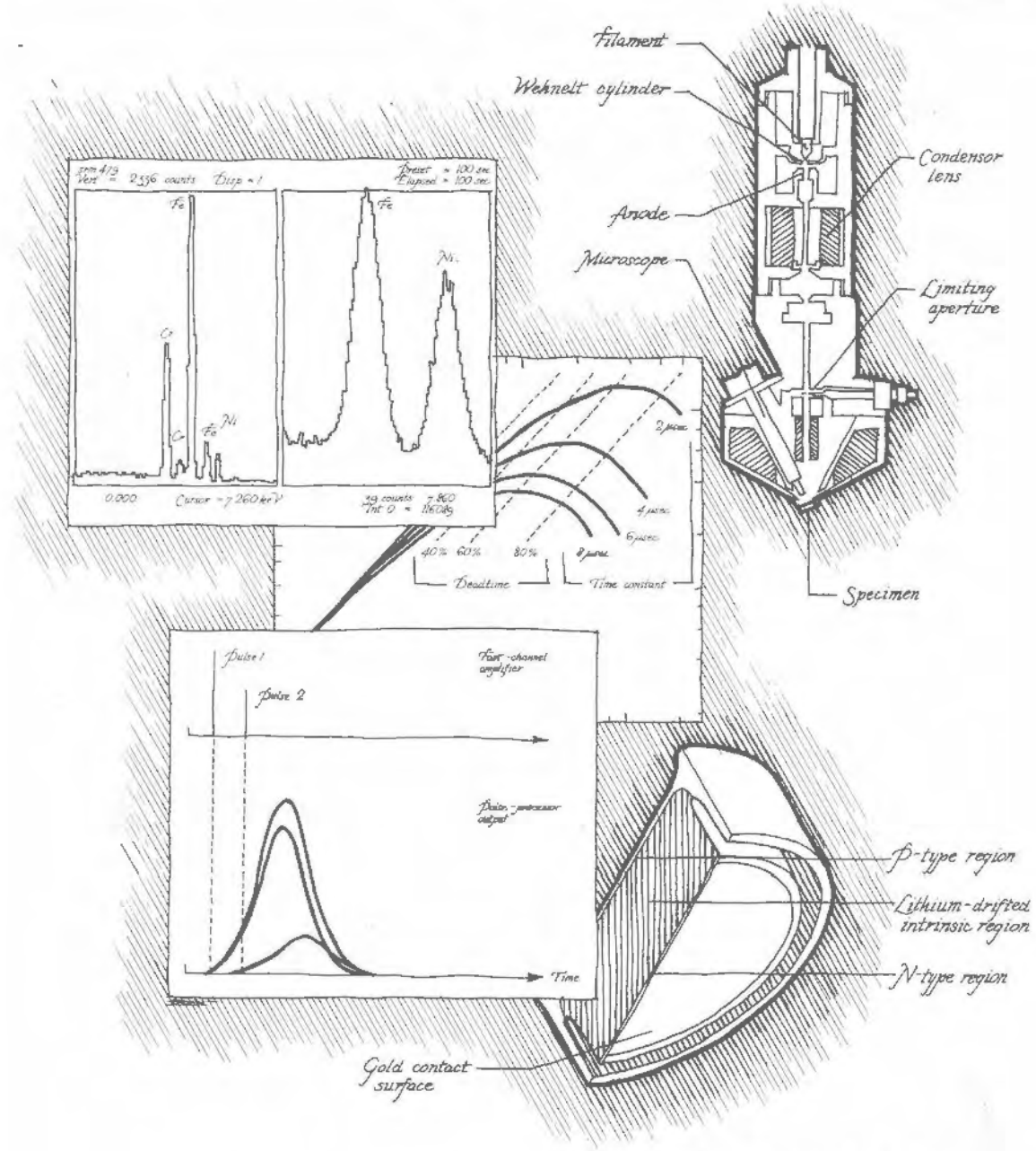
ENERGY-DISPERSIVE X-RAY MICROANALYSIS



An Introduction

ENERGY-DISPERSIVE X-RAY MICROANALYSIS

An Introduction

Kevex Instruments, Inc.
San Carlos, California

Kevex Instruments, Inc.
355 Shoreway Road
San Carlos, CA 94070

Printed in the United States of America

Contents

Preface

THIS BRIEF PRIMER on microanalysis had its origins in an introduction to the subject prepared by Robert Johnson, product manager for microanalysis at Kevex. It remains his work as much as anyone's, but many others have influenced its evolution. In particular, comments by Dave Seielstad and Dr. Carl Meltzer led to substantial rethinking of the introduction and to the aside on statistics. Dr. Rolf Woldseth also offered helpful suggestions, and his book *X-Ray Energy Spectrometry* (Kevex Corporation, 1973), now out of print, was the source of several illustrations redrawn for this work. Additional useful comments and contributions came from Christina Ellwood, Bob Fucci, John Holm, Dr. Asher Holzer, Tom Stark, Ronald Vane, and David Wherry. Finally, special thanks are due Dr. Joe Balser of the Lawrence Livermore National Laboratory, who critically reviewed the entire manuscript. Of course, these contributors and reviewers should not be held accountable for the ultimate disposition of their good advice. As final arbiter and contributor of last resort, the editor bears responsibility for omissions and errors that remain.

DOUGLAS VAUGHAN
Editor

Where the telescope ends, the microscope begins. Which of the two has the grander view?

VICTOR HUGO, *Les Misérables*

The chief result is that all the elements give the same kind of spectrum, the result for any metal being quite easy to guess from the results for the others. This shows that the insides of all the atoms are very much alike, and from these results it will be possible to find out something of what the insides are made up of.

H. G. J. MOSELEY, letter, 2 November 1913

1

INTRODUCTION

TAKEN LITERALLY, microanalysis is the analysis of "very small" samples—by whatever technique is available. Historically, however, the term has had a much narrower meaning. When electrons of appropriate energy impinge on a sample, they cause the emission of x-rays whose energies and relative abundance depend upon the composition of the sample. Using this phenomenon to analyze the elemental content of microvolumes (roughly one to several hundred cubic micrometers) is what we commonly mean by microanalysis. To narrow the topic even further, we concern ourselves here only with *energy-dispersive* microanalysis, in which the x-ray emissions are sorted electronically, rather than by means of a diffraction crystal (see the aside on page 3).

In general, microanalysis is the easiest method (and sometimes the only one) for analyzing microscopic samples. It has other advantages as well. It is sensitive to low concentrations—minimum detection limits (MDLs) are below 0.1% in the best cases and typically less than 1%; and its dynamic range runs from the MDL to 100%, with a relative precision of 1% to 5% throughout the range. Furthermore, the technique is practically nondestructive in most cases, and requirements for sample preparation are minimal.

In this cursory treatment of the subject, we can divide our task into three major parts. First, we consider the processes that follow the excitation of the sample by an electron beam. We are most interested in the process by which x-rays are emitted, but our efforts will be repaid if we also look at some of the other interactions that occur. Next, we are interested in the means by which the emitted x-rays are collected, sorted, and counted. That is, we want to know how the energetic emissions of an electron-excited sample get translated into analyzable data. Finally, we look at the analysis techniques themselves.

The process of x-ray emission is shown schematically in Figure 1-1 (we shall save the complications for later). First, an electron from, say, a scanning electron microscope, ejects an electron from an inner shell of a sample atom. The resulting vacancy is then filled by an electron from a higher-energy shell in the atom. In "dropping" to a state of lower energy, this vacancy-filling electron must give up some of its energy, which appears in the form of electromagnetic radiation. The energy of the emitted radiation, then, is exactly equal to the energy difference between the two electronic levels involved. Since this energy difference is fairly large for inner shells, the radiation appears as x-rays.

To complicate matters a bit, there are many energy levels—therefore many potential vacancy-filling mechanisms—within every atom. As a consequence,

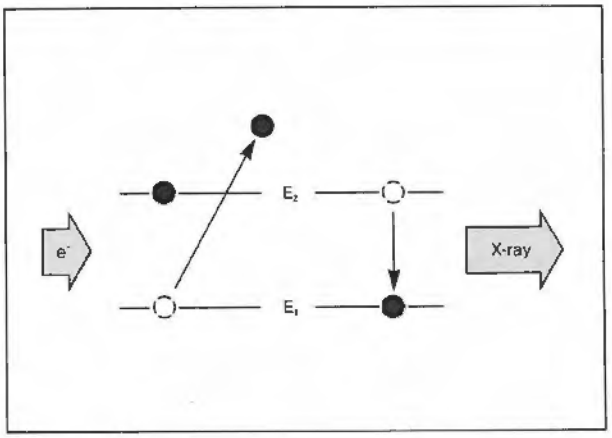


Figure 1-1. *X-ray microanalysis is based on electronic transitions between inner atomic shells. An energetic electron from an electron column dislodges an orbital electron from a shell of low energy (E_1). An electron from a shell of higher energy subsequently fills the vacancy, losing energy in the process. The lost energy appears as emitted radiation of energy $E_2 - E_1$.*

even a sample of pure iron will emit x-rays at many energies. Nonetheless, the principle is a simple one: When excited by electrons of sufficient energy, every element in a sample will emit a unique and characteristic pattern of x-rays. Furthermore, under given analysis conditions, the number of x-rays emitted by each element bears a more or less direct relationship to the concentration of that element.

Converting these x-ray emissions to analyzable data is the job of a series of electronic components (see Figure 1-2), which, in the end, produce a digital spectrum of the emitted radiation. The x-ray photon first creates a charge pulse in a semiconductor detector; the charge pulse is then converted into a voltage pulse whose amplitude reflects the energy of the detected x-ray. Finally, this voltage pulse is converted into a digital signal, which causes one *count* to be added to the corresponding channel of a multichannel analyzer. After a time, the accumulated counts from a sample produce an x-ray spectrum like the one in Figure 1-3.

Extracting quantitative information from an x-ray spectrum is complicated by the fact that the neat picture of Figure 1-1 is greatly oversimplified. The background hump in Figure 1-3, for example, arises from one of several complicating interactions. Others produce spurious peaks or cause true spectral peaks to be larger or smaller than we might expect from first principles. And, of course, peaks often overlap, making their resolution difficult. Typically, then, the quantitative analysis of an acquired spectrum comprises at least five steps: (1) accounting for spurious peaks; (2) identification of the elements giving rise to the spectrum; (3) removal of the background; (4)

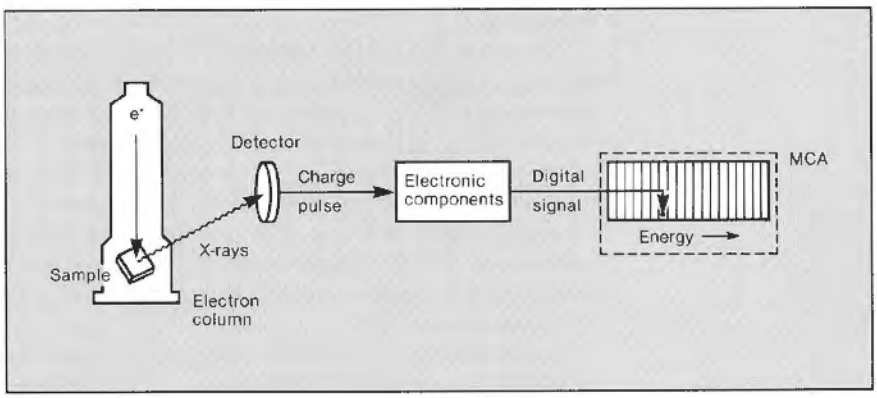


Figure 1-2. *In energy-dispersive microanalysis, each emitted x-ray produces a charge pulse in a semiconductor detector. This tiny and short-lived current is converted first into a voltage pulse, then into a digital signal reflecting the energy of the original x-ray. The digital signal, in turn, adds a single count to the appropriate channel of a multichannel analyzer (MCA).*

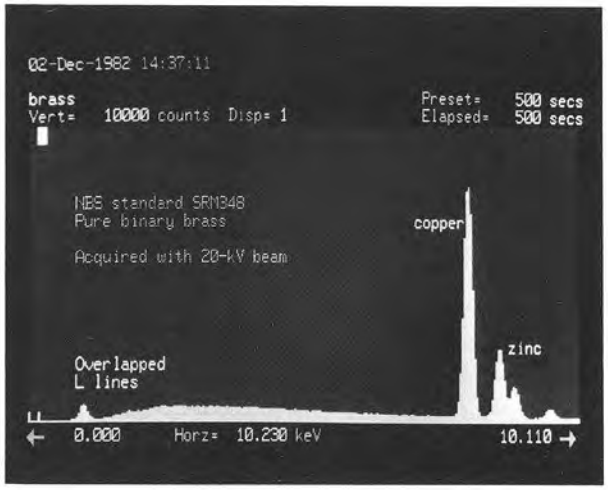


Figure 1-3. *The dominant features of a typical x-ray spectrum include major spectral peaks superimposed on a broad background. A close look at the most intense peak (labeled* copper*) reveals that the spectrum comprises a series of individual channels.*

resolution of the spectral peaks; and (5) computation of element concentrations, a process that involves accounting for interelement effects within the sample. Despite these apparent difficulties, modern systems can typically acquire and analyze a complex x-ray spectrum in a few minutes.

ASIDE: A LITTLE HISTORY

Soon after x-rays were discovered in 1895, it became apparent that x-ray energies are intimately related to the atomic structure of the substances that emit them. And since the atomic structure of each chemical element is different, it follows that each element—when stimulated to do so—emits a different pattern of x-rays. By the 1920s, these characteristic patterns had been recorded for most of the elements. Until the late forties, however, analyzing substances by stimulating and recording their x-ray emissions remained the province of the research scientist.

Then, in 1948, a prototype for the first modern commercial x-ray spectrometer was developed. Variants of this instrument remain in use today. In these instruments, a sample of unknown composition is excited by a beam of x-rays. As the excited atoms relax to their stable ground states, they emit their characteristic patterns of x-rays. These x-rays are separated into their component wavelengths by a diffraction crystal, then detected and measured. The presence of the crystal as the basis for resolving x-rays of different wavelength defines this sort of instrument as a *wavelength-dispersive spectrometer.*

In the following year, the first electron microprobe was built. The principles were the same, but the source of excitation was a beam of electrons rather than x-rays. Unlike x-ray–based systems, microprobes can examine volumes of sample as small as a cubic micrometer, but the pattern of characteristic x-rays emitted by the excited sample is, in principle, the same.

Then, in the mid-sixties, a semiconductor radiation detector was developed at the Lawrence Berkeley Laboratory that heralded the advent of *energy-dispersive x-ray spectrometry,* or *x-ray energy spectrometry (XES).* The current version of this detector, a single 3-mm-thick, 7-mm-diameter crystal of silicon, is the complete x-ray–dispersing element of a typical XES system. Together

with appropriate electronic amplifiers and signal processors, it does away with the complexities of the diffracting crystal and a physically dispersed spectrum of x-rays. Instead, the energy-dispersive system collects emitted x-rays of all wavelengths and sorts them electronically.

Since all electromagnetic radiation can be classified on the basis of its wavelength and, at the same time, can be thought of as packets of energy called photons, wavelength- and energy-dispersive techniques are measuring the same phenomenon. The equivalence is clear in Planck's equation:

$$\lambda = \frac{hc}{E}$$

where λ is the wavelength of the radiation, c is the speed of light, h is Planck's constant, and E is the energy of the radiation. Rearrangement and substitution of appropriate values then yields

$$E = \frac{12.4}{\lambda}$$

where energy is measured in kilo-electron volts and wavelength is measured in angstroms.

2

THE PHYSICS OF THE PROCESS

AN AMPLIFIED, though still brief, description of characteristic x-ray emission is the first order of business. At the same time, we shall introduce the most important of the other interactions that occur within the excited sample. For convenience, we divide the processes involved into electron-sample interactions and photon-sample interactions.

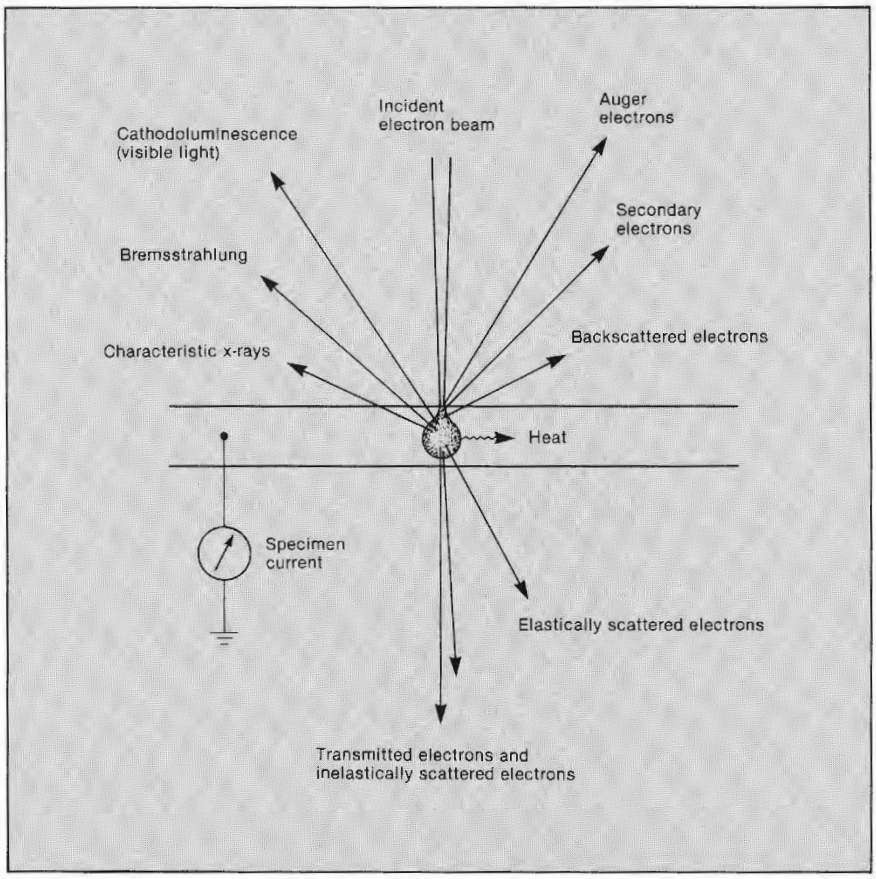


Figure 2-1. *Schematic illustration of the principal results of the interaction of an electron beam with a specimen. As suggested by the figure, Auger and secondary electrons emerge from near the surface of the sample, and elastically scattered electrons are typically scattered through larger angles than are inelastically scattered electrons.*

2.1 ELECTRON-SAMPLE INTERACTIONS

In an electron column, electrons are accelerated through an electric field, thus acquiring kinetic energy. This energy is deposited in the sample, and its dissipation yields a variety of signals for analysis, as depicted schematically in Figure 2-1. For the purposes of this discussion, the Bohr model of an atom suffices; thus, the atoms of Figure 2-2 comprise positively charged nuclei surrounded by electrons in discrete orbits, each with a well-defined energy level. The origins of several signals are shown in these simplified diagrams.

2.1.1 Secondary Electrons

The primary (exciting) electron may interact with an electron in the sample, ejecting it with some amount of kinetic energy. If the ejected electron was weakly bound, it typically emerges with only a few eVs of energy and is called a *secondary electron*. (Strictly speaking, any electron ejected from an atom in the sample is a secondary electron—whatever its energy. To the electron microscopist, however, secondary electrons are those with energies below about 50 eV.) Since they have little energy, secondary electrons can

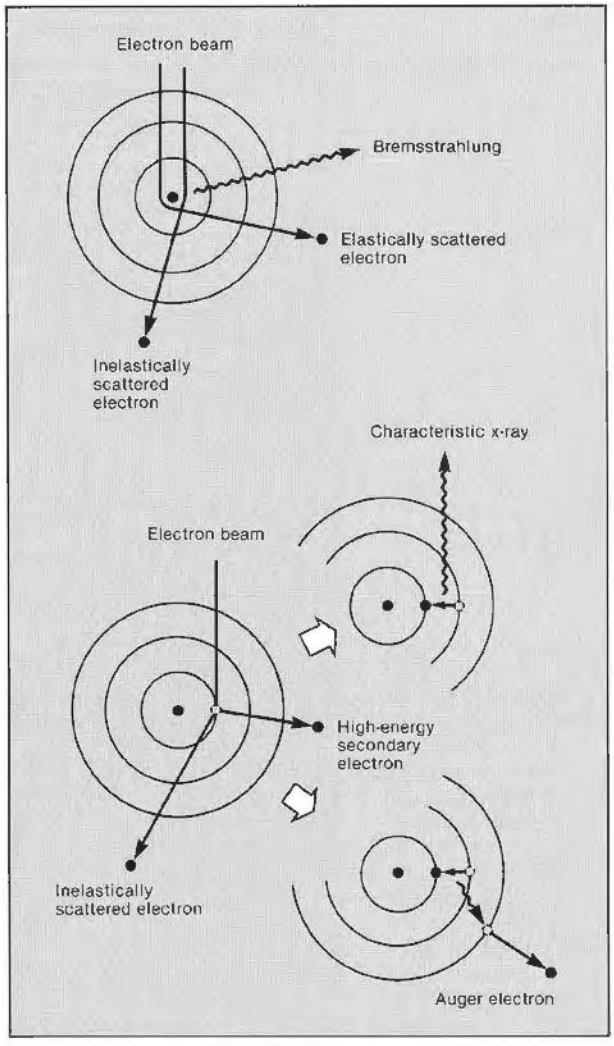


Figure 2-2. *Classical models showing the sources of several signals detected in the electron column. In the top drawing, electrons are scattered elastically and inelastically by the positively charged nucleus. The inelastically scattered electron loses energy, which appears as bremsstrahlung. Elastic scattering, which involves no energy loss, can be readily understood only by resorting to quantum mechanics. Typically, however, elastically scattered electrons (which include backscattered electrons) are scattered through larger angles than are inelastically scattered electrons. In the lower set of drawings, the incoming electron ionizes the sample atom by ejecting an inner-shell electron. Deexcitation, in turn, produces characteristic x-radiation or an Auger electron. The secondary electrons typically detected in the electron column are ejected with low energy from loosely bound states, a process not illustrated here.*

escape from the sample to be detected only if they are created near the surface. For the same reason, they are sensitive to the topography of the sample. As shown in Figure 2-3, secondary electrons created at a topographic peak have a greater chance of escaping than secondary electrons created in a topographic hole.

As the primary electrons interact with the sample, they are scattered and spread. The volume in which the primary electrons interact with the sample is generally characterized as onion shaped (Figure 2-4). Because the greatest density of secondary electrons is created by the primary beam before it has a chance to spread, they have high spatial resolution relative to other available signals. Secondary electrons carry little information about the elemental composition of the sample; however, their topographic sensitivity and high spatial resolution make them the most frequent choice for micrographic images. It is their sensitivity to topography that makes secondary electron images (SEIs) so easy to interpret visually. (Secondary electrons are also generated, sometimes efficiently, by scattered electrons outside the area of primary beam incidence. These secondary electrons add noise to the signal of interest and can degrade the quality of the SEI.)

2.1.2 Backscattered Electrons

If the primary electron interacts with the nucleus of a sample atom, it may be scattered in any direction with little loss of energy. Some of these scattered electrons will be directed back out of the sample—often after more than one scattering event—allowing them to be detected. These *backscattered electrons (BSEs)* are much more energetic than secondary electrons and so may escape from a greater depth within the sample. Therefore, compared to secondary electrons, the backscattered signal will not carry as much information about sample topography nor will it be as highly resolved in space. There is a compensating advantage, however. The main influence on the strength of the BSE signal is the mean atomic number of the sample in the interaction volume. The higher the atomic number of an atom, the greater the positive charge of its nucleus and the more likely an interaction that produces a BSE. The BSE signal therefore carries some information about sample composition.

2.1.3 X-Ray Continuum

The primary electron may also be scattered inelastically by the coulomb field of an atomic nucleus (partially screened by inner-shell electrons), thus giving up some or all of its energy. This energy may be emitted in the form of x-radiation called *bremsstrahlung* (from the German "braking radiation"). Since the primary electron can give up any amount of its energy, the energy distribution of the emitted x-rays is continuous. This component of the x-ray signal is thus often called the *continuum*.

The closer the primary electron comes to "hitting" a sample atom, the stronger the interaction and the greater the energy likely to be lost. In the

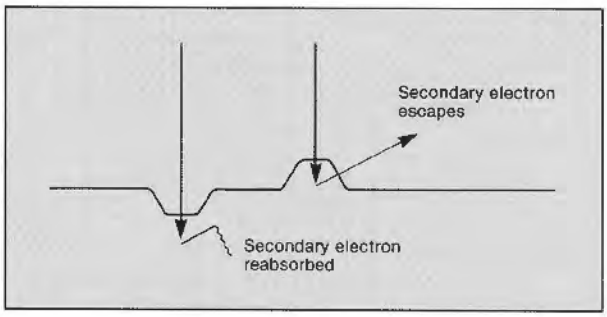


Figure 2-3. *An illustration of the topographic sensitivity of low-energy secondary electrons. Such electrons are more likely to emerge from peaks than from valleys; hence, the secondary electron signal is especially sensitive to sample surface features.*

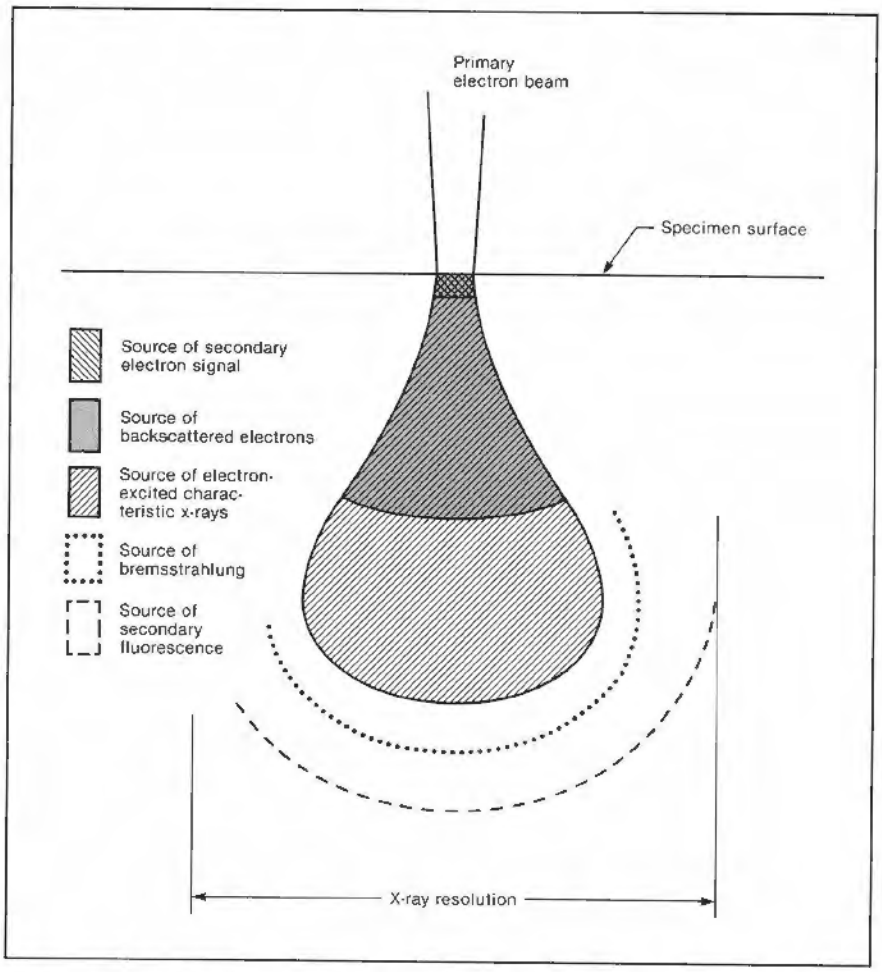


Figure 2-4. Generalized illustration of interaction volumes for various electron-specimen interactions. Auger electrons (not shown) emerge from an even thinner region of the sample surface than do secondary electrons. X-ray–excited characteristic x-rays (secondary fluorescence) emerge from deepest within the sample and have the poorest resolution. (Adapted from Reference 1.)

extreme case, the electron may give up all of its energy in a single event, which places an upper limit on the energy distribution of the continuum, namely, the accelerating voltage E_0 of the electron column. However, a wide miss is more likely than a near miss (which, in turn, is more likely than a hit); therefore, the energy distribution can be expected to climb steeply at lower energies. This simple argument yields a distribution like that shown as a dashed line in Figure 2-5. In practice, however, the escaping low-energy x-rays are preferentially absorbed in the sample and the window of the detector, leading to the observed energy distribution shown by the solid line (see also the background in Figure 1-3).

2.1.4 Characteristic X-Rays

When an electron is ejected from an inner atomic shell by interaction with a high-energy electron beam, the result is an ion in an *excited state*. Through a relaxation, or deexcitation, process, this excited ion gives up energy to return to a normal *ground state*. The most likely process in most cases is a series of transformations in each of which an electron from an outer shell "drops" into a vacancy in an inner shell. As we have seen, each drop results in the loss of a specific amount of energy, namely, the difference in energy between the vacant shell and the shell contributing the electron. This energy is given up in the form of electromagnetic radiation—x-rays in the case of high-energy transitions involving inner shells. The energy of the radiation uniquely in-

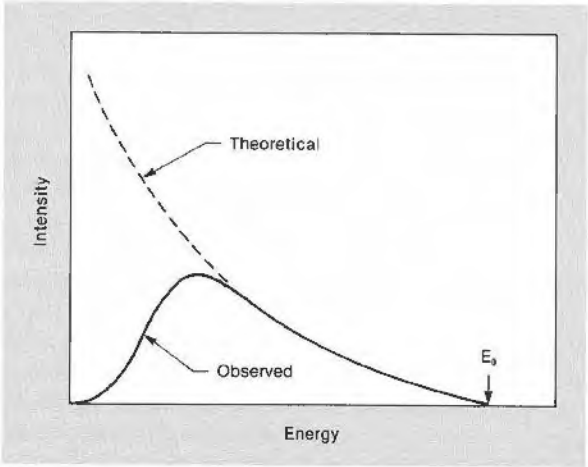


Figure 2-5. *Plot of the intensity of continuum radiation (bremsstrahlung) as a function of energy. The observed fall-off at low energies is due to x-ray absorption between the point of origin and the detector crystal.*

dicates the element from which it came, hence the name *characteristic emission*—for our purposes, *characteristic x-rays.*

X-rays travel much greater distances through the sample than electrons and therefore escape from depths at which the primary electron beam has been widely spread. Consequently, the x-ray signal has poor spatial resolution compared to the secondary electron and backscattered electron signals.

Nomenclature

Before we go on, a discussion of the nomenclature for x-ray emissions might be helpful. The lines are usually named according to the shell in which the initial vacancy occurs and the shell from which an electron drops to fill that vacancy (Figure 2-6). For instance, if the initial vacancy occurs in the *K* shell and the vacancy-filling electron drops from the adjacent shell (the *L* shell), a $K\alpha$ x-ray is emitted. If the electron drops from the *M* shell—two shells away—the emitted x-ray is a $K\beta$ x-ray. (Because of the complexity of electronic structure, the nomenclature becomes more complex when the initial vacancy occurs in higher-energy shells.) Microanalysts are generally concerned with *K*-, *L*-, and *M*-series x-rays. Therefore the common reference to KLM lines.

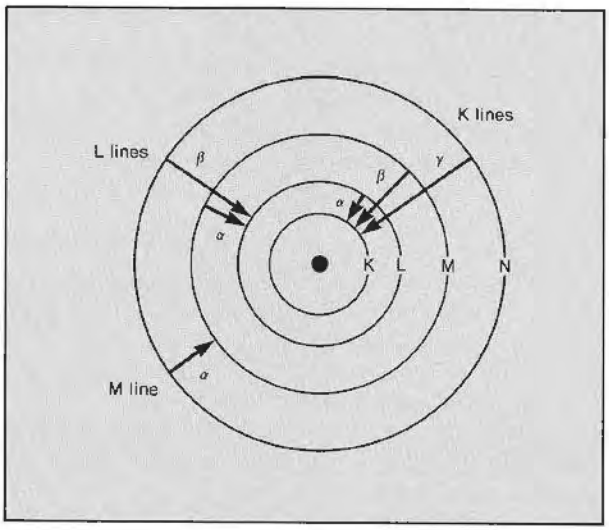


Figure 2-6. *Some line types typically observed in x-ray spectra. Each shell actually comprises several energy levels; thus, transitions are more numerous (and the nomenclature more complicated) than shown.*

Moseley's Law

The most useful property of characteristic x-rays is the variation of their energy with atomic number. This relationship is described by Moseley's law:

$$E = c_1(Z - c_2)^2$$

where E is the energy of the characteristic x-ray, Z is the atomic number, and c_1 and c_2 are constants for a given line type. Figure 2-7 illustrates this relationship. For a given line type (for instance, the $K\alpha$ lines), the x-ray energy increases with atomic number. Thus, from the energy of an x-ray emission, the atomic number of the emitter can be determined if the line type is known.

Characteristic X-Ray Intensity

The detected intensity of characteristic x-ray emissions, under given excitation conditions, is influenced by three factors. The first is atomic number—both the atomic number of the emitting atom and the average atomic number of the bulk sample. Two parameters characterize the dependency on the atomic number of the emitter. The first is the *ionization cross section*, which expresses the likelihood that an initial vacancy (an ionization) will occur under the given conditions. The second is the *fluorescent yield*, which is the probability that a vacancy, once created, will produce a characteristic x-ray. The average atomic number of the sample, on the other hand, affects the amount of energy lost to other scattering processes—energy that is thus unavailable to ionize a sample atom.

The second influence on intensity is the probability that emitted characteristic x-rays will be absorbed before they emerge from the sample. More will be said about absorption in the next section on photon-sample interac-

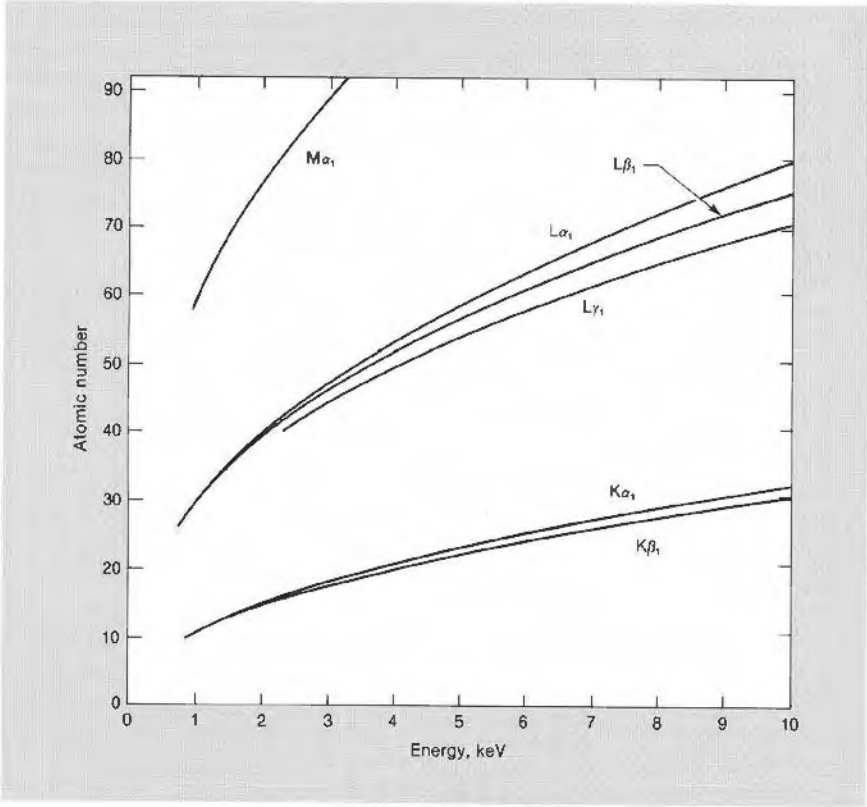


Figure 2-7. Plot of the energies of major x-ray emission lines observed below 10 keV.

tions. The final factor is secondary fluorescence, which is one result of such absorption. For example, a high-energy x-ray characteristic of element *A* may be absorbed by an atom of element *B*, thus stimulating a lower-energy emission characteristic of the second element. The presence of elements *A* and *B* in the same sample will therefore increase the intensity of characteristic emission from element *B* and decrease it from *A*. This is a common *matrix effect*—that is, an effect that depends on the sample matrix—which requires special treatment during quantitative analyses.

2.1.5 Auger Electron Emission

The atom excited by the primary electron beam often undergoes deexcitation by an alternative process. For convenience only, we discuss it here as a sequence of two independent events; in fact, it is a single process, producing an electron of characteristic energy instead of a characteristic x-ray. First, an inner-shell vacancy is filled in the usual way, producing a characteristic x-ray. Then, that x-ray is reabsorbed within the same atom, ejecting a lower-energy electron (see Figure 2-2). Therefore, the original characteristic x-ray is not detected. Instead, a secondary characteristic x-ray may be emitted as the outer vacancy is filled. More important, however, the ejected electron itself possesses an energy exactly equal to the difference between the energy of the original characteristic x-ray and the binding energy of the ejected electron. These ejected electrons are known as *Auger electrons*. They are unique among electrons emitted from the sample in that they carry specific chemical information about the atom from which they originated. Furthermore, in contrast to characteristic x-rays, Auger electrons are of very low energy and can travel only a short distance within the sample. The information they carry is therefore specific to the surface of the sample, often only the first few atomic layers.

Deexcitation of an ionized atom may occur by either the emission of an Auger electron or the emission of a characteristic x-ray. This fact is reflected in the fluorescent yield, which depends primarily on the atomic number of the excited atom. For low atomic numbers, the process of Auger emission is favored and the fluorescent yield is low. Conversely, the higher atomic numbers favor the emission of characteristic x-rays.

2.2 Photon-Specimen Interactions

As alluded to above, x-ray photons interact with sample atoms, just as electrons do. Such interactions, in fact, are the basis for x-ray–excited XES, usually referred to as *x-ray fluorescence (XRF) spectroscopy*. For our purposes, the relevant discussion concerns the interactions between the sample and characteristic or continuum x-rays, once they have been created.

2.2.1 Absorption

As an x-ray travels through the sample, it may be absorbed, giving up its energy entirely to an electron and ejecting the electron from its orbital. The likelihood that an x-ray will be absorbed in such a process depends on its energy and the energy with which the electron is bound to its nucleus. The probability of absorption increases as the x-ray energy approaches this *binding energy* from above and reaches a maximum when the x-ray energy is just greater than the binding energy. At this point, there is a discontinuity—an *absorption edge*—in the probability curve; lower-energy x-rays no longer have sufficient energy to overcome the binding energy, and the likelihood of absorption drops to a lower value. The probability of absorption then increases

again as the x-ray energy approaches the binding energy of a more loosely bound electron. As Figure 2-8 illustrates, an absorption curve for a given element includes an absorption edge for each electron shell. Each edge is denoted with the name of the electron shell and the subscript *ab;* thus, K_{ab} is the *K*-shell absorption edge.

The probability of x-ray absorption as a function of path length through the sample is given by Beer's law:

$$\frac{I}{I_0} = \exp(-\mu_m \rho d)$$

where I/I_0 is the fraction of x-rays transmitted through a thickness d of a material of density ρ. The parameter μ_m is called the *mass absorption coefficient* and is a function of the atomic number Z of the absorber and the energy E of the x-ray. For a given element, the value of μ_m can be obtained from absorption curves (similar to Figure 2-8) in which μ_m is plotted against energy, or from tables.[2] The mass absorption coefficient for a complex sample is the weighted average of the coefficients for the constituent elements.

2.2.2 Secondary Fluorescence

When an x-ray is absorbed by an atom in the sample, the absorbing atom is left in an excited state. It subsequently relaxes, emitting its own characteristic x-rays—a process called *secondary fluorescence.* Since an x-ray can be absorbed only in an interaction with an electron having a binding energy less than the energy of the absorbed x-ray, the energy of the secondary fluorescence is necessarily less than the energy of the primary x-ray. The relatively large distance an x-ray (primary or secondary) can travel through the sample and the possibility that the secondary emission process may occur at a location remote from that of primary emission, further degrade the spatial resolution of the x-ray signal.

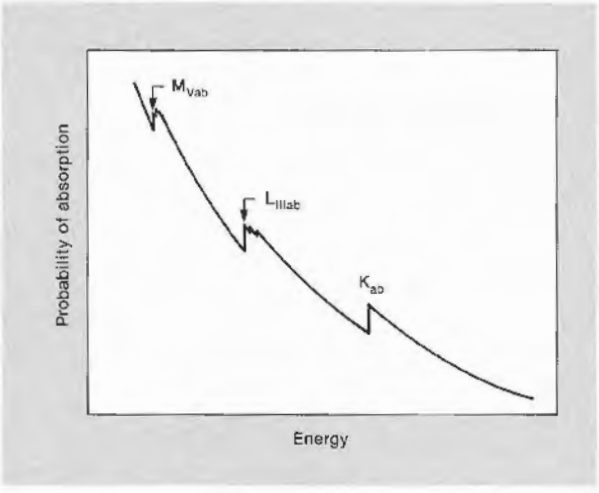


Figure 2-8. *Schematic plot of the x-ray absorption curve for a single element of high atomic weight. The qualitative features of the curve remain the same whether the vertical axis represents the probability of x-ray absorption, the stopping power of the element, or the mass absorption coefficient* μ_m.

3

THE SOURCE OF EXCITATION: THE ELECTRON COLUMN

HAVING DISPENSED with the physics of the interactions that follow electron excitation of a sample, we shall now backtrack to consider the source of the excitation. This might be regarded as an interlude in our threefold task of discussing the physics of x-ray generation, the means by which x-rays are translated into a spectrum, and the analysis of x-ray spectra.

Modern electron columns can be roughly classified in three categories. *Scanning electron microscopes (SEMs)* are the most common and are designed to provide images of high spatial resolution, usually using the secondary electron signal. The image displayed on a cathode ray tube (CRT) is created by scanning the focused electron beam in a raster pattern across some area of the sample while synchronously scanning an analogous pattern on the CRT. The CRT brightness is modulated on the basis of the intensity of the signal of interest. SEMs typically use accelerating voltages between 5 and 30 keV. Sample preparation is minimal, and spatial resolutions of the order of tens of angstroms are attainable.

A second type of column is the *electron microprobe*, though the distinction between the SEM and the microprobe is blurred in some modern instruments. Essentially, a probe is an electron column designed to deliver stable beam currents of high intensity; it may or may not have scanning and imaging capabilities. Most probes are equipped with multiple wavelength-dispersive spectrometers, and the higher beam currents are needed to provide sufficient characteristic x-ray intensities to make wavelength-dispersive analysis practical. Whereas SEMs are designed primarily to deliver high-resolution images, microprobes are intended mainly for accurate quantitative elemental analysis.

The third category of electron column is the *transmission electron microscope (TEM)*. In a TEM, the sample must be thin enough to transmit high-energy electrons. The sample is subjected to a widely dispersed and homogeneous flux of electrons, rather than a focused and scanned pinpoint of electrons. Interactions with the specimen atoms cause disturbances in this beam, which is then focused in a way analogous to the focusing of light rays in an optical microscope. The image is presented on a luminescent plate below the specimen, from which it can be photographed. TEMs are characterized by accelerating voltages between 100 and 300 keV and can provide typical spatial resolutions of a few angstroms. Sample preparation may be extensive.

A modern variation on the TEM, the STEM (scanning transmission electron microscope), combines the principles of the SEM and TEM. A finely

focused beam of electrons is scanned over the electron-transparent specimen, and the image presented on a CRT. However, the image may arise from the transmitted electron signal, as well as the signals normally imaged in the SEM. The greatest advantage of STEM analysis lies in the fact that it avoids the effects of electron beam spreading that are present in bulk sample analysis. High-resolution images can thus be acquired from signals that exhibit low resolution in bulk samples (see Figure 2-4).

3.1 BEAM CURRENT

The intensity of the emitted x-ray signal varies directly with the current of the exciting electron beam. The beam current i_p, in turn, can be expressed as

$$i_p \cong kC_s^{-2/3}\beta d_m^{8/3}$$

where k is a proportionality constant, C_s is the coefficient of spherical aberration for the final lens, β is the gun brightness, and d_m is the beam diameter at its narrowest point.[3] Several conclusions can be drawn from this equation. First, a brighter electron source always yields more current, all else being equal. This is logical, since the brightness of the electron gun is defined as the current density (current per unit area) per unit solid angle. Three electron sources are commonly available, each representing a trade-off between brightness on the one hand and economy, stability, and ease of use on the other. In order of increasing brightness, these sources are the tungsten filament, the lanthanum hexaboride (LaB_6) emitter, and the field emission gun. Second, a significant increase in count rate (which depends on i_p) can be achieved with only a small increase in beam diameter d_m. Finally, a decrease in the coefficient of spherical aberration increases beam current at a given beam diameter. Decreases in spherical aberration can be obtained by increasing the current in the final lens, thereby reducing the focal length and requiring the sample to be located closer to the objective lens. Shorter working distances are therefore desirable for increasing x-ray count rates without diminishing resolution.

Another parameter that affects beam current is the size of the objective aperture. A large aperture maximizes the beam current but decreases the depth of field for imaging purposes.

In addition to its intensity, the stability of the beam current must be considered. All conventional quantitation schemes require that the total deposited charge be known. Although techniques exist to measure beam current continuously while x-ray data are being acquired, the microanalyst usually relies on measurements of the beam current before and after data acquisition, then assumes that the current has remained stable between measurements. Therefore, beam stability is critical. Among the factors that affect stability are the age of the filament, the alignment of column elements, the cleanliness of column surfaces, and the efficiency of the column vacuum system.

3.2 ACCELERATING VOLTAGE

The accelerating voltage used in the electron column influences both the spatial resolution of the x-ray signal and the efficiency with which characteristic x-rays are excited from the sample atoms. Higher voltages produce higher-

energy electrons, which penetrate more deeply into the sample and spread more widely than low-energy electrons. The result is a degradation in resolution on the one hand, but more efficient excitation on the other. It is generally accepted that this trade-off is optimized at an overvoltage (the ratio of the accelerating voltage to the energy of the excited line) of 2½- to 3-fold.

As Figure 2-7 shows, at least one set of x-ray lines (*K*, *L*, or *M*) can be observed below 10 keV for each element. Therefore, energy-dispersive spectra are often acquired between 0 and 10 keV, though frequent use is made of higher energy ranges—usually when spectral overlaps at lower energies preclude unambiguous identifications. For instance, a notorious overlap exists between the sulfur *K* lines and the molybdenum *L* lines at about 2.3 keV. The presence of molybdenum can often be confirmed by looking for the molybdenum *Kα* lines just above 17 keV. Nonetheless, the usual acquisition range of 0 to 10 keV and the conventional overvoltage factor coincide nicely with the accelerating voltages available on modern SEMs—25 to 30 keV.

3.3 Beam-Specimen-Detector Geometry

The geometry of the system affects quantitative x-ray analysis in a variety of ways, but all of these effects arise from the fact that the microanalyst detects only a small percentage of all characteristic x-rays created. The analysis then consists in part of extrapolating backwards from the number detected to the number created, making certain assumptions about the fraction observed. System geometry strongly influences the size of this fraction.

3.3.1 Solid Angle

Perhaps the most direct relationship is the one between the solid angle Ω subtended by the detector and the x-ray detection efficiency (not to be confused with *detector* efficiency, which comes up in Section 4.1.4). The perfect detector would be a spherical one with the point of beam-specimen interaction at its center. Such a detector could be expected to detect all x-rays escaping from the specimen. For all practical, but less perfect, disk-shaped detectors, we must know the portion of the area of that hypothetical sphere that is "covered" by the detector. This portion is expressed by the solid angle of the detector, which is a function of the detector area, its orientation with respect to the point of beam-specimen interaction, and its distance from that point (Figure 3-1). In general, the largest possible detector, looking directly at the sample and located as close to the sample as possible, yields the highest detection efficiencies.

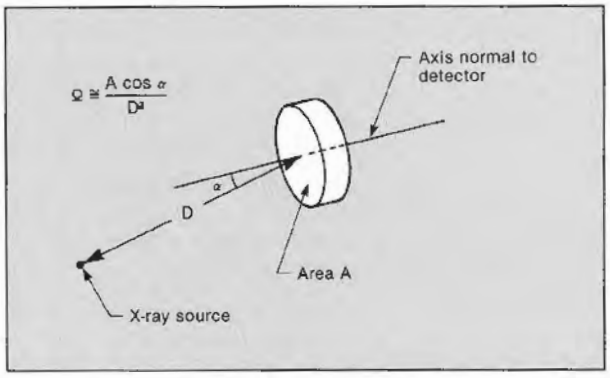


Figure 3-1. Illustration of the parameters that determine the solid angle Ω subtended by the detector at the source of x-rays. The solid angle is expressed in steradians. A hypothetical spherical detector subtends a solid angle of 4π steradians.

3.3.2 Take-off Angle

A second parameter to be considered is the take-off angle ψ. This is the angle between the sample surface and the line taken by the x-rays to the center of the detector (see Figure 3-2).

While characteristic x-rays are created throughout the volume of interaction below the surface of the sample, most quantitative approaches, for the sake of simplicity, assume that all x-rays are created at a single point within that volume of interaction. As shown by Beer's law (page 12), the likelihood of x-ray absorption depends on the length of the escape path, or absorption path, through the sample. The length of this absorption path is determined in turn by the depth of the "creation point" below the sample surface and the angle of the path with respect to the sample surface—the take-off angle.

As can be seen by looking at Figure 3-2, the larger the take-off angle, the shorter the absorption path. The relationship between ψ and x-ray absorption in the sample is shown explicitly in Figure 3-3. The fraction of x-rays transmitted increases rapidly as the take-off angle increases from 0° to 30°, but above 30° the change is less rapid. As a rule of thumb, then, take-off angles above 30° should be used. Not only do shorter absorption paths maximize count rates, but they also minimize the correction that must be applied to the data to account for x-ray absorption and secondary fluorescence in the sample. Mass absorption coefficients are still a matter of controversy and one of the major sources of uncertainty in all correction schemes.

3.3.3 Incidence Angle

A final aspect of the geometry to be considered is the angle of incidence between the electron beam and the sample surface. This parameter (often denoted as θ) affects the average depth of the interaction volume. The smaller this angle, the closer the interaction volume to the sample surface. And the closer this volume is to the surface, the shorter will be the absorption path, the greater the measured intensities, and the smaller the required absorption corrections.

3.4 VACUUM SYSTEMS AND CONTAMINATION

A last element of electron column instruments that deserves mention in a discussion of x-ray analysis is the vacuum system. All electron columns must sustain a vacuum in order to accelerate electrons. The quality of this vacuum

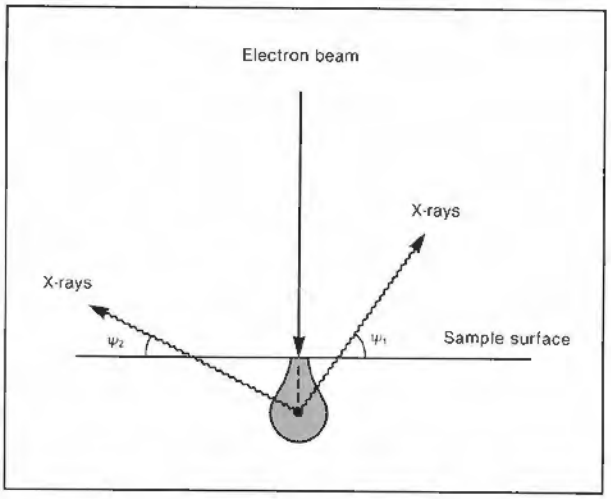


Figure 3-2. Schematic illustration of the take-off angle ψ. For a given angle of electron incidence, the length of the absorption path d *is directly proportional to csc ψ.*

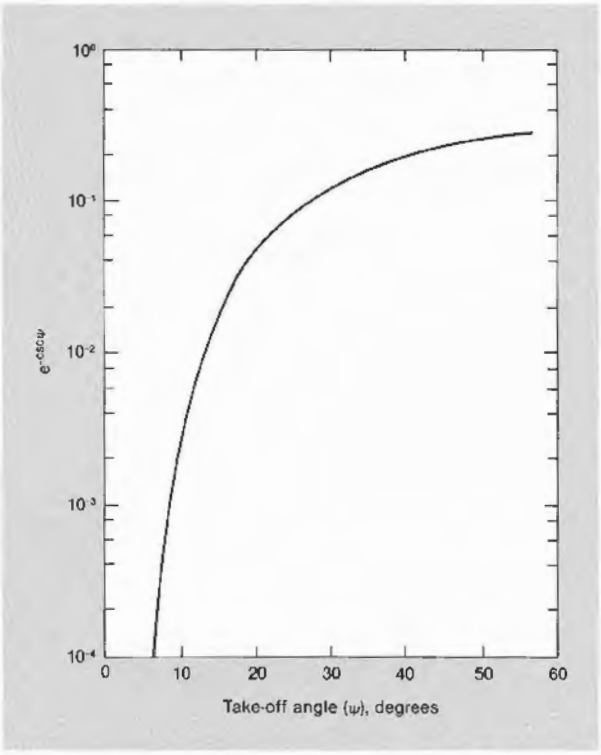


Figure 3-3. *Plot of $e^{-\csc\psi}$ as a function of take-off angle ψ. As expressed in Beer's law, the probability of x-ray absorption is directly proportional to e^{-d}, where d here is proportional to csc ψ.*

affects the quality of the x-ray analysis insofar as it affects system stability and prevents or allows the deposition of contaminants on the sample and on the x-ray detector. Energy-dispersive analysis requires a chamber vacuum of 5×10^{-5} Torr or better, a requirement met by most well-maintained vacuum systems.

Since the x-ray detector must be operated at cryogenic temperatures, it is often the coldest object in the vacuum chamber. It is therefore the first place that contaminants in the environment condense. Sometimes, one even finds visible droplets of contamination condensed on the exterior surfaces of the energy-dispersive detector/cryostat. These contaminants are particularly detrimental to performance when they accumulate on the surface of the detector window, where they absorb incoming x-rays, invalidating theoretical calculations of x-ray absorption.

Contaminants are also deposited on sample surfaces. In fact, the electron beam can actually cause such deposition. Here again, the contamination can serve as an extraneous absorber of x-rays, unaccounted for by quantitative calculations. The effects of absorption are especially pronounced in working with low-energy x-rays from light elements. (In some cases, the elements present in the contamination might be the same as those being analyzed for. The buildup of contamination can therefore actually enhance the signal originating in the sample, again introducing error into the analysis.)

When modern windowless detectors are used, contaminants in the vacuum environment can be deposited on the detector crystal itself and can cause irreversible damage. In windowless systems, particular care must be taken to assure a clean high vacuum.[4]

4

X-RAY INSTRUMENTATION

THE COMPONENTS of a typical energy-dispersive microanalysis system are shown schematically in Figure 4-1. It is the array of components from detector to multichannel analyzer that assembles the information contained in the x-ray signals into a convenient x-ray spectrum. The following paragraphs discuss these elements of the system, starting with the detector.

4.1 THE DETECTOR

All energy-dispersive spectrometers have in common a solid-state detector (Figure 4-2). For microanalysis, this detector is almost always manufactured from a single crystal of silicon. As with other semiconductors, the conductivity of silicon varies greatly, depending primarily on its purity and the perfection of its crystal lattice. In a perfect silicon crystal, there is a place for

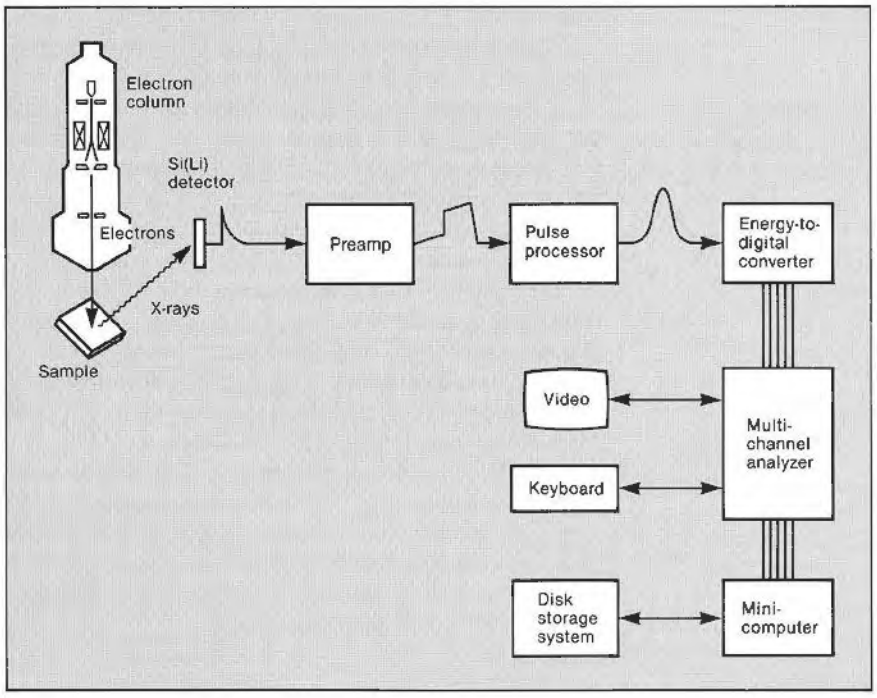


Figure 4-1. Components of a typical energy-dispersive microanalysis system. The charge pulse from the detector is converted in the preamp to a step on a voltage ramp. The pulse processor converts the signal to a well-shaped voltage pulse with an amplitude proportional to the energy of the x-ray.

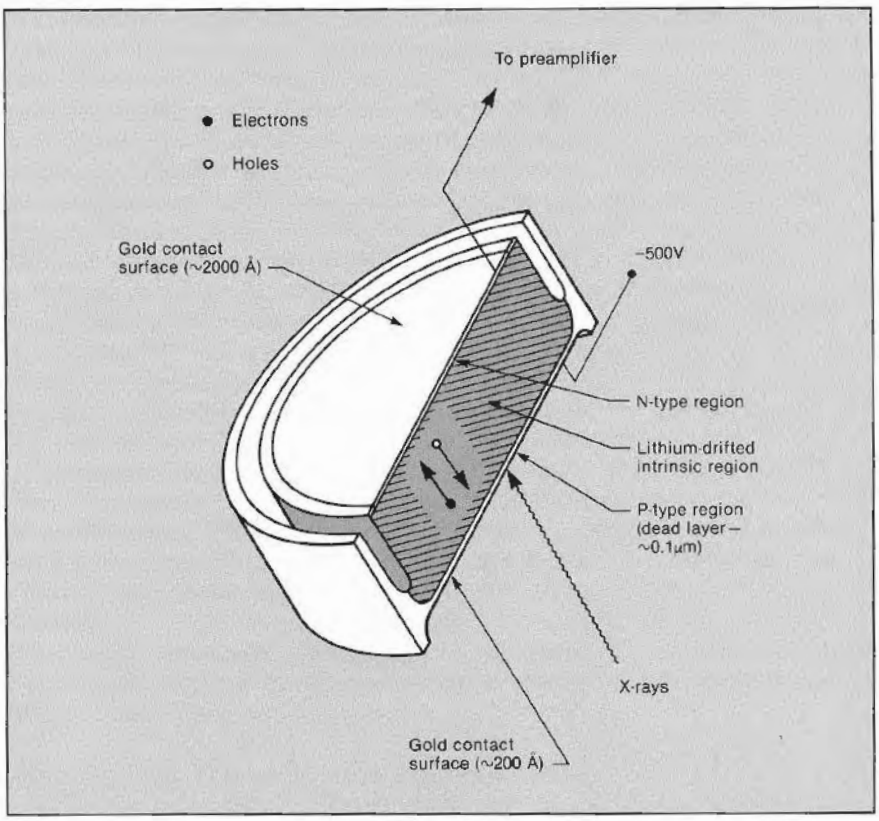


Figure 4-2. Cross section of a typical lithium-drifted silicon detector. X-rays create electron-hole pairs in the intrinsic region of the semiconductor; these charge carriers then migrate to the electrodes under the influence of an applied bias voltage.

every electron and every electron is in its place. Impurities, however, disrupt this perfect structure, creating local abundances or shortages of electrons. The resulting free electrons or *holes* may serve as charge carriers under the influence of an applied electric field. Therefore, whereas a pure and perfect crystal conducts very little current, an imperfect one allows some current to pass. Ideally, a crystal of perfect structure and the highest purity is used for x-ray detection.

4.1.1 The Physics of X-Ray Detection

The silicon atoms making up the crystal are held in the periodic structure of the crystal by a covalent bonding mechanism that essentially shares electrons among the outer orbitals of several neighboring atoms. These shared electrons are said to occupy the *valence band* of the crystal. When an x-ray enters the crystal, there is a high probability that it will be absorbed in an interaction with an electron of one of the silicon atoms, producing a high-energy photoelectron. The ejected photoelectron eventually dissipates its energy in interactions that promote valence-band electrons to the *conduction band*, leaving holes in the once-filled valence band.

Processes other than electron-hole pair generation (for example, heat generation) are involved in the dissipation of the energy deposited by the incoming x-ray. Nonetheless, a good statistical correlation exists between the amount of energy dissipated and the number of electron-hole pairs generated. On the average, 3.8 to 3.9 eV are dissipated in the creation of each electron-hole pair. This low value, relative to the energy of the x-ray (typically thousands of eVs), leads to the good statistical precision available from a silicon detector crystal.

The process of x-ray detection then becomes one of measuring the number of free charge carriers (electrons and holes) created in the crystal during the absorption of each x-ray. The crystal is operated as a reverse-bias diode under an applied voltage of 100 to 1000 volts. Any free charge created within the diode leads to a temporary increase in its conductivity. If the resulting current is integrated with respect to time, the total charge conducted is found to be directly proportional to the energy of the absorbed x-ray.

4.1.2 Leakage Current and Lithium Drifting

Even a perfect semiconductor crystal would be expected to show some residual conductivity upon which the momentary increases caused by x-ray absorption would be superimposed. Such baseline conductivity is due to the random thermal excitation of electrons across the gap between valence and conduction bands. To minimize this thermally induced background, or *leakage current*, detectors are operated at low temperatures. All detectors therefore incorporate a liquid nitrogen cooling apparatus called a *cryostat*.

As we already mentioned, imperfections and impurities in the crystalline structure of the silicon also contribute to the conductivity of the crystal and therefore to the leakage current. In fact, silicon crystals pure enough to maintain the required bias voltage are not readily fabricated. Most contain impurities that cause excess holes to be present as *extrinsic* charge carriers. It is possible, however, to "compensate" for impurities and imperfections by a process known as *lithium drifting*. In this process, lithium atoms are allowed to diffuse into the crystal to compensate for the native impurities in the crystal. The result is a lithium-drifted silicon, or Si(Li), detector.

4.1.3 Spectral Resolution

Because of the complex nature of the interaction of the x-ray with the silicon crystal lattice and the competition among various energy-dissipation processes, the charge pulses associated with the detection of identical x-rays are not necessarily equivalent in magnitude. Instead, they vary statistically about some mean value (see the aside on page 26). For a large number of pulses, the shape of the resulting distribution of values approximates a normal distribution. One indicator of the quality of a spectrometer is the width of this distribution relative to its height. This indicator is referred to as the *spectral resolution* and by convention is measured as the full width of the distribution at one-half its maximum height (FWHM). (It should be kept in mind that spectral resolution reflects the performance not only of the detector crystal but also of other components in the signal-processing chain.) Resolution is also a function of the energy of the x-ray measured. Therefore, resolution is conventionally specified for a given x-ray line and for given conditions of operation (typically the 5.9-keV manganese $K\alpha$ line, at 1000 counts per second and an 8-μsec pulse processor time constant).

4.1.4 Detector Efficiency

Consideration must also be given to detector efficiency. After successfully escaping from the sample and reaching the x-ray detector, an x-ray may remain undetected for two important reasons. First, it may not reach the detector crystal itself. Because of the requirement for high-purity detectors, the crystal must be operated in a very clean, very high vacuum. In conventional XES detectors, therefore, the crystal vacuum is maintained separately from the vacuum of the electron column. This isolation is achieved by enclosing the crystal within a tube, then sealing the end of the tube with a window of some material that is relatively transparent to the x-rays of interest. This window material is almost always beryllium.

Rolled to a thickness of 7.5 μm, beryllium withstands the pressure differ-

ential between crystal and sample environments and transmits x-rays from elements with atomic numbers 11 and greater. (X-rays with energies greater than 2 keV are transmitted by the beryllium window with nearly 100% efficiency—see Figure 4-3.) X-rays from elements lighter than sodium, however, are absorbed by the beryllium window and are therefore undetected. Only windowless or thin-window detectors transmit lower-energy x-rays. There are other barriers to x-rays as well—albeit less important ones than the beryllium window. X-rays may also be absorbed by contaminants on the window, by the conductive layer of gold on the surface of the detector crystal, or by an inevitable *dead layer* of silicon just under the gold layer.

Absorption within the window therefore limits the sensitivity of the x-ray detector to low-energy x-rays. A limit to the detection efficiency for high-energy x-rays also exists. As x-rays increase in energy, there is an increasing probability that they will pass completely through the detector crystal, escaping with at least a fraction of their original energy. The thicker the crystal, of course, the better it is at stopping high-energy x-rays. However, the detector-manufacturing process imposes a practical limit on crystal thickness, so thicknesses of 2 or 3 mm are typical. A 3-mm crystal maintains near 100% detection efficiency to almost 20 keV. Figure 4-3 also illustrates this effect of crystal thickness on detector efficiency.

4.1.5 The Dead Layer

The dead layer alluded to above is a layer at the silicon crystal surface in which neutralization has not been achieved in the lithium drifting process.

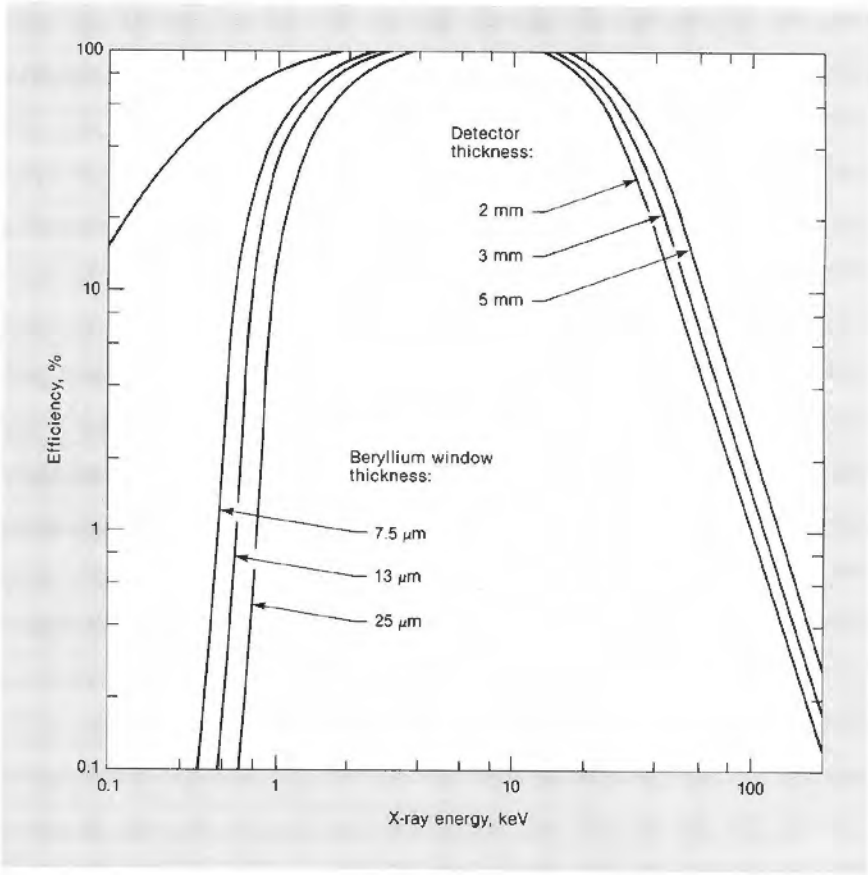


Figure 4-3. *Plot of theoretical detector efficiency as a function of x-ray energy. Curves are shown for three beryllium window thicknesses and three silicon detector thicknesses. The curve at the upper left is the theoretical efficiency of a windowless detector.*

Excess holes therefore remain. The result is the phenomenon of incomplete charge collection, or charge trapping, in which charges created as a result of x-ray absorption may be trapped in the crystal rather than being swept out by the bias voltage to be measured in the charge pulse. The size of the detected charge pulse is therefore reduced by some amount, and the x-ray is assigned some energy lower than its true energy. These reduced energy measurements appear as a "tail" on the low-energy side of the detected peak. The ratio of the FWHM to the FWTM (full width at one-tenth maximum peak height) is sometimes used as an indication of how much low-energy tailing is present.

4.1.6 Escape Peaks

In addition to charge trapping, a second phenomenon sometimes occurs near the surface of the detector crystal. The ejection of a silicon photoelectron by the incoming x-ray is sometimes followed by the emission of an x-ray characteristic of silicon—the same deexcitation process that gave rise to the x-rays in the sample. If this x-ray is subsequently absorbed in the detector crystal, it contributes appropriately to the charge pulse measured for the original x-ray. However, should the silicon x-ray escape, carrying with it a well-defined amount of energy (1.74 keV, the energy of the silicon $K\alpha$ x-ray), the energy measured for the detected x-ray will be less than the actual x-ray energy by exactly that amount. Therefore, as counts accumulate in an x-ray peak for any major constituent of the sample, an *escape peak* can be expected to appear at an energy 1.74 keV below that of the *parent peak*. This escape peak is simply the collection of counts from measurements that included escape events.

Escape peak intensities depend strongly on two parameters, the angle at which the original x-ray enters the detector crystal and the energy of the parent peak. The angle of x-ray incidence influences the average depth at which silicon x-rays are generated. Normal incidence tends to increase this average depth and thus reduce the number of escape events. At the other extreme, grazing incidence increases the number of escape events. Since it is governed by a curve like Figure 2-8, the likelihood of silicon ionization increases as the energy of the exciting x-rays approaches the silicon *K*-shell binding energy (1.84 keV) from above. Consequently, high-energy x-rays are likely to penetrate more deeply than low-energy x-rays before being absorbed. Escape events are thus most likely when the primary x-ray energy is just above 1.84 keV. No escape peaks are observed for parent peaks with energies less than 1.84 keV.

Although escape peaks will be present for all parent peaks above 1.84 keV, escape events are relatively rare under most conditions. Usually, the magnitude of an escape peak is, at most, a few percent of that of its parent peak.

4.2 PREAMPLIFIER

The next step in the signal-processing chain is the preamplifier. It is here that the current conducted by the detector crystal is integrated and amplified. An amplification circuit incorporating a field-effect transistor (FET) is the first stage. Early preamplifier designs incorporated resistive feedback at this stage; however, the electronic noise associated with this technique led to the development of alternative feedback mechanisms. Most commonly used today is a configuration referred to as pulsed optical feedback. In this design, the output of the FET is allowed to range between preestablished limits. Upon reaching the upper limit, a light-emitting diode (LED) shines on the FET and

resets the circuit, capitalizing on the photoelectric response of the transistor.

The output of the amplification circuit, then, is a voltage sawtooth comprising slowly rising linear ramps (representing the detector and FET leakage current), upon which are superimposed step increases (see the signal emerging from the preamp in Figure 4-1). The magnitude of each step is proportional to the integrated current conducted by the detector for each x-ray event. In the interest of reducing thermal and transmission noise, the FET is positioned adjacent to the detector crystal and is cryogenically cooled.

At this point, it helps to introduce the concept of analyzer *deadtime*. To reiterate the analytical problem, the analyst is asked to derive from the number of x-rays measured the concentration of the emitting element. The most straightforward way of doing this is to compare the numbers of x-rays detected from two samples (namely, from the unknown and from a standard of known composition) under identical instrument operating conditions. Therefore, in the classical analysis scheme, the number of x-rays counted from an unknown is compared to the number of x-rays counted from a standard during a given period of excitation. There are certain times, however, during which the analyzer will not record a detected x-ray. During such times, it is said to be "dead." Therefore, two measurements made for equivalent real-time periods may be compared directly only if the amount of the deadtime during those periods is assumed to be the same.

Deadtime is introduced at several stages in the signal-processing chain. One source is the brief period during which the FET is reset by the pulsed optical feedback circuit. The deadtime arising from this source varies from one acquisition to another. For example, a sample that emits 1000 10-keV x-rays per second causes roughly twice as much current to flow through the FET circuit as one that emits 1000 5-keV x-rays per second. The FET circuit should therefore reset itself twice as often and exhibit twice the deadtime. Modern analyzers incorporate deadtime-correction circuitry that automatically accounts for such variations in deadtime. Spectral acquisition is then based upon live-time seconds rather than real-time seconds.

Other preamplifier designs do not cause reset deadtime. One such design is known as dynamic charge restoration. In this scheme, the circuit is essentially reset or restored as each pulse is processed.

4.3 Pulse Processor/Amplifier

The third step in the signal-processing chain is the pulse processor or main amplifier. At this point, the step increases generated by the preamplifier are conditioned for acceptance by an analog-to-digital converter. Two methods are in common use. The first involves an initial differentiation and subsequent multiple integrations of the step signal. The result is a roughly bell-shaped voltage pulse, the height of which corresponds to the magnitude of the step input. The multiple integrations can be thought of as filters designed to remove undesirable frequency components from the signal. The desired information is carried in the dc voltage changes associated with the step outputs of the preamplifier. Any short-duration (ac) variations in the signal level constitute noise. While converting the signal to a form acceptable for digitization, it is desirable to preserve the information contained in the step changes while attenuating or filtering out any noise.

Filters can be characterized by a parameter known as the *time constant*. The larger the time constant, the less sensitive the filter to high-frequency noise

at the input. In the name of accuracy, then, it is desirable to operate at the largest possible time constant. However, the time constant is related to the length of time required for the output of the filter to reach a specified level, given an instantaneous change at the input, so it is also directly related to the time required to process each individual x-ray event. Thus, there is a trade-off between the *rate* at which x-rays can be processed (the count rate capability) and the *accuracy* with which each individual pulse can be processed (spectral resolution).

4.3.1 Time-Variant Processing

In the amplification method just described, the time constant remains the same for both the rising phase and the falling phase of each pulse. However, we can derive the information we need, namely, the height of the pulse, as soon as the pulse reaches its maximum intensity. The time during which the pulse is falling back to a zero level is essentially wasted. During this time, a subsequent pulse cannot be accepted, because it would be added to the level of the declining signal. A second method of pulse processing, known as time-variant processing, has therefore been developed to reduce this wasted time. In time-variant processing, a time constant is applied during the rising phase of the pulse that optimizes the information carried in the signal. Once the pulse maximum has been measured, the time constant is switched to a smaller value, allowing the pulse to fall off more rapidly. Time-variant processors offer a more attractive compromise between resolution and count rate, though current designs suffer some constraints in their use with electron column systems. In particular, they are sensitive to variations in count rate, which unavoidably occur during the raster scan of an inhomogeneous sample.

4.3.2 Pulse Pileup Rejection

Each signal pulse must be measured individually with reference to a zero level and cannot be measured when superimposed upon either the leading edge or the trailing edge of a nearly coincident pulse (see Figure 4-4). *Pulse pileup rejection* is the technique by which nearly coincident pulses are rejected. All pileup rejection circuits depend upon the discrimination of the beginning of a pulse in a so-called fast-channel amplifier. Given knowledge of the time

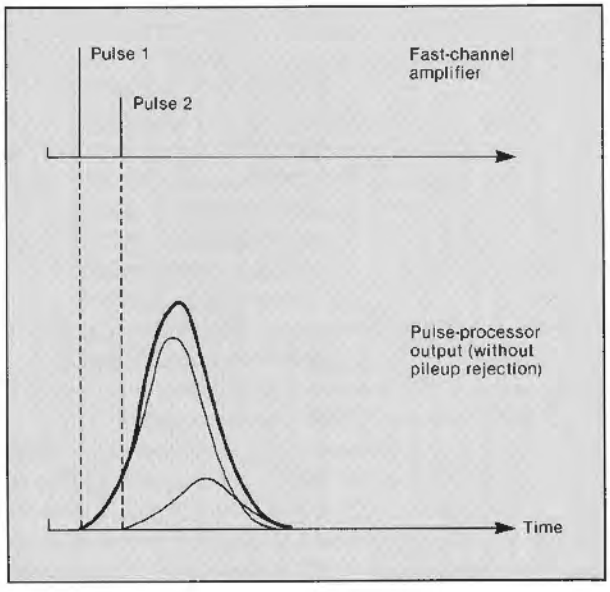


Figure 4-4. *Illustration of pulse pileup. In the case shown here, failure to discriminate (and reject) the two pulses would lead to an anomalously large pulse being digitized and stored.*

constants used in the processing amplifier, it is then possible to calculate when interfering overlaps have occurred. However, because of the requirement for fast-channel discrimination, pulse pileup rejection circuits lose their efficiency at low energies, where the amplitude of the x-ray events approaches that of noise events.

Pulse pileup rejection is another source of analyzer deadtime. In fact, because nearly coincident pulses must be rejected, an increase in the rate at which x-rays enter the detector does not necessarily result in an increase in the rate at which x-rays are accepted and processed. The higher the input rate, the greater the number of rejected pulses. As a rule of thumb, maximum throughput occurs when deadtime is about 60% of real time (Figure 4-5).

4.4 EDC AND MULTICHANNEL ANALYZER

In the energy-to-digital converter, the height of the voltage pulse from the pulse processor (which is proportional to the energy of the detected x-ray) is measured and assigned a channel number. The number of counts in that channel of the multichannel analyzer is then increased by one. The most common energy-to-digital converter used in microanalysis systems is known as a timed capacitive discharge converter. In such a converter, the voltage pulse charges a capacitor, which is then allowed to discharge at a constant rate. The time required to discharge the capacitor is a measure of the height of the voltage pulse. The multichannel analyzer, then, is the means by which the signal information is accumulated and assembled into a spectrum. In addition, control of the display and the spectrometer is usually handled though a video terminal and keyboard associated with the multichannel analyzer.

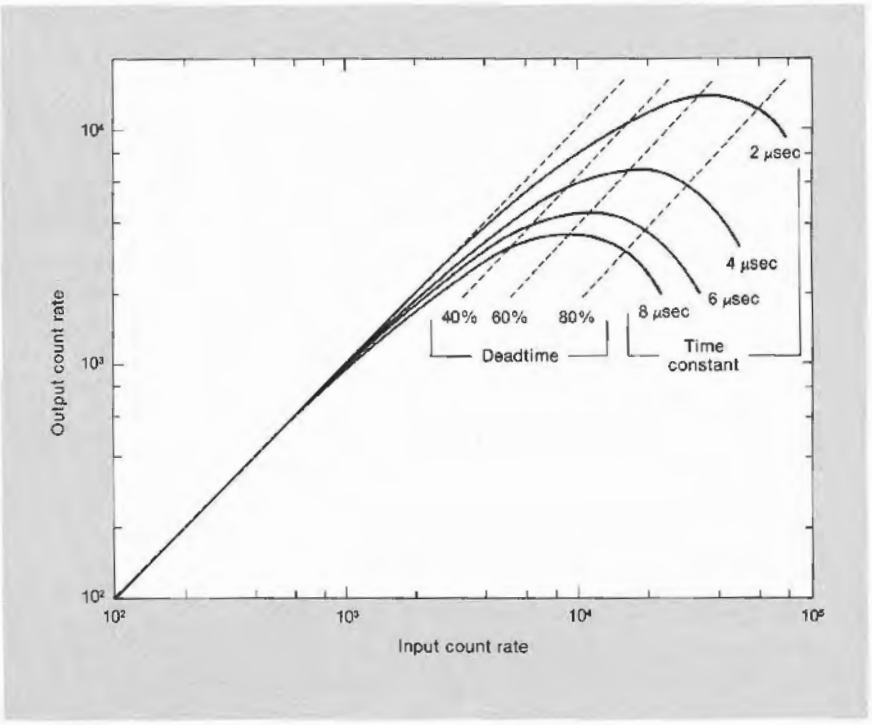


Figure 4-5. Plot of output count rate as a function of input count rate, for four pulse-shaping time constants. Because of the pulse pileup rejection circuitry, an input count rate that produces about 60% deadtime maximizes the output count rate for a given time constant.

ASIDE: STATISTICAL CONSIDERATIONS

Deriving an energy distribution for x-rays emitted by a given sample depends ultimately on assigning an energy value to each detected x-ray. The errors implicit in making this assignment are of two types. The first we call *systematic error,* which includes instrumental errors (such as errors in calibration), errors in technique, errors due to environmental effects, and errors directly attributable to the analyst performing the measurements. To some extent, this type of error is controllable, and we shall assume that it is minimized. In any case, systematic error cannot generally be evaluated by any logical, mathematical means. The second type of error, *random error,* is not controllable; however, its magnitude can be estimated from theoretical considerations.

In light of these observations, this discussion deals with random error, that is, with events of an intrinsically random nature. The processes of x-ray emission and x-ray detection both involve such events. The result is that statistics enters any discussion of microanalysis at two important points—in assigning an energy value to a spectral peak and in evaluating the intensity of that peak.

Normal Distribution and Standard Deviation

The breadth of each peak in an x-ray spectrum (for example, the one in Figure 1-3) indicates clearly that the energy of an individual x-ray cannot be measured exactly. The amount of charge the x-ray generates in the detector is vulnerable to random variations, and the electronic circuitry inevitably contributes noise to the signal. Consequently, a series of energy measurements of x-rays of energy E will form a distribution about a mean value, which we hope is very close to E.

This energy distribution can, for most practical purposes, be assumed to be an example of a *normal (or Gaussian) distribution* (Figure 4-6):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\bar{x})^2}{2\sigma^2}\right]$$

where $\bar{x}$ is the mean value (of energy in our case) and σ is the *standard deviation.* The standard deviation is an indicator of the breadth of the distri-

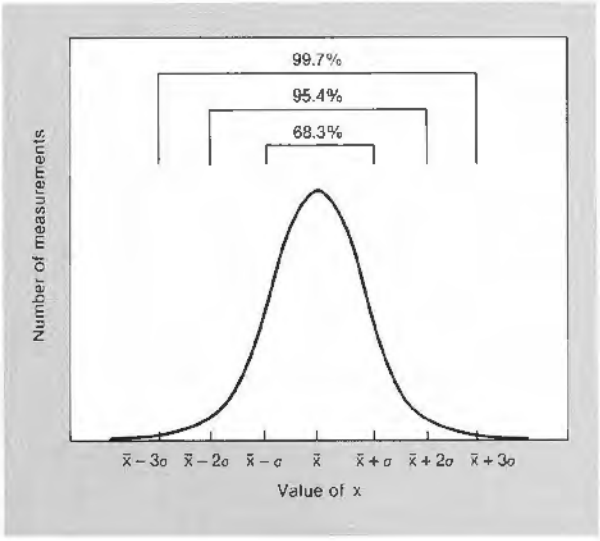


Figure 4-6. The normal distribution function. The number of measurements that fall within one, two, and three standard deviations of the mean are 68.3%, 95.4%, and 99.7% of the total, respectively.

bution. In a normal distribution with a standard deviation σ, 68.3% of all measurements of x fall between $\bar{x}-\sigma$ and $\bar{x}+\sigma$, 95.4% fall between $\bar{x}-2\sigma$ and $\bar{x}+2\sigma$, and 99.7% fall between $\bar{x}-3\sigma$ and $\bar{x}+3\sigma$. One further fact is of particular interest. Note that the value $\bar{x}$ is itself a statistical parameter. If we make a series of evaluations of $\bar{x}$, each based on N measurements of x, the values of $\bar{x}$ will themselves form a normal distribution. This "distribution of averages" is characterized by the *standard deviation of the mean*, which, for N measurements of x, can be expressed as

$$\sigma_N = \frac{\sigma}{\sqrt{N}}$$

This gives us an idea of how close a single measured value of the mean ($\bar{x}$) is to the "true" value of E. If a spectral peak has a standard deviation of 100 eV and is the result of detecting 10,000 individual x-rays, we can take σ as 100 and N as 10,000. The resulting value of σ_N is 1 eV. This gives us considerable confidence that the mean of the measured peak is very close to the true energy of the electronic transition being observed.

Counting Error

In evaluating the intensity of a spectral peak, that is, the number of x-ray counts it comprises, we encounter a source of random error even more fundamental than those we have just mentioned. The emission and subsequent detection of a characteristic x-ray can, taken together, be regarded as a statistically independent event (unrelated to past or future events), which has a fixed probability of occurring within each infinitesimal time interval δt. Under conditions such as these, the number n of x-rays detected during any finite time interval is governed by the *Poisson law:*

$$P(n) = \frac{e^{-\bar{n}}\,\bar{n}^n}{n!}$$

where $P(n)$ is the probability of detecting exactly n x-rays and $\bar{n}$ is the mean number of x-rays counted during a large number of such trials. This equation says that, for a random process occurring at a constant *average* rate, we can, in a finite time interval, only estimate the true average rate. The confidence we have in the accuracy of our estimate can be no greater than that indicated by the breadth of the Poisson distribution—a plot of $P(n)$ versus n—and the inevitable error is called the *counting error.*

The standard deviation of a Poisson distribution is

$$\sigma = \sqrt{\bar{n}}$$

and the *variance* (which we shall need later) is

$$\sigma^2 = \bar{n}$$

Furthermore, for fairly large values of $\bar{n}$, the Poisson distribution can be represented by an appropriate normal distribution. Accordingly, we can say that 68% of all measurements of n lie between $\bar{n}-\bar{n}^{1/2}$ and $\bar{n}+\bar{n}^{1/2}$, that 95% lie between $\bar{n}-2\bar{n}^{1/2}$ and $\bar{n}+2\bar{n}^{1/2}$, and so forth. Clearly, the larger the value of $\bar{n}$, the narrower the distribution relative to the mean (see Figure 4-7). This ob-

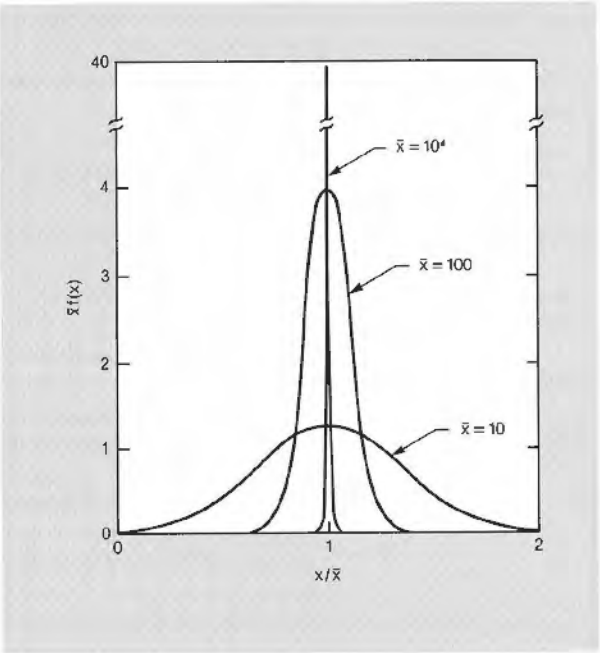


Figure 4-7. *Relative shapes for normal distributions with different means ($\bar{x}$) and with $\sigma = \bar{x}^{1/2}$. The abscissa is such that the width of each distribution is proportional to the relative precision; the ordinate is such that the integral of each is equal to unity.*

servation is reflected in the value of the *relative standard deviation:*

$$\varepsilon = \frac{\sigma}{\bar{x}}$$

or, for Poisson distributions only,

$$\varepsilon = \frac{\sqrt{\bar{n}}}{\bar{n}} = \frac{1}{\sqrt{\bar{n}}}$$

Relative error can also be expressed at higher *levels of confidence* by substituting 2σ or 3σ for σ in the equation for the relative standard deviation (see Figure 4-8).

What all this means is best shown in an example. Ignoring for a moment the problem of evaluating and removing background counts from the spectrum, let us assume that we have a spectral peak representing a single element. If that peak comprises 100 individual x-ray counts (and if we make the reasonable assumption that 100 is fairly close to $\bar{n}$), we can say, at a confidence level of 68%, that the relative counting error is no greater than

$$\varepsilon = \frac{1}{\sqrt{\bar{n}}} = \frac{1}{10} \text{ or } 10\%$$

The relative errors at 95% and 99% levels of confidence are 20% and 30%, respectively. If, on the other hand, our peak had contained 10,000 counts, the relative errors (at 68%, 95%, and 99% confidence levels) would be 1%, 2%, and 3%, respectively.

Finally, it must be borne in mind that the counting error is only a lower limit on the errors involved in quantitative estimates of element concentration based on peak intensities.

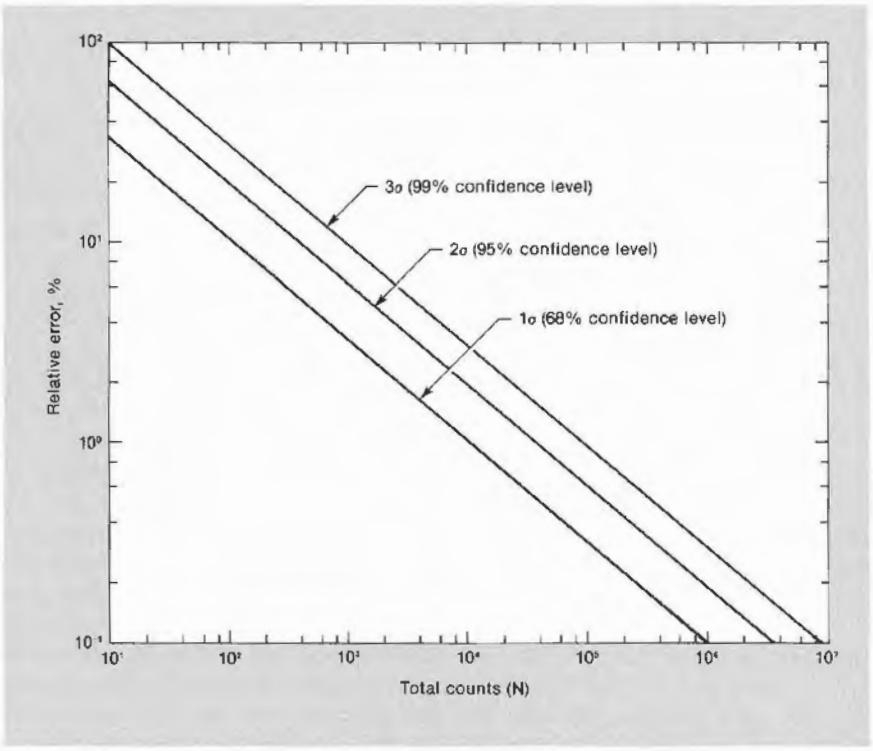


Figure 4-8. *Plot of relative error as a function of total counts for three levels of confidence.*

Minimum Detection Limits

We encounter another important statistical consideration in trace analyses, where instrumental capability is being strained in merely determining whether an element is present or not. In fact, at sufficiently low concentrations, one can only conclude that "if element *X* is present at all, its concentration must be less than. . . ." This limit is the *minimum detection limit (MDL).*

In microanalysis we are concerned with measuring the net peak intensity, that is, the intensity of the characteristic x-ray signal above the background signal. As we mentioned above, there is counting error in any measurement of peak intensity. Likewise, the background signal itself is susceptible to counting error. Therefore, the microanalyst is confronted with the problem of distinguishing between random fluctuations in the background and real peaks. Furthermore, the microanalyst must establish a confidence level to be maintained in any assertion that an element is present at the MDL. For example, a 95% confidence level would be consistent with the statement that, in a large number of observations, 95% of the observations indicating the presence of an element at the MDL reflect the actual presence of that element, whereas 5% of such observations reflect only random fluctuations in background count rate. Ninety-five percent is a typical confidence level. Conveniently, 95% confidence may be obtained if the criterion for peak presence is set as a "fluctuation" greater than two standard deviations above the expected average intensity.

The extent of the random background fluctuations can be derived from the Poisson law discussed above. If a region of interest is established, we can therefore assess the probability that the number of background counts in that region will differ from the mean by some specified amount. And again, the magnitude of the random fluctuations is a function of the total number of

counts in the region of interest; specifically, the standard deviation for the background counts is

$$\sigma_b = \sqrt{\bar{n}_b}$$

where the subscript b indicates that we are now talking about the background.

In practice, MDLs are influenced by a number of experimental factors, including instrument stability, spectral peak overlaps, and interactions within the sample matrix. However, in an ideal case—that of an unobstructed peak on a smoothly varying background—a theoretical MDL can be established. First, we need to know that the variance of the sum of, or the difference between, two values taken from statistically independent distributions is equal to the sum of the variances of the two distributions. Thus, for net counts,

$$\sigma^2_{\text{net}} = \sigma^2_b + \sigma^2_{\text{total}} = \bar{n}_b + (\bar{n}_b + \bar{n}_{\text{net}}) = 2\bar{n}_b + \bar{n}_{\text{net}}$$

where the subscript *total* refers to the total counts in a region of interest. This must be so, because the number of net counts is computed as the difference between the total number of counts and the estimated number of background counts. Now we can ask, "How many net counts must we detect to be sure (or 95% sure) that we are not merely seeing a statistical fluctuation in the background?" The answer is that the number of net counts must exceed twice the standard deviation of net counts. (Otherwise, there is at least a 5% chance that the "net counts" we observe arise merely from statistical fluctuations in the total counts and the estimated background counts.) This requires that

$$\begin{aligned} n_{\text{net}} &\geqslant 2\sigma_{\text{net}} \\ &\geqslant 2\sqrt{2\bar{n}_b + \bar{n}_{\text{net}}} \end{aligned}$$

or, assuming that measured counts (n) are close to the respective means ($\bar{n}$) and that, for small net peaks, $n_b \gg n_{\text{net}}$,

$$n_{\text{net}} \geqslant 2\sqrt{2\bar{n}_b} \cong 3\sqrt{\bar{n}_b}$$

where, once again, n_{net} is the number of computed net counts and n_b is the number of computed background counts. The MDL is the concentration corresponding to n_{net}.

Since the MDL is a function of counts, it is also a function of counting time. The size of the net peak increases linearly with acquisition time and must eventually exceed $3n_b^{1/2}$ (which increases more slowly), whatever the concentration of the element in question. Obviously, however, there is a practical limit to increasing the acquisition time as a means for lowering the MDL. Nonetheless, MDLs as low as 0.01% are feasible under certain conditions. Furthermore, the microanalytical aspect of the electron probe device yields a detection limit in terms of absolute amounts (the *mass limit*) that is very low—under the best analytical conditions, as low as 10^{-15} to 10^{-16} grams. More extensive treatments of MDLs are available in References 5 and 6.

5

ANALYSIS

THE FINAL PART of our discussion turns to analysis—the job that actually faces the microanalyst after a raw spectrum has been acquired. Most aspects of analysis are automated on modern systems, but the analyst must still make informed choices among available routines, and the options vary from system to system. As we look at a few of the popular alternatives, the tone of the discussion will become decidedly more practical.

As we saw in Figure 1-3, an energy-dispersive spectrum is usually displayed as a histogram, with the horizontal axis labeled in energy units and the vertical axis in numbers of counts or intensity. Figure 5-1 shows a portion of another x-ray spectrum, more clearly showing several typical features. The most obvious are the large characteristic peaks for iron, chromium, and nickel—the components of the sample. For each of these elements, both a $K\alpha$ and a $K\beta$ peak are present, though the nickel $K\beta$ peak is beyond the right edge of the screen. Iron and chromium escape peaks were discernible features of the original spectrum, but they have been removed and are now displayed along the baseline. The largest ones were located 1.74 keV below the corresponding $K\alpha$ parent peaks. The roughness of the overall spectrum represents channel-to-channel statistical fluctuations. Finally, all of these features are superimposed on a bremsstrahlung spectrum that falls slowly from left to right.

A feature not illustrated in the spectra of Figures 1-3 and 5-1 is the *sum*

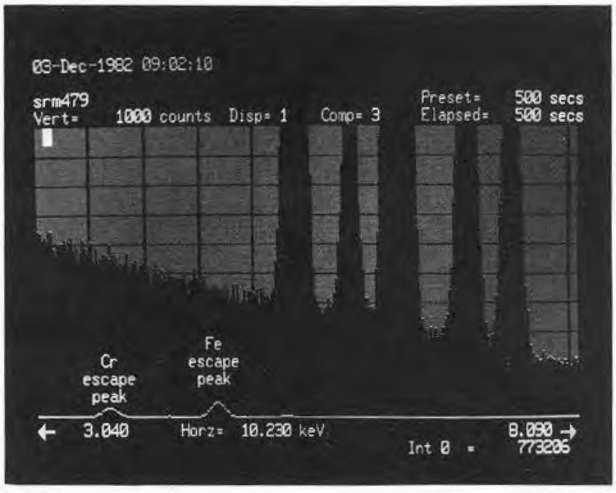


Figure 5-1. A portion of an x-ray spectrum (3.04–8.09 keV) for a stainless-steel specimen. The off-scale peaks are the Kα and Kβ peaks for iron and chromium, and the Kα peak for nickel. (The nickel Kβ peak is beyond the right edge of the display.) Escape peaks have been removed by an algorithm that calculates their positions and intensities and then adds those intensities back into the corresponding parent peaks.

peak, which results from the exact coincidence (or something very close to it) of two detected x-rays. The pulse pileup rejection circuitry has some definable limitations as to the minimum separation between two distinguishable pulses. Pulses separated by less than this minimum are processed as a single pulse with a magnitude equal to the sum of the individual pulses. In a spectrum of a pure-element sample (as in the titanium spectrum of Figure 5-2) sum peaks can sometimes be found at energies that are the sums of major peak energies. In spectra with more than a few major peaks, the combinations and permutations of x-ray energies lead to many theoretical sum peaks, and the size of each peak is consequently relatively small.

The probability of an event occurring that will produce a sum peak is proportional to the product of the count rates for the two parent peaks. Therefore, the size of a sum peak varies as the square of the count rate. Although sum peak corrections can be made on this basis, it is usually easier simply to avoid them by reducing the count rate, either by reducing the electron beam intensity or by increasing the distance between sample and detector.

5.1 QUALITATIVE ANALYSIS

Qualitative analysis is the process of identifying which elements are present in a sample. As suggested in our discussion of minimum detection limits, qualitative analysis has as its goal a statement of the form, "Elements *X*, *Y*, and *Z* are definitely present in the sample; if other elements are present, they must be present at concentrations less than the MDL." MDLs must always be kept in mind during qualitative analysis.

In its simplest form, qualitative analysis proceeds by determining the energies of peaks present in the spectrum and comparing them with a chart listing the known energies of x-ray emissions. Modern analyzers have automated this process to varying degrees, and most provide markers that can be called to the video display by atomic number or symbol. In highly automated versions, software routines detect the location of spectral peaks, compare them with tabulated energy values, check for inconsistencies (for example, an apparent $K\beta$ peak but no corresponding $K\alpha$), then print out a list of the elements

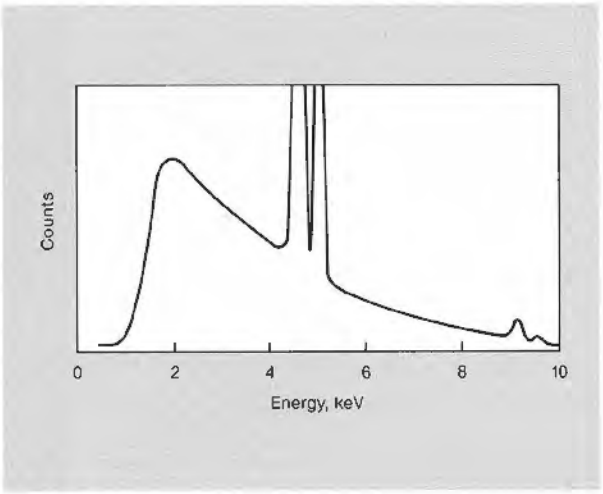


Figure 5-2. *A titanium spectrum obtained at a high input count rate. Sum peaks are visible at 9.02 keV (Kα + Kα) and 9.44 keV (Kα + Kβ). The sum peak for Kβ + Kβ is too small to be seen.*

present. In general, however, routines of this type are not intended to make sophisticated judgments, but rather to limit the number of judgments required of the user.

5.1.1 Removing Escape Peaks

Before even a qualitative identification is attempted, escape peaks should be removed from the raw spectrum. We covered the origin of these peaks in the discussion of detectors. To account for them, the analyzer computes their theoretical intensities, based on parent peak intensity, parent peak energy, and system geometry, then removes them and adds the removed counts to the parent peak. The spectrum of Figure 5-1 has been processed in this way.

5.1.2 Peak Overlap

The greatest source of error—or at least uncertainty—in qualitative analysis can be found in those spectra that contain peaks assignable to more than one element. Peaks of such elements are said to overlap. A notorious example, shown in Figure 5-3, is the overlap between barium and titanium. Titanium *K* lines appear at about 4.5 and 4.9 keV, whereas the strongest of the barium *L* lines are at about 4.5 and 4.8 keV. Confronted with the spectrum shown in Figure 5-3, the analyst can confirm the presence of barium on the basis of the multipeak pattern, but information about relative emission intensities is required to establish that titanium emissions contribute (or do not contribute) to the two most intense peaks. Ideally, element markers presented on the display carry this relative peak-height information.

This barium-titanium example also illustrates the dependence of MDLs on peak overlap. In the theoretical derivation of MDLs described on page 30, the implicit assumption was made that the peaks were to be distinguished only from a continuum background—an assumption that is obviously not valid here. Hence, the MDL for titanium in a sample containing barium is considerably higher than that in a sample not containing the interfering element.

5.1.3 Effect of Accelerating Voltage

When making qualitative determinations based on relative line intensities, it is necessary to remember the influence of accelerating voltage. An excellent example is provided by a comparison of copper spectra acquired at 10 and 20 keV (Figure 5-4). At 10 keV, the copper *K* lines at 8.04 and 8.91 keV are not efficiently excited, in contrast to the copper *L* lines at just below 1 keV. The ratio of the line intensities changes dramatically as the accelerating volt-

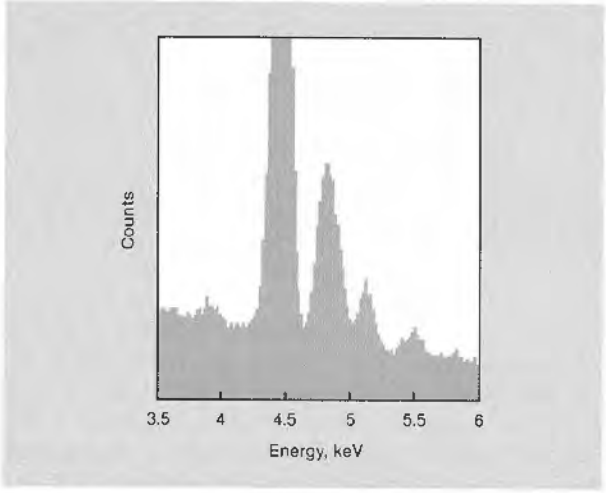


Figure 5-3. A spectrum for benitoite ($BaTiSi_3O_9$), showing the overlap of titanium K lines and barium L lines. The five-peak pattern is characteristic of barium, but the presence of titanium must be inferred from relative peak heights.

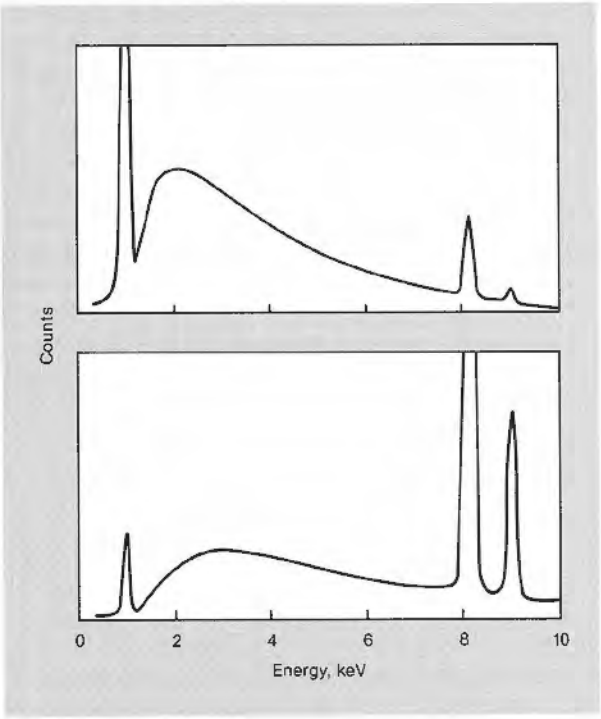


Figure 5-4. Spectra of pure copper taken at accelerating voltages of 10 keV (upper spectrum) and 20 keV (lower spectrum). At 10 keV, only the L *lines are efficiently excited* (K_{ab} = *8.98 keV).*

age is increased to 20 keV. These spectra are good illustrations of the "overvoltage rule" mentioned on page 15.

5.1.4 Line Profiles, Dot Maps, and Spatial Resolution

The results of a qualitative analysis may be presented conveniently in two graphic formats. The first is referred to as an elemental *line profile.* It is usually obtained by making multiple exposures of the electron column CRT display on a single piece of film. The first exposure creates the electron micrograph. The second singles out one line of the scanned raster for detailed examination. During the third exposure, the chosen line is scanned again very slowly, and an x-ray signal (for a selected energy interval) is acquired at each raster point. The intensity of the x-ray signal from the energy region of interest is used to modulate the deflection of the CRT beam in the y direction. In the example shown in Figure 5-5, a line profile for silicon is superimposed on a micrograph of an integrated circuit, where aluminum has been deposited on the silicon substrate.

Another useful graphic format is the *dot map* (Figure 5-6). In this technique, the brightness of the SEM CRT beam at each point on the display is modulated by the x-ray output from the element of interest. A convenient feature of many analyzers is the ability to "remove" background counts from a dot map. The emission of continuum x-rays is essentially random with respect to time, producing a random low-density spatial distribution of events on the dot map. The most commonly used technique for suppressing these background counts is to set a count-rate discriminator on the mapping module output. X-ray counts are then output only when they are received at the mapping module with a frequency above the specified level.

While on the topic of line profiles and dot maps, a final word of caution is in order with regard to x-ray spatial resolution. As we saw in Figure 2-4, the spatial resolution of secondary electrons is much higher than that of x-

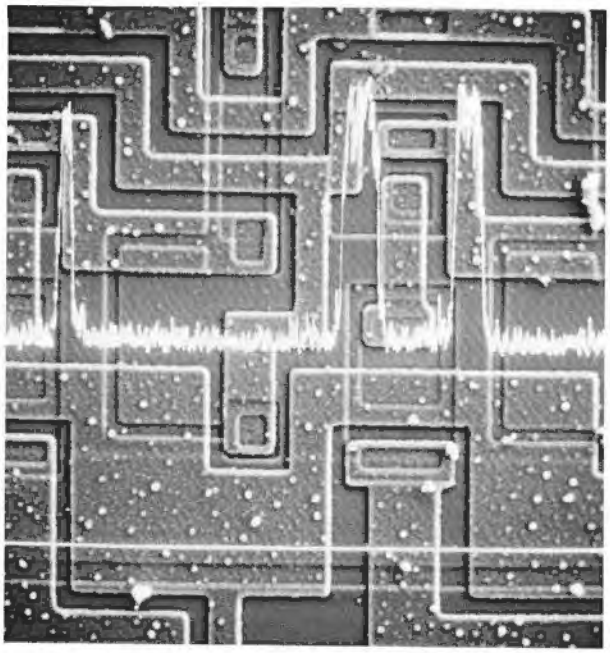


Figure 5-5. A silicon line profile for a portion of an integrated circuit. The upper trace shows the relative intensity of silicon x-rays detected as the primary beam scanned the straight line in the lower half of the photomicrograph.

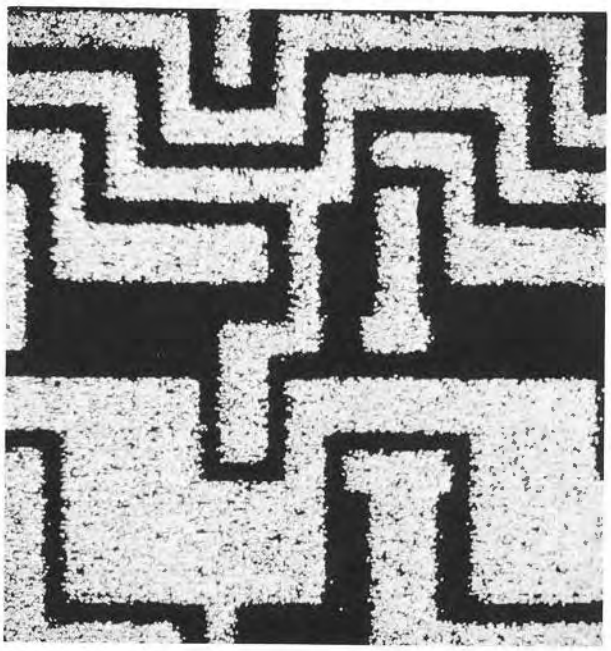


Figure 5-6. An aluminum dot map for the same region of the integrated circuit shown in Figure 5-5. The density of dots reflects the relative concentration of aluminum.

rays. As a result, a feature large enough to be seen in the secondary electron image may not be large enough to contain the entire interaction volume from which x-rays emanate. Often, the electron beam penetrates an observed feature, causing x-rays to emanate from subsurface regions or nearby regions having compositions different from that at the point of surface incidence.

Both the accelerating voltage of the electron column and the mean atomic number of the analyzed sample volume influence the spatial resolution of

the x-ray signal.[7] Figure 5-7 shows qualitatively the effects of both on the size and shape of the volume of interaction. To allow a numerical estimate of quantitative x-ray spatial resolution, the nomogram in Figure 5-8 was devised. (For the purposes of this figure, quantitative x-ray spatial resolution is defined as the diameter of the volume of interaction that yields 99% of the x-rays produced.) One technique for improving spatial resolution is sectioning the sample into slices much thinner than the depth of interaction would be in a bulk sample of similar composition. As Figure 2-4 shows, removing the lower portion of the interaction volume greatly improves spatial resolution. In fact, in thin films, the spatial resolution for x-rays approaches the diameter of the electron beam. Modern STEMs are designed to take advantage of this improvement in resolution.

5.2 QUANTITATIVE ANALYSIS

Quantitative analysis seeks to establish not only the identities of the elements present in a sample, but also their concentrations, together with an indication of the confidence that can be placed in the computed results. Assuming that a qualitative analysis has been concluded, the quantitative analysis must proceed through several phases: background removal, deconvolution of overlapped peaks, and calculation of elemental concentration.

5.2.1 Background Removal

Bremsstrahlung background can be removed in a variety of ways, ranging from simple linear interpolation to theoretical modeling. Each has its own merits and disadvantages.

The simplest and most straightforward approach to removing background from a region of interest is to represent that background as a linear inter-

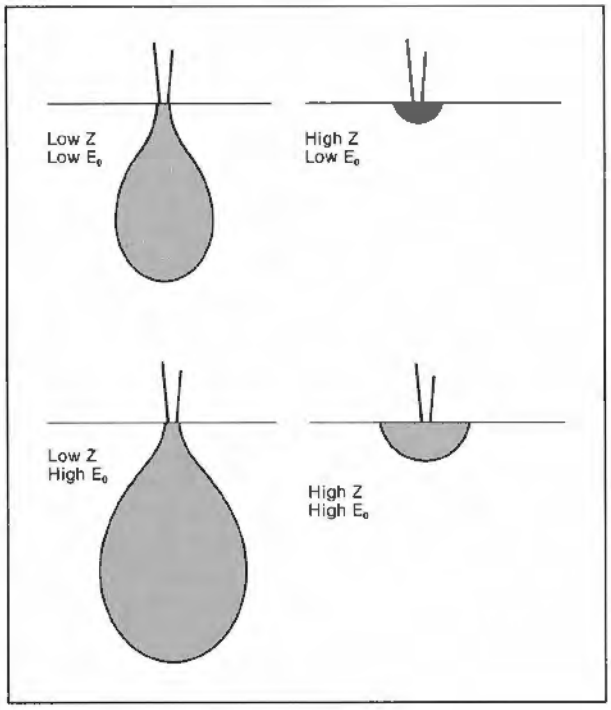


Figure 5-7. Schematic depiction of the variation of interaction volume shape with average sample atomic number (Z) and electron beam accelerating voltage (E_0).

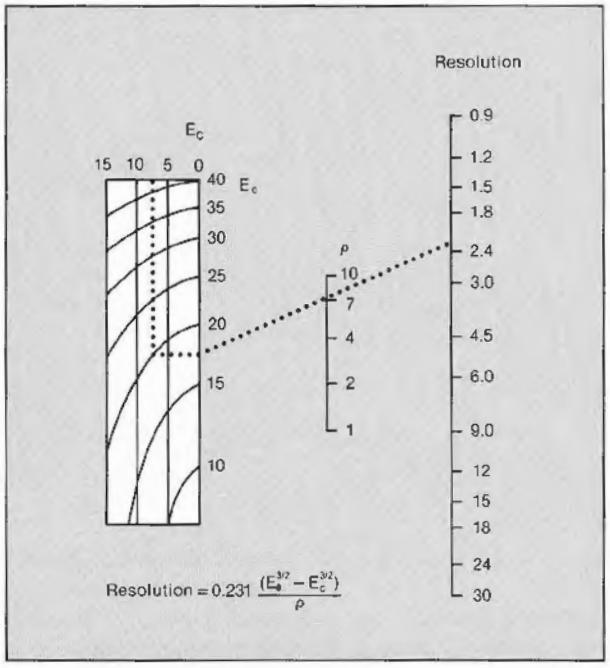


Figure 5-8. *Nomogram for calculating x-ray spatial resolution (in μm) from the accelerating voltage E_0 (in keV), the critical excitation voltage E_c (in keV), and the mean sample density ρ (in g/cc). The critical excitation voltage is numerically equivalent to the absorption edge energy for the element of interest. In this example, E_c is 7.5 keV, E_0 is 20 keV, and ρ is 7 g/cc. The diameter of the expected interaction volume is about 2.3 μm. (Adapted from Reference 8.)*

polation between background areas adjacent to the peak. This technique has the advantage of being extremely fast; however, unobstructed background adjacent to each peak is not always available, and estimates of the proper end points for the interpolation are difficult to make with any accuracy. An improved method does not require that the interpolation be linear. When using such an improved technique, the analyst chooses a set of points on the background of the spectrum, then asks the analyzer to fit some curve to the assigned points. This method is not as fast as linear interpolation, but it generally yields better background fits. Nonetheless, it still involves a great deal of subjective judgment, which can degrade the reproducibility of the results.

Finally, several available algorithms calculate a theoretical model of background shape, including absorption edges, then normalize the theoretical model to areas of the spectrum known not to include characteristic peak information (see Figure 5-9). This technique has proved reliable and highly accurate. It adequately accounts for most known phenomena associated with the continuum background. It has the disadvantage of being slower than other background removal techniques.

ASIDE: BACKGROUND FILTERING

Filtering is another technique by which background is effectively suppressed. In this case, the result often bears little resemblance to the original spectrum (see Figure 5-10), but filtered spectra are often suitable subjects for quantitative analysis. The technique relies on the distinction between the slope of the spectrum in regions where only background is present and the slope in the vicinity of characteristic peaks. If the spectrum is visualized as a function of time rather than energy, an analogy can be drawn between the background filter and frequency filters. Three frequencies of information are present: low-

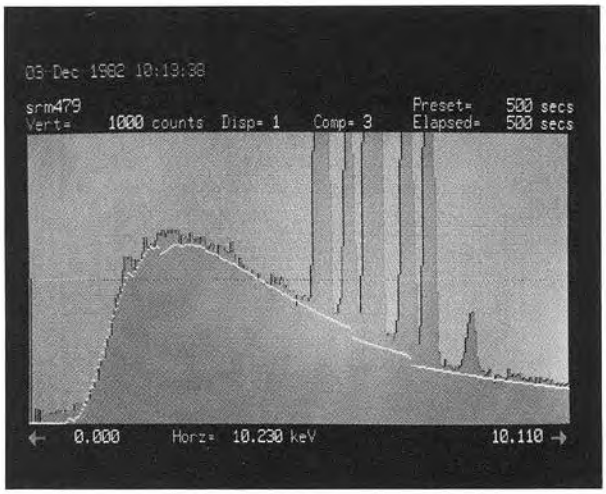


Figure 5-9. *A stainless-steel spectrum, showing the theoretical background computed by a background-modeling routine. Absorption edges for iron and chromium are clearly visible in the theoretical model.*

frequency background information, medium-frequency characteristic peak information, and high-frequency channel-to-channel statistical fluctuations. The goal is to design a filter that attenuates the low- and high-frequency components while passing the medium-frequency information with minimum perturbation. A digital filter often referred to as a *top-hat filter* has the desired effect. It produces a filtered spectrum, each channel of which contains a value obtained by "averaging" the values from several contiguous channels in the original. The average is not an arithmetic mean, but rather a weighted average that depends on the shape of the "top hat." Figure 5-11 shows an example of the use of such a filter on a Gaussian peak. It can be shown mathematically that the characteristic peak information is preserved largely intact, even though the appearance of the spectrum is significantly altered.

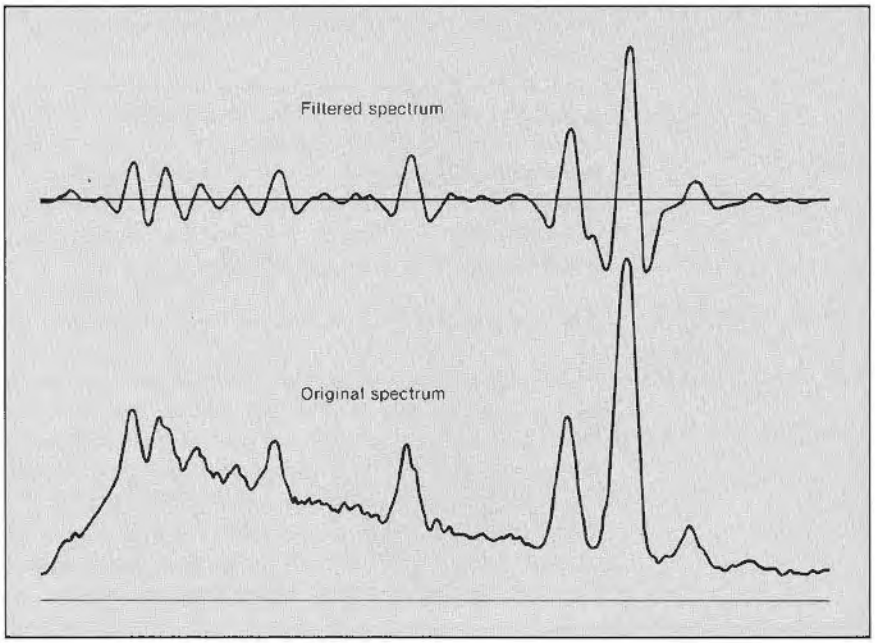


Figure 5-10. *Raw and filtered spectra of a fairly complex sample. The filter used was the top-hat filter of Figure 5-11. (Redrawn from Reference 9.)*

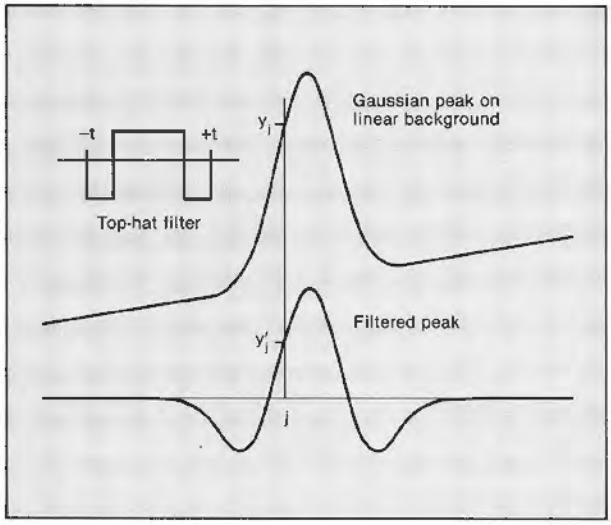


Figure 5-11. *Gaussian peak on a linear background of positive slope (upper curve) and the result of digital filtering (lower curve). For each channel* j *of the filtered peak,* y'_j *was computed as*

$$y'_j = \sum_{s=-t}^{t} f_s\, y_{j+s}$$

where f_s *is a coefficient derived from the channel* s *amplitude of the top-hat filter,* y_{j+s} *is the value in channel* j + s *of the original spectrum, and* s *ranges from* −t *to* +t. *(Adapted from Reference 9.)*

5.2.2 Deconvolution

Once spectral artifacts have been corrected for and the background removed, the peaks remaining are referred to as *net peaks*. The next step in the quantitation process is the evaluation of their intensities. This task is straightforward when there are no overlaps: A *region of interest (ROI)* for each element is defined and simply integrated. But the matter is far from simple when the peaks must first be separated. *Deconvolution* has come to mean any of several techniques used to derive the relative contributions of constituent peaks to an unresolved composite spectral peak. Again, several techniques are in current use.

Overlap Coefficients

If an x-ray peak overlaps the ROI of another element, the fraction of the peak that falls within the ROI depends only on the relative position of peak and ROI and on the standard deviation of the peak, but not on the composition of the sample (see Figure 5-12). Likewise, the fraction of the peak within its own ROI is independent of composition. Therefore, the ratio between these two fractions, called the overlap factor or the *overlap coefficient*, can be computed once, then stored for all subsequent analyses involving the same elements.[10,11]

Since each peak contributes counts to the other, an iterative approach is required in evaluating the peak intensities based on overlap coefficients. The number of counts in each ROI is measured and taken as a first estimate of the true intensity of the corresponding peak. The contribution of each peak to the other is then computed by use of the appropriate overlap coefficients. These estimated contributions are subtracted from the peaks, providing second estimates of intensity. From these revised estimates of intensity, improved values of the interfering contributions are derived, and so on.

The use of overlap coefficients, though simple and fast, has some drawbacks. Looking at Figure 5-12, one concludes that the technique is highly dependent on accurate calibration of the spectrometer. Even a minor shift in the relationship between ROI and peak position introduces errors in the peak intensity measurements. In modern instruments, this problem can be addressed with automatic calibration routines; however, these routines can be time-consuming and inconvenient. Furthermore, calibration to within a few

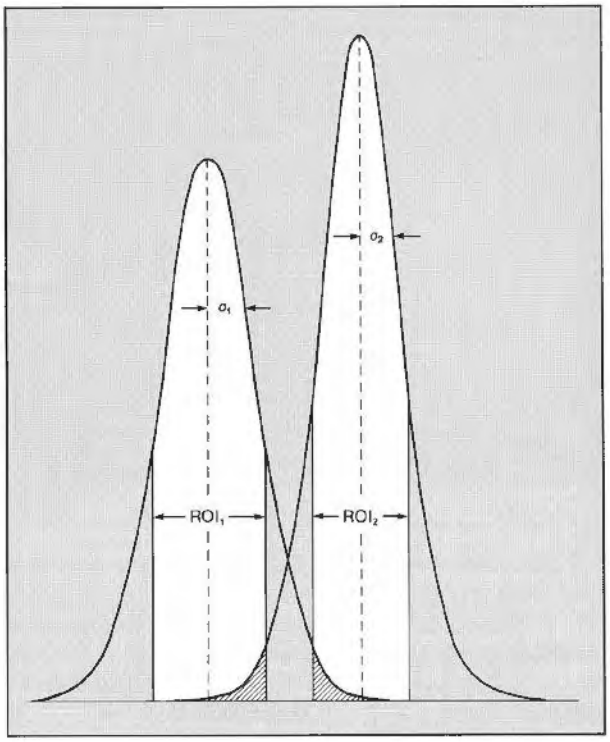


Figure 5-12. Two overlapping Gaussian peaks. Each crosshatched region represents counts contributed by one of the peaks to the total observed in the other's region of interest (ROI). The fractional size of a peak that falls within the ROI of another depends only on the relative position of peak and ROI and on the standard deviation of the peak, not on the size of the peak. (Adapted from Reference 11.)

eVs is sometimes required, and calibration shifts of this magnitude are difficult even to detect.[12] Another drawback is the difficulty of determining coefficients for peaks other than *K* lines; therefore, the accuracy of the technique is questionable for elements whose *K* lines are not easily excited.

Reference Deconvolution

Other deconvolution techniques depend on subtracting a model of the interfering peak from the analyzed composite peak. The simplest of these approaches uses as the model an acquired peak for the interfering element. This peak is most easily acquired from a pure-element standard for the element in question. This peak is then normalized to the unknown and subtracted (Figure 5-13). Normalization neglects the mutual contributions of each peak to the other; however, when the modeled peak is very much the larger of the two, the normalization error caused by the contribution of the smaller

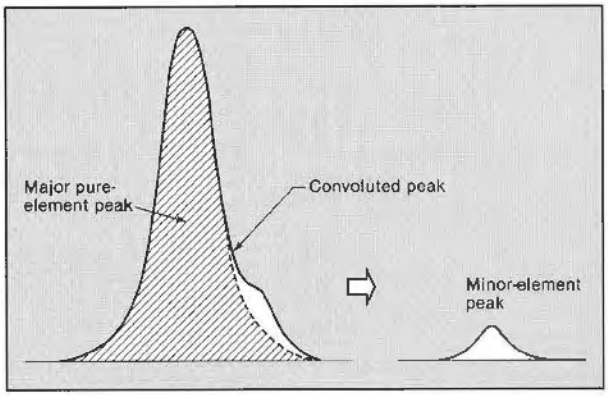


Figure 5-13. A simple example of reference deconvolution. The solid line at the left depicts a pair of overlapped peaks. A pure-element spectrum for the major component (crosshatched area) is acquired, normalized to the sample peak, and subtracted, leaving the minor overlapping peak.

peak is usually small. This normalization-and-stripping technique is also vulnerable to calibration shifts.

An elaboration of the reference deconvolution technique entails a more involved fitting of peak models to the data. The models may be computed theoretically or obtained experimentally from materials exhibiting "clean" peak structure. The most common fitting procedure adds the models together in varying proportions until the result best fits the data from the unknown. The goodness of fit is evaluated using a chi-square (χ^2) test, where χ^2 is usually evaluated as

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - x_i')^2}{x_i}$$

where x_i and x_i' are the intensities in channel i for the unknown and the fitted model, and n is the number of channels. (Before χ^2 can be evaluated, the two spectra must be normalized on the basis of total integrals.) The smaller the value of χ^2, the better the fit. The idea is that when χ^2 is minimized, the proportions in which the peak models were added together reflect the proportions in which the constituent peaks are present in the overlapped data.

Deconvolution by peak fitting has become more popular as small computers have proliferated. It is a procedure easily performed on the typical dedicated micro- or minicomputer.

Filtered Least-Squares Fitting

Another variation on peak-fitting deconvolution routines, combining background filtering and reference deconvolution in a single operation, is known as *filtered least-squares fitting (FLS)*. Peak models are first derived from acquired spectra, then the background is suppressed in each by the filtering technique discussed in the aside on page 37. The unknown spectrum is also filtered. The filtered models are then fitted to the filtered unknown to minimize χ^2. FLS methods have the advantage of speed; however, as with other techniques, they are vulnerable to calibration shifts. In fact, it has been shown that shifts as small as 0.5 eV can introduce significant errors into the deconvolution procedure.[12] Another drawback is the requirement for acquired standards spectra for each element to be deconvoluted. Moreover, the standards spectra must have clean structure, with no overlap in the ROI of the element analyzed. This criterion is not always easily met.

ASIDE: NONLINEAR TECHNIQUES

Changes in calibration are sources of error in all the peak-fitting methods we have discussed. The origins of these changes can be conveniently divided into two categories, depending on the time scale over which they are observed. *Drift* describes those long-term changes that, for the most part, can be taken care of by periodic instrument recalibration. Drift arises from such things as detector and electronics aging, changes in the environment, and so forth. *Shifts*, on the other hand, are functions of the operational characteristics of the electronics. (Most pulse processor/amplifier specifications include maximum values for peak-shift and resolution variations as functions of count rate. Typically, peak shift is specified as less than 5 eV over the usable range of count rates. Resolution may also change by several eVs.) The salient point

is that no calibration will eliminate these shifts. They are real-time variables of the experimental conditions. To take account of these problems, another fitting technique is available. This improved procedure varies not only the relative amplitudes (peak heights) contributed by each model to the final fitted model, but also the widths and mean energies of the individual models.

The shape of an x-ray peak may be described by a function of the general form

$$f(x) = A \exp\left[\frac{-(x - c)^2}{R^2}\right]$$

where A is the amplitude, c is the position of the peak centroid, and R is some measure of the peak width or resolution. (See the aside on page 26 for

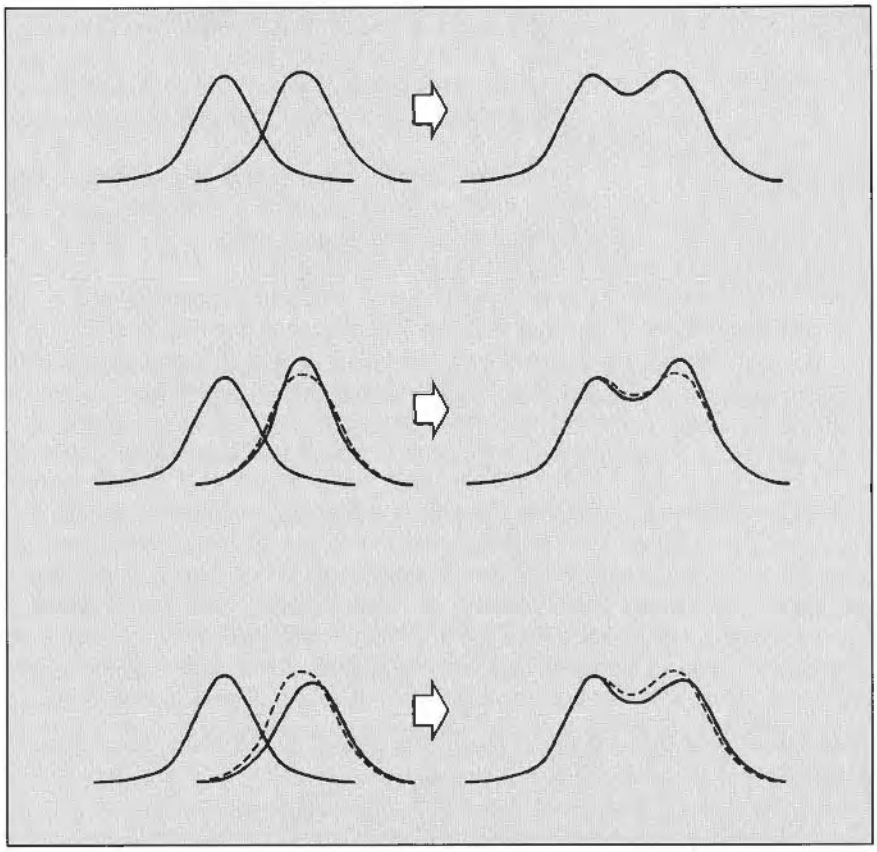


Figure 5-14. *Two examples of peak-fitting results, showing the importance of accounting for peak shifts and changes in resolution. The curve at the upper right is a composite of the two Gaussian peaks 1 and 2 shown at its left. This composite and its Gaussian components reappear as dashed lines in the remaining two examples. In both of these remaining examples, a peak-fitting routine attempted to fit the composite by varying only the amplitude of two Gaussian peaks. In both examples, the centroid position and peak width of one peak were identical to those of peak 1. Goodness of fit was measured by the value of* χ^2*. In the example illustrated by the middle pair of curves, the second peak was assigned the centroid position of peak 2, but its width (as measured by its standard deviation or its FWHM) was smaller. In the bottom pair of curves, the peak width was correctly specified, but the centroid position was shifted to the right.*

a discussion of normal distributions.) Since a variation in c or R has a nonlinear effect on the function that describes the peak, methods that allow these parameters to vary are referred to as nonlinear techniques. The importance of these variables in assuring the best possible fit can be seen in the examples of Figure 5-14. Unless the widths and positions of the model peaks can be varied, we cannot be sure of establishing the best fit.

Figure 5-15 offers another way of visualizing the situation. If we plot χ^2 versus relative amplitude for a linear fit of two overlapped peaks, we would expect to see something like the curve at the top of the figure. In this case, χ^2 is a function of a single variable. By contrast, nonlinear methods attempt to minimize a χ^2 that is a function of many variables. If, for instance, we were concerned only with relative amplitude and peak width, the problem could be portrayed graphically as the search for a minimum on a three-dimensional χ^2 surface (lower drawing of Figure 5-15). Now take the experiment one step further. Imagine still another axis—peak centroid position, for example. The problem is now one of finding a minimum on a four-dimensional surface, a solution that is mathematically accessible, though not

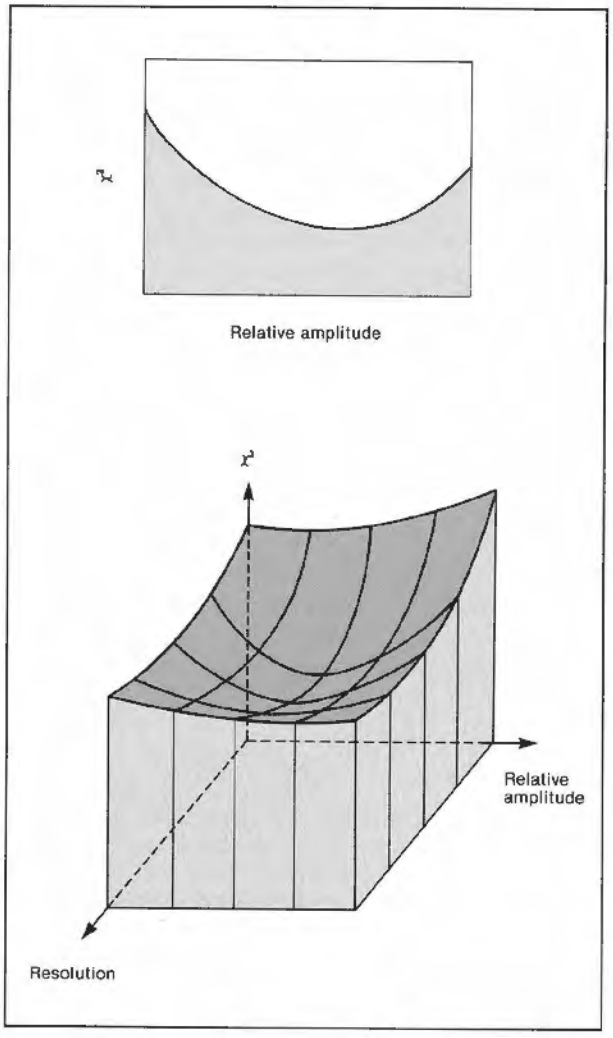


Figure 5-15. *Generalized plots of χ^2 as a function of one and two variables. Minimizing χ^2 by varying only relative peak amplitude (upper drawing) is equivalent to restricting the search to a single slice of the surface in the lower drawing.*

easily visualized. In very complex overlaps, there may be many more variables, and the surface, in general, becomes n-dimensional.

One of the mathematical techniques used in searching for a minimum in situations like this is called a *simplex search*.[13] To see how it works, we can go back to the three-dimensional case of Figure 5-15. If you imagine a triangle formed by any three points on the surface, one point will generally be higher than the other two (see Figure 5-16). Now, take the highest corner and "reflect" it across the opposite side to generate a new triangle. Reevaluate the height of each corner of the new triangle and again reflect the highest across the opposite side, and so on. It can be shown that the triangle, moved in this way, tends to migrate toward the minimum position on the surface. A simplex search on a multidimensional surface can obviously be quite time-consuming, even with a modern minicomputer; however, the judicious choice of constraints on the variables makes the problem manageable in many cases.

Nonlinear fitting procedures have the disadvantage of being slow. They are also vulnerable to errors introduced when one or more local minima exist on the χ^2 surface. These are best avoided by choosing the best available approximation as the starting point (usually derived by a preceding linear fit). On the other hand, nonlinear methods have the advantage of accuracy and a reduced vulnerability to shifts in calibration and resolution.

Other methods of deconvolution exist, but they are not in general use on minicomputer-based analyzers. Most notable among these sophisticated alternatives are Fourier transform techniques.[14] As with the various treatments of background, no single deconvolution method is best for all situations. The best general advice is that the more methods that are available, the more flexibility the investigator has in meeting the requirements of any given situation.

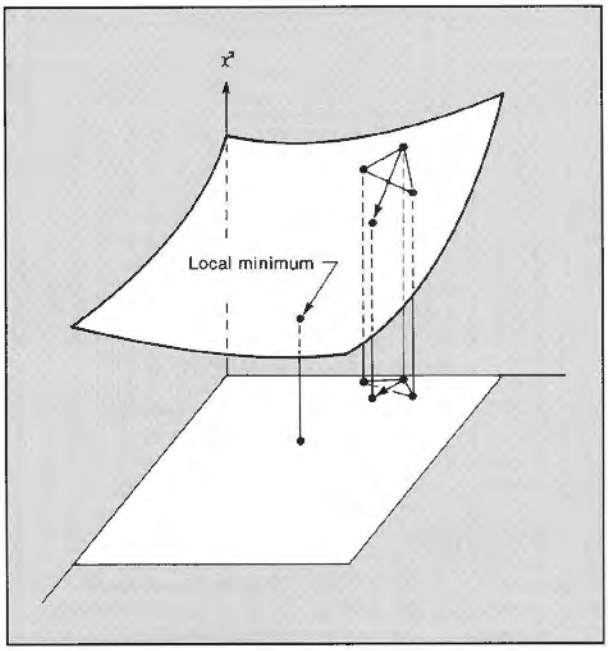


Figure 5-16. *Schematic illustration of a simplex search. The highest vertex of a triangle on the χ^2 surface is reflected across the opposite side, then the process is repeated. The triangle thus tends to migrate toward the surface minimum.*

5.2.3 Quantitative Calculations

All spectral processing discussed to this point has had a single purpose, namely, the accurate determination of the number of counts in spectral peaks associated with each element in the sample. It is assumed that these numbers in some way reflect the concentrations of the elements present. This is a reasonable assumption, but in fact the relationship between net peak counts and elemental concentrations is neither simple nor straightforward. Dealing with this nontrivial relationship is at the core of quantitative microanalysis.

ZAF Corrections

As we already know, many interactions complicate the single process we are interested in—electron-induced x-ray fluorescence. These complications can be grouped roughly into three categories: the effects of atomic number (Z), absorption within the sample and detector (A), and x-ray–induced fluorescence within the sample (F). Not surprisingly, then, most quantitative calculations center about what are called *ZAF corrections*.

The Z correction accounts for the effects of atomic number on excitation efficiency, fluorescent yield, and detector efficiency. (Excitation efficiency depends both on the ionization cross section of the element of interest and on the efficiency of competing processes. Correction factors for the *stopping power* of the sample and *backscatter loss* account for these competing processes.) The A correction reflects the likelihood that, once created within the sample, an x-ray will be absorbed before being detected. The F correction takes care of the contribution to observed peaks that arises from excitation of sample elements by x-rays generated within the sample. This x-ray–induced fluorescence, in turn, has two components: the fluorescence caused by characteristic x-rays from other elements and that caused by continuum x-rays.

The three ZAF correction terms, all of which depend strongly on the geometry of the excitation/detection system, must be integrated over a presumed pathway, taken first by the exciting electron beam, then by the emerging x-rays. One of the most serious sources of error in the ZAF-correction scheme is the simplifying assumption that all x-rays are produced at a single point within the sample. Any additional error introduced by miscalculation of the path length through the sample to and from that point dramatically diminishes the accuracy of the analysis. In modern systems, path length calculation is automatic and transparent to the user, but it demands appropriate values for all geometric parameters and for the accelerating voltage. Equally important to a successful analysis is an accurate indication of the total beam current deposited in the sample.

These ZAF corrections are applied to *k-ratios*, which, for any element in the sample, is the ratio between the number of x-rays counted in the net peak for that element and the number of x-rays counted for the same element, under the same conditions, in a sample of known concentration:

$$\text{k-ratio} = \frac{\text{net peak counts}}{\text{std counts}}$$

As a first approximation, we expect this k-ratio to be roughly equivalent to the ratio of the corresponding concentrations. For example, under given conditions of excitation, a sample composed of 50% iron should radiate about half as many iron x-rays as a pure iron sample. Because of the many complex interactions that occur, this approximation is just that—an approximation—but it does provide a starting point for the analysis. Thus, the general ZAF correction procedure starts with the assumption that k-ratios provide a good first approximation of elemental concentrations. These assumed concentra-

tions are then used to calculate corrections that should be applied to the k-ratios to account for the known effects of atomic number, absorption, and secondary fluorescence. When applied to the original k-ratios, these corrections yield a second estimate of elemental concentration. This improved estimate then serves as the basis for a recalculation of the corrections, and so on. With each iteration, the estimates change less and less, and the necessary net corrections grow smaller and smaller. The results thus converge to a value that reflects the actual concentrations.

As a footnote, a word should be added about sample preparation requirements. It is one of the great advantages of energy-dispersive microanalysis that sample preparation requirements are minimal. However, accurate quantitative measurements do place some constraints on the sample. First, the sample must be microscopically smooth. As shown in Figure 5-17, any topographic irregularities destroy the validity of the path length calculation. Second, the sample must be microscopically homogeneous. All calculations of absorption and fluorescence are based on the assumption that the material through which the x-rays pass is at every point the same as that at the point of x-ray generation. These assumptions, especially the second one, usually preclude successful quantitative analyses of microparticulates and of thin films on substrates.

Standardless Analysis

Several alternatives to the traditional ZAF analysis exist. In particular, it is possible to make certain reasonable assumptions that do away with the need for standards and empirical k-ratios.[15]

For pure-element standards, the measured α emission intensities can be expressed as

$$I \propto \varepsilon f(\chi) R \omega \left(\frac{Q}{S}\right) \beta$$

where

ε = the efficiency of the detector
$f(\chi)$ = the absorption correction
R = the backscatter-loss correction factor
ω = the fluorescent yield
Q = the ionization cross section
S = the stopping power
β = the fraction of the total emitted radiation emitted as α x-rays

All of these parameters except ε are computed in the course of calculating normal ZAF corrections, without reference to standards data. In addition, detector efficiency can be theoretically computed on the basis of available detector parameters. Therefore, theoretical pure-element intensities can be computed, then used as the basis for theoretical k-ratios. The normal iterative ZAF corrections can then proceed as usual.

Calibration Curves

Excellent quantitative results can also be obtained by establishing a calibration curve from the analysis of several samples of known composition. It is then a simple procedure to derive elemental concentration directly from an x-ray intensity in the range of the curve. The only rigid requirements are that acquisition conditions be identical for calibration standards and sample, and that the composition of the sample be similar to that of the standards.

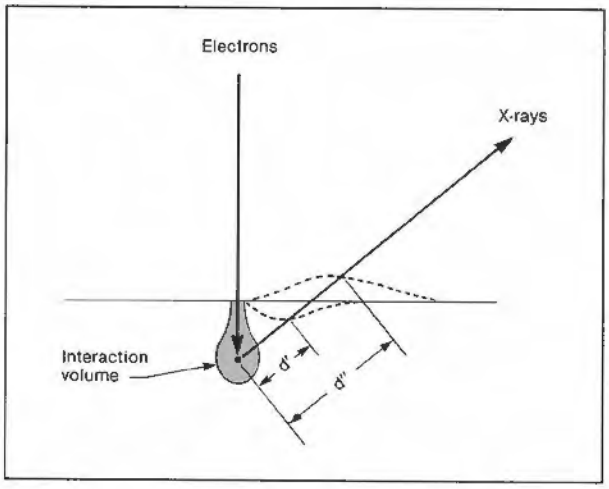


Figure 5-17. Illustration of the dependence of the absorption path length (d' and d'') on sample topography. The drawing also reflects the typical computational assumption that all x-rays originate from a single point within the interaction volume.

Oxide Analysis

Conventional beryllium-window detectors cannot detect oxygen x-rays, which are absorbed before reaching the detector. However, analysts are often interested in the presence of this element. Geologists, for example, must often analyze the stable oxides common in mineralogical specimens. This is commonly done by stoichiometry. The assumption is made that all oxygen present in the sample is in the form of oxides having known formulas. At each iterative step in the normal ZAF analysis, oxygen is included as an undetected constituent, present at the level indicated by the intensities of the signals from the other components of the oxides. For example, if Fe_2O_3 is assumed to be present, then for every atom of iron indicated by its x-ray intensity, 1½ atoms of oxygen are assumed to be absorbing and fluorescing x-rays within the sample. The ZAF corrections then take account of the assumed amount of oxygen.

Thin Films and Particles

A thin film or thin section is defined as a sample that is essentially transparent to the electron beam. Its thickness is very much less than would be the depth of penetration of the electron beam in a bulk sample of the same composition. As we discussed earlier (and as Figure 2-4 illustrates), this has implications for the spatial resolution of the x-ray signal, because the electron beam does not have the opportunity to spread. Furthermore, the thinness of these films also practically negates the effects of absorption and secondary fluorescence. This greatly simplifies the correction process, leaving only the atomic number corrections to be made. Unfortunately, this gain in simplicity must be measured against the greater demands for sample preparation.

The difficulties with particles arise from the lack of homogeneity and the undefined topography of the samples. A universally accepted method has yet to be found for dealing with either problem, but one method of accounting for the irregular topography seems to hold promise. The assumption is made that bremsstrahlung x-rays are affected in the same way as characteristic x-rays. As the particle size decreases, the length of the average absorption path becomes less. Under constant excitation conditions, more x-rays escape from a small particle than from a large one. This suggests that the intensity of the bremsstrahlung might contain useful information about the mass thickness of the volume analyzed. An analysis technique has therefore been developed that normalizes characteristic x-ray intensities to a region of the background free of characteristic x-ray peaks.[16]

Other methods of particle analysis have been proposed, ranging in complexity from the simple normalization of total results to 100%, to schemes that attempt to measure and calculate geometry and its effects for each individual particle. Analysis of particulates remains an incompletely solved problem.

Light-Element Analysis

As discussed in Section 4.1.4, the beryllium window of the traditional detector/cryostat absorbs most of the radiation from elements lighter than sodium. However, detectors are now available with x-ray–transparent windows or with no windows at all, the purpose being to make elements as light as carbon "visible" to the microanalyst. Unfortunately, these light elements still cannot be quantitatively analyzed by direct measurement of x-ray intensity with the same confidence as elements heavier than sodium. Uncertainties in the mass absorption coefficients and the difficulty of sorting out the effects of contaminants are the main hurdles to using the standard ZAF correction schemes.

References

General References

J. I. Goldstein, D. E. Newbury, P. Echlin, D. C. Joy, C. Fiori, and E. Lifshin, *Scanning Electron Microscopy and X-Ray Microanalysis* (Plenum, New York, 1981).

K. F. J. Heinrich, *Electron Beam X-Ray Microanalysis* (Van Nostrand Reinhold, New York, 1981).

Works Cited

1. T. R. Allmand and S. N. Jagger, "Electron Beam X-Ray Microanalysis Systems," Cambridge Instruments Ltd. document (no date).
2. J. Leroux and T. Thinh, *Revised Tables of X-Ray Mass Attenuation Coefficients* (Corporation Scientifique Claisse, Inc., Québec, 1977).
3. O. G. Wells, *Scanning Electron Microscopy* (McGraw-Hill, New York, 1974).
4. R. Johnson, "Applications of a Multiwindow XES Detector," *Amer. Lab.* 14(11), 93 (November 1982).
5. T. O. Ziebold, "Precision and Sensitivity in Electron Microprobe Analysis," *Anal. Chem.* **39**, 859 (1967).
6. L. A. Currie, "Limits for Qualitative Detection and Quantitative Determination," *Anal. Chem.* **40**(3), 586 (1968).
7. P. Duncumb and P. K. Shields, "The Present State of Quantitative X-Ray Microanalysis, Part I: Physical Basis," *Brit. J. Appl. Phys.* **14**, 617 (1963).
8. S. J. B. Reed, *Electron Microprobe Analysis*, (Cambridge Univ. Press, Cambridge, England, 1975).
9. J. J. McCarthy and F. H. Schamber, "Least-Squares Fit with Digital Filter: A Status Report," in K. F. J. Heinrich, D. E. Newbury, R. L. Mykelbust, and C. E. Fiori, eds., *Energy-Dispersive X-Ray Spectrometry*, NBS Special Publication 604, pp. 273-296 (1981).
10. C. E. Fiore, R. L. Mykelbust, K. F. J. Heinrich, and H. Yakowitz, "Prediction of Continuum Intensity in Energy-Dispersive Microanalysis," *Anal. Chem.* **48**(1), 172 (1976).
11. R. L. Mykelbust, C. E. Fiore, and K. F. J. Heinrich, *FRAME C: A Compact Procedure for Quantitative Energy-Dispersive Electron Probe X-Ray Analysis*, NBS Technical Note 1106 (September 1979).
12. H. Nullens, P. Van Espen, and F. Adams, "Linear and Nonlinear Peak Fitting in Energy-Dispersive X-Ray Fluorescence," *X-Ray Spectrom.* **8**(3), 104 (1979).
13. C. E. Fiore, R. L. Mykelbust, and K. Gorlen, "Sequential Simplex: A Procedure for Resolving Spectral Interference in Energy-Dispersive X-Ray Spectrometry," in K. F. J. Heinrich, D. E. Newbury, R. L. Mykelbust, and

C. E. Fiori, eds., *Energy-Dispersive X-Ray Spectrometry*, NBS Special Publication 604, pp. 233-272 (1981).

14. P. L. Ryder, "Statistical Considerations of Detectability Limits and Deconvolution in Energy-Dispersive X-Ray Spectrometry," in K. F. J. Heinrich, D. E. Newbury, R. L. Mykelbust, and C. E. Fiori, eds., *Energy-Dispersive X-Ray Spectrometry*, NBS Special Publication 604, pp. 177-191 (1981).
15. J. Colby, "ASAP Standardless Analysis: A Useful Shortcut When Absorption Effects Are Small," *Kevex Analyst*, No. 2, 3 (June 1982).
16. P. J. Statham, "Measurement and Use of Peak-to-Background Ratios in X-Ray Analysis," *Mikrochim. Acta*, Suppl. 8, 229 (1979).

Index

References in **boldface** type are to figure numbers. All other references are to page numbers. *Italics* indicate pages on which definitions appear.

Kevex Instruments, Inc.
24911 Avenue Stanford
Valencia, CA 91355
Tel. (805) 295-0019 (800) TO-KEVEX
Fax (805) 295-8714

kevex

$15