may be adopted to prepare and examine hydrated organic samples, hydrated inorganic systems, and liquid organic samples is given. A detailed examination of the topic can be found in the book by Echlin (1992).

14.2. The Properties of Liquid Water and Ice

It is important to understand the physicochemical properties of water in order to better understand the rationale behind the development of the various low-temperature preparative procedures. The chemical properties of water molecules are related to the ease by which they form hydrogen bonds with each other and with hydrophilic substances.

Water is both interactive and reactive. The different binding states of water are best thought of as a continuum. In surface-modified water, water

molecules are in transient contact with a solid surface, as distinct from perturbed or "bound" water, which is in intimate contact with hydrated ions, molecules, and macromolecules. Most of the water in hydrated systems is considered to be unperturbed or so-called "bulk" water. The binding state of water has an important bearing on its ability to solidify. The more closely liquid water is hydrogen-bonded to solutes or hydrophilic solids, the less readily it will form ice.

Depending on temperature and pressure, water can exist in several different stable crystalline forms and one metastable amorphous form. Hexagonal ice is the most favorable crystalline arrangement below 273 K, when water is transformed from a highly disorganized liquid to a highly organized solid. This phase change is accompanied by massive dimensional changes. It is the ice of snowflakes, ice cubes, frozen peas, and unfortunately, the form of ice most frequently encountered in low-temperature sample preparation.

Amorphous ice is a noncrystalline or glassy form of ice. It can be formed by rapidly cooling ($\sim 10^5$ K/s) minute quantities of aqueous solutions ($\leq 100\ \mu m^3$). Once amorphous ice is formed, it must be maintained below the glass transition temperature T_g in order to maintain the random arrangements of the water molecules and prevent devitrification and subsequent recrystallization. The T_g for pure water is 135 K, but in the presence of dissolved solutes, the temperature will be a little higher.

14.3. Conversion of Liquid Water to Ice

The conversion of liquid water to ice involves three distinct, but closely related processes that occur in the following order:

1. *Heat removal.* The low-temperature in the immediate environment of the specimen acts as a heat sink and removes thermal energy from the specimen until it is in equilibrium with the environment. The process is specimen-dependent.

2. *Nucleation.* As liquid water is cooled, the random movements of the water molecules produce clusters of the correct size and shape for the condensation of further water molecules, which form the nucleus of an ice crystal. The process whereby water provides its own nucleation sites is called homogeneous nucleation. Water invariably contains dissolved materials and particulates, which may act as catalysts for heterogeneous nucleation. The probability of heterogeneous nucleation occurring increases with the volume of water and decreases with temperature.

3. *Ice crystal growth.* Following nucleation, the ice crystals grow. The rate of growth is determined by the speed at which water molecules move to the point of crystal growth, the accommodation of water molocules at the crystal interface, and the transport of heat away from the growing interface. The rate of ice crystal growth is critically dependent on the rate at which heat can be removed from the system.

We can categorize heat removal by the rate of cooling, for example:

- *Very slow cooling:* ~0.01 K/s. Numerous nucleation events occur and few, but very large crystals develop.
- *Fast cooling:* ~1–100 K/s. More nucleation events occur and more smaller ice crystals are procuced.
- *Very fast cooling:* ~1000 K/s. Many more nucleation events occur and many more, but much smaller ice crystals develop.
- *Ultrarapid cooling:* ~100,000 K/s. So many nucleation events occur that there is virtually no time for crystallization. At this cooling rate amorphous or vitreous ice may be produced.

14.4. Specimen Pretreatment before Rapid (Quench) Cooling

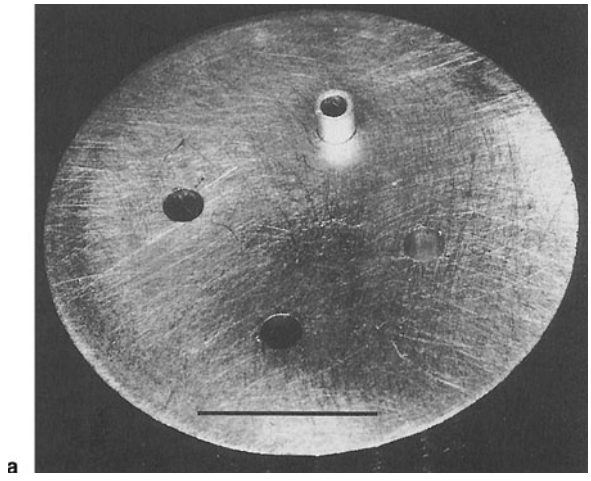
On the assumption that the aim of the preparative technique is, at best, to vitrify the specimen, at worst, to convert the water to microcrystalline ice, then the following procedures can be adopted *prior* to rapidly cooling the sample.

14.4.1. Minimizing Sample Size and Specimen Holders

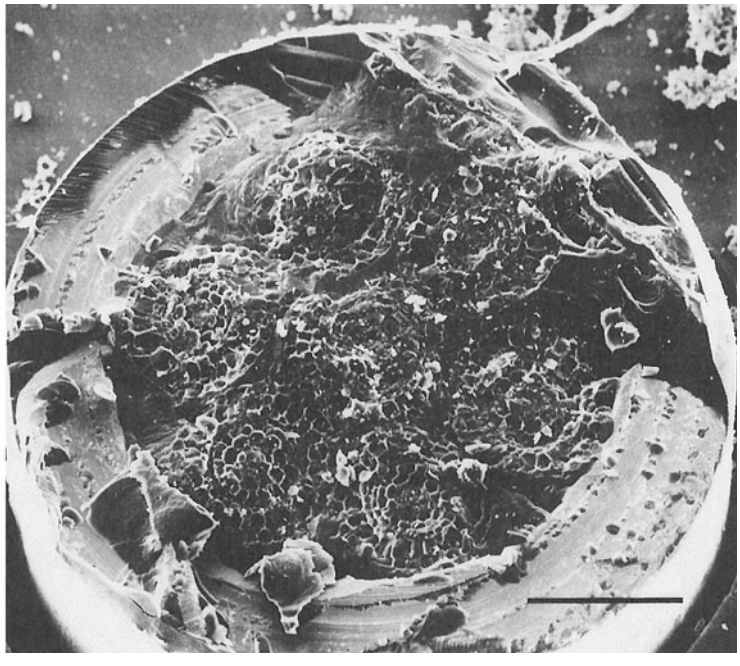
Wherever possible the surface area-to-volume ratio of the sample should be maximized and the sample made as small as possible. There are three general types of sample holders:

1. *Standard specimen holder.* The sample is fixed onto the standard (usually large) specimen holder at room temperature. This large mass can only be cooled relatively slowly with slushy nitrogen at 65 K and will result in some ice crystal damage.
2. *Large specimen holder.* Small pieces of material are quench-cooled in an appropriate liquid cryogen and, under liquid nitrogen, fixed into a vicelike holder on a larger specimen holder. The small pieces of material, although cooled rapidly, require delicate manipulation to insert them into the secondary holder.
3. *Small specimen holder.* Small samples, placed in small holders at room temperature, are quench-cooled in liquid cryogen and then fixed into larger holders under liquid nitrogen. This arrangement, which is shown in Fig. 14.2, is the preferable configuration because it allows a very small sample to be cooled rapidly.

Microorganisms, particulates, and macromolecules can be suspended either in thin films or microdroplets. Individual cells can be cultured inside 200-μm cellulose microcapillaries, which are then quench-cooled *in situ* (Tiedemann *et al.,* 1998). With larger samples the size and shape are going to be dictated by the minimum size necessary for a particular experimental procedure. For analytical studies dehydration should be avoided before rapid cooling by manipulating the sample in moist chambers.



a



b

Figure 14.2. (a) Duralium specimen holder precision drilled with tapered holes to take four small sample tubes. The tubes, which can be made from silver, beryllium, or graphite, are 1×3 mm in size with an inside diameter of 0.6 mm. The tubes are filled with the sample, quench-cooled, and then loaded into the holes under liquid nitrogen. The holder is then transferred under liquid nitrogen to the precooled cold stage of the SEM. Magnification marker = 5 mm. (b) Low-temperature (140 K) SE image of seven cross-fractured root tips of *Lemna minor* (Duckweed) loaded into a 1×3-mm silver tube as described for (a). Magnification marker = 200 μm. (Echlin, 1992.)

14.4.2. Maximizing Undercooling

The sample should be undercooled as much as possible without altering its natural state and before nucleation occurs. The amount of undercooling that may be obtained in biological material is a compromise between retaining the natural physiological processes and preventing any ice crystals from forming, for example, a piece of mammalian tissue could be slowly cooled to 280 K without any deleterious effects.

14.4.3. Altering the Nucleation Process

One approach is to apply a high pressure to the sample momentarily before it is rapidly cooled. This approach has the effect of dramatically lowering the temperature at which nucleation will occur; this particular approach is embodied in the hyperbaric cooling procedure, which is discussed in Section 14.5.

14.4.4. Artificially Depressing the Sample Freezing Point

Depressing the sample freezing point is achieved by using cryoprotectants, a group of polyhydroxy chemicals that suppress ice crystal formation. They are mainly used as antifreeze agents in biological systems and fall into two classes: *Penetrating cryoprotectants* work inside cells and tissues and include such chemicals as glycerol, sucrose (the most commonly used), dimethylsulfoxide, methanol, ethanol, and sugar alcohols. They are all quite large molecules and it is necessary to first permeabilize the cell membranes in order for them to get inside cells and tissue. *Nonpenetrating cryoprotectants* work outside cells and tissues and include such materials as polyvinylpyrrolidone and hydroxyethyl starch of varying molecular weights.

Both types of cryoprotectants are very soluble in water, forming strong interactions via their hydrogen bonding potential. They lower the equilibrium freezing point and promote undercooling before ice nucleation is initiated. They also decrease the viscosity of the medium. This in turn decreases the mobility of water molecules and slows the rate of ice crystal growth and thus promotes vitrification.

Cryoprotectants have proved to be very useful in preparing samples for structural studies even though some of the materials are toxic and have their own catalogue of artifacts. It is not advisable to use cryoprotectants in association with x-ray microanalytical studies because it is first necessary to chemically fix the sample in order for the large antifreeze agents to penetrate the cells and tissues.

14.4.5. Chemical Fixation

It might be appropriate to consider some very mild and precisely focused fixation, an invasive procedure, for special circumstances such as increased permeability to allow cryoprotectants into the specimen or

crosslinking sensitive epitopes to macromolecules for subsequent immunocytochemistry.

14.5. Quench Cooling

Quench cooling is the most critical aspect of low-temperature specimen preparation. Fortunately quench cooling is a fairly routine procedure that will give excellent, that is, ice-crystal-free, specimens. The materials used to quench-cool samples are referred to as cryogens and are either liquids or solids. Cryogens should have the following properties: (1) a low melting point and a high boiling point to minimize surface film boiling, (2) a high thermal conductivity and thermal capacity, and (3) a high density and low viscosity at the melting point. Finally, they should be safe, environmentally friendly, inexpensive, and readily available. Liquid and solid cryogens are used in different ways to achieve rapid cooling. Rapid cooling takes place through two processes: *conductive cooling,* which relies on direct contact between the sample and its surroundings, and *convective cooling,* which relies on the mass transfer of heat by circulation of the cryogen past the surface of the sample. Various cryogens and methods of quench cooling are discussed in the following sections.

14.5.1. Liquid Cryogens

The most popular liquid cooling agents are summarized as follows (Table 14.1):

1. *Liquefied organic gases* are inexpensive and readily available and are excellent primary cryogens when cooled by liquid nitrogen. There is virtually no film boiling because of the wide separation between their boiling and melting points. Liquefied organic gases such as propane or ethane are potentially explosive because their temperature is below the melting point of oxygen, which can condense onto the liquefied gas. Use of these liquefied gases should be confined to a spark-free fume hood that can be flooded with an inert gas such as nitrogen.

2. *Liquid nitrogen* is inexpensive and readily available. It is an excellent secondary cryogen, that is, it can be used to cool other cryogens. It is a poor primary cryogen because the melting point and the boiling point are so close

Table 14.1. Thermal and Physical Properties of Four Liquids and Two Solids that Can Be Used to Quench-Cool Samples

Material	Melting point (K)	Boiling point (K)	Thermal conductivity ($J\ M^{-1}\ s^{-1}\ K^{-1}$)
Ethane	90	184	0.24
Propane	84	231	0.22
Liquid nitrogen	63	77	0.13
Slushy nitrogen	63	77	0.13
Pure copper	—	—	570 (at 77 K)
Sapphire	—	—	960 (at 77 K)

together that any object placed into the liquid is quickly surrounded by a layer of poorly conducting gas due to surface film boiling.

3. *Slushy nitrogen,* a mixture of solid and liquid nitrogen, is easily made by boiling liquid nitrogen under reduced pressure. There is less film boiling because when a sample is placed into the mixture it melts the solid nitrogen before causing the liquid to boil.

4. *Liquid helium* is expensive and not readily available. It is an excellent secondary cryogen where very low temperatures are needed. It is a poor primary cryogen.

14.5.2. Solid Cryogens

Solid cryogens are used in the form of blocks of very pure materials such as gold-plated copper that have a high thermal heat capacity, that is,

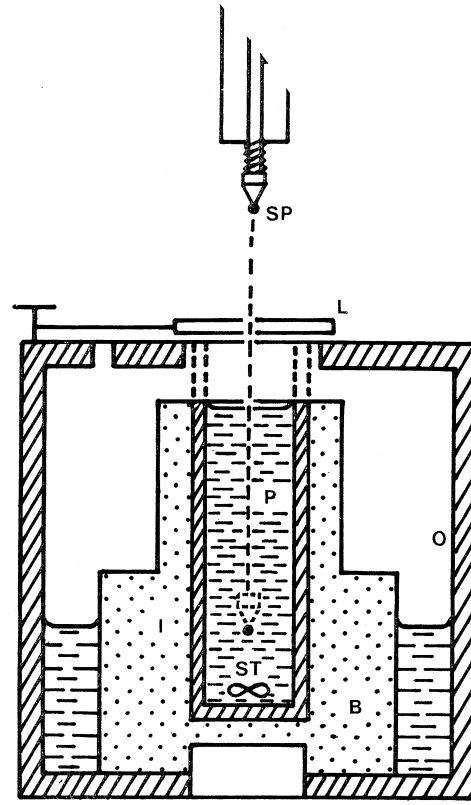


Figure 14.3. A diagram of a simple plunge-cooling device. The inner chamber sits snugly in a large block of aluminum (B) that is cooled by the primary cryogen, liquid nitrogen, in the outer chamber. The secondary cryogen (P), which may be liquid propane or liquid ethane, fills the inner chamber to within 2–3 mm from the top. The secondary cryogen is gently stirred (ST) and maintained a few degrees above its melting point. The specimen (SP), placed at the end of a spring-loaded device, is mechanically injected into the cryogen immediately after the shutter (L) is moved to one side. The specimen is quench-cooled continuously as it travels through the depth of the cryogen. (Echlin, 1992.)

they have the property of absorbing heat without any increase in temperature. The primary cryogen is either liquid helium or liquid nitrogen. In this application, it is important to keep metal surfaces very clean.

14.5.3. Methods for Quench Cooling

There are five main methods used for quench cooling. Each method works on the principle of transferring a sample at sub-ambient temperature to very low temperatures as fast as possible. Care must be taken not to precool the sample too much just before it is rapidly cooled. The five methods are discussed briefly in the following paragraphs.

1. *Immersion or plunge cooling* is a simple, inexpensive, and surprisingly effective way to quench-cool specimens and involves both conductive and convective cooling. Figure 14.3 is a diagram of a suitable device, which can be built in the laboratory. It should have a spring-loaded specimen holder to rapidly propel the sample into a secondary cryogen such as propane or ethane, kept just above its melting point by liquid nitrogen. The size and shape of the sample are important and it must travel as far as possible through the cryogen.

2. *Jet cooling* involves melting propane, which is squirted under pressure onto the side(s) of the specimen. Although this is an effective way to quench-cool thin, flat specimens by both conductive and convective cooling, the equipment is specialized and expensive.

3. *Spray and droplet cooling* involves spraying aerosol droplets onto a clean, polished copper surface cooled by liquid nitrogen. This results in very fast conductive cooling and is a good method for microorganisms and small particulates. The equipment is relatively inexpensive and can be made in the laboratory.

4. *Impact or slam cooling* is a very effective way of cooling the surface layers of quite large specimens. As Fig. 14.4 shows, the sample is propelled rapidly onto a clean, highly polished metal surface such as gold-plated, high-purity copper, cooled either by liquid nitrogen or liquid helium. The contact region of the specimen with the metal surface may show some mechanical deformation, but just below (1–2 μm) this surface there is excellent conductive cooling. The equipment is complex and moderately expensive. Modifications include cryosnappers and cryoneedles for *in vivo* sampling (Shimoni and Muller, 1998).

5. *High-pressure or hyperbaric cooling* is the final technique. An examination of the phase diagram of water shows that water will vitrify if cooled under high pressure. The high pressure, typically 2–3 kbar, diminishes crystallization and nucleation and lowers undercooling and will vitrify some specimens to a depth of 700 μm, although 250 μm is the more usually accepted figure (Sartori *et al.,* 1993). Liquid nitrogen is used as the cryogen. This is not a fast cooling procedure, but it is the only technique that will vitrify a relatively large part of the sample. The equipment is complex and expensive, although it is easy to use and gives a high throughput of well-preserved samples.

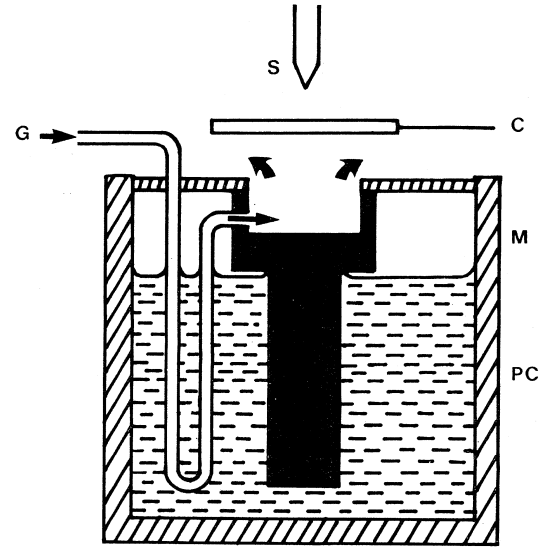


Figure 14.4. A diagram of a simple impact-cooling device. The base of a large block of high-purity copper is cooled by the primary cryogen, liquid nitrogen. The highly polished surface of the copper block is continually flushed with precooled dry nitrogen to prevent any surface frosting. The shutter (C) is moved to one side, and the specimen (S), placed at the end of a spring-loaded device, is mechanically pressed onto the cold metal surface. Although the first few micrometers of the sample may show some mechanical damage, there is excellent preservation to a depth of 15–25 μm. (Echlin, 1992.)

14.5.4. Comparison of Quench Cooling Rates

It is very difficult to compare the different cooling rates and come up with a single technique that would be best suited for all situations. The data in Table 14.2 give a general idea of the effectiveness of the five different methods used to quench-cool samples. As a generalization, samples no bigger than 50 μm in diameter would be effectively cooled (but not vitrified) by any of the methods described. Larger samples would be effectively cooled to about 25 μm from the surface. Samples no bigger than 250 μm, immersion-cooled in a mixture of melting ethane plus propane, are satisfactorily cooled, but are only vitrified to about 20 μm below their surface. Impact and jet

Table 14.2. Comparison of the Effectiveness of the Five Main Methods Used to Quench-Cool Soft Samples

Method of cooling	Mean cooling rate (10^3/s/K)	Depth of ice-crystal-free zone (μm)	Best cryogen (K)
Plunge	10–12	5–10	Ethane (93)
Spray	50–55	20–30	Propane (83)
Jet	30–35	10–15	Ethane (93)
Impact	20–25	15–25	Copper (20)
Hyperbaric	0.5	400	Liquid nitrogen (77)

cooling will vitrify to about 25 μm below the surface of a bulk specimen. Spray cooling will vitrify 40-μm-diameter droplets and suspensions, and high-pressure cooling will vitrify about 250 μm into a bulk sample.

14.6. Low-Temperature Storage and Sample Transfer

If there is a time gap between cooling the sample and any subsequent manipulations, it will be necessary to store the sample in a liquid nitrogen dewar. The quench-cooled samples should be removed from the cooling device under liquid nitrogen using liquid nitrogen-cooled tools and quickly placed into small plastic vials filled with liquid nitrogen. These vials should be properly labeled and inserted into the metal slots of the canelike rods that are a standard feature of liquid nitrogen refrigerators. Specimens should not be exposed to air even at low temperature because of the possibility of water condensation. If necessary the samples can be washed in clean liquid nitrogen filtered through cotton gauze to remove minute ice crystals. At some later stage it will be necessary to transfer the frozen sample to the next part of the preparation process. It is most important that the sample does not undergo devitrification, recrystallization, or contamination. The transfer is best achieved with the sample immersed in liquid nitrogen or in an environment of dry, cold nitrogen gas and/or *in vacuo*.

14.7. Manipulation of Frozen Specimens: Cryosectioning, Cryofracturing, and Cryoplaning

In some cases, it is only the surface features of the sample that are of interest. Figure 14.5 is a good example of how low temperature can be used to study the surface features of a very fragile specimen. Provided this surface is undamaged and uncontaminated and the sample is kept below 123 K in a clean, high-vacuum environment, the sample can be imaged directly in a SEM fitted with a cold stage.

More often, it is the inside of the sample that is of interest and it is appropriate to consider the ways we can expose the interior of a frozen specimen. Two processes, sectioning and fracturing, are physical procedures that rely on the solidified phases of water to provide the mechanical strength to otherwise soft matrices.

14.7.1. Cryosectioning

Cryosectioning is a flow process involving extensive plastic deformation of the ice as the material is thinly sliced at the tool (knife) and sample interface. The resulting sections can be viewed either in the fully frozen hydrated state or after the ice has been sublimated by freeze-drying. There are

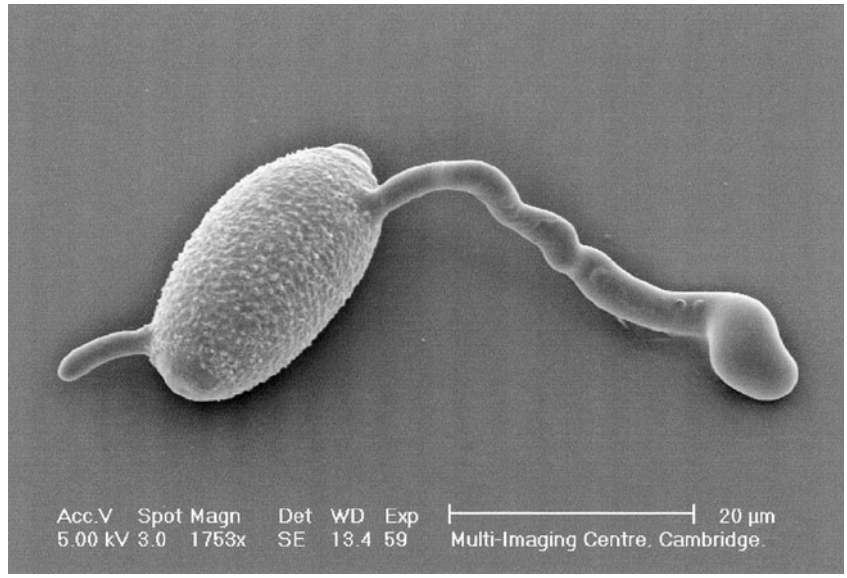


Figure 14.5. An SE image of a frozen-hydrated germinating fungal spore at 130 K. Magnification marker = 20 μm. (Courtesy of Alison Daniels, Aventis Crop Science.)

different categories of section, depending on thickness. Ultrathin sections that are 50–100 nm thick are suitable for study in the TEM. Thin sections that are 0.1–1.0 μm thick are suitable for the SEM.

Cryoultramicrotomes are used for cutting 0.1- to 1.0-μm sections. These instruments operate at temperatures between 123 and 193 K. The knives are made from glass, diamond, sapphire, and metal. Specially designed cryo-diamond knives are best for cutting thin, 0.25-μm sections, and tungsten carbide knives can be used to cut thicker, 1-μm sections.

The actual procedures for cryosectioning will not be discussed because they vary depending on the type of equipment that is used. It is useful, however, to consider certain general procedures. The frozen specimen should either be mounted on the microtome sample holder mechanically or be stuck by using a cryoglue such as melting toluene. Alternatively, the sample can be inserted into a thin-walled silver tube (1×3 mm) that is quench-frozen and then inserted into a hole in the microtome sample holder. Wherever possible, the sample mounting should be done under liquid nitrogen. The microtome sample holder and its securely fixed specimen should be transferred to the precooled working chamber of the microtome under liquid nitrogen, fixed to the cutting arm, and allowed to equilibrate to the working temperature of the cryomicrotome.

The temperature should be low, ideally 133 K, the sectioning thickness controls set to give the thinnest sections, and the cutting speed set to the lowest value. If it is not possible to cut continuous sections under these conditions, the temperature, cutting speed, and section thickness should be slowly increased until continuous, smooth sections are obtained. These values will vary with the chosen material and the success of the quench cooling process.

As a general rule, samples that are either vitrified or contain microcrystalline ice can be sectioned more thinly, more easily, and at lower temperatures. Samples containing large ice crystals appear milky, and smooth sections can be obtained only at an increased section thickness and at higher temperatures. At low cutting speeds and low temperature (233 K) there is no transient melting of the ice. Thick sections are best cut at 233–253 K, thin sections at 193–213 K, and ultrathin sections at 123–143 K.

14.7.2. Cryofracturing

Cryofracturing involves separation of the frozen specimen along a line of least resistance parallel to the applied force. A fracture face suitable for SEM and x-ray microanalysis is obtained. Figure 14.6 shows the different ways a sample can be cryofractured. Stress fractures are produced when a sharp blade is used to initiate the fracture and percussive fractures are produced when a blunt tool is pressed against the side of the specimen with sufficient force to cause it to fracture at some point below the contact point.

The salient features of the process are as follows. The sample should be held firmly below the glass transition temperature T_g and the fracturing tool should not make contact with the freshly exposed fracture face. Whereas cryosectioning produces smooth faces, cryofracturing produces undulating

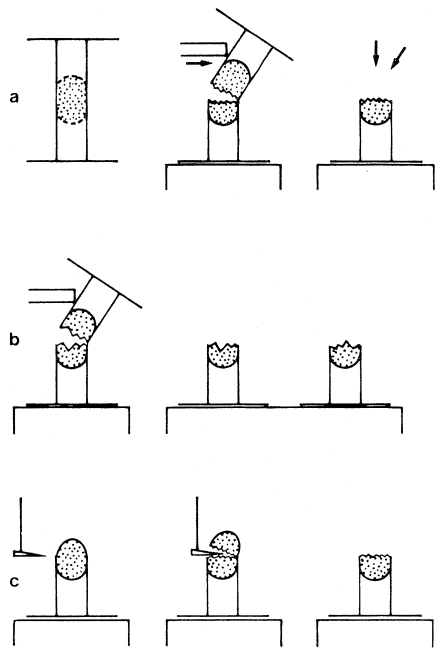


Figure 14.6. A diagram showing the different ways a sample may be cryofractured. (a) An impact fracture in which the lower half of the fracture is retained and may be etched and/or coated. (b) An impact fracture in which both the upper and the lower parts are retained to provide a pair of complementary fracture faces. (c) A sharp blade fracture that generates shear forces just ahead of the advancing cold knife. The blade should move in an upward-sweeping arc to avoid touching the freshly exposed surface. (Echlin, 1992.)

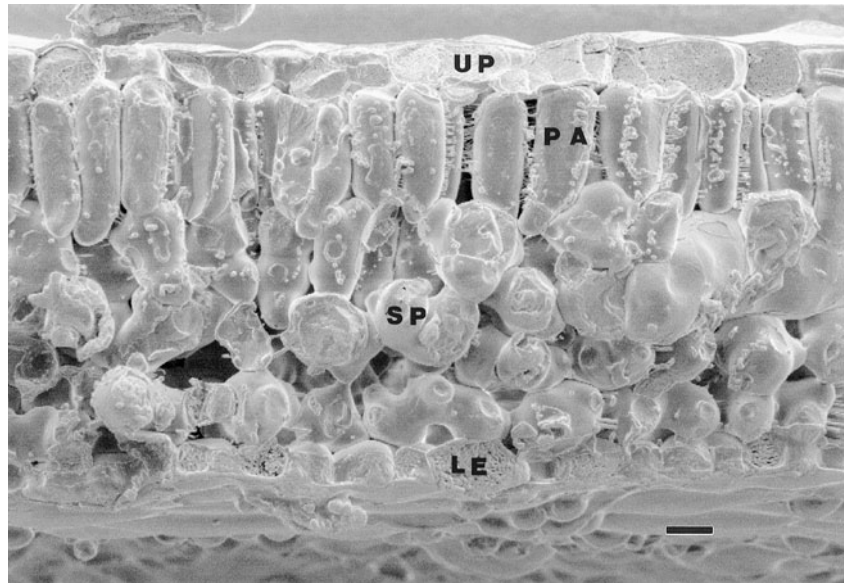


Figure 14.7. A frozen-hydrated fracture face of a green Assam tea leaf showing the upper (UP) and lower (LE) epidermal cell layers and the intervening palisade (PA) and spongy (SP) mesophyll tissue. Image recorded at 130 K, 3 keV, and 30 pA. Magnification marker = 20 μm.

faces. Brittle material will fracture with little deformation. Some ductile materials will not fracture, even at very low temperatures, but instead undergo plastic deformation.

Simple cryofracturing devices can be constructed in which a properly cooled sample is held under liquid nitrogen and continually flexed in one direction until it breaks. It is important to make sure all debris is removed from the fracture surface.

More complicated fracturing devices are usually bolted to the side of a SEM fitted with a cold stage. A properly cooled and protected sample is introduced, via an airlock, onto the precooled cold stage of the fracturing device, which is at a high vacuum. These more complicated devices allow repeated stress fracture to be made with a microtome-driven knife and ice in the fracture face can be removed by sublimation under high vacuum (etching). Once a suitable, etched or unetched, fracture face is produced, it may be lightly coated (see Chapter 15) before passing through a second airlock onto the precooled cold stage of the SEM. This type of fracturing device can be used to produce the large, low-magnification, flat fractures (Fig. 14.7) suitable for tissue identification and x-ray microanalysis (Fig. 1.8, Chapter 1) and small, high-magnification fractures (Fig. 14.8) suitable for structural studies.

14.7.3. Cryopolishing or Cryoplaning

This procedure produces very flat surfaces of frozen material suitable for x-ray microanalysis and backscattered imaging. Polished surfaces can

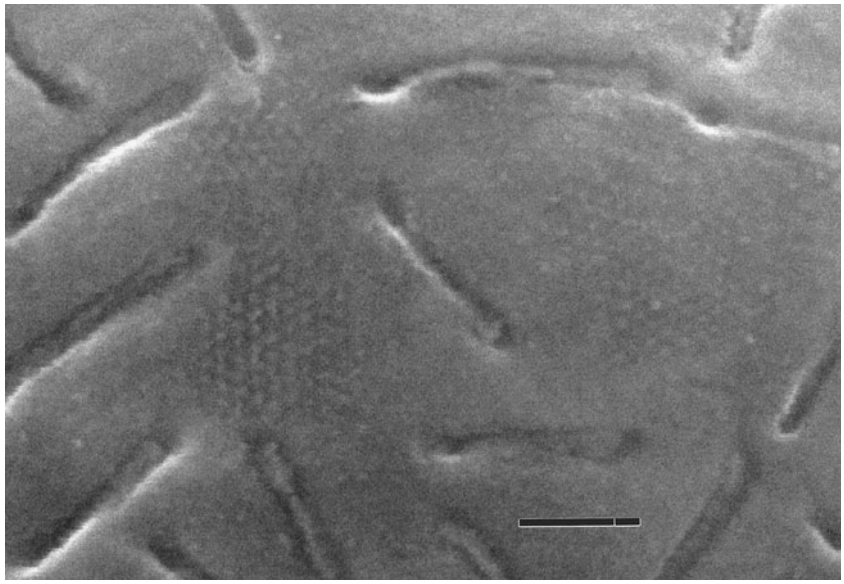


Figure 14.8. A frozen-hydrated fracture of the protoplasmic face of a yeast cell membrane. The regularly spaced protein particles are 15 nm in diameter. Image recorded at 130 K, 1 keV, and 20 pA. Magnification marker = 150 nm. (Courtesy of Tony Burgess, Multi-Imaging Centre, University of Cambridge.)

be produced by two different methods: (1) The quench-frozen sample is cryosectioned at as low a temperature as possible and the sections discarded. The remaining block face will have a very smooth face. This technique is illustrated in Fig. 14.9. (2) The frozen sample is cryofractured close to the region of interest and the undulating fracture face polished with fine alumina powder on a metal plate cooled with liquid nitrogen. Care has to be taken in transferring the polished, frozen surface to the SEM and that it is uncontaminated with either ice crystals or the polishing powder.

14.8. Ways to Handle Frozen Liquids within the Specimen

After exposing the inside of a carefully frozen sample, it is necessary to deal with the frozen liquids (usually water) within the sample. There are three possible courses of action. It is possible to *retain* the water as ice and examine and analyze the so-called frozen-hydrated samples at temperatures below T_g. Alternatively, it is possible to *sublime* the frozen water at low temperature and high vacuum and study the so-called frozen-dried samples either at ambient or at low temperature. The third option is to first *dissolve* the frozen water in organic liquids at low temperatures by a process referred to as freeze substitution and then embed the substituted material in a liquid resin that is subsequently polymerized and then sectioned.

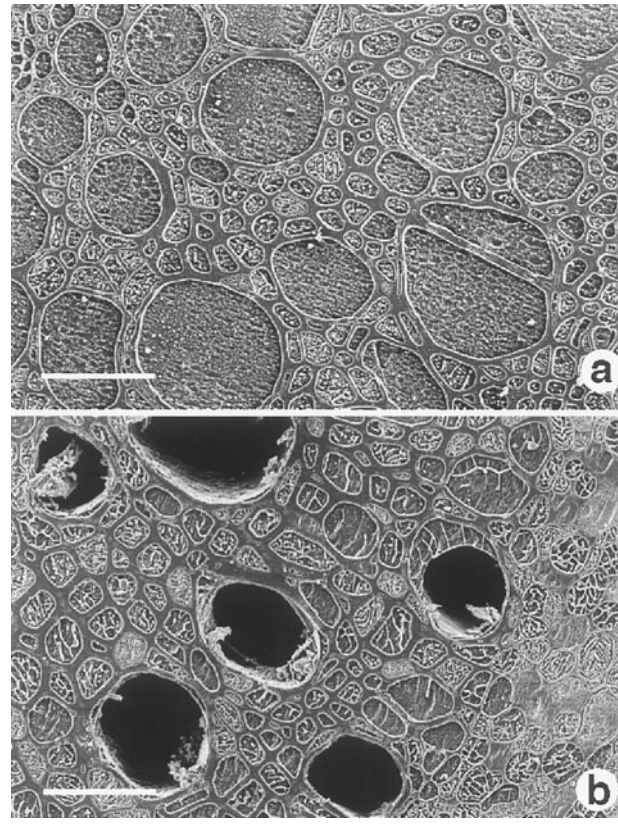


Figure 14.9. Cryoplaned transverse faces of pieces of buckwheat roots frozen *in situ* in the field. (a) Frozen at dawn: all the large xylem vessels are filled with sap. (b) Frozen at 08.30 h: many of the large vessels contain gas embolisms. Images recorded at 153 K and 7 keV. Magnification marker = 50 μm. (McCully *et al.,* 2000.)

14.8.1. Frozen-Hydrated and Frozen Samples

If the specimen is to be examined in the fully frozen state, no further preparation is needed. Care has to be taken that the liquid material (now solidified) in the sample remains below the glass transition temperature and that the surface to be studied is uncontaminated during transfer to and examination in the SEM. A variety of devices have been designed to maintain the clean and frozen state as samples are moved from where they have been prepared to the microscope cold stage. Most transfer devices have a vacuum interface with the SEM.

The fully frozen state is the best way to examine liquid and soft bulk samples, provided they have been cooled adequately. Such samples, which have not been exposed to chemicals during sample preparation, are probably the closest to their natural state. The same is true for fracture faces and sections, except that problems of beam damage make it impossible to study thin frozen sections. Provided there is no ice crystal contamination, the

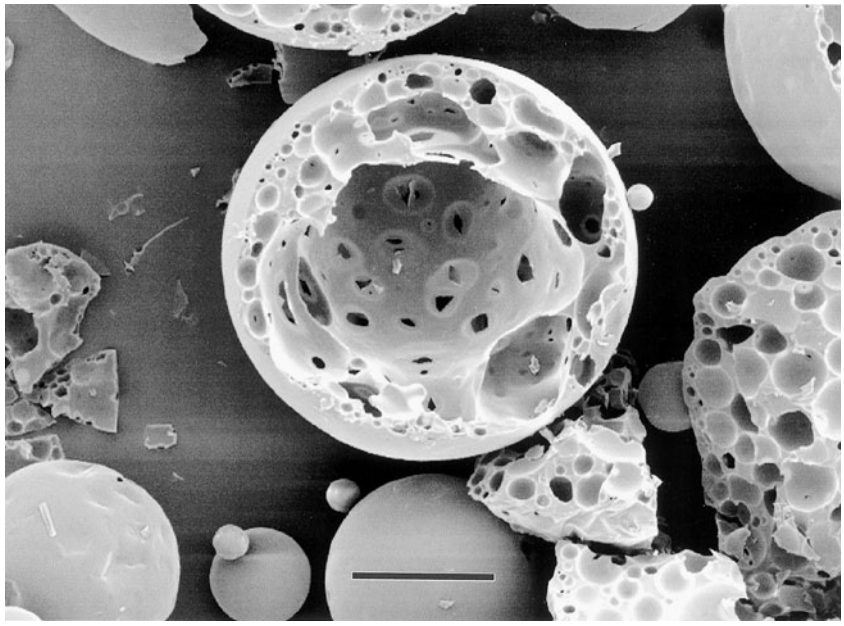


Figure 14.10. Freeze-dried instant coffee granules. Magnification marker = 3 μm. (Courtesy of Tony Burgess, Multi-Imaging Centre, University of Cambridge.)

quality of the secondary electron image of frozen intact specimens appears the same as specimens prepared at ambient temperatures.

14.8.2. Freeze-Drying

The freeze-drying procedure works on the principle that at certain temperatures and pressures, solidified liquids will subime, that is, go from the solid to the gas phase without an intervening liquid phase. Nearly all the work on freeze-drying has centered on water and it is now a well-characterized procedure for preserving food stuffs (see Fig. 14.10) and pharmaceutical agents. As a preparative technique for SEM and x-ray microanalysis, freeze-drying is more than just removing water from the sample. The water molecules play an important structural role providing support and stopping the structural macromolecules from sticking together. Removal of the "bulk" water can lead to shrinkage and removal of hydration shells around molecules can lead to molecular collapse. In a frozen-dried sample, the space once occupied by water is to all intents and purposes empty.

14.8.2.1. Physical Principles Involved in Freeze-Drying

The equipment and experimental protocols for freeze-drying are designed around the fact that there is a dynamic equilibrium between the surface of the ice (solid) and the water vapor in its immediate surroundings. The rate at which water molecules leave the solid depends on the sample

Table 14.3. Guide to the Time It Will Take to Freeze-Dry Samples at Different Temperatures and Pressures

Temperature (K)	Vapor pressure (H_2O)	Time to remove 1.0 μm of ice
193	53.6 mPa	13 sec
188	22.9 mPa	30 sec
183	9.32 mPa	1 min, 13 sec
178	3.61 mPa	3 min, 5 sec
173	1.32 mPa	8 min, 19 sec
168	457 μPa	23 min, 40 sec
163	148 μPa	1 h, 12 min, 20 sec
158	44.5 μPa	3 h, 56 min, 40 sec
153	12.4 μPa	13 h, 56 min, 40 sec

temperature and the efficiency by which the sublimed water molecules are trapped. The partial pressure of water vapor in the region immediately surrounding the sample is more important than the absolute vacuum. Although it is possible to calculate accurately the sublimation rates of pure ice as a function of temperature and saturation vapor pressure, the rates for ice in combination with other materials are less certain. Table 14.3 shows that the time it takes to sublime ice from a frozen specimen is a function of pressure and temperature.

Freeze-drying proceeds from the outside of the sample to the inside and as the sublimation occurs, the water molecules have to pass through an increasing depth of the sample, which may be 1000 times longer than for a piece of ice of the same dimension as the sample. This prolongation factor means that the drying times for ice in combination with other molecules are much longer than for the pure material. Different samples will dry at different rates.

14.8.2.2. Equipment Needed for Freeze-Drying

There are several freeze-dryers on the market, but with a little ingenuity it is possible to build one's own, provided the following points are borne in mind. Ideally, the cold platform should cool to 150 K to minimize any chance of recrystallization and heat to 300 K to sublime the last traces of water from the sample. In addition to a heater and cooler, it is important that the system be fitted with time and temperature controllers linked to a thermostat. If possible, the drying chamber should achieve a vacuum as low as 1 mPa ($\sim 10^{-5}$ torr), although some specimens can be dried at 193 K and 0.1 Pa ($\sim 10^{-3}$ torr). The drying chamber must have an efficient water trapping system, which can most conveniently be achieved by having a copper plate cooled with liquid nitrogen (77 K) and placed close (10–20 mm) to the specimen. It is useful to have an airlock in order to remove the very dry sample and transfer it to a dry environment and it should be possible to vent the whole system to atmospheric pressure with a dry inert gas. An added refinement would be to incorporate a system to allow the dried sample to be infiltrated with a liquid resin and subsequently polymerized.

Frozen sections can be freeze-dried either on the temperature-controlled platform in the microtome cold chamber or by exposing them to a gentle stream of dry nitrogen. See the accompanying CD, Enhancements 14.1 and 14.2, for detailed freeze-drying protocols.

14.8.2.3. Artifacts Associated with Freeze-Drying

Artifacts are related mainly to the fact that water plays an important structural role. The more water there is in the sample and the less the amount of structural polymers, the greater is the shrinkage. Shrinkage is not isodiametric, but is variable. If the drying temperatures and pressures are not correct, ice crystals will grow (recrystallize). This growth is only apparent as ice crystal ghosts, after the drying is complete. As the molecular shell water is removed, components of the structural matrix may aggregate and collapse. This artifact is only visible at high resolution.

14.8.3. Freeze Substitution and Low-Temperature Embedding

Although freeze substitution and low-temperature embedding are two distinct processes, they are considered together because they complement each other and are frequently used sequentially during sample preparation. Freeze substitution is a chemical dehydration procedure in which ice in frozen-hydrated samples is removed and replaced by an organic solvent. This procedure is in sharp contrast to freeze-drying discussed in the previous section, where the water is removed by purely physical means and is not replaced. Low-temperature embedding seeks to replace either the spaces once occupied by water in a frozen-dried specimen or the organic fluid that has been used in freeze substitution, by infiltrating the samples with liquid resins, which are, in turn, polymerized at low temperatures. The review article by Skepper (2000) and the paper by Monaghan *et al.* (1998) provide an excellent introduction to the subject. For more details, including experimental protocols, see Enhancement 14.3 on the accompanying CD.

14.8.3.1. Physical Principles Involved in Freeze Substitution and Low-Temperature Embedding

Freeze substitution and low-temperature embedding procedures have been developed to ensure that dehydration and embedding of hydrated bioorganic materials is achieved with minimal loss of structure and maximum retention of biological activity. These goals can be achieved provided three criteria are met during the sample preparation. The structural macromolecules must be immobilized to avoid supramolecular reorganization and denaturation. The solvent environment should remain "water-like" or polar during the entire process in order to maintain the hydration shells surrounding molecules. The amplitude of molecular thermal vibrations should be reduced to ensure biological stability during solvent exchange.

The temperatures at which the processes are carried out are critical. Ideally one would like to use an organic liquid that dissolves ice below its

Table 14.4. Melting Point and Water (Ice)-Dissolving Capacity of Organic Solvents that Can Be Used for Freeze Substitution

Compound	Melting Point (K)	Percentage of water dissolved at 193 K
Acetone	177.6	2.5
Diethyl ether	156.5	0.8
Ethanol	155.7	16.0
Methanol	179.1	32.0
Propane	83.3	0.1
Dimethoxypropane	226.0	[a]

[a] Reacts chemically.

recrystallization temperature (140 K). Unfortunately, at this temperature, the organic solutions become so viscous that they will not adequately penetrate the specimen. As Table 14.4 shows, there are only a few solvents that will dissolve ice at low temperatures and even then it takes a long time to complete the dehydration. At these same temperatures, the embedding resins solidify. The compromise is to use lower temperatures (193–173 K) for substitution and somewhat higher temperatures (253–233 K) for embedding.

14.8.3.2. Equipment Needed for Freeze Substitution and Low-Temperature Embedding

The whole process is carried out on the floor of a sealed chamber inside a liquid-nitrogen-cooled dewar. The chamber floor must be fitted with a heater and it is important to have reliable time and temperature controllers linked to a thermostat because it is necessary to change the temperature during sample processing. The chamber floor should have at least eight recesses to accommodate the 1 × 3-mm glass vials in which the substitution and embedding takes place. The chamber environment must be very dry and anoxic to enable the final process of resin polymerization to take place. The system should be fitted with an ultraviolet lamp to initiate the resin polymerization. See Enhancements 14.3–14.6 on the accompanying CD for detailed freeze substitution and low-temperature embedding protocols.

14.9. Procedures for Hydrated Organic Systems

Hydrated organic samples include most food materials, cosmetic and pharmaceutical preparations, paints, and some polymers. Such samples are generally emulsions made up of hydrophilic and hydrophobic phases and they may all be prepared using the same methods used for hydrated biological materials. In systems that contain a small amount of water, satisfactory preservation can be obtained by cooling the sample in slushy nitrogen and

for highly hydrated systems it may be appropriate to use cryoprotectants to diminish the amount of ice crystal damage.

14.10. Procedures for Hydrated Inorganic Systems

Hydrated inorganic specimens include soils, clays, cements, concretes, muds, geological samples, and, in particular, samples from the petroleum exploration industry, where the liquid phase is a mixture of oil and brine. The ambient-temperature specimen preparation methods for these types of sample have already been discussed in Chapter 11, Section 11.7. It is useful, however, to briefly review the circumstances whereby low-temperature preparative techniques can be applied to such seemingly robust samples. The paper by Fassi-Fehri *et al.* (1991) gives details of how cryoSEM can be used in the oil exploration industry. These types of sample may contain water, gas, brine, and/or oil, which may be preserved *in situ* by cooling the sample and examining it at low temperatures. Some of the same procedures that have been devised for hydrated biological materials can be used for these types of specimen. Although the samples are generally larger and high-resolution information is not usually required, it is usually possible to obtain good preservation by cooling in liquid or slushy nitrogen, after which the sample is fractured under vacuum at low temperatures. Figure 14.11

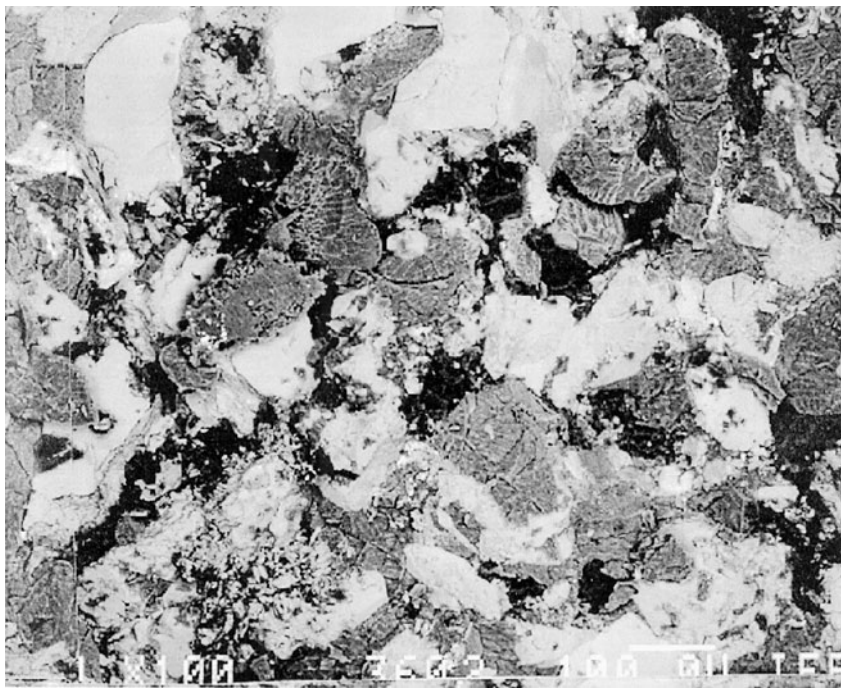


Figure 14.11. A BSE image of a frozen-hydrated fracture face of a cored piece of North Sea sandstone saturated with oil. The oil phase appears black, the brine appears mid-gray, and the parent rock appears light gray. Image recorded at 93 K. Magnification marker = 100 μm. (Robin *et al.,* 1999.)

shows an image of oil-bearing North Sea sandstone prepared using this procedure. If the sample is to be examined by x-ray microanalysis, it can be freeze-polished under liquid nitrogen and examined in the frozen-hydrated state. Alternatively, the sample can be examined in the frozen-dried state or freeze-dried, embedded in resin, and then polished by standard metallograhic procedures. The very nature of the sample precludes sectioning and there is little that could be gained by freeze substitution. For more details, including experimental protocols, see Enhancement 14.7 on the accompanying CD.

14.11. Procedures for Nonaqueous Liquids

The preparative procedures for nonaqueous liquids are usually less demanding. Although many organic liquids form a crystalline phase when cooled, the phase change is less disruptive than the changes that occur as water forms ice. If the organic liquid is known to undergo a disruptive phase change as it is cooled, then it will have to be cooled as rapidly as possible by any of the procedures discussed Section 14.5.3, with the possible exception of immersion in a liquefied organic gas. It is usually sufficient to use liquid or slushy nitrogen as the primary cryogen, after which the sample can be fractured or sectioned at low temperatures. Figure 14.12 shows an image of wax deposits in diesel fuel observed at 120 K. Freeze-drying and freeze substitution are not recommended for these types of specimen. Cooling the liquid sample on the cold stage of the microscope should be avoided because this may compromise the vacuum environment of the column.

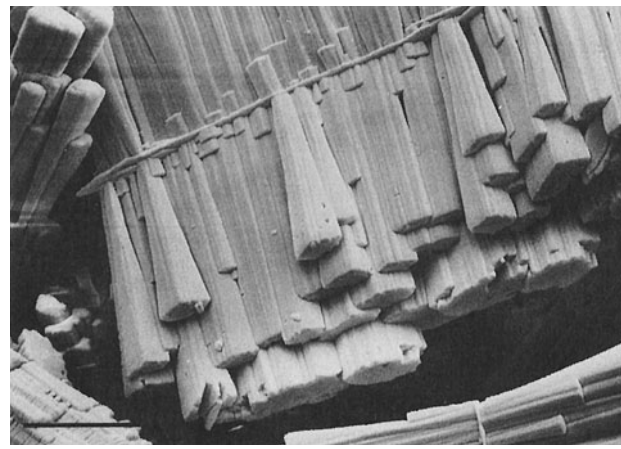


Figure 14.12. A low-temperature SE image of wax deposits filtered from diesel fuel. The wax can only be imaged at low temperatures because they contain occluded volatile solvents. Image taken at 120 K. Magnification marker = 20 μm.

14.12. Imaging and Analyzing Samples at Low Temperatures

Other than providing suitable cold stages, anticontamination plates, and transfer devices, the same SEM and x-ray microanalytical equipment can be used at both ambient and low temperatures. The papers by Marshall and Xu (1998) and Schultz *et al.* (1999) and the review article by Echlin (2001) provide an overview of the ways x-ray microanalysis can be carried out at low temperatures. Although the only real difference between low-temperature and ambient-temperature microscopy and analysis is the sample environmental there are some subtle differences in the appearance of the image and some real problems of beam damage. It is appropriate to briefly discuss these two matters at the end of this chapter on low-temperature specimen preparation.

The image quality of frozen-hydrated fracture faces, sections, and polished surfaces appears, at first glance, to have less topographic contrast because the solidified liquids have the same scattering characteristics as the natural solid phase. Image contrast can be improved dramatically by trying to sublime just the surface layers of the solidified liquids. This etching procedure is difficult to control accurately and unless care is taken, all the solidified liquid phase is removed.

Removal of the liquid phase is less important if samples are only to be imaged in the SEM. Table 14.5 shows that, when the frozen-hydrated specimens are freeze-dried, there is a dramatic mass loss from the matrix and an apparent increase in the concentration of the elements. Removal of the liquid phase will have a profound effect on any measurements of elemental concentration because of the marked change in the matrix composition. Figure 14.13 shows the differences in the size of the x-ray analytical volume between a frozen-dried and a frozen-hydrated sample.

Wherever possible, the sample should be studied in the fully frozen state because it is surprising how quickly one can recognize different parts of the sample using secondary electron imaging. With very flat or polished surfaces there is usually sufficient atomic number contrast to make use of the BSE imaging mode.

Irrespective of whether the samples are prepared at ambient temperature or low temperatures, they may be damaged and or show artifacts. An additional major problem with low-temperature sample preparation is ice crystal formation. This matter has been discussed earlier in this chapter

Table 14.5. Changes in the Peak-to-Background Ratio of a Number of Elements in Tobacco Leaf Mesophyll Cells Analyzed in the Frozen-Hydrated and the Frozen-Dried State

Hydration state	Na	Mg	P	S	K	Ca
Frozen-hydrated	0.06	0.04	0.29	0.26	0.52	0.09
Freeze-dried	0.13	0.28	0.37	0.37	0.72	0.56

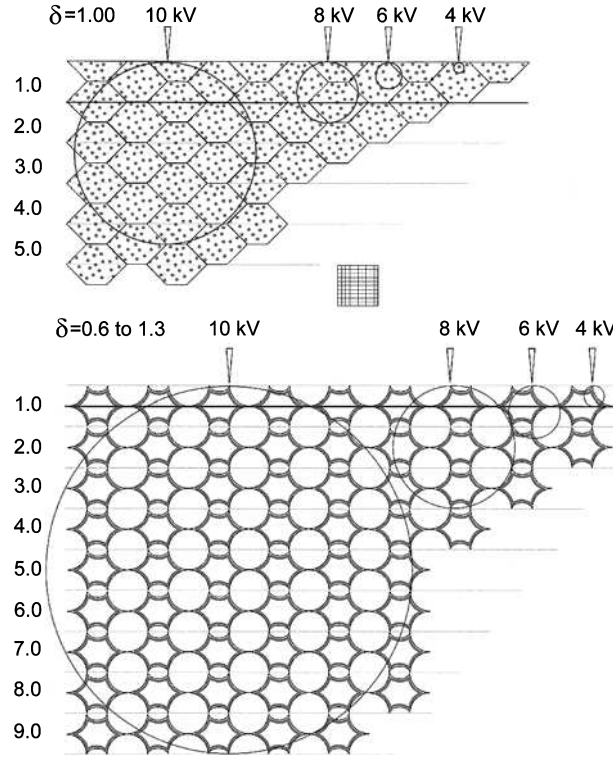


Figure 14.13. A schematic representation of the calculated size of the primary beam/specimen interactive volume in frozen-hydrated (upper) and freeze-dried (lower) generic biological tissue. The depth and width measurements are given in micrometers. The assumed mean density of the material is δ. (Echlin, 2001.)

and a number of ways have been suggested for avoiding this damage. Another problem with frozen-hydrated samples is that they are extraordinarily beam-sensitive because ice, particularly in contact with organic molecules, is a rich source of free radicals even though their diffusion rate is slowed down at low temperatures.

In many cases the beam damage is so subtle that it is assumed that the sample is undamaged. The best way to ascertain whether a problem is likely to exist is to calculate the total dose of electrons landing on the sample during examination, imaging, and analysis. Measurable and visible signs of beam damage include pits, craters, holes in the sample, and movement and melting of the sample. During x-ray microanalysis, there may be both loss of specific elements from the specimens and nonspecific mass loss from the whole sample.

References

Echlin, P. (1992). *Low Temperature Microscopy and Analysis,* Plenum Press, New York.
Echlin, P. (2001). *Microsc. Microanal.* **7,** 211.
Fassi-Fehri, O., M. Robin, and E. Rosenberg (1991). Society of Petroleum Engineers Publication 22596, Society of Petroleum Engineers, Richardson, Texas.

Marshall, A. T., and W. Xu (1998). *J. Microsc.* **190,** 305.
McCully, M. E., M. W. Shane, A. N. Baker, C. X. Huang, L. E. C. Ling, and M. J. Canny (2000). *J. Microsc.* **198,** 24.
Monaghan, P., N. Perusinghe, and M. Muller (1998). *J. Microsc.* **192,** 248.
Robin, M., R. Combes, and E. Rosenberg (1999). CryoSEM and ESEM: New Techniques to Investigate Phase Interactions with Reservoir Rocks. Society of Petroleum Engineers Publication 56829, Society of Petroleum Engineers, Richardson, Texas.
Sartori, N., K. Richter, and J. Dubochet (1993). *J. Microsc.* **172,** 55.
Schultz, M., F. Rudolf, and M.-F. Gallitelli (1999). *Microsc. Microanal.* **5,** 187.
Shimoni, E., and M. Muller (1998). *J. Microsc.* **192,** 236.
Skepper, J. N. (2000). *J. Microsc.* **199,** 1.
Tiedemann, J., H. Hohenberg, and R. Kollman (1998). *J. Microsc.* **189,**163.

15

Procedures for Elimination of Charging in Nonconducting Specimens

15.1. Introduction

The specimen can be thought of as an electrical junction into which flows the beam current i_B. The phenomena of backscattering of the beam electrons and secondary electron emission represent currents flowing out of the junction, i_{BSE} and i_{SE}, respectively. For a copper target with an incident beam energy of 20 keV, η is approximately 0.3 and δ is approximately 0.1, which accounts for 0.4, or 40%, of the beam current. The remaining beam current must flow from the specimen to ground to avoid the accumulation of charge in the junction (Thevenin's current theorem). The balance of the currents is then given by

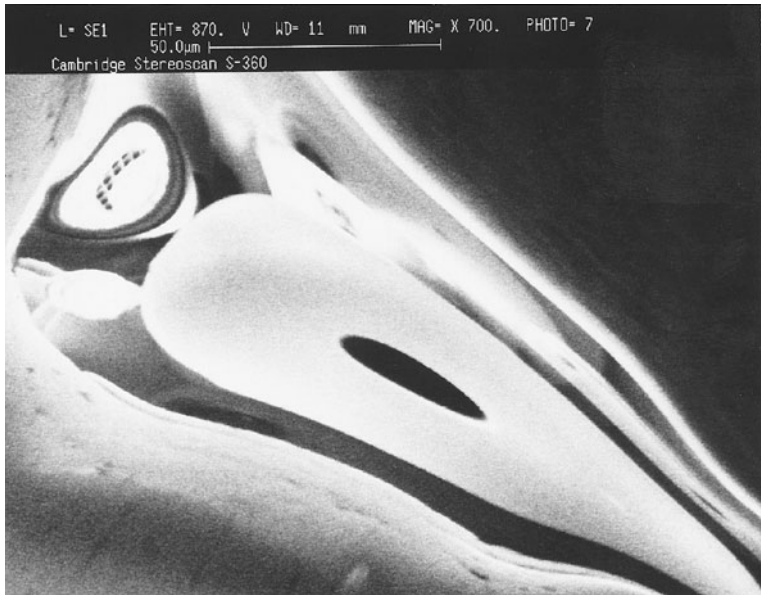
$$\sum i_{in} = \sum i_{out} \qquad (15.1)$$
$$i_B = i_{BSE} + i_{SE} + i_{SC}\,,$$

where i_{SC} is the specimen (or absorbed) current. For the example of copper, $i_{SC} = 0.6 i_B$. Thus, even with a conducting specimen such as a metal, an electrical connection must be established to conduct this substantial current from the specimen to ground (typically the specimen stage is well-grounded). Because all materials (except superconductors) have the property of electrical resistivity ρ, the specimen has a resistance R ($R = \rho\, l/A$, where l is the length of the specimen and A is the cross section). The passage of the specimen current i_{SC} through this resistance will cause a potential drop across the specimen, $V = i_{SC} R$. For a metal, ρ is typically of the order of 10^{-6} ohm-cm, so that a specimen 1 cm thick with a cross-sectional area of 1 cm^2 will have a resistance of 10^{-6} ohm, and a beam current of 1 nA (10^{-9} A) will cause a negligible potential of about 10^{-15} V to develop across the specimen. For a high-purity (undoped) semiconductor such as silicon or germanium, ρ is approximately 10^6 ohm-cm, and the

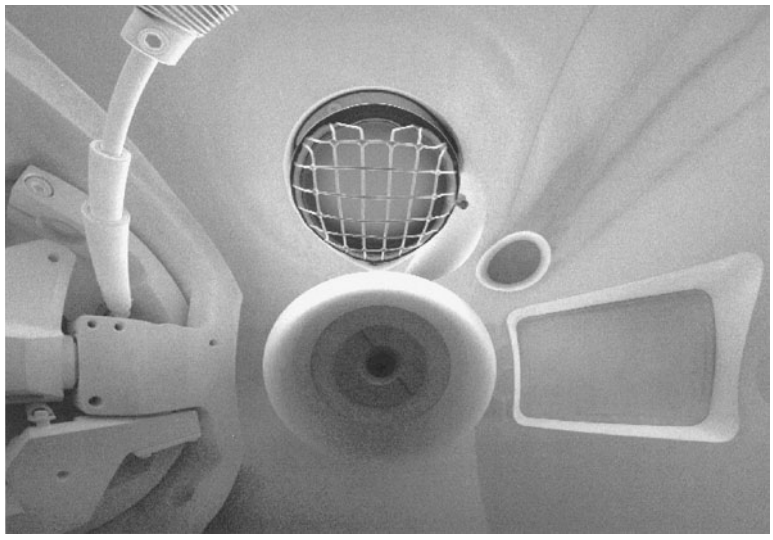
1-nA beam will cause a potential of 1 mV (10^{-3} V) to develop across the 1-cm^3 specimen, which is still negligible.

It must be noted that if the path from the specimen surface to ground is broken, the conditions for charge balance in Eq. (15.1) cannot be established. The electrons injected into the specimen by the beam will then accumulate, and the specimen will develop a high negative electrical charge relative to ground. Except for the ground connection, the specimen is otherwise well-insulated from the grounded stage. With this ground connection broken, even a good conductor such as a metal will accumulate sufficient charge to equal that of the incident beam (up to about 20 keV), with the result that the specimen will actually act like an electron mirror. The beam will be reflected before reaching the surface, so that it actually scans the inside of the specimen chamber. As shown in Fig. 15.1a, a somewhat distorted ("funhouse mirror") view of the inside of the chamber can be obtained, due in this case to the irregular electrical fields developed on the rough surface of a metal fracture. With a large, flat metal or semiconductor wafer, the electrical field from charging becomes more regular, and the resulting mirror reflection SEM image can be of sufficient quality to actually be a valuable diagnostic, such as assessing the functioning of various detectors, as shown in Fig. 15.1b.

The flow of the specimen current to ground becomes a critical problem when dealing with nonconducting (insulating) specimens. Insulating specimens include a very wide variety of materials such as plastics, polymers, elastomers, minerals, rocks, glasses, and ceramics, which may be encountered as bulk solids, porous solids or foams, particles, or fibers. Virtually all biological specimens become nonconducting when water is removed by drying, substitution with low-vapor-pressure polymers, or frozen in place. For an insulator such as an oxide, the electrical resistivity ρ is very high, 10^{12}–10^{16} ohm-cm, so that a 1-nA beam passing through the specimen would produce potentials of 1000 V or more between the beam location on the specimen surface and the grounded specimen stub. In fact the high resistivities of insulators prevent the smooth motion of the electrons injected by the beam through the specimen to ground, so that electrons accumulate in the immediate vicinity of the beam impact, raising the local potential. Such high surface potentials can severely degrade SEM images. The resulting diverse set of phenomena is referred to as "charging." Charging phenomena are incompletely understood and are usually found to be dynamic with time, a result of the time-dependent motion of the beam due to scanning action and the electrical breakdown properties of materials. The insulating specimen acts as a capacitor, so that placing the beam at a pixel causes a charge to build up with an *RC* time constant as a function of the dwell time, followed by a decay of that charge when the beam moves away. Depending on the exact material properties, especially the surface resistivity, which is often much lower than the bulk resistivity, and the beam conditions (beam energy, current, and scan rate), the injected charge may partially or even completely dissipate before the beam returns in the scan cycle. This process leads to strong effects in SEM images. A time-dependent charging situation is shown schematically in Fig. 15.2, where the surface potential at a particular



a



b

Figure 15.1. (a) Example of extreme specimen charging, in which the beam is reflected off the specimen and scans the inside of the specimen chamber, producing a highly distorted "fish-eye-lens" view of the polepiece, the Faraday cage of the E–T detector, and the chamber walls. $E_0 = 870$ V. (b) A view of the inside of the chamber of an FEI-Philips XL-30 SEM obtained using a highly charged uncoated glass sphere as a specimen. The charged sample acts as an electron mirror and enables one to see and identify the various components in the chamber. The final lens opening is at the center of the picture, with the secondary electron collector at the top. The collimator of the EDS x-ray detector is at 2 o'clock and a sealed port at 3 o'clock. Part of the stage control mechanism is at the left-hand side of the image. Although this highly distorted image is a consequence of severe charging, it can be used to check the position and connections of the various devices in the microscope chamber. (Courtesy of Tony Burgess, Multi-Imaging Centre, University of Cambridge.)

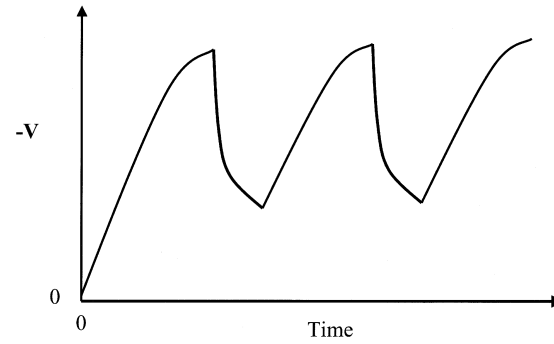


Figure 15.2. Schematic of surface potential-V (charging behavior) exhibited at a particular pixel of a insulating surface as the beam charges the region, moves away, so that discharging proceeds according to the local *RC* time constant, and then returns to begin the process again.

pixel accumulates with the dwell time and then decays until the beam returns. In more extreme behavior, the accumulated charge may cause local electrical breakdown and abruptly discharge to ground.

15.2. Recognizing Charging Phenomena

Charging phenomena cover a wide range of observed behaviors in images from practical specimens. Some of these behaviors are slight imperfections, such as unexpected brightening/darkening in images prepared with detectors that collect secondary electrons, to extreme cases, in which the incident beam is actually displaced by the charge on the specimen, resulting in complete degradation of the SEM image. Charging can also be recognized in x-ray spectra due to unexpected lowering of the Duane–Hunt limit of the x-ray continuum.

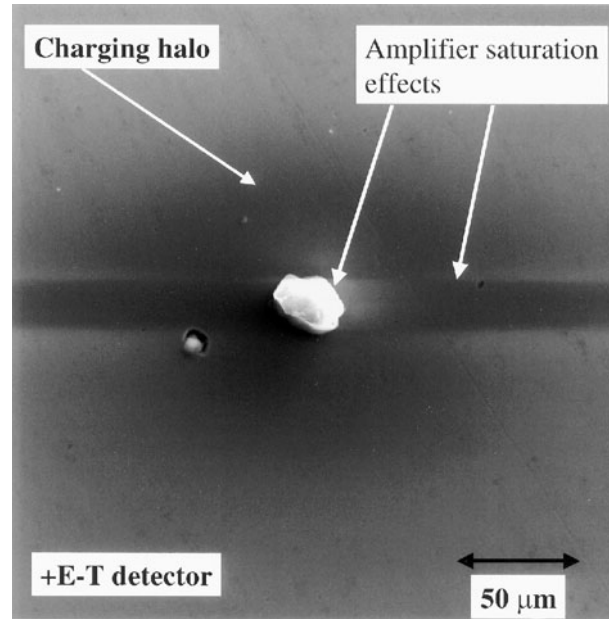
To understand the basic charging behavior, consider Fig. 3.19, which shows the behavior of the processes of backscattering and secondary electron emission as a function of beam energy. For beam energies above 5 keV, generally $\eta + \delta < 1$, so more electrons are injected into the specimen by the beam than leave as BSE and SE. The accumulation of electrons in the insulator causes the region under the beam to charge negatively. The local buildup of negative charge acts to repel the subsequent incoming beam electrons while the beam remains at that pixel. This process lowers the effective kinetic energy with which the beam strikes the surface. For most materials, especially insulators, as the beam energy is lowered, the total emission of BSE and SE increases, eventually reaching the upper crossover energy E_2, where $\eta + \delta = 1$. For beam energies below this value, the emission of S*E* can actually reach very large values for insulators, $\delta_{\max} =$ 2–20. When negative charging sufficiently lowers the effective arrival energy of the electrons into the region below E_2, the SE emission becomes very high. Such charge-enhanced secondary emission will be seen as a locally bright region in an image obtained with an SE detector, such as

the E–T detector. This situation is likely to be highly dynamic because the enhanced SE emission promptly depletes the local negative charge.

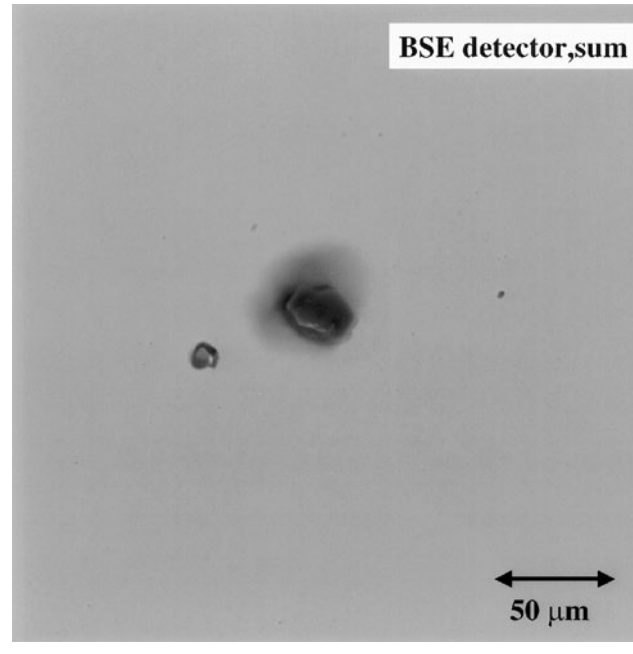
Secondary electrons are emitted with such low energies, $E_{SE} < 50$ eV, with most less than 10 eV, that local potential due to charging can have a large effect on SE collection by an external detector. The E–T detector typically imposes a potential of about +300 V on the Faraday cage at a distance of about 3 cm from the specimen, producing a field of 300 V/3 cm or 100 V/cm. SE emitted from a conducting specimen are strongly attracted to follow the field lines from the grounded specimen surface to the positively biased Faraday cage grid, and thus into the E–T detector. When an insulating specimen develops a negative charge of even tens of volts under the beam, the local field from the charged region to nearby uncharged areas of the specimen a few micrometers away or to the grounded stub a few millimeters away is likely to be much stronger (10^3–10^5 V/cm) than the field imposed by the E–T detector. Thus SE may be attracted back to the specimen surface or to the stub, lowering the effective efficiency of the E–T detector, so that such a region appears dark. Thus, the typical appearance of an isolated insulating particle undergoing charging on a conducting surface is a bright, often saturated signal (possibly accompanied by amplifier overloading effects due to signal saturation). This signal is surrounded by a dark halo that extends over surrounding conducting portions of the specimen where the local field induced by the charging causes the SE to be recollected. This type of voltage contrast must be regarded as an artifact because it interferes with and overwhelms the regular behavior of SE emission with local surface inclination that we depend upon for sensible image interpretation of specimen topography with the E–T detector. An example of this phenomenon is shown in Fig. 15.3a for an isolated insulating iron oxide inclusion in a conducting steel matrix. The local charging is so severe that our ability to discern the topography of the inclusion is almost completely eliminated, and the dark halo degrades the image of the surrounding conducting steel.

Although SE are very vulnerable to charging because of their low kinetic energy, so that even a few volts of local charging can modify an E–T image, energetic BSE from an incident beam energy with $E_0 > 5$ keV are nearly unperturbed by these small, local electric fields. Thus, shifting from the E–T detector with its large SE component to a dedicated BSE detector image can often reduce or even eliminate the effects of low-level charging on the SEM image. The charging situation encountered with the E–T image in Fig. 15.3a is eliminated in Fig. 15.3b, prepared with a dedicated BSE detector (semiconductor detector, sum mode). The topography of the inclusion can be seen with the BSE signal, along with details in the region hidden by the halo in E–T image of Fig. 15.3a.

When the specimen is insulating over its entire surface, charging can present more difficult problems. Severe charging effects can be seen in Fig. 15.4a and 15.4b, where a bare (uncoated) surface of a natural calcite ($CaCO_3$) crystal is observed with incident beam energies of 15 and 5 keV. In both images harsh white–black contrast dominates, with local bright areas due most likely to enhanced SE emission and dark areas due to depressed



a



b

Figure 15.3. Appearance of an isolated insulating particle (iron oxide inclusion) in a large conducting specimen (steel). (a) E–T detector image, $E_0 = 20$ keV. Note the extreme brightness of the particle, which destroys any capability of recognizing the particle topography, and the surrounding dark halo, where the image of the conducting steel specimen is compromised. (b) Same area imaged with a dedicated BSE detector (semiconductor detector, sum mode). The charging artifacts are eliminated, permitting visualization of the topography of the inclusion and the surrounding area of the halo.

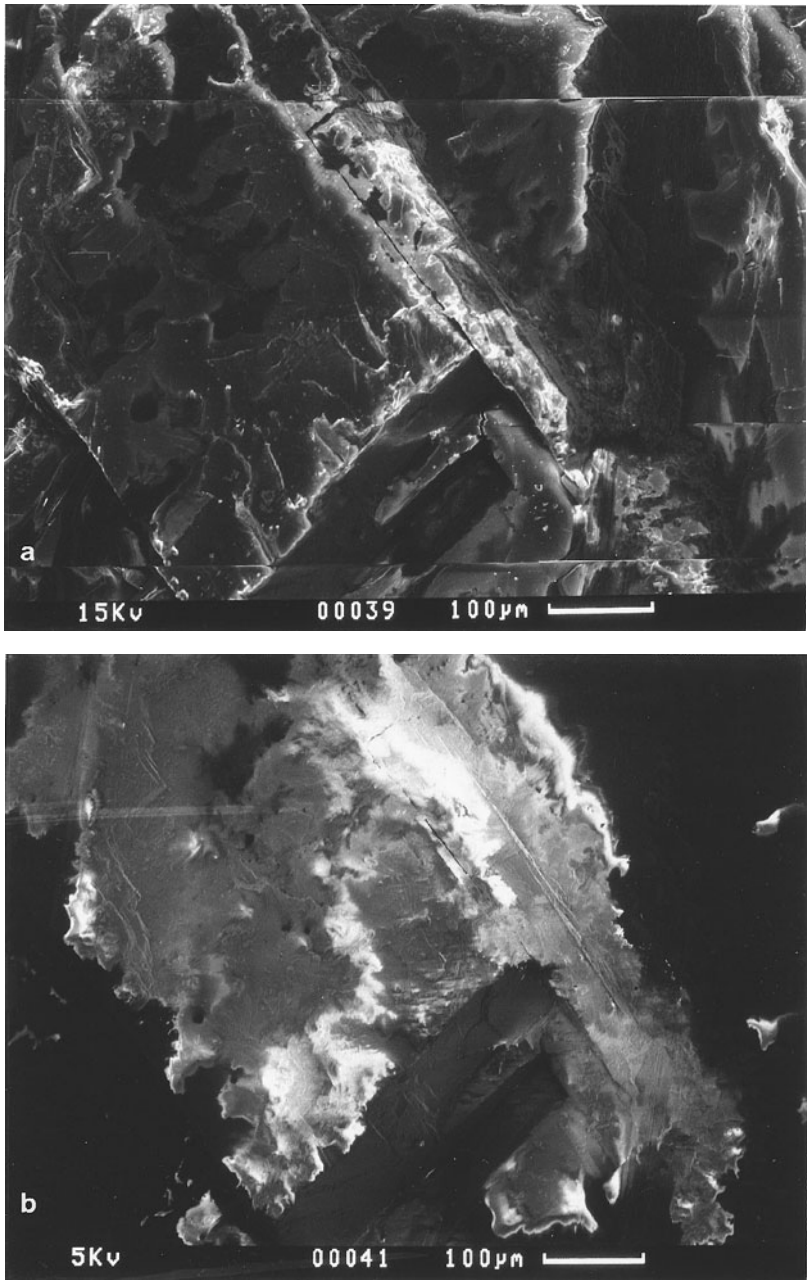


Figure 15.4. Charging observed during SEM imaging of an uncoated calcite ($CaCO_3$) crystal with an E–T detector. (a) $E_0 = 15$ keV; note the bright/dark regions and the scan discontinuities. The true shape of the crystal cannot be discerned. (b) $E_0 = 5$ keV; the scanned image is now stable for the particular selection of beam current and scan time so that no severe discontinuities are observed, but the dark/bright voltage contrast dominates the true image of topography. (c) $E_0 = 1.5$ keV; operation at the E_2 crossover point (for the particular composition and local tilt) produces a stable image in which topographic contrast showing the true features of the specimen is evident. The bright area on the tilted top surface shows that not all regions of charging can be simultaneously controlled.

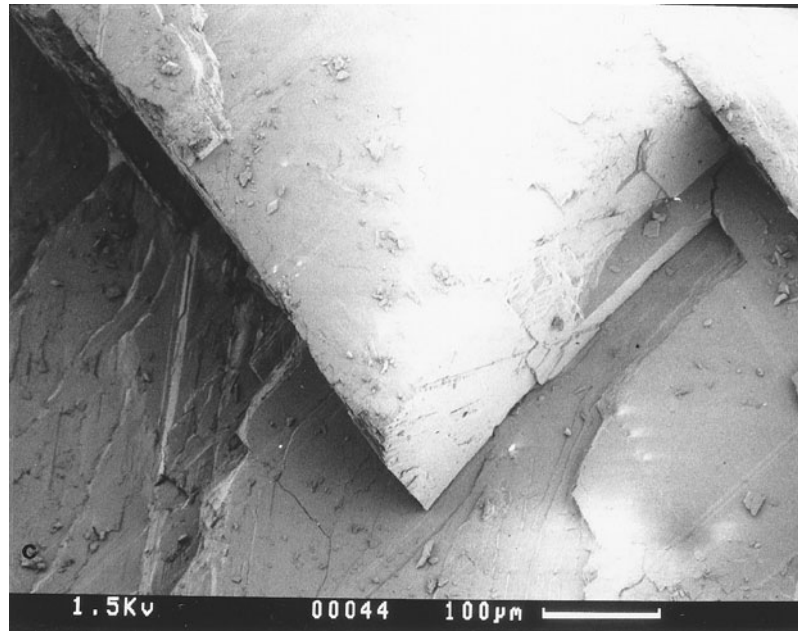


Figure 15.4. (*Continued*)

SE collection from local fields. The actual shape of the specimen is almost completely hidden in Fig. 15.4a and 15.4b. Extreme charging leading to actual beam displacement can be seen in Fig. 15.4a, where a discontinuity ("tear") in the scan occurred near the top of the field of view. BSE images obtained under these conditions were not useful.

Charging effects in images, especially E–T images with a strong SE component, can also be recognized by their dynamic nature. If charging is suspected, the SEM image should be compared over a wide range of scan speeds, for example, rapid TV-rate scanning versus slow scanning for recording a photomicrograph or a long-dwell-time digital image. An image that appears to be quite stable with reasonable, interpretable contrast can often be obtained at TV scan rates, where the beam returns to each pixel so frequently ($\sim$24 times per second) that the charging/discharging RC situation remains relatively constant. At a slow scan rate, the image may appear to be completely changed because of the large change in dynamic charging. At slow scan rates, the detector and even the beam become more vulnerable because of large swings in the charging situation at each pixel. TV-rate scanning is so useful in dealing with charging, especially low-level charging in E–T images, that it has become common to include a rapid SEM image capture system based upon a digital framestore. A single TV-rate frame usually has inadequate signal-to-noise (S/N) ratio. The framestore can record a user-defined number of TV rate scans and use pixel intensity averaging over this set of images to achieve a satisfactory signal-to-noise ratio while avoiding the slow scan rates that are much more vulnerable to dynamic charging.

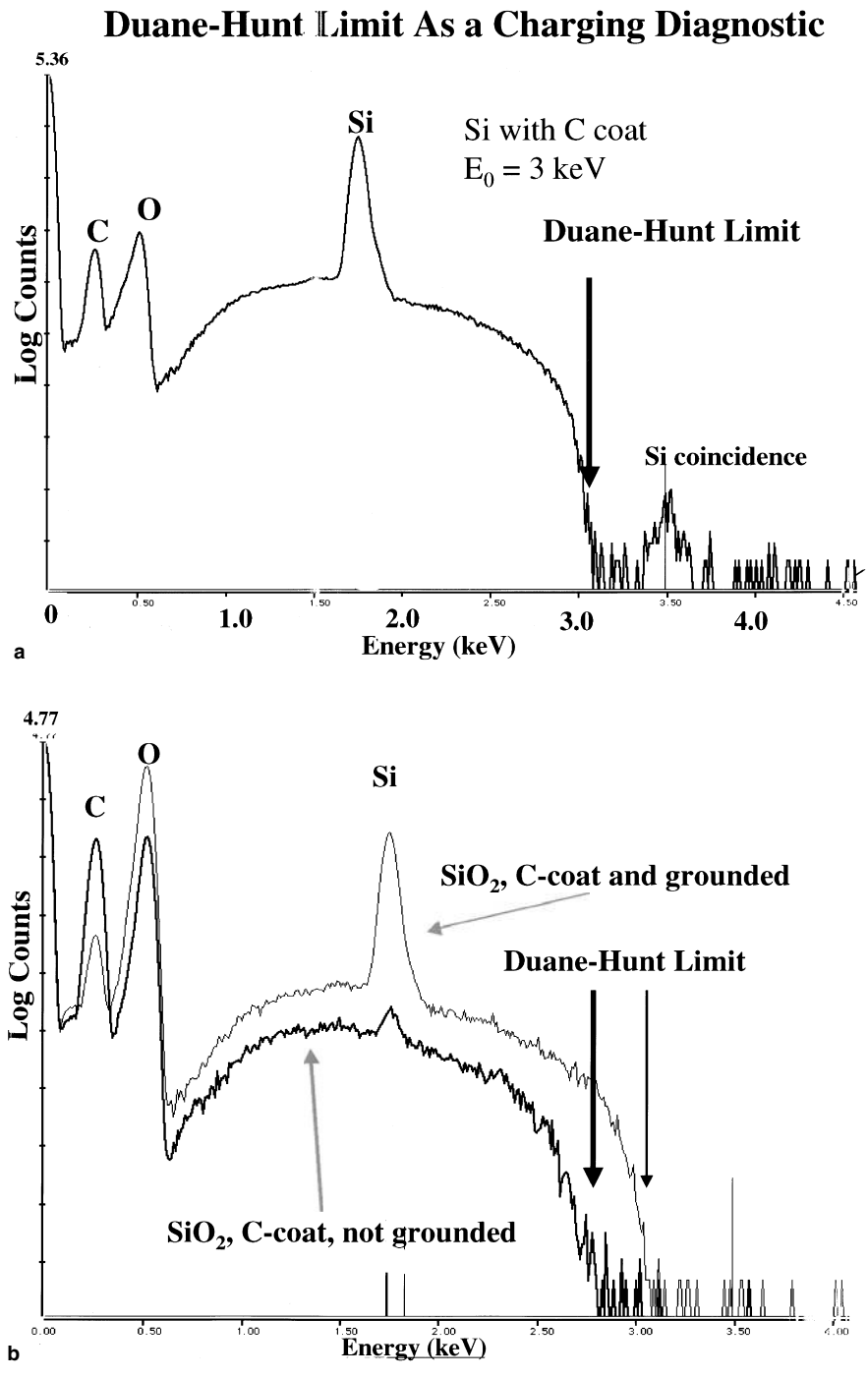


Figure 15.5. Detecting charging with an EDS x-ray spectrum: (a) Silicon coated with 3 nm of carbon, $E_0 = 3$ keV, showing a Duane–Hunt limit of 3 keV; (b) SiO_2 coated with 3 nm of carbon with the top surface properly grounded, $E_0 = 3$ keV, showing a Duane–Hunt limit of 3 keV (upper, thin trace); same specimen, but with ground connection broken (lower, thick trace), showing charging of about 500 V, as evidenced by the decrease in the Duane–Hunt limit.

Charging effects in E–T images of uncoated specimens can also be minimized by operating with E_0 selected to correspond to the E_2 upper crossover point in Fig. 3.18. The emergence of SEMs designed to operate effectively in the low-voltage region, $E_0 \leq 5$ keV, with either FEG or conventional electron sources and capable of changing the accelerating voltage by small increments, 100 V or less, makes it practical to search for the E_2. The E_2 value depends on material, the local specimen tilt (which modifies backscattering and secondary emission), and possibly the dose rate. When E_2 is found, and the procedure may require modifying more variables such as tilt in addition to E_0, charge balance is achieved and an image with minimal charge contributions is obtained, enabling the microscopist to directly observe the specimen topography, as shown in Fig. 15.4c. Note that in determining E_2, it is best to start at a low energy and work upward. If a high energy, $E_0 > 10$ keV, is chosen to start the procedure, charge may be injected so deeply into the insulator that it cannot be reached and dissipated with the beam set to the E_2 energy because the range is reduced at low beam energy.

The Duane–Hunt energy limit of the x-ray continuum (Chapter 6) provides a direct way of monitoring the charging condition of a specimen. Continuum photons of all energies up to the incident beam energy are formed. Examination of the upper limit with a calibrated EDS provides proof of the highest energy with which beam electrons enter the specimen. The true Duane–Hunt limit should first be measured with a conducting metal or semiconductor specimen. Note that because of pulse coincidence events, there will always be a few photons measured above the true Duane–Hunt limit. The true limit can be estimated with good accuracy by fitting the upper energy range of the continuum, preferably several keV in width and without characteristic peaks, and finding where it intersects zero intensity, as shown in Fig. 15.5a for a silicon target with $E_0 = 3$ keV. If the specimen charges, the charge in the specimen acts to decelerate (for negative charge) or accelerate (for positive charge) the incoming beam electrons, modifying the measured Duane–Hunt limit. This effect is shown for negative charging in Fig. 15.5b for a silicon dioxide surface coated with carbon and grounded properly, which shows the same 3-keV Duane–Hunt limit as the silicon. When the grounding connection is removed from the coated silicon dioxide, the material immediately charges by about 500 eV, as evidenced by the decrease in the Duane–Hunt limit seen in the lower spectrum in Fig. 15.5b. Note also the reduction in intensity throughout the spectrum measured under charging conditions.

15.3. Procedures for Overcoming the Problems of Charging

The bulk conductivity of the whole sample can be increased by soaking the sample in solutions of conductive salts using procedures discussed in Chapter 13, Section 13.3.5. This approach is limited to obtaining low-resolution BSE images of soft biological material. The other option is to cover the surface of the sample with a very thin film of a highly conductive

material. Both approaches are invasive in varying degrees and a judgement has to be made whether the procedure adopted will compromise, distort, or obliterate the information that it is hoped to obtain from the sample.

The best and simplest way to overcome charging problems is to deposit a thin metal layer on the surface of the sample. The prime function of this layer must be to increase the surface electrical conductivity of the sample with the additional attributes of increasing the thermal conductivity and in some cases the BSE and SE signal from the sample. Methods such as electroplating and anodization are generally impracticable because they rely on chemical solutions at high or low pH. Glow discharge coating is not a satisfactory method for applying a thin layer of carbon.

A thin metal film is defined as being between 1 and 10 nm thick. There are two procedures for applying such a coating layer under vacuum. In the evaporation method, a conductive metal is heated to its vaporization temperature in a high vacuum and the evaporated metal atoms land on the specimen surface. In the sputtering method a negatively charged conductive metal target is bombarded by positive ions in a low-vacuum gas discharge and the eroded metal atoms land on the specimen surface. These two techniques and their derivatives will be discussed separately (Sections 15.4 and 15.5).

15.4. Vacuum Evaporation Coating

Most metals, when heated in a vacuum, will evaporate once their temperature is raised to the point at which their vapor pressure exceeds 1 Pa ($\sim 10^{-2}$ torr). Many of the materials that are used for evaporative coating have a high melting point and need special heating processes. Refractory metals such as tungsten, tantalum, and iridium must be heated in a refractory crucible by means of an electron beam. Lower melting point materials, such as the noble metals, are placed either in a refractory metal container or draped over a refractory wire, which is resistance-heated. In all cases, the material is evaporated unidirectionally in the same way as a light source illuminates an object. Important Properties of selected coating materials are given in Table 10 of the Database in the Enhancement CD.

A great deal is known about how thin films are formed and this information is useful when we come to consider the way samples are coated. Figure 15.6 shows the first steps in the formation of a thin film. The first atoms to arrive on the sample drift and diffuse on the surface, losing energy, and eventually reach a critical nucleation point where they begin to bind to each other and to the sample surface. The coating procedure should aim for many small nucleation centers that will quickly grow laterally, first to form islands and then a continuous film with the minimal amount of discontinuous vertical growth. The sequence of images in Fig. 15.7 shows the changes in appearance during the growth of a thin film of silver. The variables that affect thin-film formation include the binding and the impact energy of the target material, the minimum temperature needed for evaporation, the rate of deposition, the substrate temperature, the substrate–evaporant binding energy, and the sample topography. Further details on thin-film formation can be found in Maisell and Glang (1970).

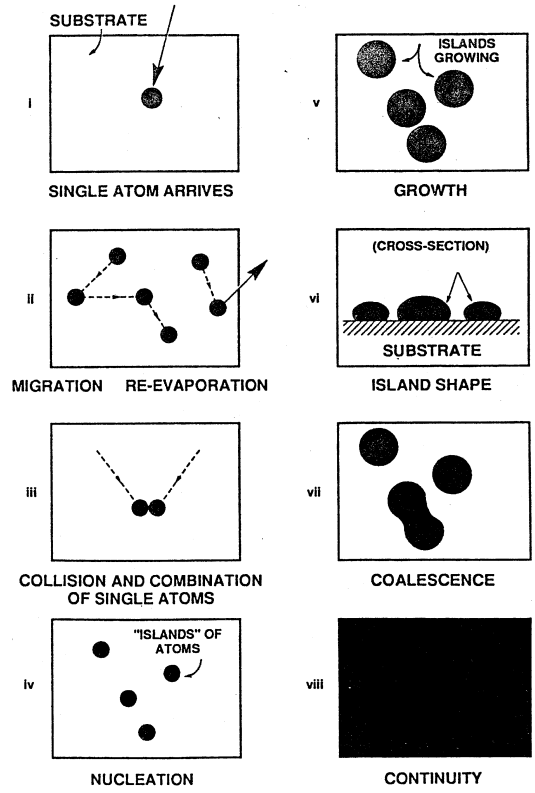


Figure 15.6. Steps in the formation of a thin film.

15.4.1. High-Vacuum Evaporation Methods

High-vacuum evaporation is one of the more popular ways of applying thin-metal and carbon films as coating layers on specimens. Figure 15.8 shows the main features of a typical coating unit. For more details, including experimental protocols, see Enhancement 15.1 on the accompanying CD.

The noble metals and their alloys are highly conductive and are suitable as a coating material for most samples to be imaged using secondary electrons. A gold:palladium 4:1 alloy is preferable to pure gold because the latter has larger particles when viewed at high magnifications. The advantage of using an alloy rather than a pure metal is that as the two different atom species land on the specimen surface, the presence of one discourages epitaxial growth of the other and thus increases the probability of obtaining a near-featureless thin film. If the sample is to be imaged using backscattered electrons, high-density metals such as iridium, platinum, and tungsten may be considered, although their conductivity is not as good as the noble metals. Carbon, although a relatively poor conductor, is one of the best coating materials for studies involving x-ray microanalysis, provided one of the elements of interest is not carbon. For specialized imaging at high resolution, it is best to use a platinum–carbon or a

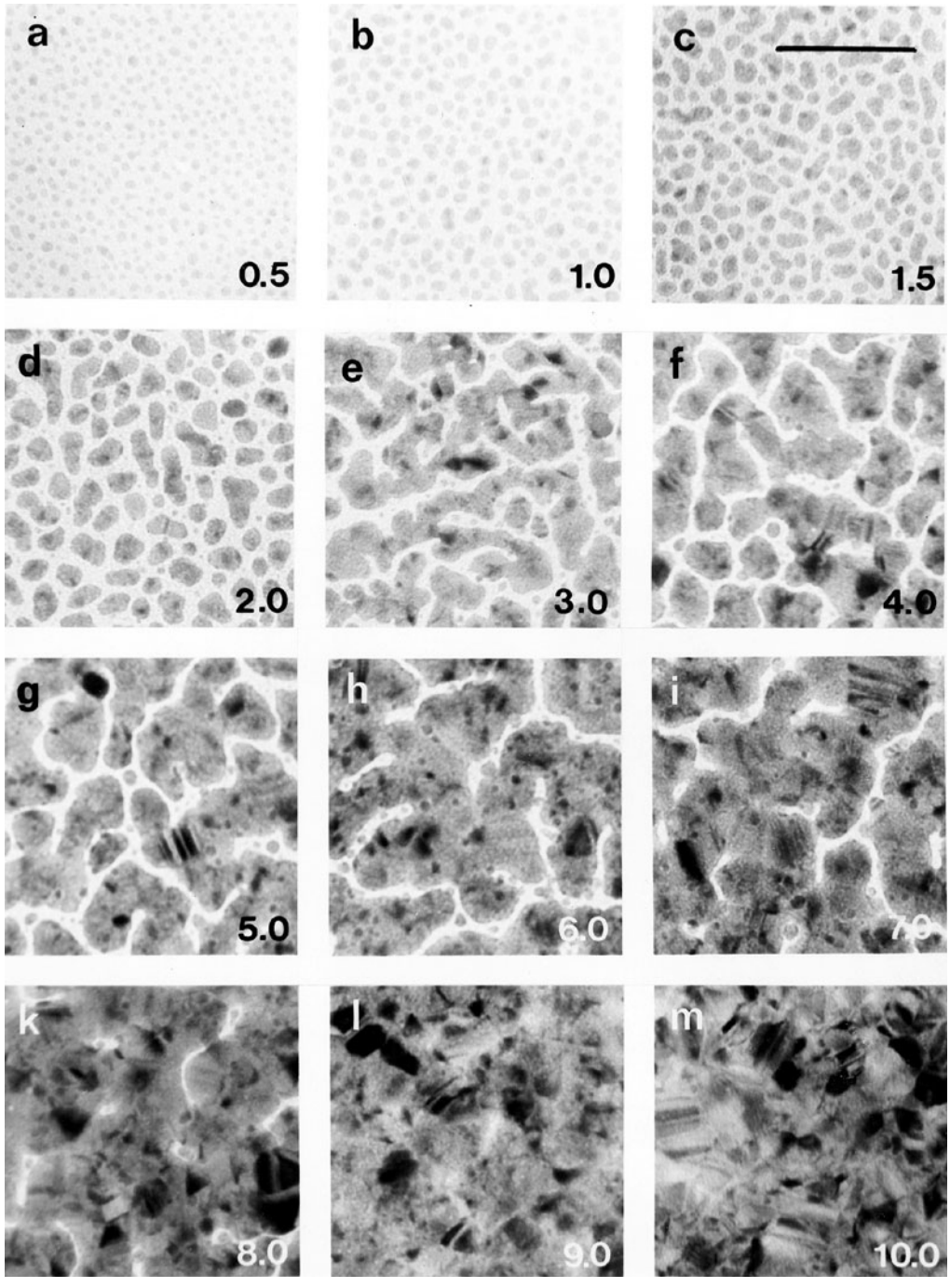


Figure 15.7. A series of high-resolution TEM images showing the stages in the growth of a silver film at 293 K. Films of increasing average mass-thickness grow in several distinct phases. After nucleation, the crystals coalesce laterally (a–e), and then smaller crystals grow on top of the existing crystals (f–h) and bridge gaps (i–l) until the film is optically continuous (m). The number in the lower right-hand corner of each image is the film thickness in nanometers. This sequence of images shows that the 3- to 5-nm films, although probably discontinuous, are satisfactory for a modern SEM operating at 5 keV. Magnification marker in (c) = 10 nm. (From Peters, 1984.)

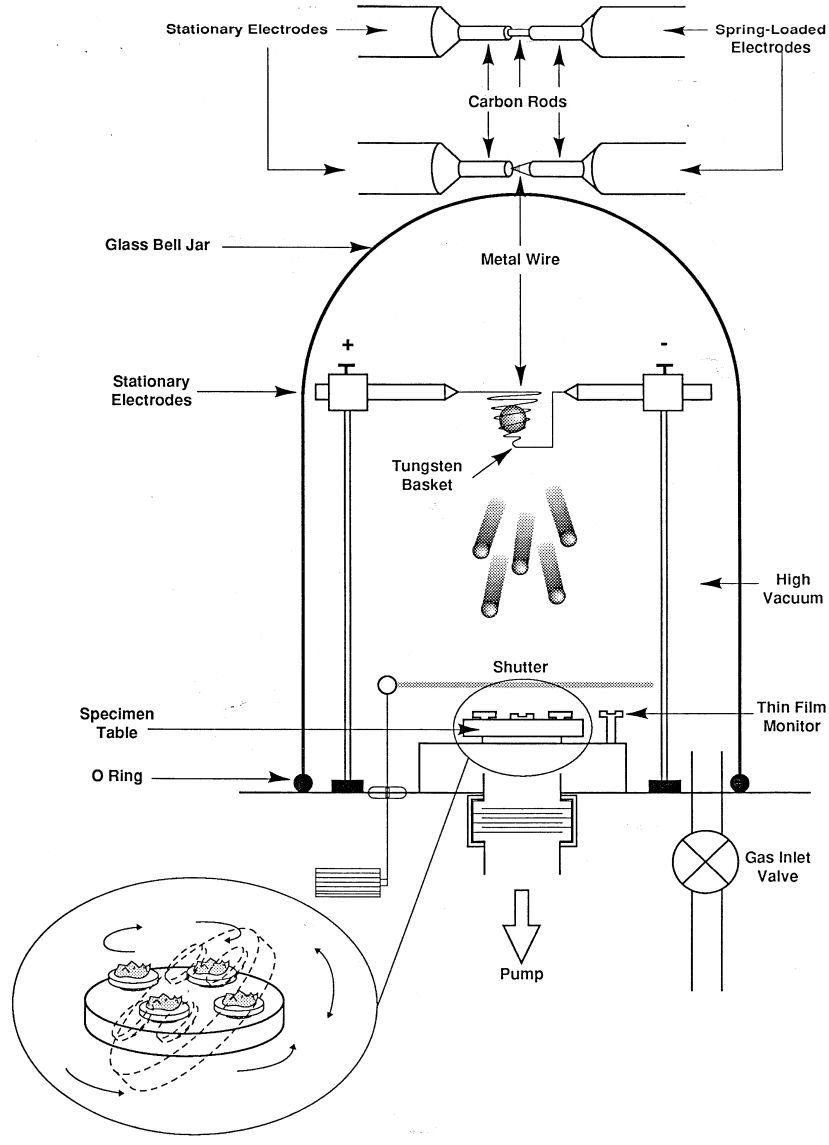


Figure 15.8. Diagram of a high-vacuum evaporative coater.

platinum–iridium–carbon mixture as the source material because the resultant coating layers are virtually devoid of any structure when viewed at high magnification (Wepf *et al.,* 1991). Table 10 in the database given in the Enhancements CD should be consulted for some of the more important properties of selected coating materials.

It is possible to evaporate a wide range of metallic target materials and, although the procedure is more time-consuming than sputter coating, it can be used as a routine procedure for coating samples for both SEM and

x-ray microanalysis. It is an excellent method for producing the very thin coating layers with small grain size needed for high-resolution studies.

Evaporative coating alloys, however, do not evaporate congruently. The lack of congruency is probably not a serious problem with gold–palladium alloys because the two evaporation temperatures are only 100 deg apart. The evaporation temperatures of the iridium–platinum–carbon mixture differ by 600 deg, which may cause separation. The evaporant travels in straight lines and may not produce even coatings on rough surfaces. This problem can be overcome by tilting and rotating the sample in a planetary motion during the coating process. The high temperatures needed to evaporate most metals may also cause thermal damage to the sample. This problem can be overcome either by placing a rapidly rotating shutter between the source and the sample or pulsing the power to the evaporation source.

15.4.2. Low-Vacuum Evaporation Methods

Although vacuum evaporation coating methods customarily operate at a pressure of between 10^{-3} and 10^{-5} Pa ($\sim 10^{-5}$–10^{-7} torr), it is possible to evaporate carbon at 1 Pa ($\sim 10^{-2}$ torr) to produce carbon substrates and carbon thin films. The carbon electrodes are assembled or carbon string is fitted between the two electrodes and the samples are placed on the specimen table. The system is first pumped to a high vacuum of 10^{-3}–10^{-4} Pa ($\sim 10^{-5}$–10^{-6} torr), backfilled several times with dry argon or nitrogen, and then pumped down to 1 Pa ($\sim 10^{-2}$ torr) and maintained at this pressure by adjusting the dry inert gas intake valve. The electrodes should be first heated to a cherry-red color to allow any contaminants to outgas. The carbon is then quickly evaporated resistively while the specimen rotates and tilts on the specimen table. The carbon atoms have a short mean free path in the poor vacuum and as they collide with the inert gas atoms, they scatter and provide an effective multidirectional coating.

This type of coating can also be carried out using a sputter coater (see Section 15.5), provided it is equipped with an accessory electrode assembly. Because most sputter coaters only reach a pressure of 1 Pa ($\sim 10^{-2}$ torr), it is necessary to thoroughly flush the chamber with inert gas before the evaporation of carbon is carried out.

The advantage of this low-vacuum evaporation method is that it provides a good coating layer for highly sculptured samples and for samples that are to be studied by x-ray microanalysis. The disadvantages are that the procedure is prone to contamination and the high evaporation temperature of carbon (3500 K) may damage heat-sensitive samples. The low-vacuum evaporation method, however, will not produce high-resolution coatings.

15.5. Sputter Coating

Sputter coating is a popular and relatively simple way of applying a coating layer of metals and their alloys to nonconductive substrates. The target material is exposed to an energized gas plasma formed from a heavy

inert gas such as argon. The target surface is eroded by the plasma and atoms are ejected and collide with residual gas molecules. Target atoms have a short free mean path and provide multidirectional coating on a stationary specimen.

The four main factors that influence the rate of sputtering are voltage, plasma current, target material, and nature of the inert gas. The voltage influences the energy of the bombarding inert gas and the plasma current has a direct effect on the rate of sputtering. High sputtering rates are obtained at high plasma energy. The binding energy of the target material has an important influence on the rate at which it is erodes. For example, the binding energy of gold and palladium is relatively low and they will sputter much faster than tungsten, which has a high binding energy. The higher the atomic number of the bombarding inert gas, the faster it will erode the sample. Thus a xenon gas plasma gives a higher sputtering rate than argon, which in turn gives a higher rate than nitrogen.

All four factors are variable and may be altered for different coating applications. Although high sputtering rates shorten the time taken to coat a sample, there is an increased chance of thermal damage and an increase in the grain size within the sputtered film. There are several different ways to sputter coat materials, including the use of plasma magnetron coaters and ion beam and penning sputter coaters. These coaters are discussed in Sections 15.5.1 and 15.5.2.

15.5.1. Plasma Magnetron Sputter Coating

Figure 15.9 shows the main features of a modern plasma magnetron sputter coater. The thin-foil metal target is the cathode and the specimen is the anode. The chamber is evacuated and backfilled with a high-purity inert gas at low pressure, and a dc voltage of between 1000 and 2000 V is applied to form a gas plasma. A permanent rare earth magnet behind the target focuses the gas plasma onto the target to give a high sputtering rate and also deflects electrons from the sample. The samples can, if necessary, be placed on a Peltier-cooled table and most models have interchangeable targets. For more details, including experimental protocols, see Enhancement 15.2 on the accompanying CD.

Although a wide range of metal targets may be used in plasma magnetron sputtering, a target of a gold–palladium alloy is the best choice for routine SEM because epitaxial growth is limited. A pure gold target may be used for low-magnification studies. For medium- to high-resolution studies, a platinum target will give a satisfactory coating layer. The noble metals all have a relatively high sputtering rate at quite low plasma currents.

For high-resolution studies, in which the coating layers must have a negligible particle size, it is necessary to use platinum, chromium, or one of the refractory metals such as tungsten or tantalum. A careful study by Stokroos *et al.* (1998) has shown that gold–palladium coatings are a good alternative to chromium and platinum for the application of the FESEM to biological tissues because of its higher yield of secondary electrons. However, this finding is strongly dependent on the type of sputtering device

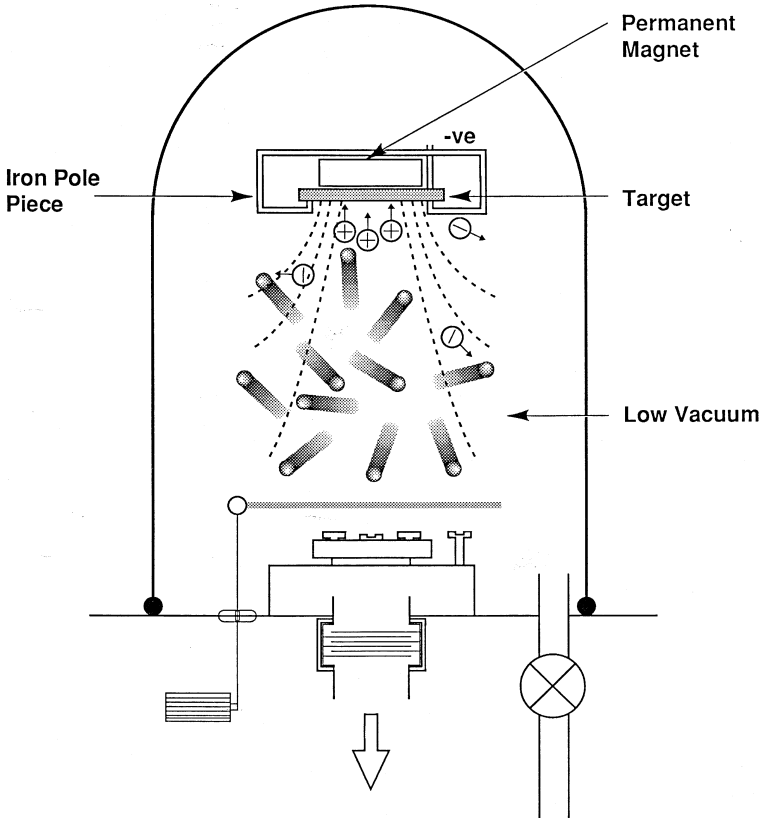


Figure 15.9. Diagram of a plasma-magnetron sputter coater.

employed. The sputtering rate of refractory metals is much slower than noble metals and it is necessary to use higher plasma currents and/or longer coating times, which increases the risk of damaging thermolabile samples. In these circumstances it is advisable to pulse the coating process.

It is possible to sputter both chromium and aluminum with a high-efficiency sputter head and very pure argon. Both metals quickly oxidize and a lot of energy is needed to break the oxygen–metal bond. The samples should first be shielded and the oxidized layer removed using high plasma current, after which the metal coating can proceed at relatively low plasma currents. Depending on the design, modern coaters will sputter Al and Cr, provided one uses very pure argon and the target material oxidation is minimized between coatings.

Carbon cannot be sputtered with the plasma magnetron coater. The carbon–carbon bond energy is high and it would be necessary to use high plasma energies to break these bonds. In addition, the sputtering would need to be carried out in a scrupulously anoxic environment.

The main advantage of plasma magnetron sputter coating is that this highly controllable and reproducible procedure can be completed in about 10 min. The coating is multidirectional and the target alloys are sputtered

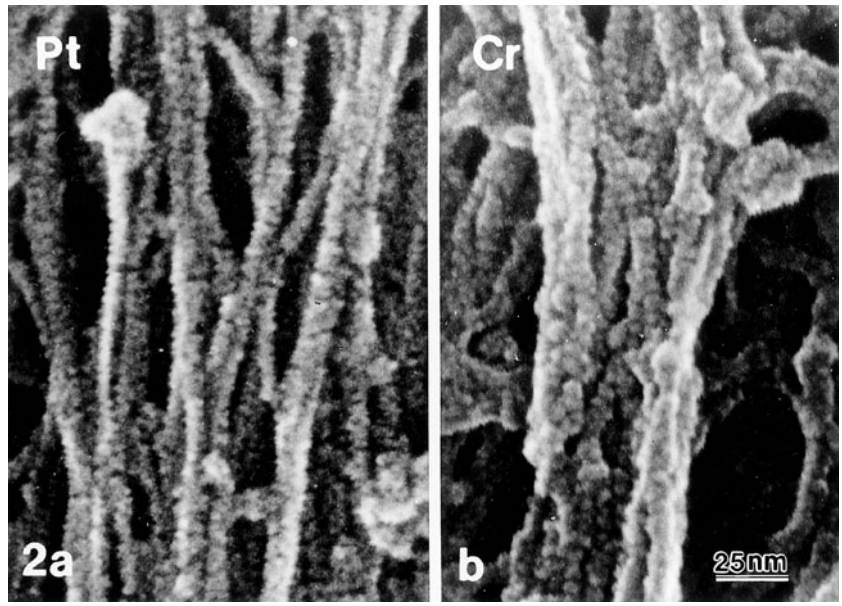


Figure 15.10. A comparison of platinum- and chromium-coated cytoskeleton from an extracted fibroblast. The surface regularities along the microfilaments in the platinum-coated material are partially obscured by the grain size of the coating material. This is not the case with the chromium-coated sample and the 5.5-nm repeat of actin subunits can be seen along individual bundles of microfilaments. Samples are freeze-dried and plasma-magnetron sputter-coated at 163 K with 1.5- to 2.0-nm mass thickness of either platinum or chromium. Magnification marker = 25 nm. (Chen *et al.,* 1995.)

congruently. As the metal atoms leave the target surface, they quickly collide either with the argon atoms or with each other. It has been estimated that metal atoms will suffer about five collisions between leaving the target and arriving on the sample surface. The result is that the metal atoms arrive at the sample surface from all directions and, with the exception of the most highly sculptured specimens, it is not necessary to rotate and tilt the sample during coating. The congruent sputtering of alloys such as gold–palladium virtually eliminates epitaxial growth of the metals and ensures a small particle size in the coating layer. Figure 15.10 shows a comparison of chromium and platinum as coatings for high-resolution images. It is now unusual to see etching and thermal damage on the sample surface and both evaporative and sputter coaters can produce thin coating layers with barely detectable particle size.

15.5.2. Ion Beam and Penning Sputtering

In both systems, the target is eroded by a high-energy argon plasma gas at low vacuum while the sample is simultaneously coated in a high-vacuum environment by means of differential apertures. Figure 15.11 shows the main features of this type of coater. The two methods differ only in the way the target is eroded. In the ion beam coater, the target is eroded by a

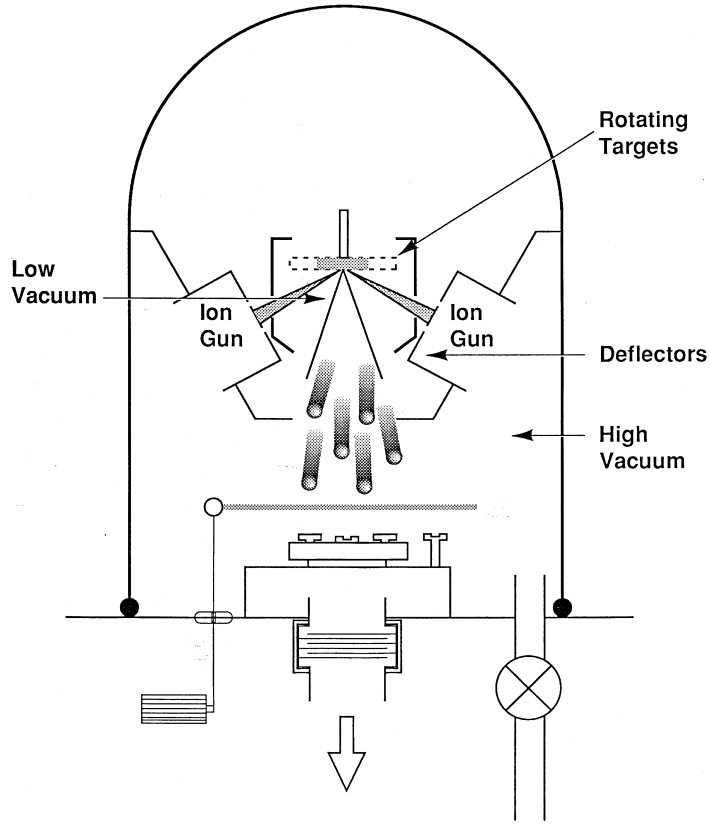
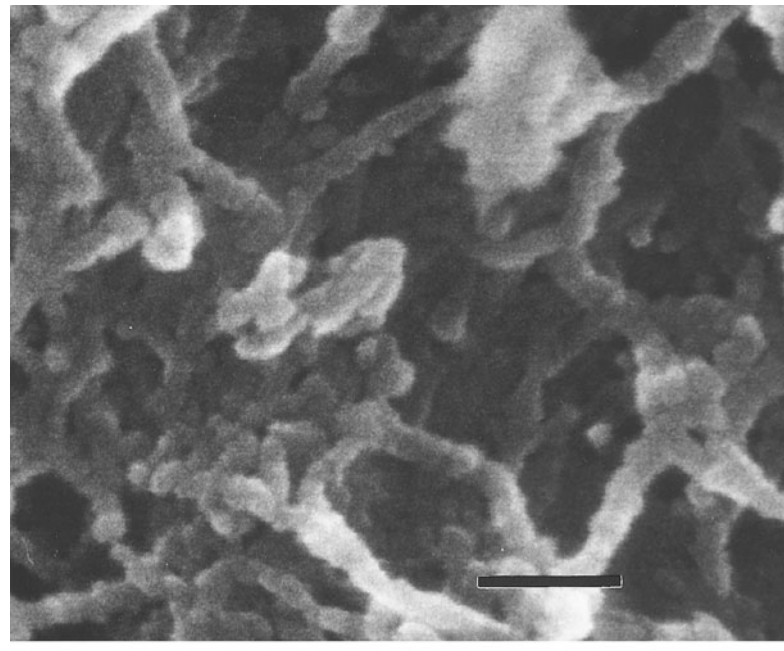


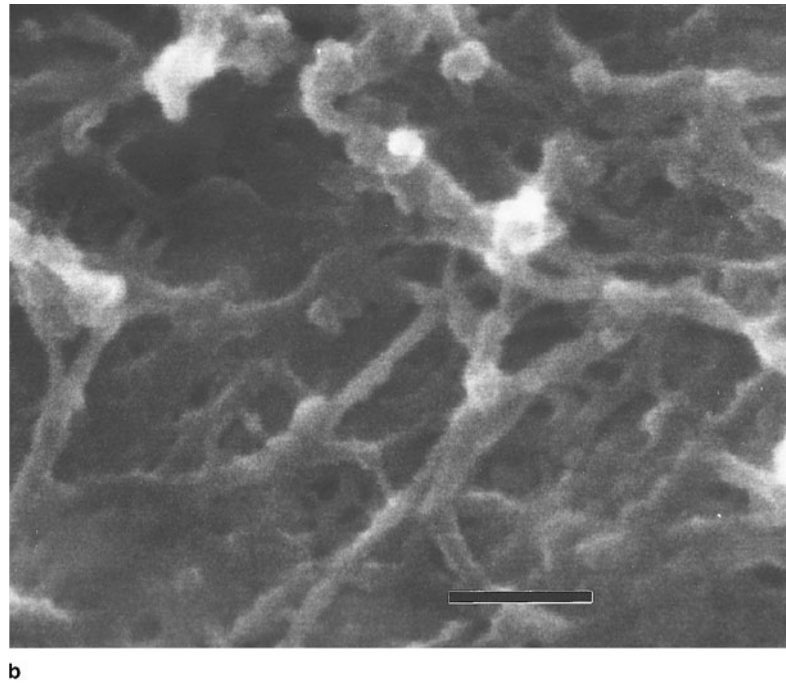
Figure 15.11. Diagram of an ion-beam sputter coater.

focused high-energy (8 keV) beam of argon ions set at right angles, which ensures that the specimen is not exposed to high-energy electrons or ions that may cause damage. In the penning coater, the target is placed inside the discharge area where the ions are generated and the specimen is shielded from the target by a small aperture. In both systems the inert gas plasma and target area are maintained at low vacuum 1 Pa ($\sim 10^{-2}$ torr) in order to maximize the target erosion rates and the sample is kept at high vacuum 10^{-4} Pa ($\sim 10^{-6}$ torr) to minimize in-flight nucleation of the target atoms.

The positive ions and electrons generated in the gas plasma are deflected by large electrodes, which are placed immediately outside the low-vacuum chamber. This process ensures that a virtually pure stream of metal atoms is directed onto the specimen. The mean free energy of the target atoms is an order of magnitude greater than that which can be obtained by thermal evaporation and this energy is not reduced by scattering with residual argon atoms. The primary energy of the target atoms coupled with a low deposition rate ensures that the metal atoms are deposited onto the substrate with little lateral movement and that these anchoring atoms act as nucleation centers from which the thin film is formed with minimum vertical



a



b

Figure 15.12. A comparison between (a) platinum-ion-beam sputter-coated and (b) osmium-plasma-coated samples. The fine filaments of biological material have been coated either with 6 nm of platinum or 3 nm of osmium. There appear to be no decoration artifacts on the osmium-coated sample. Magnification marker = 100 nm. (Osawa and Nozaka, 1998.)

growth. Because the mean free path of the target atoms is much longer than the target–substrate distance, the scattering advantage of conventional plasma magnetron sputter coating is lost and sculptured specimens must be rotated and tilted during coating.

The two images in Fig. 15.12 shows a comparison between a sample of biological material that has been ion-beam sputter-coated and osmium-plasma-coated. The coating process causes minimal heating and etching effects and a wide range of targets, including carbon, may be used a source material. The ion beam and penning sputtering equipment is rather expensive due to the need to have a very clean vacuum system.

15.6. High-Resolution Coating Methods

High-quality evaporation and plasma magnetron sputter coating methods are suitable for most samples, provided the correct target materials are used and the coating is carried out in a clean and orderly manner. The basic aim of all coating methods, and especially for high-resolution studies, is to ensure the thinnest possible continuous film over the sample surface with the smallest grain size. The granularity of the film limits the resolution, electrical conductivity, and mechanical properties of the thin metal layer. It is appropriate to consider briefly some of the factors that influence high-resolution imaging in the SEM.

The coating layer must be thinner than the range of the electrons that carry the signal, usually the SE_I. For this reason, the coating layer must be 1–2 nm thick and of sufficient density to provide image contrast and conductivity. The coating procedure should aim at producing a thin continuous layer with very small grain size. Large grain size will reduce resolution and thick films are unstable and will heat up, distort, and obscure surface features.

Lowering the sample temperature will ensure that much of the thermal energy of the arriving atoms is transferred immediately to the specimen, thus lowering the mobility of the atoms and causing them to stay at the point of impact. In addition, the vacuum system should be made as clean as possible by using an oil-free pumping system, and by strategically placing anticontamination devices and always backfilling the system to atmospheric pressure using a dry inert gas. The paper by Woodward and Zasadinski (1996) contains information that may prove useful when designing a high-resolution coating protocol.

Evaporation coaters can be adapted to take advantage of the low grazing incidence and rotary shadowing methods used in TEM, which use target materials of platinum, tungsten, and a mixture of platinum–iridium–carbon to obtain 1-nm layers. The technique devised by Hermann *et al.* (1988) involves moving an electron-gun-heated evaporation source of chromium or germanium back and forth through 90° during coating while maintaining the sample at 190 K and rotating it at 115 rpm. It is possible to obtain continuous thin films of between 0.5 and 1.0 nm on flat surfaces. Another simple procedure is to immediately apply a thin (1 nm) coating of carbon

after coating the sample with a thin layer of metal in order to reduce recrystallization in the thin metal film. This procedure is also useful when using chromium and some refractory metals that oxidize very quickly; for example, 60% of a chromium film may be chromium oxide.

For the very finest coating layers, ion beam and penning sputtering give the best results. This type of sputtering prevents metal atoms that travel from the target to the sample from aggregating, and they arrive at the surface as single atoms. In contrast, as plasma-magnetron-sputtered and -evaporated metal atoms travel from the target to the sample, they will aggregate and arrive at the surface as very small groups of atoms. For very high resolution imaging, this aggregation process may be sufficient to degrade resolution.

Tanaka (1994) has shown that when subliming osmium tetroxide vapor is introduced into the negative glow phase of a sputter coater, the molecules are broken down and the positively charged osmium atoms land on the specimen to give a very fine conducting layer of osmium metal. Great care has to be taken when using this coating procedure because of the toxicity of osmium tetroxide. A specially designed osmium coater is now commercially available and includes the appropriate features to enable osmium tetroxide to be handled safely. Figure 15.12b (from Osawa and Nozaka, 1998) and Fig. 15.13 (from Akahori *et al.,* 2000) show that this method can be used to provide high resolution. A more recent study by Osawa *et al.* (1999) demonstrated that despite its extraordinarily high density, osmium coating

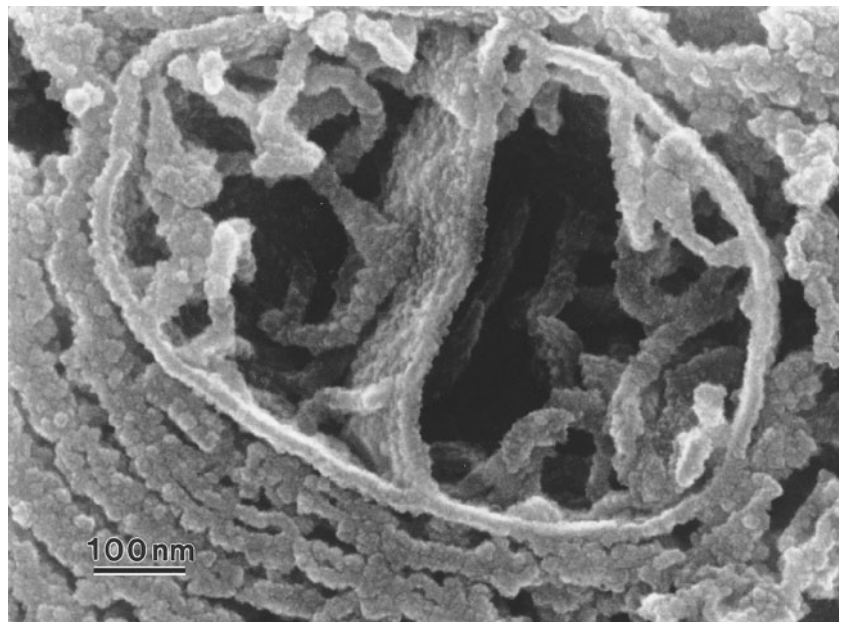


Figure 15.13. A freeze-cracking preparation of a rat liver mitochondrion coated with 5 nm of osmium metal using a hollow-cathode plasma vapor device. Magnification marker = 100 nm. (Akahori *et al.,* 2000.)

Table 15.1. Properties of Elements that May Be Considered as Coating Materials for Elements $Z = 8$–20 that Are to Be Subjected to X-Ray Microanalysis

Physical property	C	Al	Cr
Density (kg m^{-3})	2300	2700	7200
Thermal conductivity (W cm^{-1} K^{-1})	1.29	2.37	0.93
Resistivity (micro-ohm-cm)	3500	2.83	13.0
Evaporation temperature (K)	3500	1275	1450
X-ray energy (eV)			
$K\alpha$ peak	277	1487	5415
$K\beta$ peak			5946
$L\alpha$ peak			598

can be used in conjunction with BSE imaging of colloidal-gold-labeled antibodies.

15.7. Coating for Analytical Studies

The equipment used in coating for analytical studies is the same as the equipment discussed in the previous sections. The only differences center on the choice of the target material. The coating layer should not contribute to the x-ray spectrum and must be sufficiently thin not to mask or overlap the emitted x-ray photons or attenuate the incoming electrons. For example, gold should not be used as a coating material in x-ray studies of light elements because the M lines of gold overlap the K lines of both phosphorus and sulfur and will strongly absorb other light element x-rays.

For elements $Z = 8$–20, there are only three elements that can be used routinely as coating materials. Pertinent details for C, Al, and Cr are given in Table 15.1. The production of thin films of carbon is a routine procedure in an evaporative coater and the only disadvantage is that the high evaporation temperature may damage delicate specimens. There are no substantive K-peak overlaps with carbon. Thin films of both aluminum and chromium can be produced by both evaporative and sputter coaters. There are minimal peak overlaps with aluminum, but the L line of chromium is close to the $K\alpha$ line of oxygen.

15.8. Coating Procedures for Samples Maintained at Low Temperatures

This subject is considered in some detail in the book by Echlin (1992) and additional details can be found in the papers by Chen *et al.* (1995) and Walther and Muller (1999). The coating procedures are the same as those described elsewhere in this chapter, but with the added complication that the sample must be maintained at or below the glass transition temperature T_g of water (130 K). This added requirement also means that the coating procedure must be carried out in a completely anhydric environment and

that the sample is not subjected to any thermal stress. Low-temperature coating can be achieved routinely with plasma magnetron sputter coater fitted to one of the commercially available cryopreparation units (see Chapter 14). A gold–palladium target can be used for routine SEM studies and chromium or platinum for high-resolution work.

15.9. Coating Thickness

The key rule regarding coating thickness is to apply the minimal thickness in order to obtain the maximum information from the specimen. As the sequence of images in Fig. 15.14 shows, if the coating layer is too thick, it will obscure the surface detail. If the coating layer is too thin or discontinuous, the specimen will charge and the sample may be damaged.

The thickness of a layer deposited by thermal evaporation can be estimated from the amount of the starting material that will be consumed. In high-vacuum thermal evaporation, one can assume that all the vapor molecules leave each part of the evaporant surface without preferred direction and pass to the substrate surface without colliding with residual gas molecules. By allowing for the vapor incident angle at the specimen, and assuming that all the incident vapor molecules have the same condensation coefficient, the coating thickness distribution can be calculated. The following formula can be applied to a flat, nonrotating surface oriented at an angle θ to the source direction, and can be used to calculated the thickness from the quantity of material to be evaporated:

$$T = \frac{M}{4\pi R^2 \rho} \cos\theta, \tag{15.2}$$

where T is the thickness in centimeters, M is the mass of the evaporant in grams, ρ is the density in g/cm^3, R is the distance from source to specimen in centimeters, and θ is the angle between the normal to the specimen and the evaporation direction. With Eq. (15.2), the thickness can be calculated to approximately $\pm 50\%$ if the geometry of the evaporation rig and substrate is known and if the material in question evaporates slowly and does not alloy with the heater. It is assumed that evaporation is from a point source and occurs uniformly in all directions. The uncertainty in T arises in part because of uncertainty in the sticking coefficient of the evaporant or the substrate and the deviation in film density from bulk density. The efficiency is material-dependent and is rarely better than 75%. The thickness calculated by Eq. (15.2) must therefore be reduced by multiplying by the sticking efficiency (approximately 0.75). As an example, consider a carbon coater that operates by completely evaporating a reduced carbon rod section 1 mm in diameter and 3 mm long. The mass M in this carbon cylinder is $\pi r^2 l \rho = \pi\ (0.05\,\text{cm})^2(0.3\,\text{cm})(2\text{g/cm}^3)$, or 0.0047 g, taking an assumed density of 2 g/cm^3 for amorphous carbon. If the substrate is at a distance of 10 cm from the source and perpendicular to the direction of evaporation ($\theta = 0$, so $\cos\theta = 1$), the thickness calculated with Eq. (15.2) is $T = 18.8$ nm.

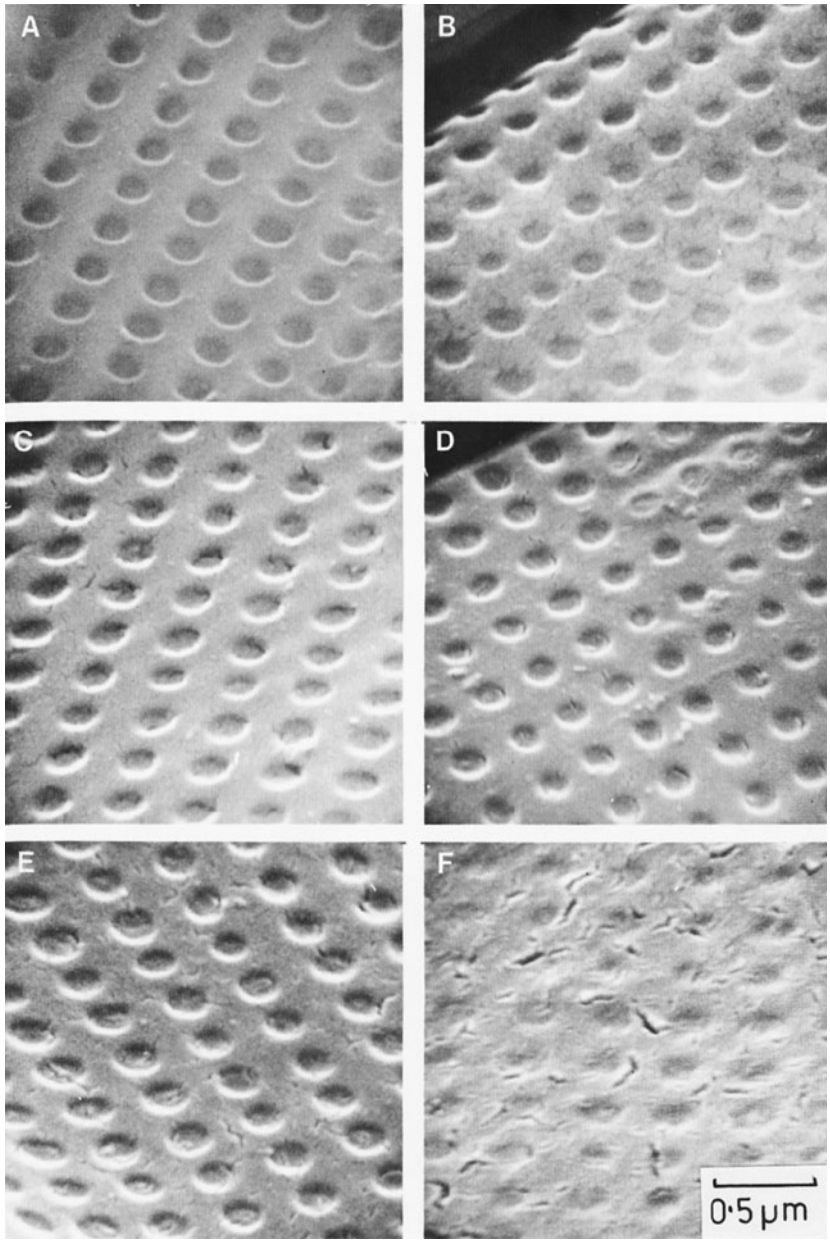


Figure 15.14. A series of images of acid-cleaned diatom silica frustules evaporative-coated with an increasing mass thickness of gold: (a) 2.5, (b) 5, (c) 10, (d) 20, (e) 50, and (f) 100 nm. Note that optimal image information is seen at between 2.5 and 5.0 nm. Above this thickness, the surface is progressively obscured. magnification marker = 0.5 μm.

A quick and easy method for estimating coating thickness is to look at the change in the color of the surface of a small piece of white photocopying paper half covered with a razor blade and placed close to the sample. A suitable coating of carbon will be deposited on the specimen if the exposed piece of paper is a pale chocolate color. A shiny light-gray color will indicate that too much carbon has been deposited. A satisfactory layer of gold would be indicated by a pale-bronze color on the paper. Too much gold will have been deposited if the paper is a distinct shiny gold color. For gold–palladium and aluminum, a mid-gray color will indicate that sufficient material has been deposited onto the specimen.

Although this procedure is very subjective, it can provide a useful and inexpensive way of determining whether the sample is likely to have been coated adequately. A more accurate, but still not definitive measure of coating thickness can be obtained by using a quartz crystal thin-film monitor.

Thin-film monitors can be used in all types of coating equipment to calculate film thickness during coating, by measuring changes in the vibration frequency of a planar quartz crystal as the average mass thickness of the coating material increases. The monitor is able to relate the measured average mass thickness to the average metric thickness that is needed. The relationship between what is measured on a flat crystal and what lands on a tilting and rotating rough sample is difficult to establish. For most applications, the actual film thickness measured by the thin-film monitor is less important than the thickness of the coating layer that provides the optimal information about the specimen.

15.10. Damage and Artifacts on Coated Samples

The sample may be damaged by excessive thermal radiation from an evaporative source. This is observed as melting of crystals, as pitting in fine filaments, and as cracks in the specimen surface. There is rarely any visible damage in samples that have been plasma-magnetron sputter-coated. However at very high resolution, small pits and discontinuities may be seen in otherwise smooth surfaces. In all instances where the sample is damaged, it is necessary to recoat a new sample using less energy.

The most obvious artifact is seen when the sample surface is obscured by a thick layer of coating material. Other artifacts are less obvious and are usually only seen at high magnifications. The sample may become contaminated with particulate material as a result of out-gasing and backstreaming in the vacuum system. Contamination is also indicated if the sample image has discontinuous black regions that do not disappear if the beam accelerating voltage is decreased. In low-temperature microscopy, residual water vapor molecules may go to hydrophilic surface features on the specimen and in turn may favor preferential condensation and nucleation of the coating atoms. Under conditions of high beam current, a thick metal coating layer may melt and distort the specimen surface. Finally, in roughly sculptured samples that are evaporatively coated, self-shadowing can occur and some parts of the sample may receive more coating than others. The more

exposed parts of the sample will show exaggerated changes in dimension. This artifact is minimized if the sample is sputter-coated and/or tilted and rotated during the coating procedure.

15.11. Summary of Coating Guidelines

In order to obtain very high resolution images in the SEM, a 0.5- to 0.8-nm film is recommended using a penning/ion beam coater with targets of platinum, tantalum, iridium, tungsten, or chromium. Alternatively, one can use electron beam evaporation to apply the same thickness of Pt–Ir–C (Wepf *et al.,* 1991). To obtain medium- to high-resolution images in the SEM, a 2- to 4-nm film of either a gold–palladium alloy, platinum, or chromium (for backscattered images) is recommended using a plasma magnetron sputter coater and a high-purity argon plasma. For routine SEM images, that is, up to a magnification of 20,000×, one can use a plasma magnetron sputter coater to apply a 3- to 5-nm layer of gold–palladium. For low-voltage SEM images, one should try imaging without a coating layer. If this does not work, a plasma magnetron coater should be used to apply a 2- to 3-nm layer of gold–palladium. Finally, for x-ray microanalytical studies a 3- to 5-nm layer of evaporated carbon or a 3- to 5-nm layer of a plasma magnetron sputter-coated conductive metal should be applied.

These suggestions are only guidelines and it is important to experiment with materials, film thickness, and mode of application. Highly sculptured specimens will need more coating material than flat specimens and samples examined at high voltages will need more coating than those examined at low voltages.

References

Akahori, H., M. Handa, H. Yoshida, and Y. Kozuka (2000). *J. Electron Microsc.* **49,** 735.
Chen, Y., V. E. Centonze, A. Verkhovsky, and G. G. Borisy (1995). *J. Microsc.* **179,** 67.
Echlin, P. (1992). *Low Temperature Microscopy and Analysis,* Plenum Press, New York.
Hermann, R., J. Pawley, T. Nagatani, and M. Muller (1988). *Scanning Microsc*. **2,** 1215.
Maisell, L. I. and R. Glang, eds. (1970). *Handbook of Thin Film Technology,* McGraw-Hill, New York.
Osawa, T., and Y. Nozaka (1998). *J. Electron Microsc*. **47,** 273.
Osawa, T., Y. Yoshida, F. Tsuzuku, M. Nozaka, M. Takashio, and Y. Nozaka (1999). *J. Electron Microsc.* **48,** 665.
Peters, K. R. (1984). In *Electron Beam Interactions with Solids for Microscopy, Microanalysis, and Microlithography* (D. F. Kyser, H. Niedrig, D. E. Newbury, and R. Shimizu, eds.), SEM, Inc., O'Hare, Illinois, p. 363.
Stokroos, L., D. Kalicharan, J. J. L. Van der Want, and W. L. Jongebloed (1998). *J. Microsc*. **189,** 79.
Tanaka, A. (1994). *J. Electron Microsc*. **43,** 177.
Walther, P., and M. Muller (1999). *J. Microsc*. **196,** 279.
Wepf, R., M. Amrein, U. Burkli, and H. Gross (1991). *J. Microsc*. **163,** 51.
Woodward, J. T., and J. A. Zasadinski (1996). *J. Microsc*. **184,** 157.

Subject Index