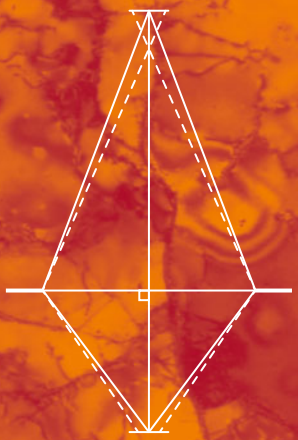
R.F. Egerton



Physical Principles of Electron Microscopy

An Introduction to TEM, SEM, and AEM

Second Edition

Springer

Physical Principles of Electron Microscopy

R.F. Egerton

Physical Principles of Electron Microscopy

An Introduction to TEM, SEM, and AEM

Second Edition

Springer

R.F. Egerton
Department of Physics
University of Alberta
Edmonton, AB
Canada

ISBN 978-3-319-39876-1 ISBN 978-3-319-39877-8 (eBook)
DOI 10.1007/978-3-319-39877-8

Library of Congress Control Number: 2016940809

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

To Maia

Preface

This book originates from an undergraduate physics course at the University of Alberta, whose aim was to illustrate various principles of electricity and magnetism, optics, and modern physics and to show that these topics have practical application in engineering and scientific research. Finding a textbook for the course was always a challenge; most electron microscopy texts overwhelm the non-specialist student, or they concentrate on practical skills rather than fundamental principles. In any event, Physical Principles of Electron Microscopy seems to have found a niche in the Springer microscopy collection; it has sold quite well around the world and was translated into Russian in 2010. Special features of this book include a discussion (in Chap. 1) of the development of electron microscopy and its relationship with other kinds of imaging, and an introduction (in Chap. 2) to the fundamentals of electron optics.

In 2015, the book was used as the basis of a new undergraduate course at Zhejiang University (Hangzhou, China), and I have used that event as an opportunity to update the text and introduce several new topics of current interest. Chapter 7 now includes sections on environmental TEM, radiation damage, electron tomography, and time-resolved microscopy. I have expanded the historical and scientific references, in the form of Additional Reading sections at the end of each chapter.

Although this book contains numerous equations, the mathematics is restricted to simple algebra, trigonometry, and calculus. SI units are employed throughout. I have used *italics* for emphasis and **bold** characters to mark technical terms when they first appear. Since it provides a more intuitive description at an elementary level, I use classical physics (treating the electron as a particle) whenever possible, even in connection with phase-contrast images. Supplementary material relating to this book, including sample examination questions (some with answers), can be found under educational materials at the website www.tem-eels.ca/.

I am indebted to several colleagues for reading the text and commenting on its accuracy, including Peter Crozier, Peter Hawkes, Chuanhong Jin, Marek Malac, Paul Midgley, Frances Ross, John Spence, Masashi Watanabe, Nestor Zaluzec, and Yimei Zhu.

Edmonton, Canada Ray Egerton
April 2016

Contents

Chapter 1
An Introduction to Microscopy

Microscopy involves the study of objects that are too small to be examined by the unaided eye. In the SI (metric) system of units, the sizes of these objects are expressed in terms of sub-multiples of the meter, such as a **micrometer** (1 μm = 10^{-6} m, also called a *micron*) or a **nanometer** (1 nm = 10^{-9} m). Older books use the Angstrom unit (1 Å = 10^{-10} m), not an official SI unit but convenient for specifying the distance between atoms in a solid, which is generally in the range 2–3 Å. To describe the wavelength of fast-moving electrons or their behavior inside an atom, we can use even smaller units, such as the **picometer** (1 pm = 10^{-12} m).

The diameters of several small objects of scientific or general interest are listed in Table 1.1.

1.1 Limitations of the Human Eye

Our concept of the physical world comes largely from what we *see* around us. For most of recorded history, this has meant observation using the human eye, which is sensitive to radiation within the **visible region** of the electromagnetic spectrum: light with wavelength in the range 300–700 nm. The eyeball contains a fluid whose **refractive index** ($n \approx 1.34$) is substantially different from that of air ($n \approx 1$). As a result, most of the refraction and focusing of the incoming light occurs at its curved front surface, the **cornea**; see Fig. 1.1.

In order to focus on objects located at *different* distances (referred to as accommodation), the eye incorporates an elastically deformable **lens** of slightly higher refractive index ($n \sim 1.44$) whose shape and focusing power are controlled by eye muscles. Together, the cornea and lens of the eye behave like a single glass lens of variable focal length, forming a **real image** on the curved **retina** at the back of the eyeball. The retina contains photosensitive receptor cells that send electrochemical signals to the brain, the strength of each signal representing local intensity in the image. However, the photochemical processes in the receptor cells work over

R.F. Egerton, *Physical Principles of Electron Microscopy*,
DOI 10.1007/978-3-319-39877-8_1

Table 1.1 Approximate sizes of some common objects and the smallest magnification M* required to distinguish them, according to Eq. (1.5)

Object	Typical diameter D	M* = 75 μm/D
Grain of sand	1 mm = 1000 μm	None
Human hair	150 μm	None
Red blood cell	10 μm	7.5
Bacterium	1 μm	75
Virus	20 nm	4000
DNA molecule	2 nm	40,000
Typical atom	0.2 nm = 200 pm	400,000

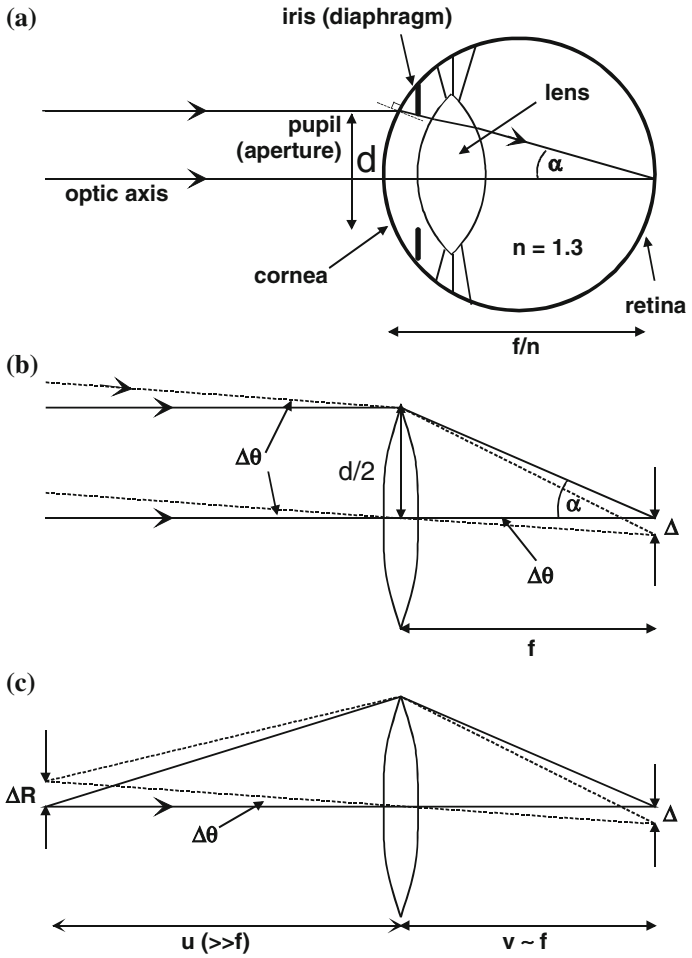


Fig. 1.1 **a** A physicist's simplified view of the human eye, showing two light rays focused to a single point on the retina. **b** Equivalent thin-lens ray diagram for a *distant* object, showing *parallel* light rays arriving from opposite ends (*solid* and *dashed lines*) of the object and forming an image that in air would occur at a distance f (the focal length) from the thin lens. **c** Ray diagram for a *nearby* object (object distance u = 25 cm, image distance v slightly below f)

a limited range of image intensities, so the eye controls the amount of light reaching the retina by varying the **aperture** of the eye (its effective diameter d, typically 2–8 mm), also known as the **pupil.** This aperture takes the form of a circular hole in the **diaphragm** (or **iris**), an opaque disk located between the lens and the cornea, as shown in Fig. 1.1.

The **spatial resolution** of the retinal image which determines how *small* an object can be if it is to be distinguished from a similar adjacent object, is determined by three factors: the size of the receptor cells, imperfections of focusing (known as **aberrations**), and **diffraction** of light at the entrance pupil of the eye. Diffraction cannot be explained using a particle view of light (geometrical or ray optics); it requires a wave interpretation (physical optics), according to which any image is actually an interference pattern formed by light rays that take different paths to reach the same point in the image.

In the simple situation shown in Fig. 1.2, parallel light (of wavelength λ) from a distant point source strikes an opaque diaphragm containing a circular aperture whose radius a subtends an angle α at the center of a white viewing screen. Light passing through the aperture illuminates the screen in the form of a circular pattern with diffuse edges (a **disk of confusion**) whose intensity profile is an oscillating but highly damped **Airy function** whose radius, measured to the first minimum, is Δx; see Fig. 1.2. Light arriving at an angle θ (from a second displaced source) forms a displaced Airy disk and according to the **Rayleigh criterion**, the two sources are just resolved if the displaced Airy maximum coincides with the first minimum of the undisplaced source. Diffraction theory gives $\sin\theta = (3.83/a)(2\pi/\lambda) = 0.61(\lambda/a)$ and for $\lambda \ll a$, the use of a small-angle approximation: $\sin\theta \approx \tan\theta = \Delta x/v$ gives:

$$\Delta x \approx v \sin\theta = 0.61(\lambda/a)v \approx 0.6\lambda/\alpha \tag{1.1}$$

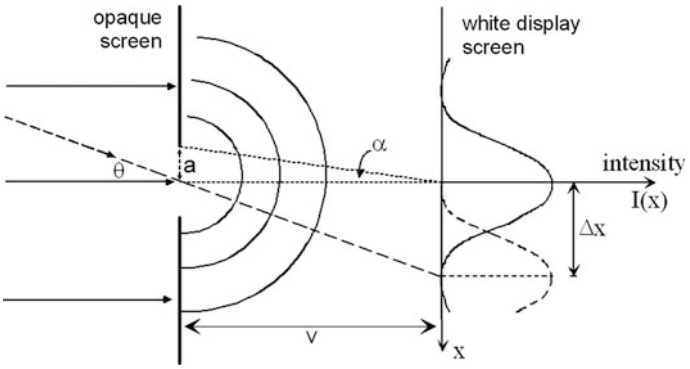


Fig. 1.2 Diffraction of light by a slit or by a circular aperture. Waves spread out from the aperture and fall on a white screen to produce a disk of confusion (*Airy disk*) whose intensity distribution I (x) is shown by the graph on the *right*. Light from a second source produces a displaced *Airy disk*, which is separately resolved if it satisfies Eq. (1.1)

where $a = v \tan \alpha \approx v \alpha$ is assumed, and α is measured in radians. While the Rayleigh criterion is somewhat arbitrary, Eq. (1.1) gives an intensity distribution (the sum from both sources) that dips by about 19 % in the middle, which is typical of what the eye can detect in the presence of electrical noise.

Equation (1.1) can be applied to the eye by referring to Fig. 1.1b, which depicts an equivalent image formed in air at a distance f from a single focusing lens. For wavelengths in the middle of the visible region, $\lambda \approx 500$ nm and if we take $d = 2a \approx 4$ mm and $f \approx 2$ cm, $\tan \alpha \approx a/f = 0.1$ and from Eq. (1.1) the diffraction-limited image resolution is $\Delta x \approx (0.6)(500\ \text{nm})/0.1 = 3\ \mu\text{m}$.

Imperfect focusing (aberration) of the eye contributes a roughly *equal* amount of image blurring, which we therefore take as 3 μm, and the receptor cells of the retina have diameters in the range 2–6 μm (mean value ≈ 4 μm). It seems that evolution has refined the eye up to the point where further improvements would give relatively little improvement in overall resolution, relative to the diffraction limit Δx imposed by the wave nature of light.

To a reasonable approximation, the three contributions to the retinal-image blurring can be combined **in quadrature** (by adding squares), treating them in a similar way to the statistical quantities involved in error analysis. Using this procedure, the total image blurring Δ is given by:

$$(\Delta)^2 = (3\,\mu\text{m})^2 + (3\,\mu\text{m})^2 + (4\,\mu\text{m})^2. \tag{1.2}$$

leading to $\Delta \approx 6$ μm as the total blurring in the retinal image. This value also corresponds to an *angular* blurring for distant objects (see Fig. 1.1b) of

$$\Delta\theta \approx (\Delta/f) \approx (6\,\mu\text{m})\,/\,(2\,\text{cm}) \approx 3 \times 10^{-4}\text{rad} \approx (1/60)\,\text{degree} = 1\,\text{minute of arc} \tag{1.3}$$

Distant objects (or details within objects) will be separately distinguished if they subtend angles larger than this. Accordingly, early astronomers were able to determine the positions of bright stars to within a few minutes of arc, using only a dark-adapted eye and simple pointing devices. To see greater detail in the night sky, such as the faint stars within a galaxy, a *telescope* was needed to provide *angular* magnification.

Changing the shape of the lens in an adult eye alters its overall focal length by only about 10 %, so the *closest* object distance for a focused image on the retina is $u \approx 25$ cm. At this distance, an angular resolution of 3×10^{-4} rad corresponds (see Fig. 1.1c) to an object separation of:

$$\Delta R \approx (\Delta\theta)u \approx 0.075\,\text{mm} = 75\,\mu\text{m} \tag{1.4}$$

Since $u \approx 25$ cm is the smallest object distance for clear vision, we can take $\Delta R = 75$ μm as the diameter of the *smallest* object that can be resolved by the unaided eye, known as its **object resolution** (its spatial resolution in the *object* plane).

As illustrated in Table 1.1, there are many interesting objects *below* this size, but to see them we need an optical device with **magnification factor** M (>1); in other words, a **microscope**. To barely resolve an object of diameter D, we need a magnification M^* such that the *magnified* diameter (M^* D) at the object plane is equal to the *object* resolution ΔR (≈75 μm) of the eye. In other words:

$$M^* = (\Delta R)/D \tag{1.5}$$

Values of this minimum magnification are given in the right-hand column of Table 1.1.

1.2 The Light-Optical Microscope

Light microscopes were developed in the early 1600s and some of the best observations were made by Anton van Leeuwenhoek, using a tiny glass lens placed very close to the object and close to the eye; see Fig. 1.3. By the late 1600s, this Dutch scientist had observed blood cells, bacteria, and structure *within* the cells of animal tissue, all revelations at that time. This simple one-lens device had to be positioned accurately and held very still, making long-term observation extremely tiring.

For routine use, it is more convenient to use a **compound microscope**, containing at least two lenses: an **objective** (placed close to the *object* being magnified) and an **eyepiece** (placed close to the *eye*). By using an objective of very short focal length or by using more lenses, the magnification M of a compound microscope can be increased indefinitely. However, a large value of M does not guarantee that objects of very small diameter D will be resolved; besides satisfying Eq. (1.5), we must ensure that aberrations and diffraction *within the microscope* are sufficiently low.

Nowadays, the aberrations of a light-optical instrument can be made vanishingly small, by grinding the lens surfaces to a correct shape or by spacing the lenses so that their aberrations are compensated. But even with such aberration-corrected optics, the spatial resolution of a compound microscope is limited by *diffraction* at the objective lens. This effect depends on the diameter (aperture) of the lens, just as in the case of diffraction at the pupil of the eye or at a circular hole in an opaque screen.

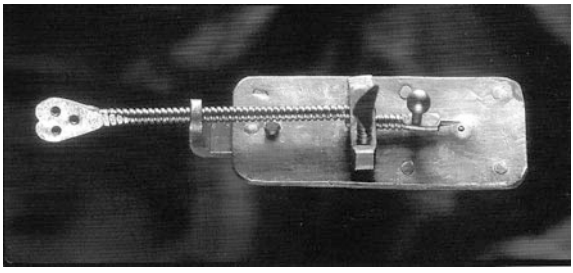


Fig. 1.3 One of the single-lens microscopes used by Van Leeuwenhoek. The adjustable pointer was used to center the eye on the optic axis of the lens and thereby minimize image aberrations. Courtesy of the FEI Company

For a large-aperture lens ($\alpha \approx 1$ rad), Eq. (1.1) predicts a resolution limit of just over half the wavelength of light, as first deduced by Abbé in 1873. For light in the middle of the visible spectrum ($\lambda \approx 0.5$ μm), the best-possible object resolution is therefore about 0.3 μm.

This value is a substantial improvement over the resolution (≈75 μm) of the unaided eye. But to achieve such resolution, the microscope must magnify the object to a diameter greater than ΔR, so that the *overall* resolution is determined by diffraction in the microscope rather than by the eye's own limitations. This requires a microscope magnification of $M \approx (75\,\mu\text{m})/(0.3\,\mu\text{m}) = 250$. A larger value ("empty magnification") does not significantly improve the sharpness of the magnified image and in fact reduces the **field of view**, the area within the object plane that can be simultaneously viewed in the image.

Light-optical microscopes are widely used in scientific research and come in two basic forms. The **biological** microscope (Fig. 1.4a) requires an optically transparent specimen, such as a thin slice (section) of animal or plant tissue. Daylight or light from a lamp is directed via a lens or mirror through the specimen and into the microscope, resulting in a magnified image on the retina or within an attached camera. Variation of light intensity (**contrast**) within the image occurs because different parts of the specimen *absorb* light to differing degrees. By using **stains** (light-absorbing chemicals that are absorbed preferentially within certain regions of the specimen) the contrast can be increased; the image of a tissue section may then reveal individual components (**organelles**) within each biological cell. Because light travels *through* the specimen, this instrument could equally well be called a **transmission** light microscope; it is used also by geologists, who are able to

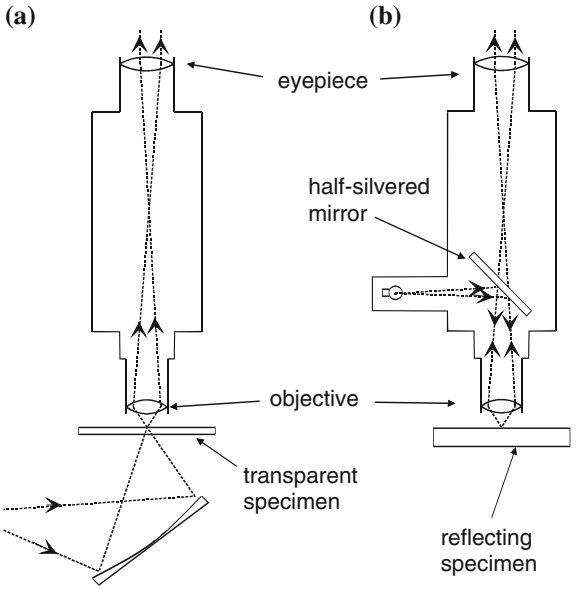


Fig. 1.4 Schematic diagrams of **a** a biological microscope, which images light transmitted through the specimen, and **b** a metallurgical microscope, which uses light (often from a built-in illumination source) reflected from the specimen's surface

prepare rock specimens that are thin enough (below 0.1 μm thickness) to be optically transparent.

The **metallurgical** microscope (Fig. 1.4b) is used for examining metals and other materials that are not easily made thin enough to be optically transparent. Here, the image is formed by light *reflected* from the surface of the specimen. Since perfectly smooth surfaces would provide little or no contrast, the specimen is usually immersed for a few seconds in a chemical **etch**, a solution that preferentially attacks certain regions to leave an uneven surface whose reflectivity varies from one location to another. The microscope then reveals the internal microstructure of crystalline materials, such as the different phases present in a metal alloy. Most etches preferentially dissolve the region between individual crystallites (grains) of the specimen, where atoms are less closely packed, leaving a grain-boundary groove that is visible as a dark line, as in Fig. 1.5. The metallurgical microscope can therefore be used to determine the grain shape and grain size of metals and alloys.

As we have seen, the resolution of a light-optical microscope is limited by diffraction. As implied by Eq. (1.1), one possibility for *improving* resolution (*reducing* Δx, and therefore Δ and ΔR) is to decrease the wavelength λ of the radiation. A simple option is to use an **oil-immersion** objective lens: a drop of a transparent liquid (refractive index *n)* is placed between the specimen and the objective, so the light being focused (and diffracted) has a reduced wavelength: λ/n. Using cedar oil (n = 1.52) allows a 34 % improvement in resolution.

Greater improvement comes from using **ultraviolet** (UV) radiation, meaning wavelengths in the range 100–400 nm. The light source can be a gas-discharge lamp and the final image is viewed on a phosphor screen that converts the UV to

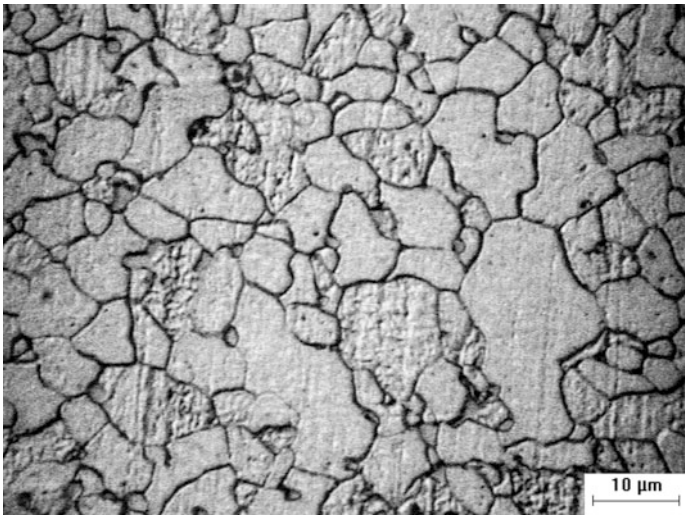


Fig. 1.5 Light-microscope image of a polished and etched specimen of X70 pipeline steel, showing *dark lines* representing the grain boundaries between ferrite (bcc iron) crystallites. Courtesy of Dr. D. Ivey, University of Alberta

visible light. Since ordinary glass strongly absorbs UV radiation, the focusing lenses must be made from a material such as quartz (transparent down to 190 nm) or lithium fluoride (transparent down to about 100 nm).

1.3 The X-ray Microscope

Because of their short wavelength, x-rays offer the possibility of even better spatial resolution but cannot be focused by transparent lenses, since the refractive index of solid materials is close to that of air (1.0) at x-ray wavelengths. Instead, x-ray focusing relies on devices that make use of *diffraction* rather than refraction.

Hard X-rays have wavelengths below 1 nm and are diffracted by planes of atoms in a crystal, whose spacing is of similar dimensions. In fact, such diffraction is routinely used to determine the atomic structure of solids. X-ray microscopes more commonly use **soft x-rays**, with wavelengths in the range 1–10 nm. These x-rays are diffracted by structures having a periodicity of several nm, such as thin-film multilayers that act as focusing mirrors, or **zone plates**, which are essentially diffraction gratings with circular symmetry (Fig. 5.23) that focus **monochromatic** x-rays (those of a single wavelength); see Fig. 1.6.

Unfortunately, such focusing devices are less efficient than the glass lenses used in light optics. Also, laboratory x-ray sources are relatively weak (x-ray diffraction patterns are recorded over minutes or hours). This situation prevented the practical realization of an x-ray microscope until the development of a more intense radiation source: the **synchrotron**, in which electrons circulate at high speed in vacuum within a **storage ring**. Guided around a circular path by strong electromagnets, their centripetal acceleration results in the emission of **bremsstrahlung x-rays**. Devices called undulators or wobblers can also be inserted into the ring; an array of magnets

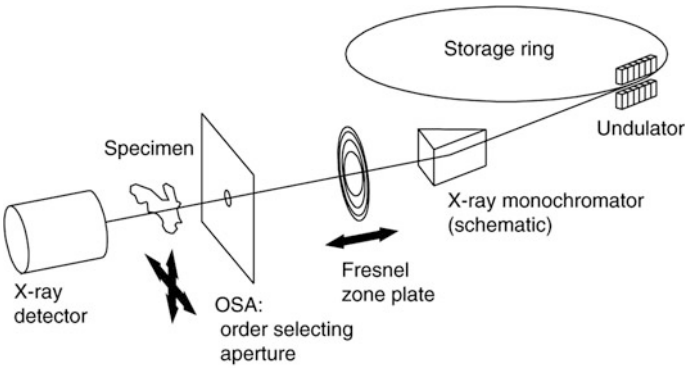


Fig. 1.6 Schematic diagram of a scanning transmission x-ray microscope (STXM) attached to a synchrotron radiation source. The monochromator transmits x-rays with a narrow range of wavelengths, and these monochromatic rays are focused onto the specimen by means of a Fresnel zone plate. From Neuhausler et al. (1999), courtesy of Springer-Verlag

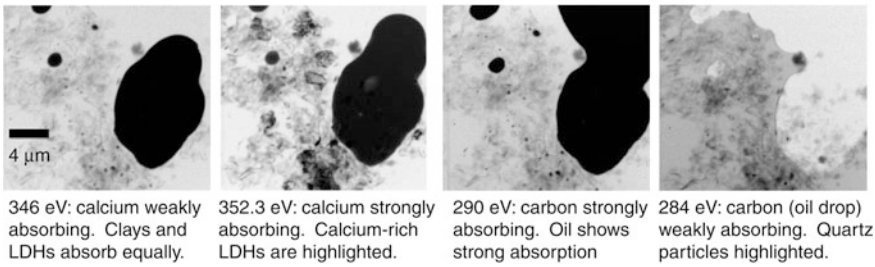


Fig. 1.7 Scanning transmission x-ray microscope (STXM) images of a clay-stabilized oil–water emulsion. By changing the photon energy, different components of the emulsion become *bright* or *dark* and can be identified from their known x-ray absorption properties. From Neuhausler, U., Abend, S., Jacobsen, C. and G. Lagaly G. (1999) Soft x-ray spectro-microscopy on solid-stabilized emulsions. *Colloid. Polym. Sci*. **277**:719–726, courtesy of Springer-Verlag

causes further deviation of the electron from a straight-line path and produces a strong bremsstrahlung effect, as shown in Fig. 1.6. Synchrotron x-ray sources are large and expensive but their radiation has a variety of uses and several dozen are in operation throughout the world.

An important feature of the x-ray microscope is that it can be used to study hydrated (wet or frozen) specimens such as biological tissue or water/oil emulsions, surrounded by air or a water-vapor environment during their observation. In this case, x-rays in the wavelength range 2.3–4.4 nm are used (photon energy between 285 and 543 eV), the so-called **water window** in which hydrated specimens appear relatively transparent. Contrast in the x-ray image arises because different regions of the specimen absorb the x-rays to differing extents, as illustrated in Fig. 1.7. The resolution of these images is typically 30 nm, limited by zone-plate aberrations and an order of magnitude larger than the x-ray wavelength.

As we will see, the specimen in an *electron* microscope is usually in a dry state, surrounded by a high vacuum. Unless the specimen is cooled well below room temperature or enclosed in an environmental cell (Sect. 7.1), any water quickly evaporates into the surroundings.

1.4 The Transmission Electron Microscope

Early in the 20th century, physicists discovered that material particles such as electrons possess a wavelike character. By analogy with Einstein's photon description of electromagnetic radiation, Louis de Broglie proposed that the electron wavelength is given by

$$\lambda = h/p = h/(mv) \tag{1.6}$$

where $h = 6.626 \times 10^{-34}$ Js is the Planck constant; p, m, and v represent the momentum, mass, and speed of the electron. For electrons emitted into vacuum

from a heated filament and accelerated through a potential difference of 50 V, $v \approx 4.2 \times 10^6$ m/s and $\lambda \approx 0.17$ nm. Since this wavelength is close to atomic dimensions, such "slow" electrons are strongly diffracted from the regular array of atoms present at the surface of a crystal, as first observed by Davisson and Germer (1927).

Raising the accelerating potential to 50 kV, the electron wavelength shrinks to about 5 pm (0.005 nm) and such higher-energy electrons can penetrate distances of several microns (μm) into a solid. If this solid is crystalline, the electrons are diffracted by planes of atoms inside the material, as happens for x-rays. It is therefore possible to form a **transmission electron diffraction** pattern from electrons that have passed *through* a thin specimen, as first demonstrated by Thomson (1927). Later it was realized that if these transmitted electrons could be focused, their very short wavelength would allow the specimen to be imaged with a spatial resolution far superior to that of the light-optical microscope.

The focusing of electrons relies on the fact that, in addition to their wavelike character, they behave as negatively charged particles and are therefore deflected by electric or magnetic fields. This principle was utilized in early cathode-ray tubes, television display tubes, and computer screens. In fact, the first electron microscopes made use of the technology developed for radar applications of cathode-ray tubes. In a **transmission electron microscope** (TEM), electrons penetrate a *thin* specimen and are then imaged by the appropriate lenses, in broad analogy with the *biological* light microscope (Fig. 1.4a).

Some of the first development work on electron lenses was done by Ernst Ruska in Berlin. By 1931 he had observed his first transmission image (magnification = 17) of a metal grid, using the two-lens microscope shown in Fig. 1.8. His electron lenses were short coils carrying a direct current that produced a magnetic field directed along the optic axis. In 1933, Ruska added a third lens and obtained images of cotton fiber and aluminum foil, with a resolution somewhat better than that of a light microscope.

Similar microscopes were built by Marton and co-workers in Brussels, who by 1934 produced the first images of the nuclei within biological cells. These early TEMs used a horizontal sequence of lenses, as in Fig. 1.8, an arrangement that was

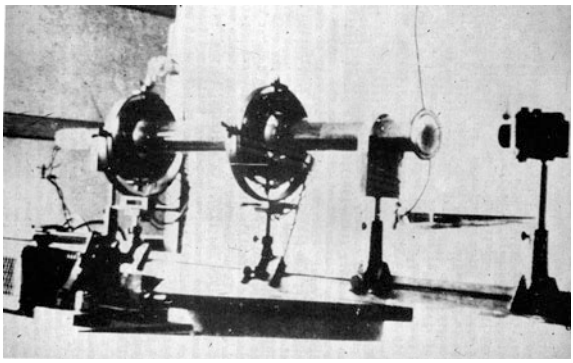


Fig. 1.8 Early photograph of a horizontal two-stage electron microscope (Knoll and Ruska 1932). This material is used by permission of Wiley-VCH, Berlin

abandoned after it was realized that precise alignment of the lenses along the optic axis is critical to obtaining the best resolution. Stacking electron lenses in a vertical column allows good alignment to be maintained for a longer time because gravitational forces act *parallel* to the optic axis, making slow mechanical distortion (creep) less troublesome.

In 1936, the Metropolitan Vickers company embarked on commercial production of a TEM in the United Kingdom. However, the first regular production came from the Siemens Company in Germany; their 1938 prototype achieved a spatial resolution of 10 nm with an accelerating voltage of 80 kV; see Fig. 1.9.

Some early TEMs used a gas discharge as the source of electrons but this was soon replaced by a V-shaped filament made from tungsten wire, which emits electrons when heated in vacuum. The vacuum was generated by a mechanical pump together with a diffusion pump, originally made of glass and containing boiling mercury. The electrons were accelerated by applying a high voltage

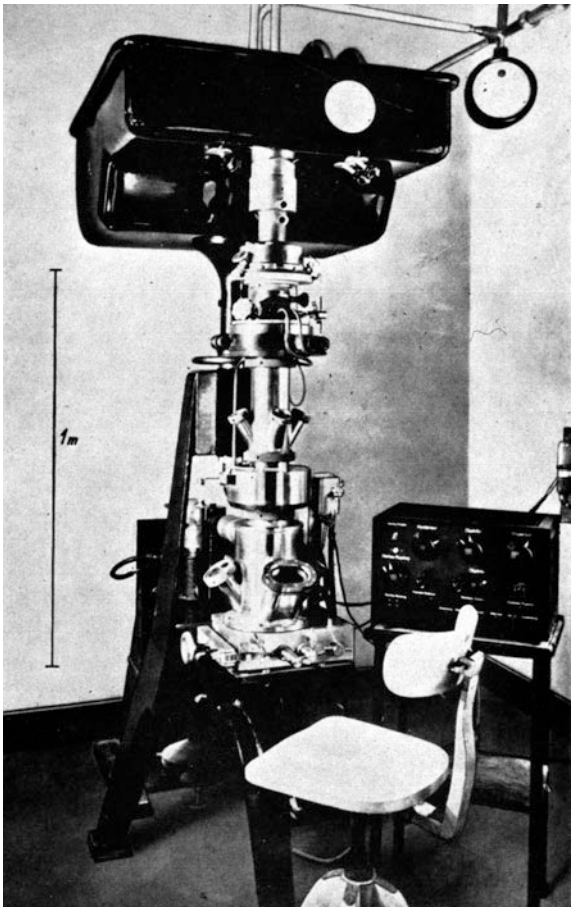


Fig. 1.9 First commercial TEM from the Siemens Company, employing three magnetic lenses that were water-cooled and energized by batteries. The objective lens had a focal length down to 2.8 mm at 80 kV, giving an estimated resolution of 10 nm

generated by an electronic oscillator circuit and a step-up transformer. As the transistor had not yet been invented, the oscillator circuit used vacuum-tube electronics. In fact, vacuum tubes were used in the high-voltage circuitry of television receivers up to the 1980s since they were less easily damaged by the voltage spikes that occurred whenever there was a high-voltage discharge. Vacuum tubes were also used to control and stabilize the dc current applied to the electron lenses of the early TEMs.

Although companies in USA, Holland, UK, Germany, Japan, China, USSR, and Czechoslovakia have manufactured transmission electron microscopes, competition eventually reduced their number to three: the Japanese Electron Optics Laboratory (JEOL) and Hitachi in Japan, and Philips/FEI in Holland/USA.

Recent development of the TEM is illustrated by the two instruments shown in Fig. 1.10. Like most modern instruments, the JEOL Grand ARM (Fig. 1.10a) uses integrated circuits and digital control; at 300 kV accelerating voltage and with lens-aberration correction, it provides a resolution of 63 pm. The FEI TITAN Themis (Fig. 1.10b) can reach similar resolution at 300 kV and is available in an enclosed version (Fig. 1.10c) that makes it less sensitive to environmental fluctuations.

The TEM has proved very useful for examining the ultrastructure of metals. For example, crystalline defects named dislocations were first predicted by theorists, in response to the fact that metals deform under much lower forces than expected for a perfect crystalline array of atoms. They were first observed in TEM images of aluminum; one of M.J. Whelan's original micrographs is reproduced in Fig. 1.11. Note the increase in resolution compared to the light-microscope image of Fig. 1.5;

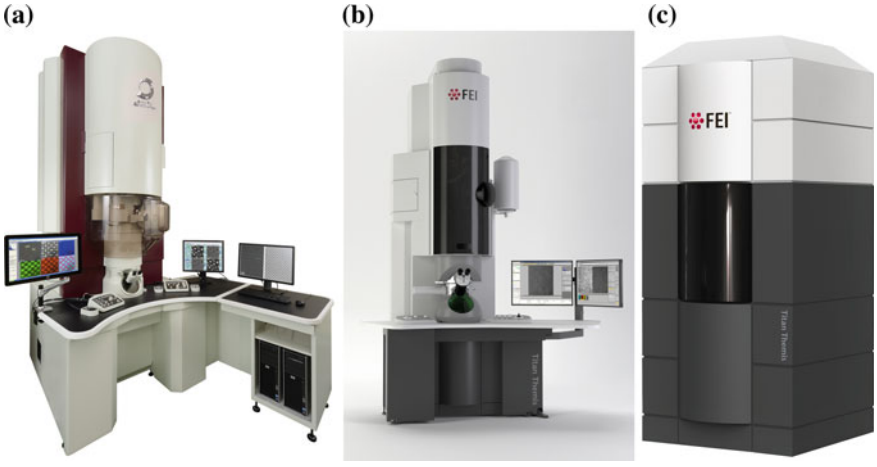


Fig. 1.10 Two recent transmission electron microscopes: **a** the JEOL Grand ARM (300 kV, 63 pm resolution with aberration correction), **b** the FEI TITAN Themis, which can reach 60 pm resolution at 300 kV if fitted with aberration correction, and whose 5.4 mm objective-lens polepiece gap allows in-situ experiments

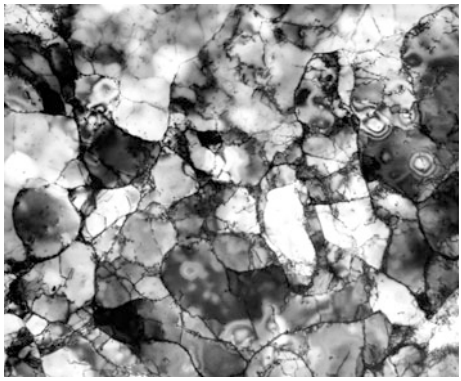


Fig. 1.11 TEM diffraction-contrast image ($M \approx 10{,}000$) of polycrystalline aluminum. The crystallites (*grains*) show up with different brightness levels; low-angle boundaries and dislocations are visible as *dark lines* within each crystallite. *Circular* fringes (*top-right*) represent local changes in specimen thickness. Courtesy of M. J. Whelan, Oxford University

fine detail can now be seen within each metal crystallite. With a modern TEM (resolution below 0.2 nm), individual atomic planes or columns of atoms can be distinguished, as discussed in Chap. 4.

The TEM has been equally helpful in the life sciences, for example for examining plant and animal tissue, bacteria, and viruses. Figure 1.12 shows images of mouse-liver tissue obtained using transmission light and electron microscopes. Cell membranes and a few internal organelles are visible in the light-microscope image but the higher spatial resolution of the TEM reveals structure within the organelles.

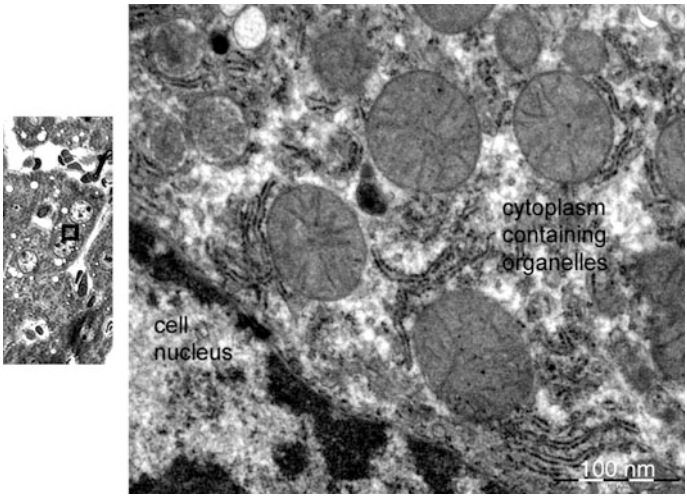


Fig. 1.12 TEM image (*right*) of a stained specimen of mouse-liver tissue, corresponding approximately to the small square area within the light-microscope image on the *left*. Courtesy of R. Bhatnagar, Biological Sciences Microscopy Unit, University of Alberta

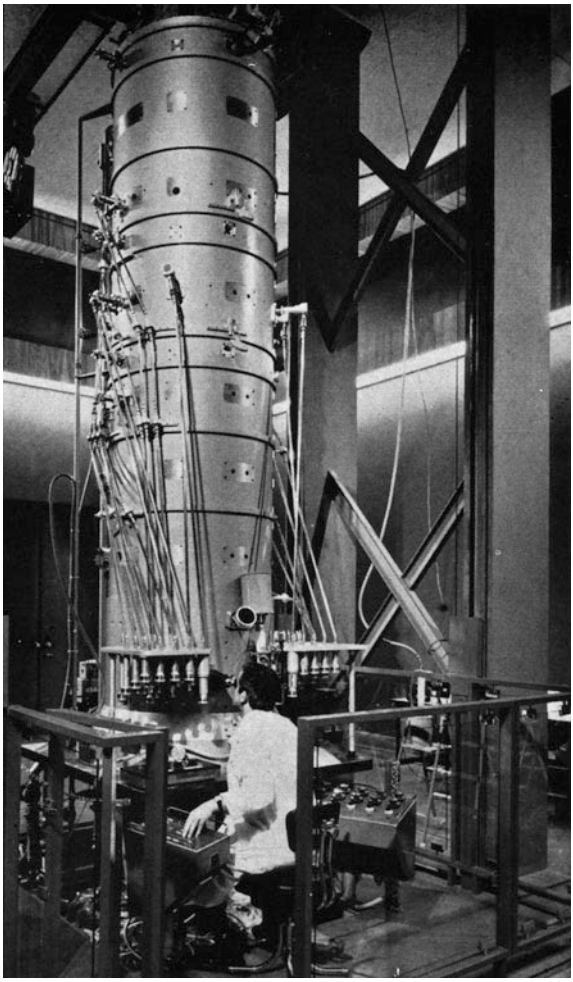


Fig. 1.13 A 3 MV HVEM constructed at the C.N.R.S. Laboratories in Toulouse and in operation during the 1970s. To focus the high-energy electrons, large-diameter lenses were required and the TEM column was so high that long control rods were needed between the operator and the moving parts (for example, to provide specimen motion). Courtesy of G. Dupouy, personal communication

Although most modern TEMs use electron-accelerating voltages between 60 and 300 kV, high-voltage instruments (HVEMs) have been constructed with voltages up to several MV; see Fig. 1.13. One motivation was the fact that increasing the electron energy (and therefore momentum) decreases the de Broglie wavelength of electrons and therefore the diffraction limit to spatial resolution. However, technical problems of voltage stabilization have prevented HVEMs from achieving their theoretical resolution. A few are still in regular use and are useful for looking at thicker specimens, since electrons of very high energy can penetrate further into a solid (1 μm or more) without excessive scattering.

One of the original hopes for the HVEM was to study the behavior of living cells. By enclosing a specimen within an environmental chamber, water vapor can be supplied to keep the cells hydrated. But high-energy electrons are a form of ionizing radiation, similar to x-rays or gamma rays in their ability to ionize atoms

and produce irreversible chemical changes. In fact, a focused beam of electrons represents a radiation flux exceeding that at the center of an exploding nuclear weapon. Not surprisingly, it was found that TEM observation kills living tissue in much less time than needed to record a high-resolution image.

1.5 The Scanning Electron Microscope

One limitation of the TEM is that, unless the specimen is made very thin, electrons are strongly scattered or even absorbed in the specimen, rather than being transmitted. This constraint provided an incentive to develop electron microscopes capable of examining thick (so-called **bulk**) specimens. The general aim was to produce an electron-beam instrument equivalent to the metallurgical light microscope but with better spatial resolution.

Electrons can indeed be "reflected" (backscattered) from a bulk specimen, as in the original experiments of Davisson and Germer (1927). But another possibility is that the incoming (**primary**) electrons supply energy to the *atomic* electrons that are present in a solid, which are then released as **secondary electrons**. These electrons are emitted with a range of energies, making it hard to focus them into an image by electron lenses. However, there is an alternative mode of image formation that uses a **scanning** principle: primary electrons are focused into a small-diameter **electron probe** that is scanned across the specimen by electrostatic or magnetic fields that change the direction of the incident beam. By scanning simultaneously in two perpendicular directions (**raster** scanning), a square or rectangular area of specimen can be covered and an image of this area formed by collecting secondary electrons released from each local region of the specimen.

The raster-scan signals were also used to deflect the beam inside a **cathode-ray tube** that contained an electron gun producing a narrow beam of electrons, which were then focused onto a phosphor screen to convert the electrons into visible light. The secondary-electron signal was amplified and applied to the electron gun in order to modulate the beam intensity. Because of the exact synchronism with the motion of the electron beam on the specimen, the light emitted by the phosphor represented a secondary-electron image of the specimen. This scanning method generates an image serially (point by point) rather than simultaneously (as in the TEM or light microscope) and the same principle is employed in all video imaging.

The first **scanning electron microscope** (SEM) based on secondary emission was developed at the RCA Laboratories in New Jersey, under wartime conditions. Some of the early prototypes utilized a field-emission electron source (discussed in Chap. 3) whereas later models used a heated filament, the electrons being focused onto the specimen by electrostatic lenses. An early-version FAX machine was used for image recording; see Fig. 1.14. The spatial resolution was estimated to be 50 nm, an order of magnitude better than the light-optical microscope.

Further SEM development occurred after the Second World War, when Charles Oatley and colleagues launched a research and construction program in the

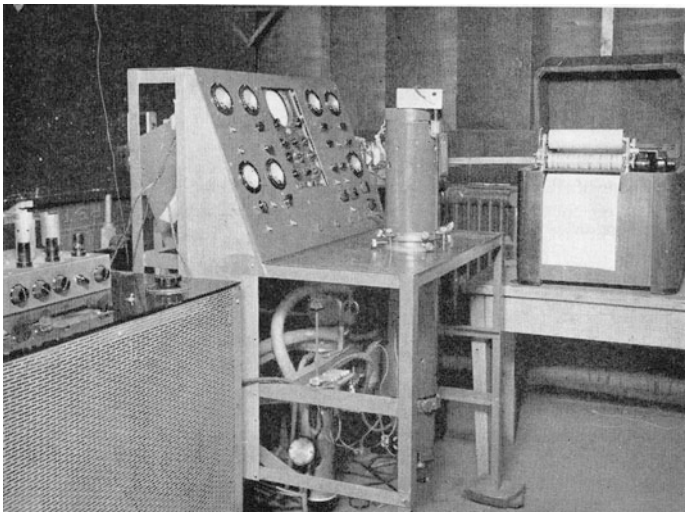


Fig. 1.14 Scanning electron microscope at RCA Laboratories (Zworykin et al. 1942) that used electrostatic lenses and vacuum-tube electronics (as in the amplifier on the *left*). An image was produced on the facsimile machine visible on the *right-hand side* of the picture. This material is used by permission of John Wiley & Sons, Inc

Engineering Department at Cambridge University. Their first SEM images were obtained in 1951 and a commercial model (built by the AEI Company) was delivered to the Pulp and Paper Research Institute of Canada in 1958. Sustained commercial production was initiated by the Cambridge Instrument Company in 1965 and there are now about a dozen SEM manufacturers worldwide. Figure 1.15 shows an example of a modern instrument. Image information is stored in the computer that controls the SEM and the images appear on a flat-panel display screen.

A modern SEM provides an image resolution typically between 1 and 10 nm, not as good as the TEM but greatly superior to the light microscope. In addition, SEM images have a relatively large **depth of focus**: specimen features that are not exactly in the plane of focus still appear sharp and distinguishable. As we shall see, this characteristic results from the fact that electrons in the SEM (and the TEM) travel very close to the optic axis, their angular deviation being at most a few degrees.

1.6 Scanning Transmission Electron Microscope

The fine-probe/scanning technique can also be used with a thin (transmission) specimen, the result being a **scanning-transmission electron microscope** (STEM). Rather than secondary electrons, it is usual to record the primary electrons that are

Fig. 1.15 Hitachi-SU5000 scanning electron microscope. This instrument uses a Schottky-emission source and provides a resolution of 1.2 nm at an accelerating voltage of 30 kV or 3 nm at 1 kV (2 nm in deceleration mode). Variable pressure mode (10–300 Pa) is an option

scattered in a particular direction and emerge from the beam-exit surface. The first STEM was constructed by von Ardenne in 1938, by adding scanning coils to a TEM. Nowadays many TEMs can also operate in scanning mode, making them dual-mode (**TEM/STEM**) instruments.

In order to compete with a conventional TEM in terms of spatial resolution, the electrons must be focused into a probe of sub-nm dimensions. For this purpose, the hot-filament electron source is often replaced by a field-emission source, in which electrons are released from a sharp tungsten tip under the influence of an intense electric field. This was the arrangement used by Crewe and co-workers in Chicago, who in 1965 constructed a **dedicated STEM** that operated only in the scanning mode. A field-emission gun requires ultrahigh vacuum (UHV), meaning ambient pressures of around 10^{-8} Pa. After five years of development, this instrument produced the first-ever images of single atoms. These images were obtained using a high-angle annular dark-field (HAADF) detector that collects electrons scattered through relatively large angles. As shown in Chap. 4, this scattering is proportional to Z^2, where Z is the atomic number, so heavy atoms appear as bright dots on a dark background; see Fig. 1.16.

Nowadays atomic-scale resolution is available in a typical TEM, operated either in the conventional (fixed-beam) mode or in scanning mode. A crystalline specimen can be oriented so that its atomic columns lie parallel to the incident-electron beam, so that *columns* of atoms are imaged; see Fig. 1.17.

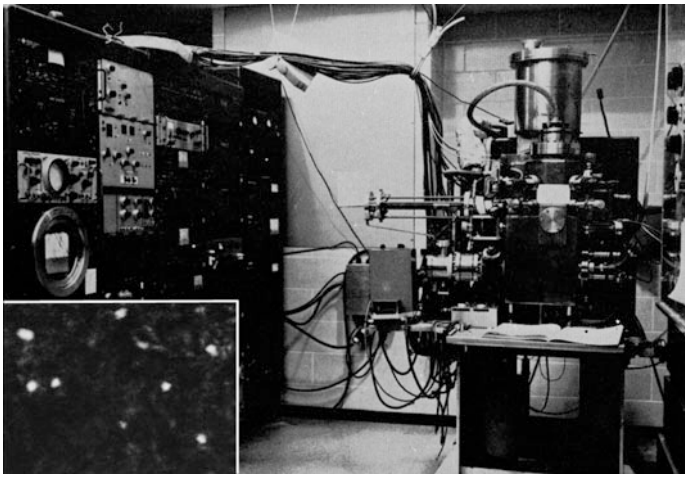


Fig. 1.16 Photograph of the Chicago STEM and (*bottom-left inset*) an image of mercury atoms on a thin-carbon support film. Courtesy of Dr. Albert Crewe (personal communication)

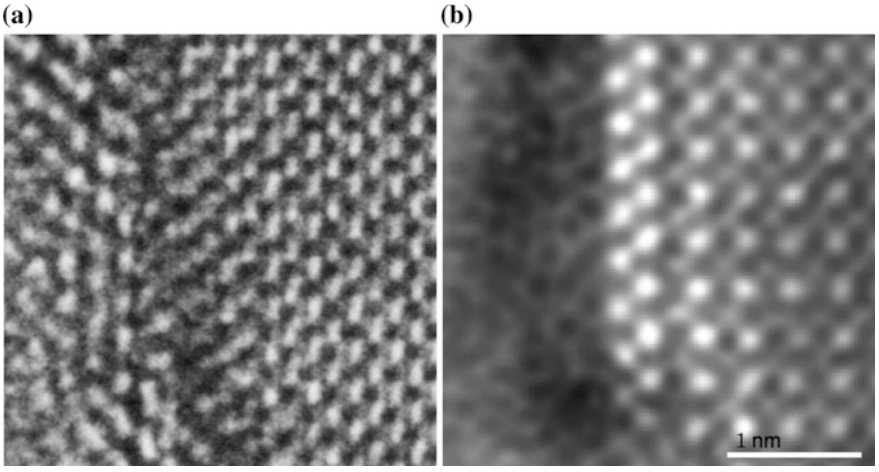


Fig. 1.17 Atomic columns in a strontium titanate crystal, imaged **a** in a TEM using phase contrast and **b** using STEM mode with a high-angle annular dark-field (HAADF) detector. Bismuth atoms have segregated to an interface and are visible as the *brightest dots* in the STEM image. Image courtesy of M. Kawasaki and M. Shiojiri, published in Phil Mag A, 81, 245–260 (2001) by Taylor and Francis Ltd., see http://www.tandfonline.com

1.7 Analytical Electron Microscopy

The previous images provide high-resolution information about the *structure* of a specimen but there is often a need for chemical information, such as the local chemical composition. For this purpose, we need some response from the specimen that is sensitive to the precise atomic number Z of the atoms. As Z increases, the nuclear charge increases, drawing electrons closer to the nucleus and changing their energy. Outer (valence) electrons are affected by the chemical bonding with nearby atoms but the **inner-shell** electrons are not, so their energies can provide an indication of the nuclear charge and therefore atomic number.

When an inner-shell electron makes a transition from a higher to a lower energy level, the atom can emit an x-ray photon whose energy ($hf = hc/\lambda$) is equal to the difference in the two quantum levels. This property is utilized in an x-ray tube, where primary electrons bombard a solid target (the anode) and excite inner-shell electrons to a higher energy, followed by the de-excitation process in which **characteristic** x-rays are generated. In a similar way, the primary electrons entering a TEM, SEM, or STEM specimen cause x-ray emission, and by identifying the wavelengths or photon energies of these x-rays, we can perform **elemental analysis**. Nowadays, an x-ray spectrometer is a common attachment to a TEM, SEM, or STEM, and the instrument becomes an **analytical electron microscope** (AEM).

Other forms of AEM make use of Auger electrons emitted from the specimen with characteristic energies, or the primary electrons themselves after they have traversed a thin specimen and have lost characteristic amounts of energy. We will examine each of these options in Chap. 6.

1.8 Low-Energy and Photoelectron Microscopes

In the low-energy electron microscope (LEEM), electrons are first accelerated to an energy of typically 20 keV, deflected through 60° or 90° by a magnetic prism, and focused into the back-focal plane of an objective lens. This "immersion lens" produces parallel illumination at the specimen and *decelerates* the electrons to a kinetic energy within the range 1–100 eV because the specimen is held at a potential only slightly less negative than the electron-source potential. Electrons penetrate the specimen by only a fraction of a nanometer; those that are elastically backscattered are accelerated back to 20 keV in the objective lens, deflected away from the source by the magnetic prism and form a low-energy diffraction pattern (LEED) that contains information about the atomic spacing on the specimen surface. By selecting a particular diffraction spot, the surface can be imaged to reveal atomic steps (see Fig. 1.18a), differences in atomic spacing, chemical composition, or (with spin-polarized electrons) magnetic properties. Because of the extreme surface sensitivity, sub-monolayer adsorbed films can be detected; to prevent contamination layers from having a predominant effect, the specimen must be

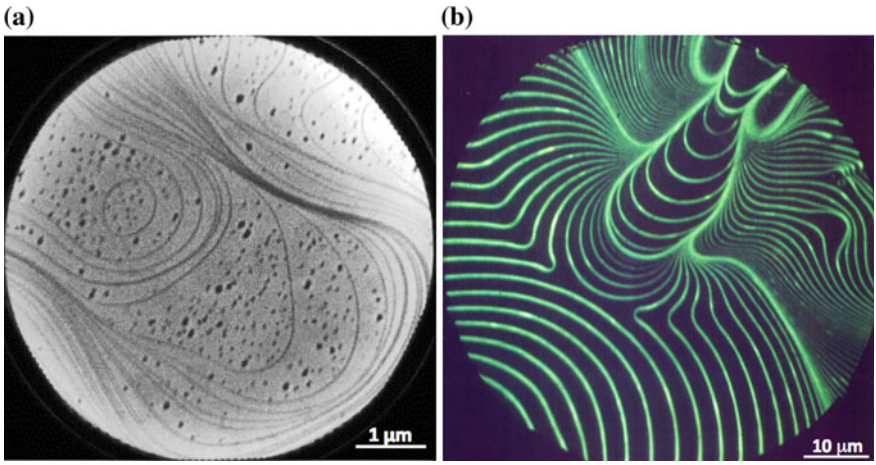


Fig. 1.18 **a** LEEM image showing Co-monolayer islands (*dark areas*) and monatomic steps or step bunches (*dark lines*) on a tungsten (110) surface, taken from a Co-growth movie at an average coverage of 0.05 monolayer. **b** PEEM image of monatomic steps and step bunches on a molybdenum (110) surface decorated with sub-monolayer copper. The contrast is due to the work function difference between Cu and Mo. Images courtesy of Prof. Ernst Bauer, Arizona State University

prepared and kept in ultrahigh vacuum (UHV). In aberration-corrected instruments, a spatial resolution of 2 nm can be achieved.

The photoelectron electron microscope (PEM or PEEM) also uses an immersion objective lens, through which low-energy electrons are accelerated, but these electrons are produced by irradiating the specimen surface with UV light (e.g. from a mercury discharge lamp: UVPEEM) or soft x-rays (usually from a synchrotron: XPEEM). Although this radiation may penetrate a considerable distance, the photoelectrons have low energies and they escape to form an image only if produced very close to the specimen surface. The photoelectron yield is sensitive to differences in work function (see Fig. 1.18b), so PEEM is surface-sensitive and requires UHV. Incorporating an energy filter in XPEEM allows increased chemical sensitivity. With lens-aberration correction, the spatial resolution can be better than 10 nm. Pulsed synchrotron radiation allows the possibility of observing fast processes on surfaces (time-resolved PEEM).

1.9 Field-Emission and Atom-Probe Microscopy

The **field-emission microscope** (FEM), invented as early as 1936, consists of a metal tip of small radius ($r \sim 100$ nm) at a negative potential (1–10 kV) and a fluorescent screen, both in ultrahigh vacuum. The intense electric field ($\sim 10^{10}$ V/m) at the tip results in field emission of electrons, which tunnel out of the

surface potential barrier and travel along field lines to produce an intensity pattern on the screen that depicts atomic planes with different barrier height (work function). This pattern is also an image of the tip, of magnification $L/r \sim 10^5$–10^6, where L is the tip-screen distance. The spatial resolution (about 2 nm) is limited by the wavelength of the emitted electrons.

The **field-ion microscope** (FIM) is similar except that a low-pressure imaging gas (usually He) replaces the UHV and the tip (cooled by liquid nitrogen) is at a positive potential. Field ionization at the tip produces positive gas ions that travel to the screen, giving an image of the tip that now has *atomic resolution*. In further developments, the fluorescent screen was replaced with a position-sensitive time-of-flight mass spectrometer, and voltage or laser pulses were applied to the tip to allow atoms to be individually field-evaporated and mass-analyzed. The end result is **atom-probe tomography** (APT), in which the three-dimensional structure of the tip is displayed graphically on a computer screen, with chemical elements separately identified; see Fig. 1.19. APT instruments are manufactured by the Cameca company and have become a useful research tool since focused-ion-beam (FIB) instruments are now available to produce the necessary needlelike specimens.

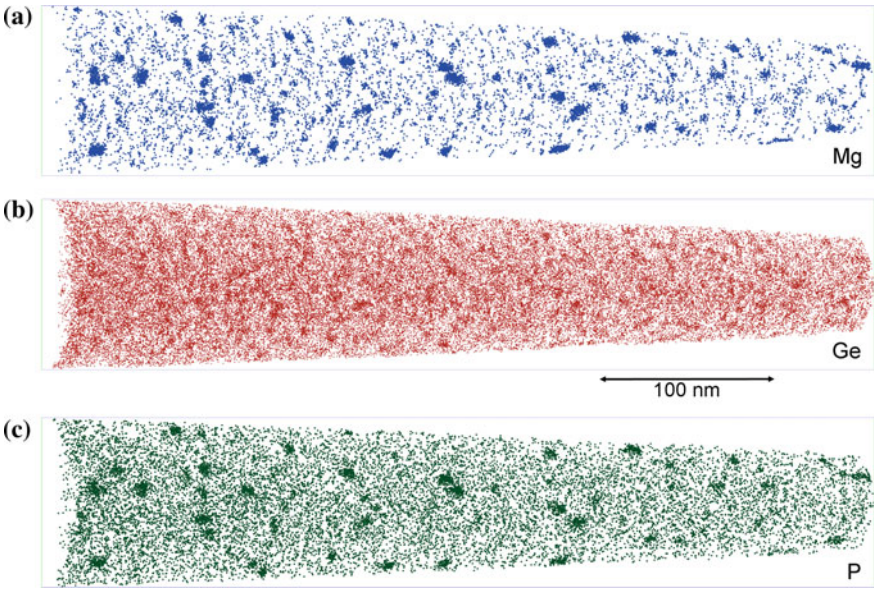


Fig. 1.19 Atom-probe tomography (APT) images of ion-doped silica (an active optical fiber), showing the distributions of **a** magnesium, **b** germanium, and **c** phosphorus in a needle-shaped specimen. Images courtesy of D.J. Larson, D.A. Reinhard and H. Francois-Cyr (CAMECA Instruments Inv.)

1.10 Scanning-Probe Microscopes

The raster method of image formation is employed in a **scanning-probe microscope** (SPM), where a sharply-pointed tip (the probe) is *mechanically* scanned in close proximity to the surface of a specimen, in order to sense some local property. The first such device to achieve really high spatial resolution was the **scanning tunneling microscope** (STM), where a metal tip is brought within 1 nm of the specimen and a small potential difference (≈1 V) is applied between the two. Provided the specimen is electrically conducting, electrons travel between the tip and the specimen through the process of **quantum-mechanical tunneling**. This phenomenon is a consequence of the wavelike character of electrons and is analogous to the leakage of visible-light photons between two glass surfaces brought within 1 μm of each other, sometimes called frustrated internal reflection.

Maintaining a tip within 1 nm of a surface (without touching) requires great mechanical precision, an absence of vibration, and the presence of a feedback mechanism to correct for mechanical and thermal drift. Since the tunneling current increases dramatically with decreasing tip-sample separation, a motorized gear system can be set up to advance the tip toward the sample (in the *z*-direction) until a pre-set tunneling current (e.g. 1 nA) is detected; see Fig. 1.20a. The tip-sample gap is then about 1 nm and fine *z*-adjustments can be made with a piezoelectric drive: a ceramic crystal that changes its length when voltage is applied. If the gap were to decrease, due to thermal expansion for example, the tunneling current would start to increase, raising the voltage across a series resistor; see Fig. 1.20a. However, this voltage change can be amplified and applied to the piezo *z*-drive so as to *decrease* the gap and return the current back to its original value. Such an arrangement is called **negative feedback** because information (about the gap length) is *fed back*, in this case to the electromechanical system that acts to keep the tip-sample separation constant.

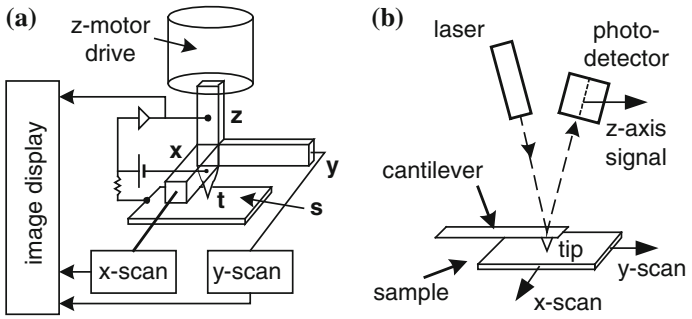


Fig. 1.20 **a** Principle of the scanning tunneling microscope (STM); **x**, **y**, and **z** represent piezoelectric drives, **t** is the tunneling tip, and **s** is the specimen. **b** Principle of the scanning force (or atomic force) microscope. In the *x*- and *y*-directions, the tip is stationary and the sample is raster-scanned by piezoelectric drives

To perform STM imaging, the tip is raster-scanned across the surface of the specimen in *x*- and *y*-directions, again using piezoelectric drives. With the negative-feedback mechanism active, the tip-sample gap remains constant and the tip moves in the *z*-direction in precise synchronism with undulations of the surface (the specimen **topography**). This *z*-motion is represented by variations in the *z*-piezo voltage, which can therefore be used to generate an intensity-modulated image on a display screen (as in the SEM) or stored in computer memory as a topographical image.

One remarkable feature of the STM is its high spatial resolution: better than 0.01 nm in the *z*-direction, resulting from the fact that the tunneling current is a strong (exponential) function of the tunneling gap. It is also possible to achieve high resolution (<0.1 nm) in the *x*- and *y*-directions, even with a tip that is not atomically sharp. The explanation is again in terms of the strong dependence of tunneling current on gap length: most electrons tunnel from tip atoms that are closest to the specimen, rather than from atoms only slightly further away. As a result, the STM can sometimes be used to study the structure of surfaces with single-atom resolution, as illustrated in Fig. 1.21.

Problems can nevertheless occur due to the existence of "multiple tips", resulting in false features (**artifacts**) in an image. Scan times can be long if the scanning is done slowly enough to keep the tip-specimen gap constant. As a result, *atomic-resolution* images are often recorded in variable-current mode, where (once the tip is close to the sample) the feedback mechanism is turned off and the tip is scanned over a *short* distance parallel to the sample surface; changes in tunneling current are then displayed in the image. In this mode, the field of view is limited; for a larger distance of travel, the tip would crash into the specimen, damaging the tip or the specimen or both.

Fig. 1.21 Hydrogen-passivated Si (111) surface, imaged in an STM with a tip potential of −1.5 V. The background hexagonal structure represents H-covered Si atoms, while the two prominent *white* patches arise from dangling bonds where the H atoms have been removed (their *non-circular* appearance indicates an irregularly shaped tip). Courtesy of Jason Pitters and Bob Wolkow, National Institute of Nanotechnology, Canada

Typically the STM head is small, a few cm in dimensions; small size minimizes temperature variations (and therefore thermal drift) and forces mechanical vibrations (resonances) to higher frequency, where they are more easily damped.

The STM was originally developed at the IBM Zurich Laboratory (Binnig et al. 1982) and earned two of its inventors the 1986 Nobel Prize in Physics (shared with Ernst Ruska, for his development of the electron microscope). It quickly inspired other types of scanning-probe microscopes, such as the **atomic force microscope** (AFM) in which a sharp tip (at the end of a cantilever) is brought close enough to a specimen that it essentially touches the surface and senses an interatomic force. For many years, this principle had been applied to measure the roughness of surfaces or the height of surface steps, with a height resolution of a few nm. But in the 1990s, the instrument was refined to provide near-atomic resolution.

Initially the *z*-motion of the AFM tip and cantilever was detected by locating an STM tip immediately above. Nowadays the motion is often monitored by observing the angular deflection of a reflected laser beam, while the specimen is scanned in the *x*- and *y*-directions; see Fig. 1.18b. AFM cantilevers can be made in large quantities from materials such as silicon nitride, using the same kind of photolithography process as used for semiconductor integrated circuits, so they are easily replaced when damaged or contaminated. As with the STM, scanning-force images must be examined critically to avoid interpreting artifacts (such as multiple-tip effects) as real structure.

The mechanical force is repulsive if the tip is in direct contact with the sample but attractive at a small distance above, where the tip senses a Van der Waals force; either regime can be used to provide images. Alternatively, a 4-quadrant photodetector can sense torsional motion (twisting) of the AFM cantilever that represents a sideways frictional force, giving an image that represents the local coefficient of friction. With a modified tip, the magnetic field of a sample can be monitored, allowing the direct imaging of magnetic data-storage media materials. New applications and modes of operation are continually being discovered.

Although it is not so easy (compared to STM) to obtain atomic resolution, the AFM has the advantage that it does not require a conducting sample. It can operate with tip and specimen immersed in a liquid such as water, making the instrument valuable for imaging biological specimens. This versatility, combined with the high resolution and relatively moderate cost, has enabled the scanning probe microscope to take over some of the applications previously reserved for the SEM and TEM. However, mechanically scanning a large area of specimen is time-consuming, whereas an electron-beam microscope can rapidly zoom in and out by changing magnification. Also, it is difficult to provide universal elemental analysis in the AFM. An STM can be used in a spectroscopy mode but the information obtained relates to the *outer-shell* electron distribution and is less directly linked to chemical composition. Except in special cases, a scanning-probe image represents only the surface properties of a specimen and not the *internal* structure that can be viewed using a TEM.

Figure 1.22a shows a typical AFM image, presented in a conventional way in which local changes in image brightness represent variations in surface height

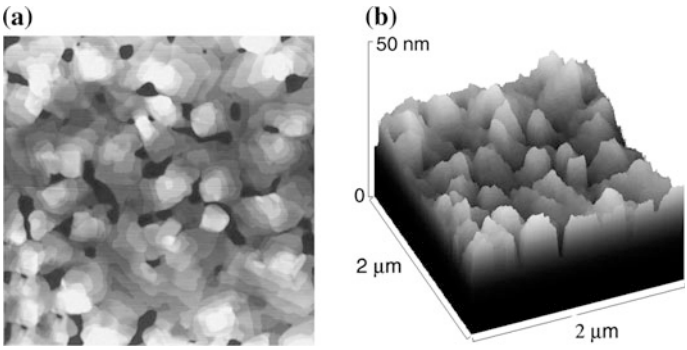


Fig. 1.22 AFM images of a vacuum-deposited thin film of the organic semiconductor pentacene: **a** brightness-modulation image, in which abrupt changes in image brightness represent steps (*terraces*) on the surface, and **b** *y*-modulation image of the same area, providing a direct impression of the surface topography. Courtesy of Hui Qian, University of Alberta

(obtained through motion of the tip in the *z*-direction). However, scanning-probe images are often presented in *y*-modulation mode, in which the *z*-motion of the tip generates a *y*-deflection in the display, perpendicular to the line-scan direction. This procedure gives a three-dimensional effect, equivalent to viewing the specimen surface at an oblique angle rather than perpendicular to the surface. The distance scale of this *y*-modulation is often magnified relative to the scale along the *x*- and *y*-scan directions, exaggerating height differences and making them more easily visible; see Fig. 1.22b. Brightness- and *y*-modulation are sometimes combined, with the addition of color, to create dramatic images.

1.11 Further Reading

Historical references quoted in this Chapter include:

G. Binnig, H. Rohrer, H., C. Gerber and E. Weibel: (1982) *Surface studies by scanning tunneling microscopy*. Phys. Rev. Lett. **49** (1982) 57–60.

C. Davisson and L.H. Germer: *Diffraction of electrons by a crystal of nickel*, Phys. Rev. **30** (1927) 705–740.

M. Knoll and E. Ruska: (1932). *Contribution to geometrical electron optics* (in German). Annalen der Physik 5 (1932) 607.

G.P. Thomson: (1927) *Diffraction of cathode rays by thin films of platinum*. Nature **120** (1927) 802.

V.K. Zworykin, J. Hillier and R.L. Snyder: (1942). *A scanning electron microscope.* ASTM Bull. **117** (1942) 15.

X-ray Microscopy by V.E. Cosslett and W.C.Nixon (Cambridge University Press, 2014, ISBN: 9781107654655), first published in 1960, describes physical principles and early techniques.

Other information about the history of microscopy can be readily found by searching the internet. More recent books, at an advanced level, include the following:

The Science of Microscopy, edited by P.W. Hawkes and J.C.H. Spence (Springer, 2007, ISBN 10: 0-387-25296-7) is a two-volume 1265-page text containing chapters written by experts and dealing with many kinds of electron, optical and scanning-probe microscopy.

Surface Microscopy with Low Energy Electrons by Ernst Bauer (Springer, 2014; ISBN 978-1-4939-0953-3) is available as an eBook and deals with all aspects of LEEM and PEEM. Dr. Bauer's earlier review is in Rep. Prog. Phys. 57 (1994): 895–938. (doi:10.1088/0034-4885/57/9/002). A general and historical introduction to LEEM and PEEM is given by O.H. Griffith and W. Engel in Ultramicroscopy, **36** (1991) 1–28.

Atom Psrobe Tomography: The Local Electrode Atom Probe by M.K. Miller and R.G. Forbes (Springer, 2014; ISBN-10: 978-1-4899-7429-7) is a 423-page book dealing with APT theory and modern practice.

Not discussed in this chapter are instruments that form images with material particles other than electrons. The **helium-ion microscope** provides scanning-mode secondary-electron images with a resolution down to 0.3 nm and with virtually no damage to the specimen. One of its limitations is that elemental analysis by emitted x-rays is not possible. Although the ions are accelerated to a kinetic energy of 30 keV, their mass is 3700 times larger than that of an electron and their speed is below the orbital speed of inner-shell electrons, meaning no x-ray production as explained by D.C. Joy in *Helium Ion Microscopy: Principles and Applications* (Springer, 2013), ISBN: 978-1-4614-8659-6 (Print) 978-1-4614-8660-2 (Online).

However, hydrogen ions (protons) are the basis of an elemental analysis technique called proton-induced x-ray emission (**PIXE**), which can detect very low elemental concentrations: below 10 parts per million. This high sensitivity arises because the background between the characteristic x-ray peaks is very low; the bremsstrahlung process that causes this background (see Sect. 6.2) is much less than with incident electrons because protons are nearly 2000 times more massive and are deflected much less by the electrostatic field of atomic nuclei. A proton energy of 1–5 MeV is typical and 1μm spatial resolution is possible.

Detailed information about ion-beam techniques is given in *Handbook of Modern Ion Beam Materials Analysis*, ed. Y. Wang and M. Nastasi (Materials Research Society, 2nd edition, 2010), ISBN: 9781605112176.

Chapter 2
Electron Optics

Chapter 1 represented an overview of various forms of microscopy, carried out using light, electrons, or mechanical probes. In each case, a microscope forms an enlarged image of the original object (the specimen) in order to show its internal or external structure. Before considering in more detail different aspects of electron microscopy, we will first examine some very general concepts involved in image formation. These ideas were derived during the development of visible-light optics but have a range of applications that is much wider.

2.1 Properties of an Ideal Image

Clearly, an optical image bears a close resemblance to the corresponding object, but what does this mean? What properties should the image have in relation to the object? The answer to this question was provided by the Scottish physicist James Clark Maxwell, who also developed the equations that underlie electrostatic and magnetic phenomena, including electromagnetic waves. In a journal article in 1858, he stated the requirements of a perfect imaging as follows:

1. For each point in the object, there is an ***equivalent*** point in the image.
2. The object and image are geometrically ***similar***.
3. If the object is ***planar*** and perpendicular to the optic axis, so is the image.

Besides defining the desirable properties of an image, Maxwell's principles are useful for categorizing the **image defects** that occur when (in practice) the image is *not* ideal. To see this, we will discuss each rule in turn.

Rule 1 states that for each point in the object we can define a corresponding point in the image. In many forms of microscopy, the connection between these two points is made by some particle (e.g., electron or photon) that leaves the object and ends up at the corresponding image point. The particle is conveyed from object to

R.F. Egerton, *Physical Principles of Electron Microscopy*,
DOI 10.1007/978-3-319-39877-8_2

image through a focusing device (some form of lens) and its trajectory is referred to as a **ray** path. One particular ray path is called the **optic axis**; if no mirrors are involved, the optic axis is a straight line passing through the center of the lens or lenses.

How closely Maxwell's first rule is obeyed depends on several properties of the lens. For example, if the focusing *strength* is incorrect, the image formed at the plane of observation will be **out-of-focus**; particles leaving a single object point arrive anywhere within a circular region, a **disk of confusion**. But even if the focusing power is appropriate, a real lens may produce a disk of confusion because of lens **aberrations**: particles having different energies, or which take different paths after leaving the object, arrive displaced from the "ideal" image point. The image then appears blurred, indicating a loss in spatial resolution.

Rule 2: If we consider object points that form a pattern, their equivalent points in the image should form a *similar* pattern, rather than being distributed at random. For example, any three object points occur at the corners of a triangle and the equivalent points in the image define a triangle which is similar in the geometric sense: it contains the same angles. The image triangle may have a different orientation; for example, it could be **inverted** (relative to the optic axis, as in Fig. 2.1) without violating Rule 2. Or the separations of the three image points may differ from those in the object by a **magnification factor** M, in which case the image is **magnified** (if $M > 1$) or **demagnified** (if $M < 1$).

Even when the light image formed by a glass lens appears similar to the object, a close inspection may reveal the presence of **distortion**. This effect is most easily observed if the object contains straight lines, which appear as curved lines in the distorted image.

The presence of distortion is equivalent to a variation of M with *position* in the object or image: **pincushion** distortion corresponds to M increasing with radial distance away from the optic axis (Fig. 2.2a), **barrel** distortion corresponds to

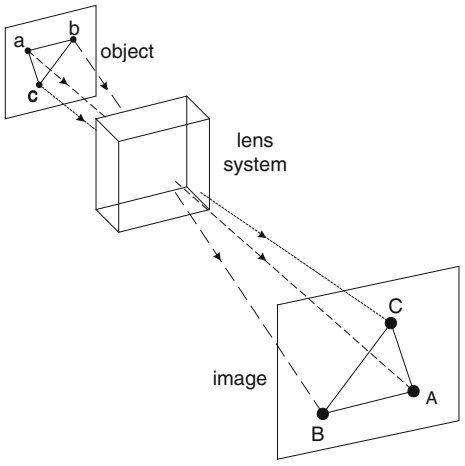


Fig. 2.1 A triangle imaged by an ideal lens, with magnification and inversion. Image points A, B, and C are *equivalent* to the object points a, b, and c, respectively

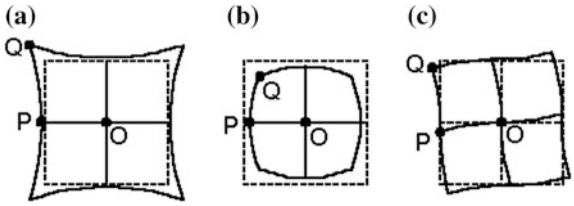


Fig. 2.2 **a** A square mesh (*dashed lines*) imaged with pincushion distortion (*solid curves*); magnification *M* is higher at point Q than at point P. **b** Image showing barrel distortion, with *M* at Q lower than at P. **c** Image of a square, showing spiral distortion; the counterclockwise rotation is higher at Q than at P

M decreasing away from the axis (Fig. 2.2b). As we will see, electron lenses may cause a *rotation* of the image and if this rotation changes with distance from the axis, the result is **spiral** distortion (Fig. 2.2c).

Rule 3: Images usually exist in two dimensions and occupy a *flat* plane. In fact, lenses are usually evaluated using a flat test chart and the sharpest image should ideally be produced on a flat screen. But if the focusing power of a lens depends on the distance of an object point from the optic axis, different regions of the image are brought to focus at different distances from the lens. The optical system then suffers from **curvature of field** and the sharpest image would be formed on a surface that is curved rather than planar. Cameras and telescopes have sometimes compensated for this defect by having a curved recording plane.

2.2 Imaging in Light Optics

Because electron optics involves many of the same concepts as light optics, we will first review some basic optical principles. Glass lenses are used to focus light, based on the property of **refraction**: deviation in direction of a light ray at a boundary where the **refractive index** changes. Refractive index is inversely related to the *speed* of light, which is $c = 3.00 \times 10^8$ m/s in vacuum (and almost the same in air) but c/n in a transparent material (such as glass) of refractive index n. If the angle of incidence (between the ray and a line drawn perpendicular to the interface) is θ_1 in material 1 (e.g., air, $n_1 \approx 1$), the corresponding angle θ_2 in material 2 (e.g. glass, $n_2 \approx 1.5$) differs by an amount given by Snell's law:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \tag{2.1}$$

Refraction can be demonstrated by means of a glass prism; see Fig. 2.3. The total deflection angle α, due to refraction at the air/glass *and* the glass/air interfaces, is independent of the thickness of the glass, but increases with increasing prism angle ϕ. For $\phi = 0$, corresponding to a flat sheet of glass, there is no overall angular deflection ($\alpha = 0$).

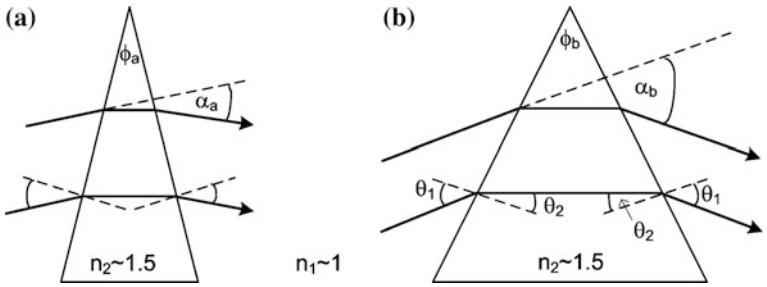


Fig. 2.3 Light refracted by a glass prism **a** of small angle ϕ_a and **b** of large angle ϕ_b. Note that the angle of deflection α is independent of the glass thickness but increases with increasing prism angle

A **convex lens** can be regarded as a prism whose angle increases with distance away from the optic axis. Therefore rays that arrive at the lens far from the optic axis are deflected more than those that arrive close to the axis (the so-called **paraxial** rays), even though the lens gets thinner towards its edge. To a first approximation, the deflection of a ray is proportional to its distance from the optic axis (at the lens), as needed to make rays starting from an on-axis object point *converge* towards a single (on-axis) point in the image (Fig. 2.4) and therefore satisfy the first of Maxwell's rules for image formation.

Figure 2.4 shows only rays that originate from a *single point* in the object, which happens to lie on the optic axis. In practice, rays originate from all points in the two-dimensional object and travel at various angles relative to the optic axis. A diagram showing *all* of these rays would be highly confusing and in practice it is sufficient to represent just a few selected rays, as in Fig. 2.5.

One special ray travels *along* the optic axis. Since it passes through the center of the lens, where the prism angle is zero, this ray does not deviate from a straight line. Similarly, an oblique ray that leaves the object at a distance x_o from the optic axis

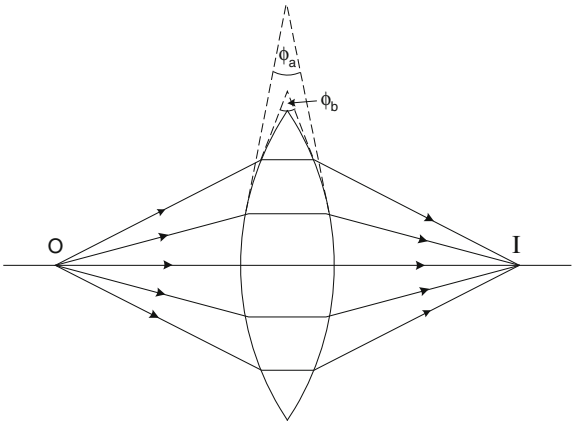


Fig. 2.4 A convex lens focusing rays from axial object point O to an axial image point I

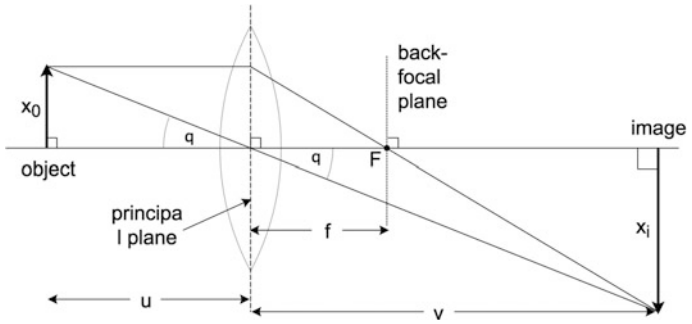


Fig. 2.5 A thin-lens ray diagram, in which bending of light rays is imagined to occur at the mid-plane of the lens (*dashed vertical line*). These special rays define the focal length *f* of the lens and the location of the back-focal plane (*dotted vertical line*)

and passes through the *center of the lens*, remains unchanged in direction. A third special ray leaves the object at distance x_o from the axis but travels *parallel to* the optic axis. The lens bends this ray towards the optic axis, so that it crosses the axis at point F, a distance *f* (the **focal length**) from the center of the lens. The plane that is perpendicular to the optic axis and passes through F is known as the **back-focal plane** of the lens.

In drawing these ray diagrams, we are using **geometric optics** to depict the image formation. Light is represented by rays rather than waves and so we ignore any diffraction effects, which would require **physical optics**.

It is convenient to assume that the bending of light rays takes place at a single plane (known as the **principal plane**) perpendicular to the optic axis (dashed line in Fig. 2.5). This assumption is reasonable if the lens is *physically* thin, so it is known as the **thin-lens approximation**. Within this approximation, the object distance *u* and the image distance *v* (both measured from the principal plane) are related to the focal length *f* by the **thin-lens equation**:

$$1/u + 1/v = 1/f \tag{2.2}$$

We can define image magnification as the ratio of the lengths x_i and x_o measured perpendicular to the optic axis. Since the two triangles defined in Fig. 2.5 are similar (both contain a right angle and the angle θ), the ratios of their horizontal and vertical dimensions must be equal. In other words,

$$v/u = x_i/x_o = M \tag{2.3}$$

From Fig. 2.5, we see that if a single lens forms a **real image** (one that could be viewed by inserting a screen at the appropriate plane), this image is *inverted*, equivalent to a 180° rotation about the optic axis. If a second lens is placed beyond this real image, the latter acts as an *object* for the *second* lens, which produces a *second* real image that is upright (non-inverted) relative to the original object.

The location of this second image is given by applying Eq. (2.2) with appropriate new values of u and v, while the additional magnification produced by the second lens is given by Eq. (2.3). The *total* magnification (between second image and original object) is then just the *product* of magnification factors of the two lenses.

If the second lens is placed *within* the image distance of the first, a real image may not occur. Instead, the first lens is said to form a **virtual image**, which acts as a virtual *object* for the second lens (having a *negative* object distance u). In this situation, the first lens produces *no* image inversion. A familiar example is a magnifying glass held within its focal length of the object; there is no inversion and the only real image is that produced on the retina of the eye, which the brain *interprets* as upright.

Most glass lenses have *spherical* surfaces (sections of a sphere) because these are the easiest to make by grinding and polishing. Such lenses suffer from **spherical aberration**, meaning that rays arriving at the lens at larger distances from the optic axis are focused to points that differ slightly from the focal point of the paraxial rays. Each image point then becomes a disk of confusion and the image produced on any given plane is blurred (reduced in resolution, as discussed in Chap. 1). *Aspherical* lenses have their surfaces tailored to the precise shape required for ideal focusing (for a given object distance) but are more expensive to produce.

Chromatic aberration arises when the light being focused has more than one wavelength present. A common example is white light, which contains a continuous range of wavelengths between its red and blue components. Because the refractive index of glass varies with wavelength (called **dispersion**, since it allows a glass prism to *separate* the component colors of the white light), the focal length f and the image distance v are slightly different for each wavelength present. Again, each image point is broadened into a disk of confusion and image sharpness is reduced.

2.3 Imaging with Electrons

Electron optics has much in common with light optics. We can imagine individual electrons leaving an object and being focused into an image, similar to visible-light photons. Because of this analogy, each electron trajectory is often referred to as a *ray* path.

To obtain the equivalent of a convex lens for electrons, the angular deviation of an electron ray must increase as its displacement from the optic axis increases. We cannot rely on refraction by a material such as glass, since electrons are strongly scattered and absorbed after entering a solid. However, we can take advantage of the fact that the electron has an electrostatic charge and is therefore deflected by an electric field. Alternatively, we can utilize the fact that the electrons in a beam are moving; the beam is equivalent to an electric current in a wire, and can be deflected by applying a magnetic field.

Electrostatic lenses

The simplest example of an electric field is the uniform field produced between two parallel conducting plates when a constant voltage is applied between them. An electron entering such a field would experience a constant force, regardless of its trajectory (ray path). This arrangement is suitable for *deflecting* an electron beam, but not for focusing.

The simplest electrostatic *lens* consists of a circular conducting electrode (disk or tube) connected to a negative potential and containing a circular hole (aperture) centered about the optic axis. An electron passing along the optic axis is repelled equally in all **azimuthal** directions (*around* the optic axis) and therefore suffers no deflection. But an off-axis electron is repelled more strongly by the negative charge that lies closest to it and is therefore deflected back towards the axis, as in Fig. 2.6. To a first approximation, the deflection angle is proportional to displacement from the optic axis and a point source of electrons is focused to a single image point.

A practical form of electrostatic lens (known as a **unipotential** or *einzel* lens, since electrons enter and leave it at the same potential) uses additional electrodes placed before and after, to limit the extent of the electric field produced by the central electrode, as illustrated in Fig. 2.6. Note that the electrodes, and therefore the electric fields that provide the focusing, have cylindrical or **axial symmetry**, which ensures that the focusing force depends only on *radial* distance of an electron from the axis and is independent of its *azimuthal* direction around the axis.

Electrostatic lenses were used in cathode-ray tubes and television picture tubes, to ensure that the electrons emitted from a heated filament were focused back into a small spot on the phosphor-coated inside face of the tube. Although some early electron microscopes used electrostatic optics, modern electron-beam instruments

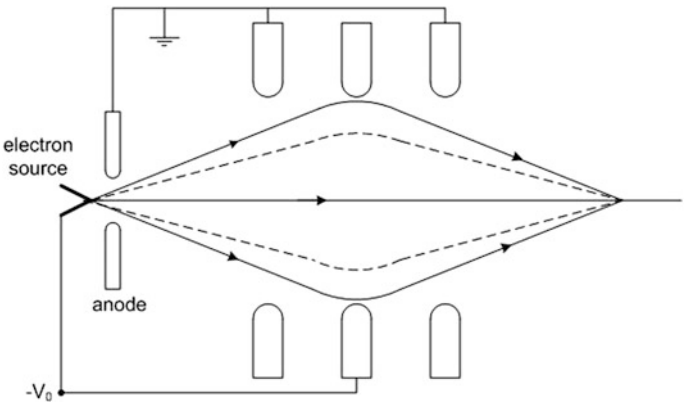


Fig. 2.6 Electrons emitted from an electron source, accelerated through a potential difference V_0 towards an anode, and then focused by a unipotential electrostatic lens. The electrodes, seen here in cross-section, are circular disks containing round holes (apertures) with a common axis, the optic axis of the lens

use electromagnetic lenses that do not require high-voltage insulation and have somewhat lower aberrations.

Magnetic Lenses

To focus an electron beam, an electromagnetic lens generates a magnetic field by passing a direct current through a coil containing many turns of wire. As in the electrostatic case, a *uniform* field (applied perpendicular to the beam) would deflect but not focus a parallel beam. To achieve focusing, we require a field with rotational symmetry, similar to that of the einzel lens, and such a field is generated by the short coil illustrated in Fig. 2.7a. As an electron passes through this non-uniform magnetic field, the force acting on it varies in both magnitude and direction, so the force must be represented by a vector quantity $\boldsymbol{F}$. According to the electromagnetic theory,

$$\boldsymbol{F} = -e(\boldsymbol{v} \times \boldsymbol{B}) \tag{2.4}$$

In Eq. (2.4), $-e$ is the negative charge of the electron, $\boldsymbol{v}$ is its velocity vector, and vector $\boldsymbol{B}$ represents the magnetic field: its magnitude B (the **induction**, measured in Tesla) and its direction. The symbol × indicates a cross-product or **vector product** of $\boldsymbol{v}$ and $\boldsymbol{B}$, a mathematical operator that gives Eq. (2.4) the following two properties.

1. The direction of $\boldsymbol{F}$ is *perpendicular* to both $\boldsymbol{v}$ and $\boldsymbol{B}$. Consequently, $\boldsymbol{F}$ has *no* component in the direction of motion, implying that the electron speed v (the *magnitude* of the velocity $\boldsymbol{v}$) remains constant at all times. But since the

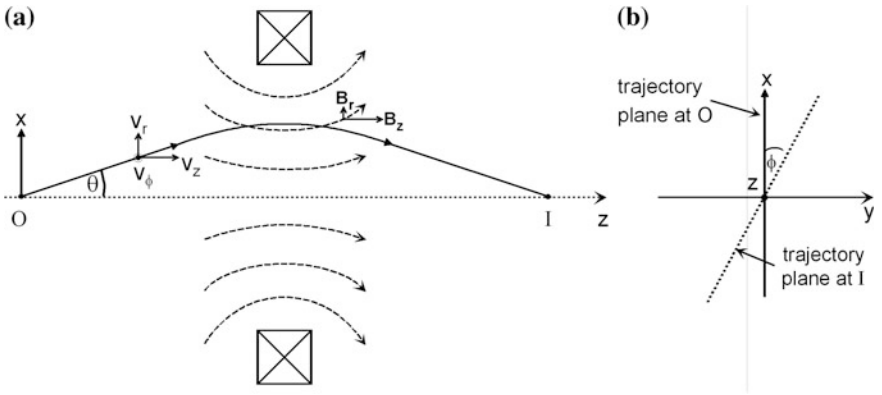


Fig. 2.7 **a** Magnetic flux lines (*dashed curves*) produced by a short coil, seen here in cross-section, together with the trajectory of an electron from an axial object point O to the equivalent image point I. **b** View along the optic axis (z-direction) showing the rotation ϕ of the plane of the electron trajectory, which is also the azimuthal rotation angle for an extended image produced at I

directions of $\boldsymbol{B}$ and $\boldsymbol{v}$ change as the electron moves through the field, the direction of $\boldsymbol{F}$ changes also.

2. The magnitude F of the force is given by:

$$F = e\,v\,B\sin(\varepsilon) \tag{2.5}$$

where ε is the instantaneous angle between $\boldsymbol{v}$ and $\boldsymbol{B}$ at the position of the electron. Because B changes continuously as an electron passes through the field, so does F. Note that for an electron traveling along the coil axis, $\boldsymbol{v}$ and $\boldsymbol{B}$ are both always in the axial direction, giving $\varepsilon = 0$ and $F = 0$ at every point, implying no deviation of the ray path from a straight line. Therefore the symmetry axis of the magnetic field is the optic axis.

For non-axial trajectories, the motion of the electron is more complicated. It can be analyzed in detail by using Eq. (2.4) in combination with Newton's second law ($\boldsymbol{F} = m\,\mathrm{d}\boldsymbol{v}/\mathrm{d}t$) and such analysis is simplified by considering $\boldsymbol{v}$ and $\boldsymbol{B}$ in terms of their vector components. Although we could take components parallel to three perpendicular axes (x, y, and z), it makes more sense to recognize that the magnetic field here possesses axial (cylindrical) symmetry. We therefore use cylindrical coordinates: z, r (=radial distance away from the z-axis) and ϕ (=azimuthal angle, representing the direction of the radial vector **r** relative to the plane of the initial trajectory).

In Fig. 2.7a, v_z, v_r, and v_ϕ are the axial, radial, and tangential components of electron velocity, while B_z and B_r are the axial and radial components of magnetic field. Equation (2.5) can then be rewritten to give the tangential, radial, and axial components of the magnetic force on an electron:

$$F_\phi = -e(v_z B_r) + e(B_z v_r) \tag{2.6a}$$

$$F_r = -e(v_\phi B_z) \tag{2.6b}$$

$$F_z = e(v_\phi B_r) \tag{2.6c}$$

We now trace the path of an electron that starts from an axial point O and enters the field at an angle θ relative to the symmetry (z) axis, as defined in Fig. 2.7. As the electron approaches the field, the main component is B_r and the predominant force comes from the term ($v_z\,B_r$) in Eq. (2.6a). Since B_r is negative (field lines *approach* the z-axis), this contribution ($-e\,v_z\,B_r$) to F_φ is positive, meaning that the tangential force F_φ is clockwise, viewed along the $+z$ direction. As the electron approaches the center of the field ($z = 0$) the magnitude of B_r decreases but the second term $e(B_z\,v_r)$ in Eq. (2.6a), which is also positive, increases. So as a result of both terms, the electron starts to spiral within the field, acquiring an increasing tangential velocity v_φ directed *out* of the plane of Fig. (2.7a). As a result of this tangential component, a new force F_r starts to act on the electron. According to Eq. (2.6b), this force is negative (towards the z-axis), so we have a focusing action: the non-uniform magnetic field acts like a convex lens.

Provided that the radial force F_r towards the axis is large enough, the radial motion of the electron will be *reversed* and the electron will approach the z-axis; v_r becomes negative and the *second* term in Eq. (2.6a) becomes negative. After the electron passes the $z = 0$ plane (the center of the lens), the field lines start to diverge so that B_r becomes positive and the *first* term in Eq. (2.6a) also becomes negative. As a result, F_φ becomes negative (reversed in direction) and the tangential velocity v_φ falls, as shown in Fig. 2.8c; by the time the electron leaves the field, its spiraling motion is reduced to zero. However, the electron is now traveling in a plane that has been *rotated* relative to its original (x–z) plane, as shown in Fig. 2.7b.

This rotation effect is *not* depicted in Fig. 2.7a or in other ray diagrams in this book, where for convenience we plot the radial distance r of the electron (from the axis) as a function of its axial distance z. This common convention allows the use of two-dimensional rather than three-dimensional diagrams; by deliberately ignoring the rotation effect, we can draw ray diagrams that resemble those of light optics. Even so, it is important to remember that when an electron passes through an axially symmetric magnetic lens, its motion has a rotational component.

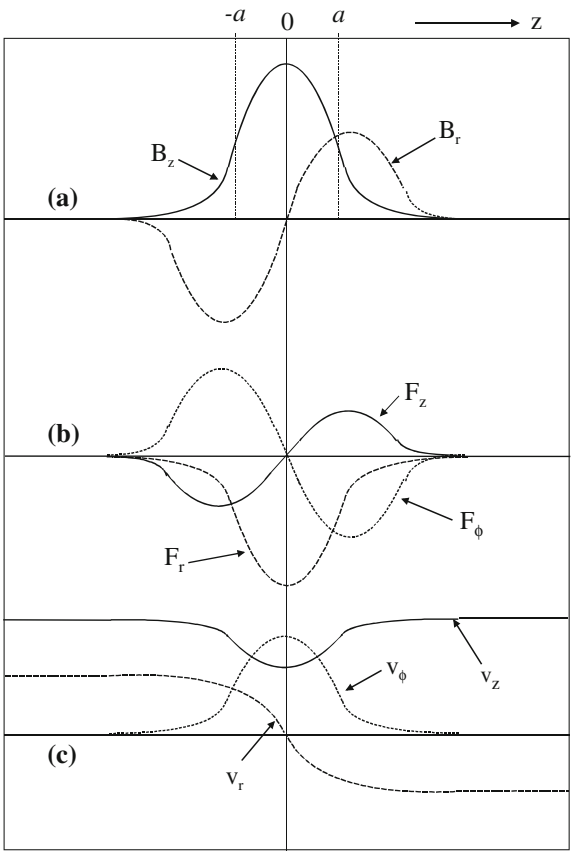


Fig. 2.8 Qualitative behavior of the radial (r), axial (z), and azimuthal (ϕ) components of **a** magnetic field, **b** force on an electron, and **c** the resulting electron velocity, as a function of the z-coordinate of an electron going through the electron lens shown in Fig. 2.7a

Since the overall speed v of an electron in a magnetic field remains constant, the appearance of tangential and radial components of velocity implies that v_z must decrease, as depicted in Fig. 2.8c. This is in accord with Eq. (2.6c), which predicts the existence of a third force F_z that is negative for $z < 0$ (because $B_r < 0$) and therefore acts in the $-z$ direction. After the electron passes the center of the lens, z, B_r and F_z all become positive and v_z increases back to its original value. The fact that the total speed v remains constant contrasts with the case of the einzel electrostatic lens, where an electron slows down as it passes through the retarding field.

We have seen that the radial component of magnetic induction B_r plays an essential part in electron focusing. If a long coil (solenoid) were used to generate the field, B_r would be present only in the **fringing field** at either end. (The uniform field inside the solenoid can focus electrons radiating from a point source but not a broad beam of electrons traveling parallel to its axis). So rather than using an *extended* magnetic field, we should make the field as short as possible by partially enclosing the current-carrying coil by ferromagnetic material such as soft iron, as shown in Fig. 2.9a. Due to its high permeability, the iron conducts most of the magnetic flux lines. However, the magnetic circuit contains a **gap** filled with nonmagnetic material, so that magnetic flux is forced to appear within the internal **bore** of the lens. The magnetic field experienced by an electron can be increased and further localized by the use of ferromagnetic (soft iron) **polepieces** having small internal diameter, as illustrated in Fig. 2.9b. These polepieces are machined to high precision to ensure that the magnetic field has the high degree of axial (rotational) symmetry needed for good focusing.

A cross-section through a typical magnetic lens is shown in Fig. 2.10, where the optic axis is now vertical. A typical electron-beam instrument contains several lenses and stacking them vertically (in a lens **column**) provides a mechanically robust structure in which the weight of each lens acts parallel to the optic axis. There is then no tendency for the column to gradually bend under its own weight, leading to lens **misalignment**. The strong magnetic field (up to about 2 T) in each lens gap is generated by a relatively large coil that contains many turns of wire and may carry a few amps of direct current. To remove heat generated in the coil, due to its ohmic resistance, water flows into and out of each lens. Water cooling ensures

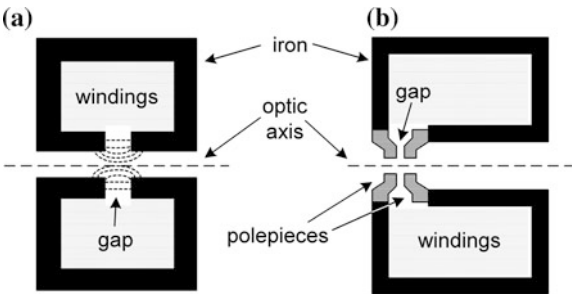


Fig. 2.9 **a** Use of ferromagnetic soft iron to concentrate magnetic induction within a small volume. **b** Use of ferromagnetic polepieces to further concentrate the field

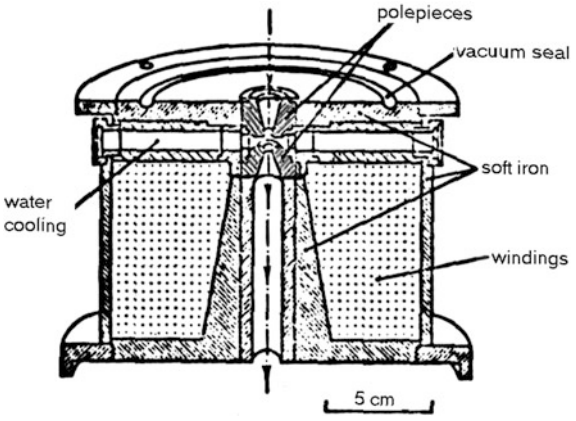


Fig. 2.10 Cross-section through a magnetic lens whose optic axis (*dash-dot line*) is vertical

that the temperature of the lens column reaches a stable value, so that thermal expansion (which could result in column misalignment) is minimized. Temperature changes are further reduced by controlling the temperature of the cooling water, using a refrigeration system to circulate water through the lenses (in a closed cycle) and remove the heat generated.

Rubber **o-rings** (of circular cross-section) provide an airtight seal between the interior of the lens column (under vacuum) and the exterior, which is at atmospheric pressure. The absence of air in the vicinity of the electron beam is essential to prevent collisions and scattering of the electrons by gas molecules. Some internal components (such as apertures) must be located close to the optic axis but adjusted in position by external controls. Sliding o-ring seals or thin-metal bellows are used to allow this motion, while preserving the internal vacuum.

2.4 Focusing Properties of a Thin Magnetic Lens

The use of ferromagnetic polepieces results in a focusing field that extends only a few mm along the optic axis, so that to a first approximation the lens can be considered thin. Deflection of the electron trajectory then occurs close to a single principal plane, allowing thin-lens formulas such as Eqs. (2.2) and (2.3) to be used to describe the optical properties of the lens.

The thin-lens approximation also simplifies analysis of the motion of a particle of charge e and mass m, leading to a compact expression for the **focusing power** (reciprocal of focal length) of a magnetic lens:

$$1/f = \left[e^2/(8mE_0)\right] \int B_z^2 \, \mathrm{d}z \tag{2.7}$$

Since we are considering electrons, $e = 1.6 \times 10^{-19}$ C and $m = 9.11 \times 10^{-31}$ kg; E_0 represents the kinetic energy of the electrons passing through the lens, expressed in Joule and given by $E_0 = (e)(V_0)$ where V_0 is the potential difference used to accelerate the electrons from rest. The integral ($\int B_z^2\, dz$) can be interpreted as the total area under a graph of B_z^2 plotted against distance z along the optic axis, B_z being the z-component of magnetic field (in Tesla). Because the field is non-uniform, B_z is a function of z and depends also on the lens current and polepiece geometry.

There are two simple cases in which the integral in Eq. (2.7) can be solved analytically. One of these corresponds to the assumption that B_z has a constant value B_0 in a region $-a < z < a$ but falls abruptly to zero outside this region. The total area is then that of a rectangle and the integral becomes $2aB_0^2$. This rectangular distribution is unphysical but would approximate the field produced by a long solenoid of length $2a$.

For a typical electron lens, a more realistic assumption is that B increases smoothly towards its maximum value B_0 (at the center of the lens) according to a symmetric *bell-shaped* curve described by the **Lorentzian** function:

$$B_z = B_0/\left(1 + z^2/a^2\right) \tag{2.8}$$

As seen by substituting z = a in Eq. (2.8), a is the half-width at half maximum (HWHM) of a plot of B_z versus z, meaning the distance (away from the central maximum) at which B_z falls to *half* of its maximum value; see Fig. 2.8a. Alternatively, $2a$ is the *full* width at half maximum (FWHM) of the field: the length (along the optic axis) over which the field exceeds $B_0/2$. If Eq. (2.8) is a good approximation, the integral in Eq. (2.7) becomes $(\pi/2)aB_0^2$ and the focusing power of the lens is:

$$1/f = (\pi/16)\left[e^2/(mE_0)\right]aB_0^2 \tag{2.9}$$

As an example, we can take $B_0 = 0.3$ T and $a = 3$ mm. If the electrons entering the lens have been accelerated from rest by applying a voltage $V_0 = 100$ kV, we have $E_0 = eV_0 = 1.6 \times 10^{-14}$ J. Equation (2.9) then gives the focusing power $1/f = 93$ m^{-1} and focal length $f = 11$ mm. Since f turns out to be less than twice the full width of the field ($2a = 6$ mm) we might question the accuracy of the thin-lens approximation in this case. More exact theory (Reimer and Kohl 2008) shows that the thin-lens formula underestimates f by about 14 % for these parameters. For larger B_0 and a, Eqs. (2.7) and (2.9) become unrealistic (see Fig. 2.13 later). In other words, strong lenses have to be treated as *thick* lenses, for which (as in light optics) the mathematical description is more complicated.

In addition, our thin-lens formula for $1/f$ is based on *non-relativistic* mechanics, in which the mass of the electron is assumed to be equal to its rest mass. The relativistic increase in mass, predicted by Einstein's Special Relativity, can be incorporated by replacing E_0 by E_0 (1 + V_0/1022 kV) in Eq. (2.9). This modification increases f by about 1 % for each 10 kV of accelerating voltage, that is by 10 % for $V_0 = 100$ kV, 20 % for $V_0 = 200$ kV, and so on.

Although only approximate, Eq. (2.9) allows us to see how the focusing power of a magnetic lens depends on the strength and spatial extent of the magnetic field, and on the kinetic energy, charge, and mass of the particles being imaged. Since kinetic energy E_0 appears in the denominator of Eq. (2.7), the focusing power decreases as the accelerating voltage is increased. As might be expected, electrons are deflected less in the magnetic field as their momentum increases.

Since B_0 is proportional to the current supplied to the lens windings, changing this current allows the focusing power of the lens to be varied. This ability to vary the focal length means that an electron image can be focused by adjusting the lens current, without mechanical motion. But it also implies that the lens current must be highly stabilized (typically to within a few parts per million) to prevent *unwanted* changes in focusing power, which would cause the image to drift out-of-focus. In *light optics*, change in f can only be achieved mechanically: by changing the curvature of the lens surfaces (in the case of the eye) or by changing the spacing between elements of a compound lens, as in the zoom lens of a camera.

When discussing the action of a magnetic field, we saw that an electron passing through a lens executes a spiral motion. As a result, the plane containing the exit ray is rotated through an angle ϕ relative to the plane containing the incoming electron. By again making a thin-lens approximation ($a \ll f$) and assuming a Lorentzian field distribution, the equations of motion can be solved to give:

$$\phi = \left[e/(8mE_0)^{1/2}\right] \int B_z \mathrm{d}z = \left[e/(8mE_0)^{1/2}\right] \pi a B_0 \tag{2.10}$$

Using $B_0 = 0.3$ T, $a = 3$ mm, and $V_0 = 100$ kV as before: $\phi = 1.33$ rad $= 76°$, so the rotation is significant. Note that this rotation is *in addition* to the inversion about the z-axis that occurs when a real image is formed (Fig. 2.5). In other words, the rotation of a *real* electron image, relative to the object, is actually $\pi \pm \phi$ radians.

Note that the image rotation would reverse (e.g., go from clockwise to counter-clockwise) if the current through the lens windings were reversed, as B_z would be reversed in sign. On the other hand, reversing the current has no effect on the focusing power, since the integral in Eq. (2.7) involves B_z^2, which is always positive. In fact, all of the terms in Eq. (2.7) are positive, implying that we cannot use an axially symmetric magnetic field to produce the electron-optical equivalent of a diverging (concave) lens having negative focal length.

Equations (2.7)–(2.10) apply equally well to the focusing of other charged particles, such as protons and ions, provided e and m are replaced by the appropriate charge and mass. Equation (2.7) shows that, for the same lens current and kinetic energy (same accelerating potential), the focusing power of a magnetic lens is much less for these particles, whose mass is thousands of times larger than that of the electron. For this reason, **ion optics** commonly involves *electrostatic* lenses.

2.5 Comparison of Magnetic and Electrostatic Lenses

Because the electrostatic force on a charged particle is parallel (or in the case of an electron, antiparallel) to the field and since axially symmetric fields have no tangential component, electrostatic lenses offer the convenience of no image rotation. Their low weight and near-zero power consumption have made them attractive for space-based equipment, such as planetary probes.

Electrostatic lenses are also used in focused-ion-beam (FIB) machines, which are commonly used for the fabrication of nm-scale devices and for the preparation of TEM specimens; see Sect. 4.10. Ions such as Ga+ are accelerated by voltages of typically 5–30 kV and can remove atoms of the specimen by sputtering. Because of the large mass of these ions compared to that of an electron, magnetic lenses would be inefficient. Electrostatic lenses can generate a probe with a diameter of a few nm, to provide scanning-mode images of the specimen using secondary electrons or secondary ions.

If the voltage V_0 that accelerates electrons is also used as the voltage applied to an electrostatic lens (as in Fig. 2.6), the lens focusing becomes insensitive to small voltage changes. If V_0 increases, the electrons travel faster but require a higher lens voltage to focus them, so the effect of the voltage change cancels to first order. Electrostatic lenses were used in some early electron microscopes because their high-voltage supplies were prone to slow change (drift) and fluctuating components (ripple). Nowadays, high-voltage supplies can be stabilized to one part in a million, although this requires careful and expensive manufacture.

The fact that a voltage comparable to the accelerating voltage V_0 must be applied to an electrostatic lens means that insulation and safety problems become severe for $V_0 > 50$ kV. Because higher accelerating voltages permit better image resolution, magnetic lenses came to be preferred for electron microscopy. Magnetic lenses also provide somewhat lower aberrations, for the same focal length, thereby improving the image resolution.

As we will see later in this chapter, lens aberrations are also minimized by making the focal length of a lens small, implying an *immersion* objective lens with the specimen present *within* the lens field. This situation gives rise to problems in the case of ferromagnetic specimens, which distort the local field of a magnetic lens and prevent good imaging (unless the objective excitation is turned off, which impairs the image resolution). However, introducing *any* specimen into the field of an electrostatic lens would change its field distribution, making an electrostatic immersion lens impractical.

With the advent of spherical-aberration correction (Sect. 2.7), an immersion lens of small focal length becomes less essential to good resolution, so electrostatic lenses begin to look more attractive, at least at lower voltages where insulation is less problematic. Chromatic aberration becomes limiting at low electron energies but can also be corrected; see Sect. 2.7.

Table 2.1 Comparison of electrostatic and electromagnetic lens designs

Advantages of electrostatic lenses	Advantages of magnetic lenses
No image rotation	Lower lens aberrations
Lightweight, consume no power	No high-voltage insulation needed
Highly stable voltage unnecessary	Can function as an immersion lens

Especially if used with ions, which can contaminate an optical column with sputtered atoms, electrostatic lenses may require periodic cleaning to avoid electrostatic discharge across their insulator surfaces. The differences between electrostatic and magnetic optics are summarized in Table 2.1.

2.6 Defects of Electron Lenses

In the case of a microscope, the most important focusing defects are lens *aberrations*, since they reduce the spatial resolution of the image, even when it has been optimally focused. We will discuss two kinds of **axial aberration** that lead to image blurring even for object points that lie *on the optic axis*.

Spherical Aberration

The focusing of light by a glass lens with spherical surfaces is imperfect: a **marginal** ray that is refracted at the edge of the lens is focused more strongly than a **paraxial** ray that travels close to the optic axis. A similar effect occurs with electron lenses, magnetic or electrostatic.

Spherical aberration can be defined by means of a diagram that shows electrons arriving at a thin lens after traveling parallel to the optic axis but not necessarily along it; see Fig. 2.11. Paraxial rays (represented by dashed lines in Fig. 2.11) are brought to a focus F, a distance f from the center of the lens, at the **Gaussian** image plane. Electrons arriving at a distance x from the axis are focused to a different point F_1, located a shorter distance f_1 from the center of the lens.

We might expect the axial shift in focus ($\Delta f = f - f_1$) to depend on the initial x-coordinate of the electron and on the degree of imperfection of the lens focusing. Without knowing the details of this imperfection, we can represent the x-dependence in terms of a power series:

$$\Delta f = c_2 x^2 + c_4 x^4 + higher\ even\ powers\ of\ x \tag{2.11}$$

with c_2 and c_4 as unknown coefficients. Note that odd powers of x have been omitted: provided the magnetic field that focuses the electrons is rotationally symmetric, the deflection angle α should be identical for electrons that arrive with coordinates $+x$ and $-x$, as in Fig. 2.11. This would not be the case if terms involving x or x^3 were present in Eq. (2.11).

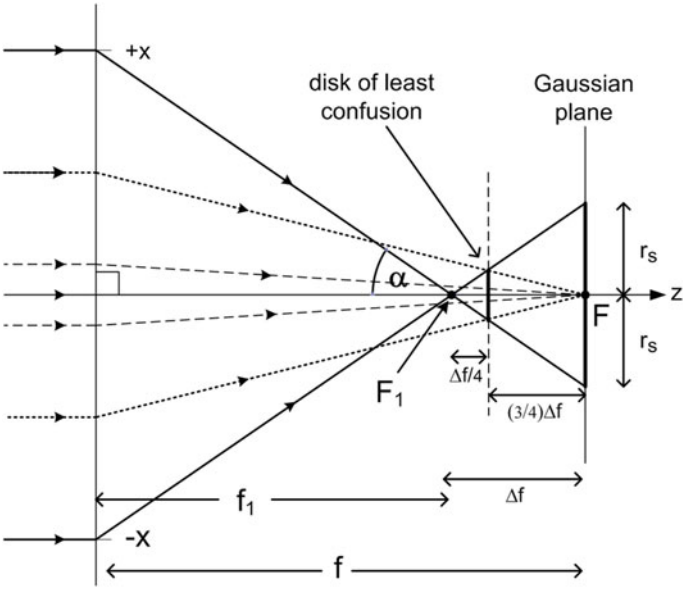


Fig. 2.11 Definition of the disk of confusion (at the Gaussian image plane) arising from spherical aberration, in terms of the focusing of parallel rays by a thin lens

From the geometry of the large right-angled triangle in Fig. 2.11,

$$x = f_1 \tan\alpha \approx f \tan\alpha \approx f\,\alpha \tag{2.12}$$

Here we have assumed that $x \ll f$, taking the angle α to be small, and that $\Delta f \ll f$, supposing spherical aberration to be a *small* effect for electrons that deviate by no more than a few degrees from the optic axis. These approximations are reasonable for typical electron-accelerating voltages.

When non-paraxial electrons arrive at the Gaussian image plane, they are displaced radially from the optic axis by an amount r_s given by:

$$r_s = \Delta f \tan\alpha \approx \Delta f\,\alpha \tag{2.13}$$

where we once again assume that α is small. Since small α implies small x, we can to a first approximation neglect powers higher than x^2 in Eq. (2.11) and combine this equation with Eqs. (2.12) and (2.13) to give:

$$r_s \approx [c_2\,(f\alpha)^2]\alpha = c_2\,f^2\alpha^3 = C_s\,\alpha^3 \tag{2.14}$$

where we have combined c_2 and f into a single constant C_s, the **coefficient of spherical aberration** of the lens. Because α (in radian) is dimensionless, C_s has the dimensions of length.

Figure 2.11 illustrates a limited number of off-axis electron trajectories. More typically, we have a broad entrance beam of circular cross-section, containing

electrons arriving at the lens with a continuous range of radial displacement within the x–z plane (that of the diagram), within the y–z plane (perpendicular to the diagram), and within all intermediate planes that contain the optic axis. Due to the rotational symmetry, all these electrons arrive at the Gaussian image plane *within* the disk of confusion (radius r_s). The angle α now represents the *maximum* angle of the focused electrons, which might be determined by the internal diameter of the lens bore or by a circular aperture placed in the optical system.

Figure 2.11 is directly relevant to a scanning electron microscope (SEM), where the objective lens focuses a near-parallel beam into an electron probe of very small diameter at the specimen. Because the spatial resolution of a secondary-electron image cannot be better than the probe diameter, spherical aberration might be expected to limit the spatial resolution to a value of the order of $2r_s$. In fact, this conclusion is too pessimistic: if the specimen is advanced towards the lens, the illuminated disk gets smaller and at a certain location (represented by the dashed vertical line in Fig. 2.11), its diameter has a minimum value ($=r_s/2$) corresponding to the **disk of least confusion**. Advancing the specimen any closer to the lens would make the disk larger, due to contributions from medium-angle rays, shown dotted in Fig. 2.11.

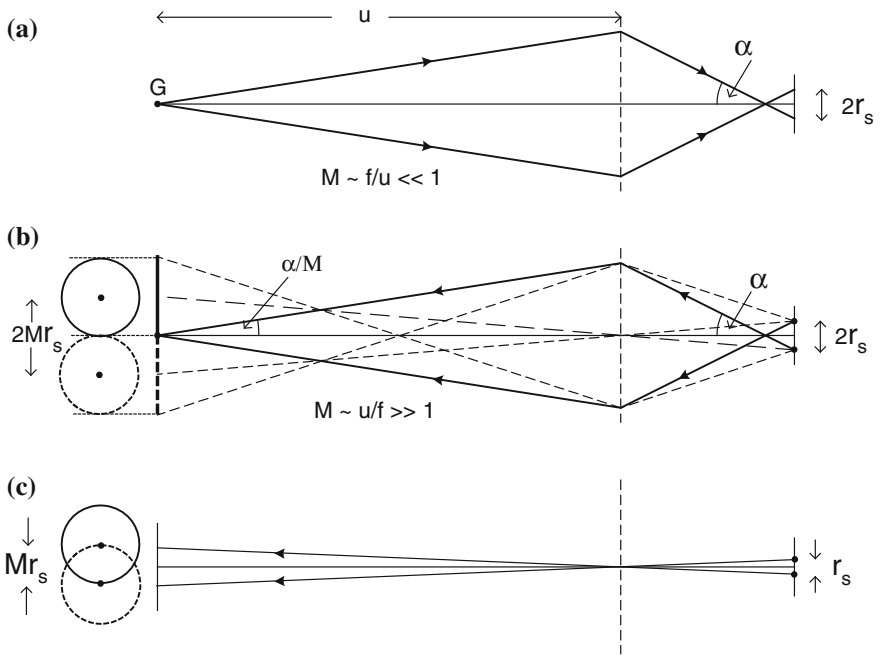


Fig. 2.12 **a** Ray diagram as in Fig. 2.11 but with an object distance u large but finite, a situation that applies to demagnification of an electron source G by an SEM objective lens. **b** Equivalent diagram with the solid rays reversed, showing two image disks of confusion arising from object points whose separation is $2r_s$. **c** Same diagram but with object-point separation reduced to r_s such that the two points are barely resolved in the image, according to the Rayleigh criterion

In the case of a transmission electron microscope (TEM), a relatively broad beam of electrons arrives at the specimen and an objective lens simultaneously images each object point. Figure 2.11 can be adapted to this case by first imagining the lens to be weakened slightly, so that F is the Gaussian image of an object point G located at a large but finite distance u from the lens, as in Fig. 2.12a. Although r_s is still given approximately by Eq. (2.14), this diagram depicts large *de*magnification ($M \ll 1$). To make it relevant to a TEM objective ($M \gg 1$), we can simply reverse the ray paths, a procedure that is permissible in both light and electron optics, resulting in Fig. 2.12b. By adding extra dashed rays as shown, Fig. 2.12b illustrates how electrons emitted from two points a distance $2r_s$ apart are focused into two magnified disks of confusion in the image (on the left) of radius Mr_s and separation $2Mr_s$. Although these disks touch at their periphery, the image would still be recognizable as representing two separate point-like objects in the specimen. If the separation between the object points is now reduced to r_s (as in Fig. 2.12c), the disks overlap enough to satisfy the Rayleigh criterion for resolution (Sect. 1.1). We can therefore take r_s as the spherical-aberration limit to the **point resolution** of a TEM objective lens. In microscopy we always specify blurring in the *object plane* because it is the object resolution that we care about.

Spherical aberration occurs in TEM lenses *after* the objective but is much less significant. This situation arises from the fact that each lens *reduces* the maximum angle of electrons (relative to the optic axis) by a factor equal to its magnification (as illustrated in Fig. 2.12b) and the spherical-aberration blurring depends on the *third power* of this angle, according to Eq. (2.14).

So far, we have said nothing about the *value* of the spherical-aberration coefficient C_s. On the assumption of a Lorentzian (bell-shaped) field, C_s can be calculated from a somewhat-complicated formula (Glaser 1952). Figure 2.13 shows the calculated C_s and focal length f as a function of the maximum field B_0, for

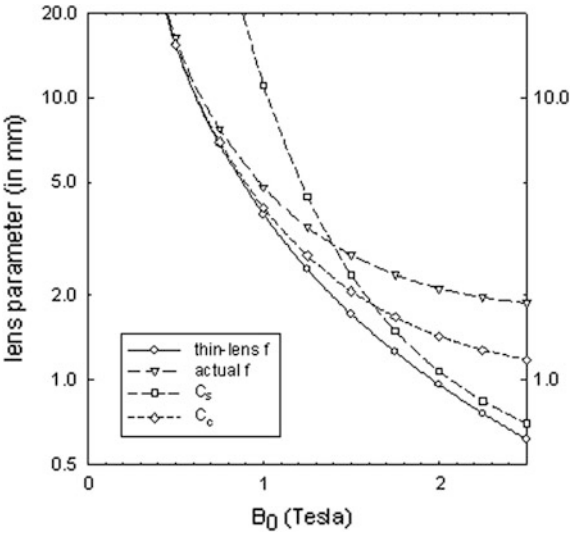


Fig. 2.13 Focal length and the coefficients of spherical and chromatic aberration for a magnetic lens containing a Lorentzian field with peak field B_0 and half-width a = 1.8 mm, focusing 200 keV electrons. Thin-lens values were calculated from Eq. (2.7), other values taken from Glaser (1952)

200 kV accelerating voltage and a field half-width of $a = 1.8$ mm. The thin-lens formula, Eq. (2.9), is seen to be quite good at predicting the focal length of a weak lens (low B_0) but gives too low a value for a strong lens. The aberration coefficient C_s exceeds the actual focal length for a weak lens but is less than the focal length for a strong lens (and closer to f given by the thin-lens formula). For a typical TEM objective lens, we might take $f = 2$ mm and $C_s \approx 0.5$ mm; for a point resolution $r_s < 1$ nm, the maximum angle of the electrons (relative to the optic axis) must satisfy: $C_s\alpha^3 < r_s$, giving $\alpha \approx 10^{-2}$ rad = 10 mrad. This low value justifies our use of small-angle approximations in the preceding analysis.

The basic physical properties of a magnetic field dictate that the spherical aberration of a rotationally symmetric electron lens *cannot* be eliminated by careful design of the lens polepieces. However, spherical aberration can be *minimized* by using a strong objective lens in the TEM. The shortest possible focal length is determined by the maximum field ($B_0 \approx 2.6$ T) obtainable, which is limited by magnetic saturation of the lens polepieces. In practice, f and C_s can be made small enough to achieve atomic resolution in suitably thin specimens. Further improvement requires the use of an aberration corrector to reduce C_s, as discussed in Sect. 2.7.

Chromatic Aberration

In light optics, chromatic aberration occurs when there is a spread in the wavelength of the light passing through a lens, combined with a variation of refractive index with wavelength (dispersion). In the case of an electron, the de Broglie wavelength depends on the particle momentum, and therefore on its kinetic energy, while Eq. (2.7) shows that the focusing power of a magnetic lens varies inversely with the kinetic energy E_0. So if electrons are present with different kinetic energies, they will be focused at different distances from a lens; at an image plane, there will be a *chromatic* disk of confusion rather than a point focus. The spread in kinetic energy can arise from several causes.

(1) Variations in the energy of the electrons emitted from the source. For example, electrons emitted by a thermionic source have a **thermal spread** of the order of kT, where T is the temperature of the emitting surface, due to the statistics of the electron-emission process.
(2) Fluctuations in the potential V_0 applied to accelerate the electrons. Although high-voltage supplies are stabilized as well as possible, there is always some drift (slow variation) and ripple (alternating component) in the accelerating voltage, and therefore in the kinetic energy eV_0.
(3) Energy loss due to **inelastic scattering** in the specimen, a process in which energy is transferred from an electron to the specimen. This scattering is a statistical process: not all electrons lose the same amount of energy, resulting in an energy spread within the transmitted beam. Because the TEM *imaging* lenses focus electrons *after* they have passed through the specimen, inelastic scattering will cause chromatic aberration in the magnified image.

We can estimate the radius of the *chromatic* disk of confusion by the use of Eq. (2.7) and thin-lens geometric optics, setting aside spherical aberration and other

lens defects. Consider an axial point source P of electrons (distance u from the lens) that is focused to a point Q in the image plane (distance v from the lens) for electrons of energy E_0 as shown in Fig. 2.14. Since $1/f$ increases as the electron energy decreases, electrons whose kinetic energy is $E_0 - \Delta E_0$ will have a smaller image distance $v - \Delta v$ and arrive at the image plane a radial distance r_i from the optic axis. If the angle β of the arriving electrons is small,

$$r_i = \Delta v \tan\beta \approx \beta\, \Delta v \tag{2.15}$$

As in the case of spherical aberration, we need to know the x-displacement of a second point object P′ whose disk of confusion partially overlaps the first, as shown in Fig. 2.14. Following again the Rayleigh criterion, we take the required displacement in the image plane to be equal to the disk radius r_i, corresponding to a displacement *in the object plane* of $r_c = r_i/M$. The image magnification M is given by:

$$M = v/u = \tan\alpha/\tan\beta \approx \alpha/\beta \tag{2.16}$$

From Eqs. (2.15) and (2.16), we have:

$$r_c \approx \beta\Delta v/M \approx \alpha\Delta v/M^2 \tag{2.17}$$

Assuming a thin lens, $1/u + 1/v = 1/f$ and taking derivatives of this equation (for a fixed object distance u) gives: $0 + (-2)\, v^{-2}\, \Delta v = (-2)\, f^{-2}\, \Delta f$, leading to:

$$\Delta v = \left(v^2/f^2\right)\Delta f \tag{2.18}$$

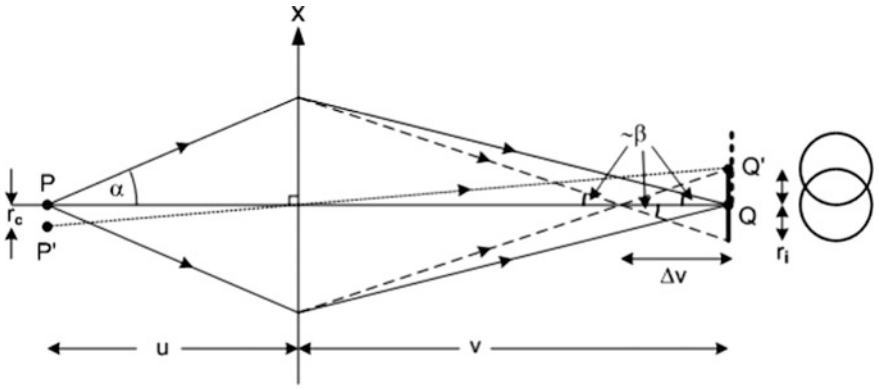


Fig. 2.14 Ray diagram illustrating the change in focus and the disk of confusion resulting from chromatic aberration. With two object points, the image disks overlap; the Rayleigh criterion (about 15 % reduction in intensity between the current-density maxima) is satisfied when the separation PP' in the object plane is given by Eq. (2.20)

For $M \gg 1$, the thin-lens equation, $1/u + 1/(Mu) = 1/f$, implies that $u \approx f$ and $v \approx Mf$, so Eq. (2.18) becomes $\Delta v \approx M^2\,\Delta f$ and Eq. (2.17) gives:

$$r_c \approx \alpha \Delta f \tag{2.19}$$

From Eq. (2.7), the focal length of the lens can be written as $f = A\,E_0$ where A is independent of electron energy, so taking derivatives gives: $\Delta f = A\,\Delta E_0 = (f/E_0)\,\Delta E_0$. The loss of spatial resolution due to chromatic aberration is therefore:

$$r_c \approx \alpha f(\Delta E_0/E_0) \tag{2.20}$$

More generally, a **coefficient of chromatic aberration** C_c is defined by the equation:

$$r_c \approx \alpha C_c(\Delta E_0/E_0) \tag{2.21}$$

Our analysis has shown that $C_c = f$ in the thin-lens approximation. A more exact (thick-lens) treatment gives C_c slightly smaller than f for a weak lens and closer to $f/2$ for a strong lens; see Fig. 2.13. As in the case of spherical aberration, chromatic aberration cannot be eliminated through lens design but is minimized by making the lens as strong as possible (large focusing power, small f), by using an angle-limiting aperture (restricting α), and by using a high electron-accelerating voltage (large E_0) as Eq. (2.21) indicates. Chromatic effects can also be minimized by reducing ΔE_0 through the use of an electron-beam monochromator, as discussed in Sect. 2.7.

Axial Astigmatism

So far, we have assumed complete rotational symmetry of the magnetic field that focuses the electrons. In practice, lens polepieces cannot be machined with perfect accuracy and the polepiece material may be slightly inhomogeneous, resulting in local variations in relative permeability. In either case, the departure from rotational symmetry will cause the magnetic field at a given radius r from the z-axis to depend on the *plane of incidence* of an incoming electron (i.e., on its **azimuthal** angle ϕ, viewed along the z-axis). According to Eq. (2.9), this difference in magnetic field will give rise to a difference in focusing power; the lens is said to suffer from **axial astigmatism**.

Figure 2.15a shows electrons leaving an *on-axis* object point P at equal angles to the z-axis but traveling in the x–z and y–z planes. They cross the optic axis at different points, F_x and F_y that are displaced along the z-axis. In Fig. 2.15a, the x-axis corresponds to the *lowest* focusing power and the perpendicular y-direction corresponds to the *highest* focusing power.

In practice, electrons leave P with *all* azimuthal angles and at all angles (up to α) relative to the z-axis. At the plane containing F_y, the electrons lie within a **caustic figure** that approximates to an ellipse whose long axis lies parallel to the x-direction. At F_x they lie within an ellipse whose long axis points in the y-direction. At some intermediate plane F, the electrons define a circular disk of confusion of radius R, rather than a single point. If that plane is used as the image (magnification M), astigmatism will limit the point resolution to a value R/M.

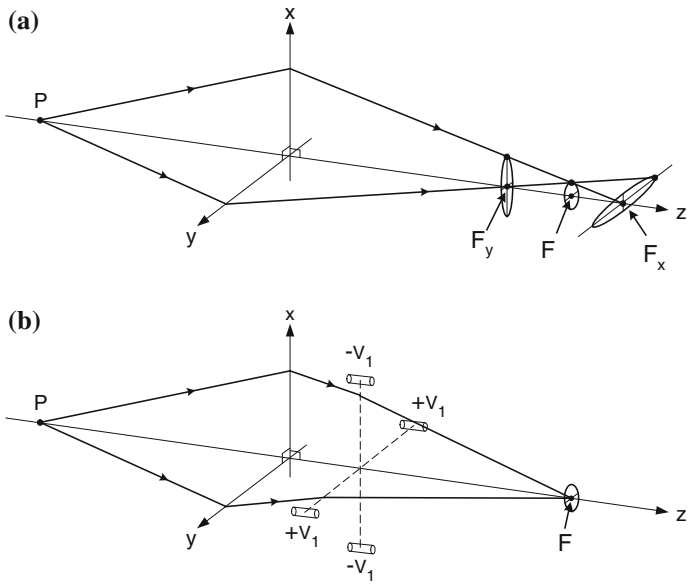


Fig. 2.15 **a** Rays leaving an axial image point, focused by a lens (with axial astigmatism) into ellipses centered on F_x and F_y or into a circle of radius R at some intermediate plane. **b** Use of an electrostatic stigmator to correct for the axial astigmatism of an electron lens

Axial astigmatism also occurs in the human eye, when there is a lack of rotational symmetry. It can be *corrected* by wearing lenses whose focal length varies with azimuthal direction by an amount just sufficient to compensate for the azimuthal variation in focusing power of the eyeball. In *electron* optics, the device that corrects for astigmatism is called a **stigmator** and it takes the form of a weak **quadrupole lens**.

An *electrostatic* quadrupole consists of four electrodes, in the form of short conducting rods aligned parallel to the z-axis and located at equal distances along the $+x$, $-x$, $+y$, and $-y$ directions; see Fig. 2.15b. A power supply generating a voltage $-V_1$ is connected to the two rods that lie in the x–z plane, so electrons traveling in that plane are repelled *towards* the axis, resulting in a *positive* focusing power (convex-lens effect). A potential $+V_1$ is applied to the other pair, which therefore *attracts* electrons traveling in the y–z plane and provides a *negative* focusing power in that plane. By adding the stigmator to an astigmatic lens and choosing V_1 appropriately, the focal lengths (of the entire system) in the x–z and y–z planes can be made equal; the two foci F_x and F_y are brought together to a single point and the axial astigmatism is eliminated, as in Fig. 2.15b.

In practice, we cannot predict which azimuthal direction will correspond to the smallest or largest focusing power of an astigmatic lens. Therefore the *direction* of the stigmator correction must be adjustable, as well as its strength. One way of achieving this is to mechanically rotate the quadrupole around the z-axis. A more

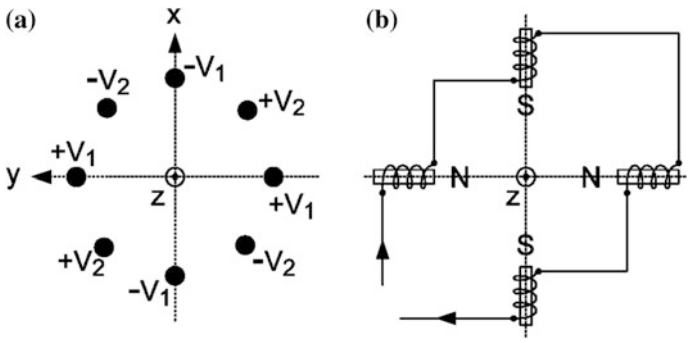


Fig. 2.16 **a** Electrostatic stigmator, viewed along the optic axis. **b** Magnetic quadrupole, which is the basis of an electromagnetic stigmator

convenient arrangement is to add four more electrodes, connected to a second power supply that generates potentials of $+V_2$ and $-V_2$, as in Fig. 2.16a. Varying the magnitude and polarity of the two voltage supplies is equivalent to varying the strength and orientation of a single quadrupole, without the need for mechanical rotation.

A more common form of stigmator consists of a *magnetic* quadrupole: four short solenoid coils with their axes pointing towards the optic axis. The coils that generate the magnetic field are connected in series and carry a common current I_1. They are wired so that similar (north or south) magnetic poles face each other, as in Fig. 2.16b. The magnetic force on an electron is perpendicular to the magnetic field, but in other respects, a magnetic stigmator acts similarly to an electrostatic one. Astigmatism correction could be achieved by adjusting the current I_1 and the azimuthal orientation of the quadrupole, but in practice a second set of four coils is inserted at 45° to the first and carries a current I_2 that can be varied independently. Independent adjustment of I_1 and I_2 enables the astigmatism to be corrected without mechanical rotation.

The stigmators found in electron-beam columns are *weak* quadrupoles, designed to correct for *small* deviations in focusing power of a much stronger lens. *Strong* quadrupoles are used in synchrotrons and nuclear-particle accelerators, to focus high-energy electrons or other charged particles. Their focusing power is positive in one plane and negative (diverging, equivalent to a concave lens) in the perpendicular plane. However, a series combination of *two* quadrupole lenses can result in an overall convergence in both planes, without image rotation and with less power dissipation than required by an axially symmetric lens. This last consideration favors the use of quadrupoles for focusing ions or in situations where the electron optics must be compact.

In light optics, the surfaces of a glass lens can be machined with sufficient accuracy that *axial* astigmatism is negligible. But for rays coming from an *off-axis* object point, the lens appears elliptical rather than round, so *off-axis* astigmatism is unavoidable.

In electron optics, this kind of astigmatism is less significant because the electrons are confined to small angles relative to the optic axis (to avoid excessive spherical and chromatic aberration). Another off-axis aberration, called *coma*, is of importance if a TEM is to achieve its highest possible resolution, and can be minimized by a suitable lens-alignment procedure.

Distortion and Curvature of Field

In an undistorted image, the distance R of an image point from the optic axis is given by $R = M\,r$, where r is the distance of the corresponding object point from the axis and the image magnification M is a constant. The presence of distortion changes this ideal relation to:

$$R = M\,r + C_{\mathrm{d}} r^3 \tag{2.22}$$

where C_{d} is a constant. If $C_{\mathrm{d}} > 0$, each image point is displaced outwards, particularly those further from the optic axis, and the entire image suffers from pincushion distortion (Fig. 2.2c). If $C_{\mathrm{d}} < 0$, each image point is displaced inward relative to the ideal image and barrel distortion is present (Fig. 2.2b).

As might be expected from the third-power dependence in Eq. (2.22), distortion is related to spherical aberration. In fact, a rotationally symmetric electron lens (for which $C_{\mathrm{s}} > 0$) will give $C_{\mathrm{d}} > 0$ and pincushion distortion. Barrel distortion is produced in a two-lens system in which the second lens magnifies a *virtual* image produced by the first lens. In a multi-lens system, it is therefore possible to combine the two types of distortion to achieve a distortion-free image.

In the case of magnetic lenses, a third type of distortion arises from the fact that the image rotation ϕ may depend on the distance r of the object point from the optic axis. This spiral distortion was illustrated in Fig. 2.2c. Again, compensation is possible in a multi-lens system.

For most purposes, distortion is a less serious lens defect than aberration, since it does not result in a loss of image detail. In fact, it may not be noticeable unless the microscope specimen contains straight-line features. In some TEMs, distortion is observable when the final (projector) lens is operated at a reduced current (giving larger C_{d}) to achieve low overall magnification.

Curvature of field is not a serious problem in the TEM or SEM, since the angular deviation of electrons from the optic axis is small. This results in a large *depth of focus* (the image remains acceptably sharp as the viewing plane is moved along the optic axis), as discussed in Chap. 3.

2.7 Aberration Correctors and Monochromators

As early as 1936, Otto Scherzer proved that any electron lens will suffer from spherical and chromatic aberration if it has rotational symmetry, produces a real image, uses static (time-independent) focusing fields, and involves no on-axis

electrostatic charge. Since then, numerous scientists have aimed to construct aberration-free systems by violating one or other of these constraints. Reflecting electrons from an electrode connected to a potential higher than the electron-source potential represents a special case of time dependence (the electron reverses its velocity vector) and such electrostatic mirrors have been used to correct the aberrations of low-energy and photoelectron microscopes (Rempfer et al. 1997). But for the TEM and SEM, success has depended on the use of multipole lenses that involve non-axial electric or magnetic fields.

Multipole Optics

The stigmator used to correct for axial astigmatism (Figs. 2.15 and 2.16) is a weak **quadrupole** lens, consisting of *four* elements that generate a symmetric pattern of electric or magnetic fields, perpendicular to the optic axis. *Strong* quadrupoles (used in pairs) are used in the spectrometer optics (Sect. 6.9) and for focusing high-energy electrons or heavier particles in particle accelerators. Electrostatic or magnetic **dipoles** contain *two* electrodes that generate a roughly uniform field perpendicular to the optic axis, and are used as beam deflectors (without focusing) for all kinds of charged particles. **Sextupoles** (also known as hexapoles) and **octupoles** are employed for correcting lens aberrations.

Figure 2.17a shows the general construction of a magnetic octupole lens, where iron-cored solenoid coils generate magnetic fields whose north and south poles alternate. Figure 2.17b shows the direction of the magnetic force on an electron along different radial directions. Along the dashed diagonal lines, this force is towards the optic axis (compressive) and is proportional to r^3, as needed to cancel the effect of spherical aberration (r represents radial distance from the optic axis).

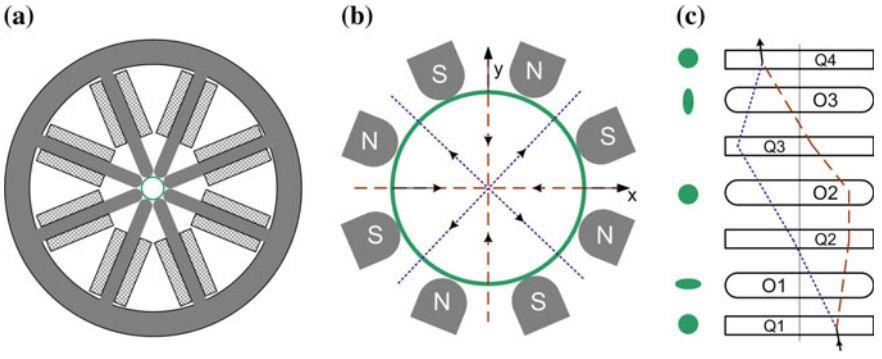


Fig. 2.17 **a** Construction of an octupole lens. Ferromagnetic material is shown in gray, solenoid-coil windings are cross-hatched. **b** Enlargement of the central region, showing the polarities of the magnetic poles and the direction of the magnetic force on an electron traveling in the vacuum contained within the circular beam tube. **c** Quadrupole-octupole spherical-aberration corrector, showing cross-sectional profiles of the beam at various planes and the path taken by the electrons traveling in two perpendicular (x–z and y–z) planes

Along the dotted lines the force is away from the axis and the aberration is increased.

In the octupole-quadrupole corrector (Fig. 2.17c) a magnetic quadrupole Q1 is used to stretch the beam into a thin ellipse oriented along one of the compressive directions (dashed ray) so that an octupole O1 can cancel spherical aberration along that direction, with little effect in other directions. A second quadrupole Q2 restores the beam into a circular shape, allowing a second octupole O2 to correct for fourfold astigmatism. A third quadrupole Q3 then stretches the beam into an ellipse elongated in a perpendicular direction, so that an appropriately oriented octupole O3 can correct for spherical aberration in that direction. A final quadrupole Q4 restores the beam to a circular cross-section, with spherical aberration corrected. This form of corrector is used in STEM instruments manufactured by the Nion Company.

Another form of spherical-aberration corrector uses magnetic sextupoles. A *short* sextupole would give radial forces proportional to r^2 (and therefore the same at coordinates +r and -r) but two *long* sextupoles (oriented 60° apart) provide an appropriate form of correction. This design is manufactured by the CEOS and FEI companies and has been fitted to many commercial TEMs, while JEOL and Hitachi have developed their own sextupole designs. CEOS also manufactures a device containing magnetic and electrostatic multi-poles that can correct for chromatic as well as spherical aberration, a procedure that becomes necessary for obtaining atomic resolution at accelerating voltages below 50 kV.

As an alternative to reducing C_c, chromatic aberration can be controlled by reducing the energy spread ΔE of the electron beam, using an electron-beam **monochromator**, which is an electron spectrometer (see Sect. 6.9) followed by an energy-selecting slit. In the **Wien filter**, a magnetic field B_y and electrostatic field E_x are applied, perpendicular to each other and to the optic axis, such that the net force F on an electron is zero at some precise electron speed v given by:

$$F = (-\mathrm{e})E_x - (-\mathrm{e})B_y v = 0 \tag{2.23}$$

These electrons travel in a straight line and pass through a narrow slit; those moving at other speeds are deflected sideways and are absorbed by the slit blades. The Wien filter is usually placed just below the electron gun (where the electron speed v is small, allowing higher sensitivity) and is known as a gun monochromator; it can reduce ΔE to about 0.1 eV. Other spectrometer designs have been used for a monochromator, giving energy widths down to 0.01 eV (see Sect. 6.9).

2.8 Further Reading

As stated in Section 2.2, many aspects of electron optics are similar to those of geometrical optics, which is treated in many undergraduate textbooks. There is not much information available in the literature on the practical aspects of electron optics;

nowadays such expertise resides with commercial microscope manufacturers and is likely to be kept confidential. However, there are several books that deal with the electron-optic principles. For example:

Principles of Electron Optics by P.W. Hawkes and E. Kasper (Springer, 1996; ISBN: 978-0-12-333340-7) is a three-volume treatise dealing with electron-optical elements (including magnetic and electrostatic lenses, mirrors, and beam deflection systems) as well as electron scattering and image formation and interpretation, with an emphasis on underlying theory.

Handbook of Charged Particle Optics, ed. J. Orloff (CRC Press, 2nd edition, 2008: ISBN-13: 978-1420045543) contains chapters written by SEM, STEM, and FIB experts and includes aberration correction, which is also treated in *Aberration-Corrected Imaging in Transmission Electron Microsc*opy by R. Erni (Imperial College Press, 2nd edition, 2015; ISBN-13: 978-1783265282).

An easy-to-read account of aberration correction using an electrostatic mirror is given by G.F. Rempfer, D.N. Delage, W.P. Skoczlas and O.H. Griffith in Microscopy and Microanalysis 3 (1997) 14-27: *Simultaneous correction of spherical and chromatic aberrations with an electron mirror: an optical achromat*.

Figure 2.13 is based on information given in *Foundations of Electron Optics* written (in German) by W. Glaser (Springer-Verlag, Vienna, 1952). For further insight into the operation of a magnetic lens, see *Transmission Electron Microscopy* by L. Reimer and H. Kohl (5th edition, Springer-Verlag, 2008).

Maxwell's rules for image formation are contained in the article: *On the general laws of optical instruments,* published in Quarterly Journal of Pure and Applied Mathematics 2 (1858) 233-246. Fermat's least-time principle provides an interesting alternative view of image formation by a glass lens, as discussed at http://tem-eels.ca/education/education_index.html.

Chapter 3
The Transmission Electron Microscope

As we saw in Chap. 1, the TEM is capable of recording magnified images of a thin specimen, typically with a magnification in the range 10^3–10^6. In addition, the instrument can be used to produce electron-diffraction patterns, useful for analyzing the properties of crystalline specimens. This overall flexibility is achieved with an electron-optical system containing an **electron gun** (which produces the beam of electrons) and several magnetic lenses, stacked vertically to form a lens column. It is convenient to divide the instrument into three sections, which we will first define and then discuss in some detail separately.

The **illumination system** comprises the electron gun, together with two or more condenser lenses that focus the electrons onto the specimen. Its design and operation determine the diameter of the electron beam (often called the illumination) at the specimen and the intensity level in the final TEM image.

The **specimen stage** allows specimens to be held stationary, or else intentionally moved, and to be inserted or withdrawn from the TEM. Its mechanical stability is an important factor in determining the spatial resolution of the TEM image.

The **imaging system** contains at least three lenses that together produce a magnified image (or a diffraction pattern) of the specimen on a fluorescent screen, for immediate viewing or for recording by an electronic camera system. How the imaging lenses are operated determines the magnification of the TEM image, while their design specifications largely determine the spatial resolution that can be obtained from the microscope.

In the discussion, we will assume that our electron lenses suffer from spherical and chromatic aberration. If present, an aberration corrector eliminates only spherical aberration of the objective-lens pre- or post-field, and only the lower-order components. The remaining higher-order aberrations still limit the divergence angle of a beam, but to an angle typically 3–5 times larger than without correction applied.

R.F. Egerton, *Physical Principles of Electron Microscopy*,
DOI 10.1007/978-3-319-39877-8_3

3.1 The Electron Gun

The electron gun produces a beam of electrons whose kinetic energy is high enough to allow them to pass through thin areas of the TEM specimen. The gun consists of an electron source, also known as the **cathode** since it is held at a high negative potential, and an electron-accelerating region. There are several types of electron sources, each operating on a different physical principle, as we now discuss.

3.1.1 Thermionic Emission

Figure 3.1 shows a thermionic electron gun. The electron source is a V-shaped ("hairpin") filament made of tungsten wire, spot-welded to straight-wire leads mounted in a ceramic or glass socket, allowing the filament assembly to be easily exchanged when the filament eventually burns out. A direct (dc) current heats the filament to about 2700 K, at which temperature tungsten emits electrons into the surrounding vacuum by the process known as **thermionic emission**.

The process of thermionic emission is illustrated by an electron-energy diagram (Fig. 3.2), in which the vertical axis represents the energy E of an electron and the horizontal axis represents the distance z from the tungsten surface. Within the tungsten, the electrons of highest energy lie at the top of the **conduction band**, close to the **Fermi energy** E_F. These conduction electrons carry the electrical current within a metal and they normally cannot escape from the surface because E_f is an amount ϕ (the **work function**) below the **vacuum level**, which represents the energy of a stationary electron located some distance outside the surface. As shown in Fig. 3.2, the electron energy does not change *abruptly* at the metal/vacuum interface; when an electron leaves the metal, it generates lines of electric field that

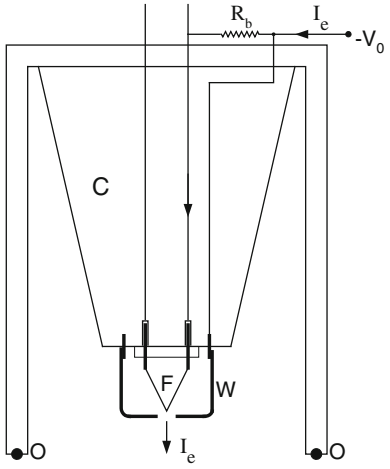


Fig. 3.1 Thermionic electron gun containing a tungsten filament F, Wehnelt electrode W, ceramic high-voltage insulator C and an o-ring seal O to the lower part of the TEM column An autobias resistor R_b (actually located inside the high-voltage generator, as in Fig. 3.6) is used to generate a potential difference between W and F, thereby controlling the electron-emission current I_e. *Arrows* denote the direction of *electron* flow that gives rise to the emission current

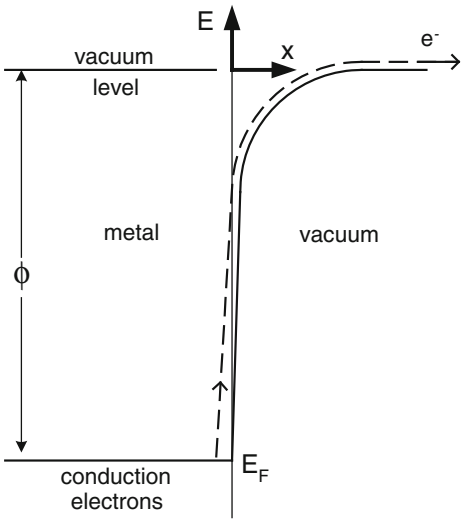


Fig. 3.2 Electron energy-band diagram of a metal, for the case where no electric field is applied to its surface. The process of thermionic emission of an electron is indicated by the *dashed line*

terminate on positive charge (reduced electron density) at the metal surface (see Appendix A.1). This charge provides an electrostatic force towards the surface that weakens only *gradually* with distance. Therefore the electric field and its associated potential (and the potential energy of the electron) also fall off gradually outside the surface.

Raising the temperature of the cathode causes the nuclei of its atoms to vibrate with increased amplitude. Because the conduction electrons are in thermodynamic equilibrium with the atoms, they share this thermal energy and a small proportion of them achieve energies *above* the vacuum level, enabling them to escape across the metal/vacuum interface.

The rate of electron emission can be represented as a current density J_e (in A/m^2) at the cathode surface and is given by the **Richardson law**:

$$J_e = A\,T^2\,\exp(-\phi/kT) \tag{3.1}$$

In Eq. (3.1), T is the absolute temperature (in K) of the cathode and A is the **Richardson constant** ($\approx 10^6$ Am^{-2}K^{-2}), which depends to some degree on the cathode material but not on its temperature; k is the Boltzmann constant (1.38×10^{-23} J/K) and kT is *approximately* the mean thermal energy of an atom (or of a conduction electron, if measured relative to the Fermi level). The work function ϕ is conveniently expressed in **electron volts** (eV) of energy and must be converted to Joules by multiplying by $e = 1.6 \times 10^{-19}$ for use in Eq. (3.1). Despite the T^2 factor, the *main* temperature dependence in this equation comes from the exponential function. As T is increased, J_e remains very low until kT approaches a few percent of the work function. The temperature is highest at the *tip* of the V-shaped filament (furthest from the leads, which act as heat sinks) so most of the emission occurs in the immediate vicinity of the tip.

Tungsten has a high cohesive energy and therefore a high melting point (≈3650 K) and a low vapor pressure, allowing it to be maintained at a temperature of 2500–3000 K in vacuum. Despite its rather high work function ($\phi = 4.5\,\text{eV}$), ϕ/kT can be made sufficiently *low* to provide adequate electron emission. Tungsten is an electrically conducting metal that can be heated by passing a current through it. It is chemically stable at high temperatures and does not combine with the residual gases that are sometimes present in the relatively poor vacuum (pressure $>10^{-3}$ Pa) in a thermionic electron gun. Chemical reaction would lead to contamination ("poisoning") of the emission surface, causing a change in the work function and the emission current.

An alternative strategy is to employ a material with a *low* work function, which does not need to be heated to such a high temperature. The preferred material is lanthanum hexaboride (LaB_6; ϕ = 2.7 eV), fabricated in the form of a short rod (about 2 mm long and less than 1 mm in diameter) sharpened to a tip, from which the electrons are emitted. The LaB_6 crystal is heated to 1400–2000 K by mounting it between wires or onto a carbon strip through which a current is passed. The conducting leads are mounted on pins set into an insulating base whose geometry is identical with that used for a tungsten-filament source, therefore the two types of electron sources are mechanically interchangeable. Unfortunately, lanthanum hexaboride becomes poisoned if it combines with traces of oxygen, so a better vacuum (pressure $<10^{-4}$ Pa) is needed in the electron gun.

Compared to a tungsten filament, the LaB_6 source is relatively expensive but lasts longer, provided it is brought to and from its operating temperature *slowly* to avoid thermal shock and resulting mechanical fracture. It provides comparable emission current from a smaller cathode area, enabling the electron beam to be focused onto a smaller area of the specimen. The resulting higher current density provides a brighter image at the viewing screen or camera of the TEM, which is particularly important at high image magnification.

Another important component of a thermionic electron gun (Fig. 3.1) is the **Wehnelt** cylinder, a metal electrode that can be easily removed (to allow changing the filament or LaB_6 source) but which normally surrounds the filament completely except for a small (<1 mm diameter) hole through which the electron beam emerges. The purpose of the Wehnelt electrode is to control the emission current of the electron gun, and for this purpose, its potential is made *more negative* than that of the cathode. This negative potential prevents electrons from leaving the cathode *unless* they are emitted from a region very near its tip, which is located immediately above the hole in the Wehnelt where the electrostatic potential is less negative. Increasing the magnitude of the negative bias reduces both the emitting area and the emission current I_e.

Although the Wehnelt bias could be provided by a voltage power supply, it is usually achieved through the autobias (or self-bias) arrangement shown in Figs. 3.1 and 3.6. A bias resistor R_b is inserted between the filament and the negative high-voltage supply ($-V_0$) that is used to accelerate the electrons. Since the electron current I_e emitted from the filament F must pass through the bias resistor, a potential difference (I_eR_b) is developed across it, making the filament less negative than the

Wehnelt. Changing the value of R_b provides a convenient way of intentionally varying the Wehnelt bias and therefore the emission current I_e. A further advantage of this autobias arrangement is that, if the emission current I_e starts to increase spontaneously (for example due to upward drift in the filament temperature T) the Wehnelt bias becomes more negative, canceling out most of the increase in electron emission. This is another example of the *negative-feedback* concept discussed in Sect. 1.7.

The dependence of the electron-beam current on the filament heating current is shown in Fig. 3.3. As the heating current is increased from zero, the filament temperature eventually becomes high enough to give a small emission current. At this point, the Wehnelt bias ($-I_eR_b$) is *in*sufficient to control the emitting area according to the feedback mechanism just described, so I_e increases *rapidly* with filament temperature T, as expected from Eq. (3.1). As the filament temperature is further increased, the negative-feedback mechanism starts to limit the beam current, which then becomes approximately *independent* of filament temperature and is said to be **saturated**. The filament heating current (which is adjustable by the TEM operator) should never be set higher than the value required for current saturation. Higher values give very little increase in beam current and would result in a *decrease* in source lifetime, due to evaporation of W or LaB_6 from the cathode. The change in I_e shown in Fig. 3.1 can be monitored from an emission-current meter or by observing the brightness of the TEM screen, allowing the filament current to be set correctly. If the beam current needs to be changed, this is done using a *bias-control* knob that selects a different value of R_b, as indicated in Fig. 3.3.

3.1.2 Schottky Emission

Thermionic emission can be increased by applying an electrostatic field to the cathode surface. This field lowers the height of the potential barrier (that keeps electrons inside the cathode) by an amount $\Delta\phi$ (see Fig. 3.4), the so-called

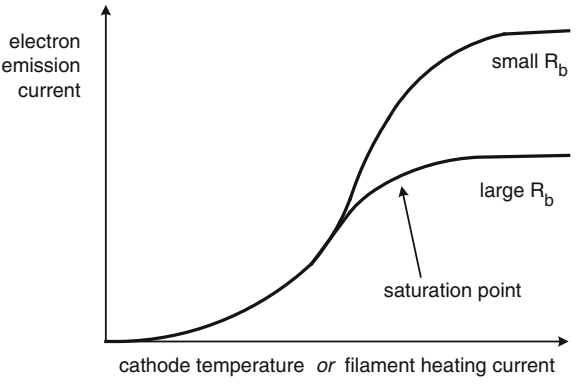


Fig. 3.3 Emission current I_e as a function of cathode temperature, for thermionic emission from an autobiased (tungsten or LaB_6) cathode

Schottky effect. As a result, the emission-current density J_e is increased by a factor $\exp(\Delta\phi/kT)$, typically a factor of 10 as demonstrated in Appendix A.1.

A Schottky source consists of a pointed crystal of tungsten welded to the end of V-shaped tungsten filament. The tip is coated with zirconium oxide (ZrO) to provide a low work function (≈2.8 eV) and needs to be heated to only about 1800 K to provide adequate electron emission. The tip protrudes about 0.3 mm *outside* the hole in the Wehnelt, so an *accelerating* field exists at its surface, created by an extractor electrode biased *positive* with respect to the tip. Because the tip is very sharp, electrons are emitted from a very small area, resulting in a relatively high current density ($J_e \approx 10^7$ A/m^2) at the surface. ZrO is easily poisoned by ambient gases, so the Schottky source requires a vacuum substantially better than that of a W or LaB_6 source.

3.1.3 Field Emission

If the electrostatic field at a tip of a cathode is increased sufficiently, the *width* (horizontal in Fig. 3.4) of the potential barrier becomes small enough to allow electrons to escape *through* the surface potential barrier by quantum-mechanical tunneling, a process known as **field emission**. We can estimate the required electric field as follows. The probability of electron tunneling becomes *high* when the barrier width w is comparable to de Broglie wavelength λ of the electron. This wavelength is related to the electron momentum p by $p = h/\lambda$ where $h = 6.63 \times 10^{-34}$ Js is the Planck constant. Because the barrier width is smallest for electrons at the top of the conduction band (see Fig. 3.4), they are the ones most likely to escape. These electrons (at the Fermi level of the cathode) have a speed v of the order of 10^6 m/s and a wavelength $\lambda = h/p = h/mv \approx 0.5$ nm.

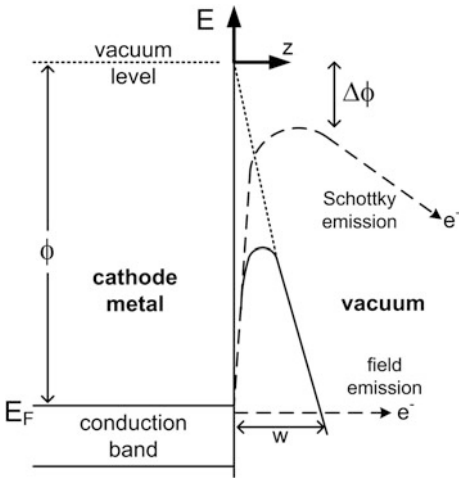


Fig. 3.4 Electron-energy diagram of a cathode, with both moderate (≈10^8 V/m) and high (≈10^9 V/m) electric fields applied to its surface; the corresponding Schottky and field emission of electrons are shown by *dashed lines*. The *upward vertical axis* represents the potential energy E of an electron (in eV) relative to the vacuum level, therefore the downward direction represents electrostatic potential (in V)

As seen from the right-angled triangle in Fig. 3.4, the electric field E_s that gives a barrier width (at the Fermi level) of w is $E_s = (\phi/e)/w$. Taking $w = \lambda$ (which allows high tunneling probability) and $\phi = 4.5\,\text{eV}$ for a tungsten tip, so that $(\phi/e) = 4.5$ V, gives $E_s = 4.5/(0.5 \times 10^{-9}) \approx 10^{10}$ V/m. In fact, such a huge value is not necessary for field emission. Due to their high speed v, electrons arrive at the cathode surface at a fast rate and adequate electron emission can be obtained with a tunneling probability of the order of 10^{-2}, requiring a surface field of the order of 10^9 V/m.

This high electric field is obtained by replacing the Wehnelt cylinder (used in thermionic emission) by an extractor electrode maintained at a *positive* potential $+V_1$ *relative to* the tip. If we approximate the tip as a sphere whose radius is r, much less than the distance to the extractor electrode, we can use the electrostatic formula for an isolated sphere: $E_s = KQ/r^2$ to relate the surface electric field E_s to the charge Q on the tip, $K = 1/(4\pi\varepsilon_0)$ being the Coulomb constant. This electric field is also the *potential gradient* just outside the tip: $E_s = -\mathrm{d}V/\mathrm{d}r$, so *by integration* we obtain the potential of the tip (relative to the extractor electrode) as $V_1 = KQ/r = E_s\, r$. Making the radius of curvature r of the tip very small enables the required local field $E_s = V_1/r$ to be produced with practical values of V_1 (a few thousand volts).

The tip is made sufficiently sharp by electrolytically dissolving the end of a short piece of tungsten wire. Electrons are emitted from an extremely small area ($\approx r^2$), resulting in an effective source diameter below 10 nm, which would allow the electrons to be focused by a *single* demagnifying lens into an "electron probe" of sub-nanometer dimensions.

Since thermal excitation is not required, a field-emission tip can operate at room temperature, hence the name *cold* field emission gun (CFEG). As there is no evaporation of tungsten during normal operation, the tip can last for many months or even years before replacement. It is heated ("flashed") from time to time to remove adsorbed gases, which affect the work function and cause the emission current to become unstable. Even so, cold field emission needs *ultra-high* vacuum (UHV: pressure $\approx 10^{-8}$ Pa) to achieve stable operation, which requires a more elaborate vacuum system and increases the cost of the instrument.

3.1.4 Comparison of Electron Sources; Source Brightness

Operating conditions and performance criteria for the four types of electron sources are compared in Table 3.1. The numbers in this table are approximate but can be taken as typical for the normal operation of a TEM or an SEM.

Although the available emission current I_e decreases as we go from a tungsten-filament to a field-emission source, the effective source diameter d_s and the emitting area ($A_s = \pi d_s^2/4$) decrease by a larger factor, resulting in an *increase* in the current density J_e. More importantly, there is an increase in a quantity known as the electron-optical **brightness** B_s of the source, defined as the current density divided by the *solid* angle Ω over which electrons are emitted:

Table 3.1 Operating parameters of four types of electron sources[a]

Type of source	Tungsten thermionic	LaB_6 thermionic	Schottky emission	Cold field emission
Material:	W	LaB_6	ZrO/W	W
φ (eV)	4.5	2.7	2.8	4.5
T (K)	2700	1800	1800	300
E (V/m)	low	low	$\approx 10^8$	$>10^9$
J_e (A/m^2)	$\approx 10^4$	$\approx 10^6$	$\approx 10^7$	$\approx 10^9$
B_s (A/m^2/sr)	$\approx 10^9$	$\approx 10^{10}$	$\approx 10^{11}$	$\approx 10^{12}$
d_s (μm)	≈40	≈10	≈0.02	≈0.01
Vacuum (Pa)	$<10^{-2}$	$<10^{-4}$	$<10^{-7}$	$\approx 10^{-8}$
Lifetime (h)	≈100	≈1000	$\approx 10^4$	$\approx 10^4$
ΔE (eV)	1.5	1.0	0.5	0.3

[a]φ is the work function, T the temperature, E the electric field, J_e the current density, and β the electron-optical brightness at the cathode; d_s is the effective (or virtual) source diameter and ΔE is the energy spread of the emitted electrons

$$B_s = I_e/(A_s\Omega) = J_e/\Omega \tag{3.2}$$

Solid angle is the three-dimensional equivalent of angle in two-dimensional (Euclidean) geometry. Whereas an angle in radians is defined by $\theta = s/r$, where s is the arc length and r is the arc radius (see Fig. 3.5a), *solid* angle is measured in **steradians** and defined as $\Omega = A/r^2$ where A is the *area* of a section of a sphere at distance r (see Fig. 3.5b). Dividing by r^2 makes Ω a dimensionless number, just like θ .

If electrons were emitted in *every* possible direction from a point source, the solid angle of emission would be $\Omega = A/r^2 = (4\pi r^2)/r^2 = 4\pi$ steradian. In practice, Ω is small and is related in a simple way to the half-angle α of the emission cone. For small α, the area of the curved end of the cone in Fig. 3.5c approximates to that of a flat disk of radius s, giving:

$$\Omega \approx \left(\pi s^2\right)/r^2 = \pi\alpha^2 \tag{3.3}$$

Since image broadening due to spherical aberration of an electron lens increases as α^3, the ability of a demagnifying lens to create a high current density within a small-diameter probe is helped by keeping α small, implying an electron source of high brightness. In the *scanning* electron microscope, for example, we can achieve better spatial resolution by using a Schottky or field-emission source, whose brightness is considerably larger than that of a thermionic source (see Table 3.1).

Equation (3.3) can also be used to define an electron-optical brightness B at *any* plane in an electron-optical system where the beam diameter is d and the convergence semi-angle is α. This concept is particularly useful because β retains the same

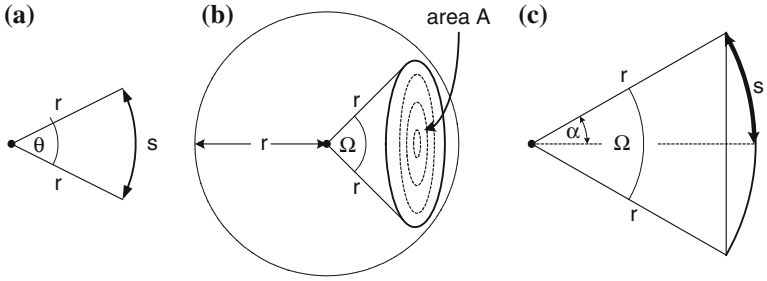


Fig. 3.5 **a** Two-dimensional diagram defining angle θ in radians. **b** Three-dimensional diagram defining solid angle Ω in steradians. **c** Cross-section through (**b**) showing the relationship between solid angle Ω and the corresponding half-angle α of the cone

value at each image plane, known as the principle of **brightness conservation**. In other words,

$$B = J_e/\Omega = I\left[\pi(d/2)^2\right]^{-1}\left(\pi\,\alpha^2\right)^{-1} = B_s = \text{constant} \tag{3.4}$$

Brightness conservation remains valid when the system contains an aperture whose diaphragm absorbs some of the electrons. Although the beam current I is reduced at the aperture, the solid angle of the beam is reduced in the same proportion. If there are *no* apertures, I remains constant and Eq. (3.4) implies that the product αd is the same at each plane.

The last row of Table 3.1 contains typical values of the energy spread ΔE for the different types of electron guns: the variation in kinetic energy of the emitted electrons. In the case of thermionic and Schottky sources, this energy spread is a reflection of the statistical variations in thermal energy of electrons within the cathode, which depends on the cathode temperature T. In the case of a field-emission source, energy spread occurs because some electrons are emitted from energy levels (within the tip) that are below the Fermi level. In both cases, ΔE increases with emission current I_e (known as the Boersch effect) because of the electrostatic interaction between electrons, particularly at "crossovers" where the beam has a small diameter and the electron separation becomes small. Larger ΔE leads to increased chromatic aberration (Sect. 2.4.2) and a loss of image resolution in both the TEM and SEM.

3.2 Electron Acceleration

After emission from the cathode, electrons are accelerated to their final kinetic energy E_0 under the influence of an electric field parallel to the optic axis. This field is generated by applying a potential difference V_0 between the cathode and an

anode, a round metal plate containing a central hole vertically below the cathode, through which the beam of *accelerated* electrons emerges. Many of the accelerated electrons are absorbed in the anode plate and only around 1 % passes through the hole, so the beam current in a TEM may be only 1 % of the emission current from the cathode.

To produce electron acceleration, the anode needs to be positive *relative to* the cathode. This situation is most conveniently arranged by having the anode (and the rest of the microscope column) at *ground* potential and the electron source at a high negative potential ($-V_0$). The cathode and its control electrode are attached to the bottom of a high-voltage insulator (Fig. 3.1), which is made of a ceramic (metal-oxide) material. To deter electrical breakdown, the gun insulator has a smooth surface and is long enough to withstand the applied voltage, which may be 100 kV or more.

Since the thermal energy kT is small ($\ll$1 eV), we can take the kinetic energy (KE) of an electron to be *zero* before acceleration and its potential energy (PE) as the product of its electrostatic charge ($-e$) and the source potential ($-V_0$). After acceleration, the KE of the electron is E_0 and its PE is zero. The total energy (KE + PE) is conserved, so that:

$$0 + (-e)\,(-V_0) = E_0 + (-e)\,(0) \tag{3.5}$$

The final kinetic energy of the electron (in J) is therefore $E_0 = (e)V_0$ but expressed in electron volts, it is $(e)V_0/e = V_0$. In other words, the electron energy in eV is *equal* to the magnitude of the accelerating voltage. A similar statement *would* apply to the acceleration of protons (charge = $+e$) but *not* to alpha particles (charge = $+2e$) or any ion whose charge differs from $\pm e$.

According to classical physics (Newtonian mechanics), we could calculate the speed v of an accelerated electron by equating E_0 to $mv^2/2$. However, this simple expression becomes inaccurate for any object whose speed is a significant fraction of the speed of light in vacuum (c). In that case, we can make use of Einstein's Special Theory of Relativity, according to which the energy E of a material object is re-defined so as to include a rest-energy component: m_0c^2, where m_0 is the familiar *rest mass* used in classical physics. If defined in this way, $E = mc^2$, where $m = \gamma m_0$ is the *relativistic* mass and $\gamma = 1\Big/(1 - v^2/c^2)^{1/2}$ is a relativistic factor that represents the increase in mass with increasing speed. In other words,

$$E = mc^2 = (\gamma\, m_0)c^2 = E_0 + m_0c^2 \tag{3.6}$$

and the *general* formula for the kinetic energy E_0 becomes:

$$E_0 = (\gamma - 1)\, m_0\, c^2 \tag{3.7}$$

Applying Eq. (3.7) to an accelerated electron, we find that its speed v reaches a significant fraction of c for the acceleration potentials V_0 commonly used in a TEM,

Table 3.2 Speed v and fractional increase in mass (γ) of an electron (rest mass m_0) for five values of the accelerating potential V_0

V_0 (kV) or E_0(keV)	γ	v/c	$m_0v^2/2$ (keV)
60	1.12	0.45	51
100	1.20	0.55	77
200	1.39	0.70	124
300	1.59	0.78	154
1000	2.96	0.94	226

The last column illustrates how the classical expression for kinetic energy fails at high particle energy

as shown in the third column of Table 3.2. Comparison of the first and last columns of this table indicates that use of the classical expression $m_0v^2/2$ would result in a substantial underestimation of the kinetic energy of the electrons, especially at higher accelerating voltages.

The accelerating voltage ($-V_0$) is supplied by an electronic high-voltage (HV) generator, connected to the electron gun by a thick, well-insulated cable. The high potential is derived from the secondary winding of a step-up transformer whose primary is connected to an electronic oscillator circuit that provides an (ac) output proportional to an applied (dc) input voltage; see Fig. 3.6. Since the oscillator operates at low voltage, the large potential difference V_0 appears between the primary and secondary windings of the transformer. Consequently, the secondary winding must be separated from the transformer core by insulating material of sufficient thickness; it cannot be tightly wrapped around an electrically conducting soft-iron core, as in many low-voltage transformers. Because of this less-efficient magnetic coupling, the transformer operates at a frequency well above mains frequency (60 or 50 Hz). Accordingly, the oscillator output is a high-frequency (≈10 kHz) sine wave, whose amplitude is proportional to the input voltage V_i (Fig. 3.6).

To provide direct (rather than alternating) high voltage, the current from the transformer secondary winding is **rectified** by means of a series of solid-state diodes, which allow electrical current to pass only in one direction, and **smoothed** to remove its alternating component (ripple). Smoothing is achieved largely by the HV cable, which has enough capacitance between its inner conductor and its grounded outer sheath to short-circuit the ac-ripple component to ground at the transformer operating frequency.

Because the output of the oscillator circuit is proportional to its input V_i and rectification is also a linear process, the magnitude of the accelerating voltage V_0 is given by:

$$V_0 = G\,V_i \tag{3.8}$$

where G is a large amplification factor (or gain) that depends on the design of the oscillator and step-up transformer. Changing V_i (by altering the reference voltage $V+$ in Fig. 3.6) allows V_0 to be intentionally changed, for example, from 100 to 200 kV. However, any slow change in G caused by drift in the oscillator or diode

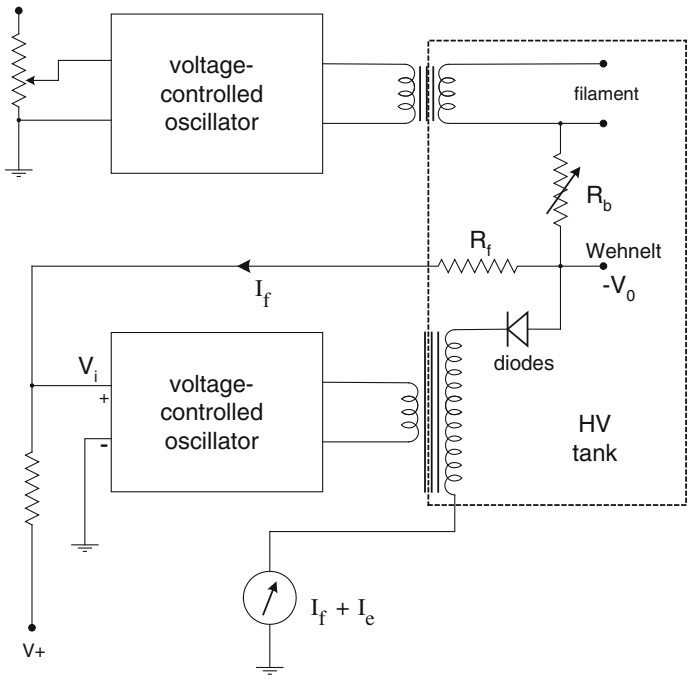


Fig. 3.6 Schematic diagram of a TEM high-voltage generator. The ground connection of the high-voltage transformer contains a meter that reads the sum of the emission and feedback currents. The *arrow* indicates the direction of electron flow (opposite to conventional current) from the feedback resistor R_f. Components within the *dashed rectangle* operate at high potential

circuitry, or any change in the emission current I_e for example, would cause V_0 to drift from its original value. Such HV instability would lead to unwanted changes in focusing due to chromatic aberration. To stabilize the high voltage, a **feedback resistor** R_f is connected between the HV output and the oscillator input, as in Fig. 3.6. If G were to increase slightly, V_0 would increase proportionally but the resulting increase in the feedback current I_f would drive the input of the oscillator more negative, canceling the effect of the increase in G. In this way, the high *voltage* is stabilized by negative feedback, in a similar way to stabilization of the emission *current* by the bias resistor.

To provide adequate insulation of the high-voltage components in the HV generator, they are immersed in transformer oil (also used in HV power transformers) or in a gas such as sulfur hexafluoride (SF_6) at a few atmospheres pressure. The high-voltage "tank" also contains the bias resistor R_b and a transformer to supply the heating current for a thermionic or Schottky source. Since the source operates at high voltage, this second transformer must also have good insulation between its primary and secondary windings. Its primary is driven by a second voltage-controlled oscillator, whose input is controlled by a potentiometer that the TEM operator adjusts to change the filament temperature; see Fig. 3.6.

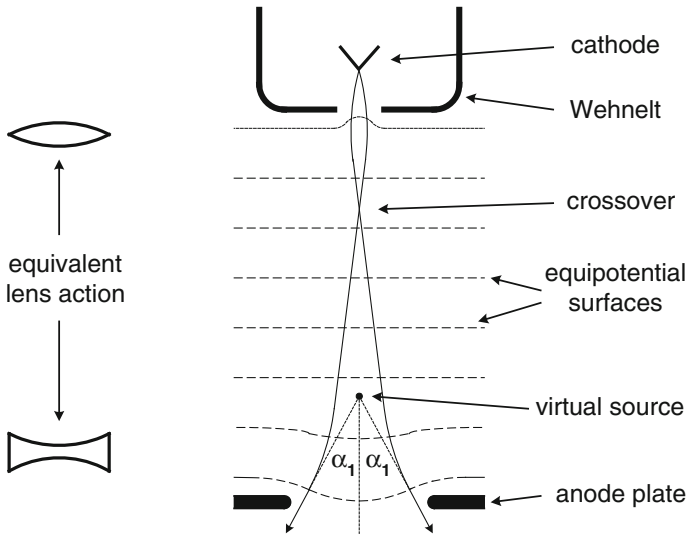


Fig. 3.7 Lens action within the accelerating field of an electron gun, between the electron source and the anode. Curvature of the equipotential surfaces around the hole in the Wehnelt electrode constitutes a converging electrostatic lens (equivalent to a convex lens in light optics) whereas the non-uniform field just above the aperture in the anode creates a diverging lens (the equivalent of a concave lens in light optics)

Although the electric field that accelerates the electrons is primarily along the optic axis, the electric-field lines *curve* in the vicinity of the hole in the Wehnelt control electrode; see Fig. 3.7. This curvature arises because the electron source (just above the hole) is less negative than the electrode, by an amount equal to the Wehnelt bias. Field curvature results in an *electrostatic* lens effect that is equivalent to a weak *convex* lens, bringing the electrons to a focus (crossover) just below the Wehnelt. Similarly, electric field lines curve above the hole in the anode plate, giving the equivalent of a *concave* lens and a diverging effect on the electron beam. As a result, electrons entering the lens column *appear* to come from a **virtual source**, whose diameter d_s is typically 40 μm and divergence semi-angle α_1 (relative to the optic axis) about 1 mrad (0.06°) in the case of a thermionic source.

3.3 Condenser-Lens System

The TEM must be capable of producing a high-magnification (e.g., $M = 10^5$) image of a specimen on a fluorescent screen, whose diameter may be as large as 15 cm. To ensure that the screen image is not too dim, most of the electrons that pass through the specimen should fall within this diameter, which is equivalent to a diameter of (15 cm)/M = 1.5 μm at the specimen. But for viewing larger areas of specimen, the

final-image magnification might need to be as low as 2000, requiring an illumination diameter of 75 μm at the specimen. In order to achieve the required flexibility, the condenser-lens system must contain at least two electron lenses.

The first condenser (C1) lens is a *strong* magnetic lens, with a focal length f of typically 2 mm. Using the *virtual* electron source (diameter d_s) as its object, C1 produces a *real* image of diameter d_1. Because the lens is located 20 cm or more below the object, its object distance is $u \approx 20\ \text{cm} \gg f$ and the image distance is $v \approx f$ according to Eq. (2.2). The magnification factor of C1 is therefore $M = v/u \approx f/u \approx (2\ \text{mm})/(200\ \text{mm}) = 1/100$, corresponding to *de*magnification by a factor of a hundred. For a W-filament electron source, $d_s \approx 40\ \mu\text{m}$, giving $d_1 \approx M\ d_s = (0.01)(40\ \mu\text{m}) = 0.4\ \mu\text{m}$. In practice, the C1 lens current can be adjusted to give *several* different values of d_1, using a control that is often labeled "spot size".

The second condenser (C2) lens is a *weak* magnetic lens ($f \approx$ several cm) that provides little or no magnification ($M \approx 1$) but allows the diameter of illumination (d) at the specimen to be varied continuously over a wide range. The C2 lens also contains the **condenser aperture** (hole in the condenser diaphragm) whose diameter D can be changed to control the **convergence semi-angle** α of the illumination, the *maximum* angle that electrons (arriving at the specimen) deviate from the optic axis.

The case of **fully focused illumination** is shown in Fig. 3.8a. An image of the electron source is formed *at* the specimen plane (image distance v_0) and the illumination diameter at that plane is therefore $d_0 = M\ d_1$ ($\approx d_1$ if object distance $u \approx v_0$). This condition provides the *smallest* illumination diameter (below 1 μm), as required for high-magnification imaging. Because the condenser aperture is located close to the principal plane of the lens, the illumination convergence angle is given by $2\alpha_0 \approx D/v_0 \approx 10^{-3}$ rad = 1 mrad for $D = 100\ \mu\text{m}$ and $v_0 = 10$ cm.

Figure 3.8b shows the case of **underfocused illumination**, where the C2 lens current has been *decreased* so that an image of the electron source is formed *below* the specimen, at a larger distance v from the lens. The specimen plane no longer contains an image of the electron source, so the diameter of illumination at that plane is not determined by the source diameter but by the value of v. Taking $v = 2v_0$, for example, simple geometry gives the convergence semi-angle *at the image* as $\theta \approx D/v \approx \alpha_0/2$ and the illumination diameter as $d \approx (2\theta)(v - v_0) \approx \alpha_0 v_0 = 50\ \mu\text{m}$. As shown by the dashed lines in Fig. 3.8b, an electron arriving at the center of the specimen at the previous angle α_0 relative to the optic axis (as in Fig. 3.8a) would have to originate from a region *outside* the demagnified source and since there are no such electrons, the new convergence angle α of the illumination must be smaller than α_0. Using the brightness- conservation theorem, Eq. (3.4), the product (αd) must be the same at the new image plane and at the specimen, giving $\alpha = \alpha_0\ (d_0/d) \approx (0.5\ \text{mrad})(1\ \mu\text{m}/50\ \mu\text{m}) \approx 0.010$ mrad. This low value illustrates how defocusing the illumination gives almost *parallel* incident illumination. This condition is useful for recording electron-diffraction patterns in the TEM or for maximizing the contrast in images of crystalline specimens, and is achieved by defocusing the C2 lens or using a small C2 aperture, or both.

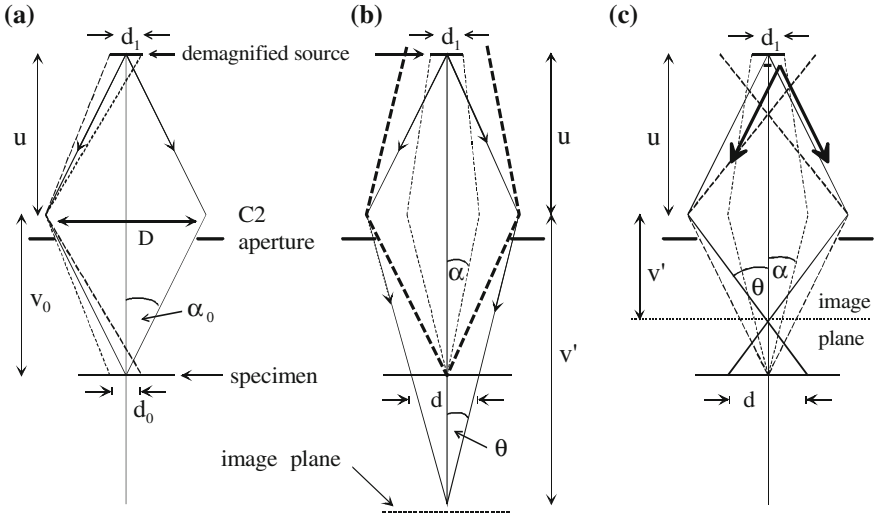


Fig. 3.8 Operation of the second condenser (C2) lens; *solid rays* represent electrons emitted from the *center* of the C1-demagnified source. **a** Fully focused illumination whose diameter d_0 is comparable to the diameter d_1 of the demagnified source (see *dashed rays*) and whose convergence angle ($2\alpha_0$) depends on the diameter of the C2 aperture. **b** Underfocused illumination whose diameter d depends on the image distance v' and whose convergence angle α depends on v' and on d_1. **c** Overfocused illumination, also providing large d and small α

The situation for **overfocused illumination**, where the C2 current has been increased so that the image occurs *above* the specimen plane, is shown in Fig. 3.8c. Relative to the fully focused condition, the illumination diameter d is again increased and the convergence semi-angle α *at the specimen plane* is reduced in the same proportion, in accordance with the brightness theorem. Note that this low convergence angle occurs despite an increase in the *beam* angle θ at the electron-source image plane. The convergence angle of the illumination is always defined in terms of the variation in angle of electrons that arrive *at a single point* in the specimen.

Figure 3.9 summarizes these conclusions in terms of the current-density profile at the specimen plane, which is directly observable (with the radial distance magnified) as an intensity variation in the TEM image. To provide independent control over illumination diameter and convergence angle, some TEMs use more than two condenser lenses, but the general principles remain the same.

3.3.1 Condenser Aperture

The condenser aperture is the small hole in a metal diaphragm located just below the polepieces of the C2 lens. In order to center the aperture on the optic axis, the

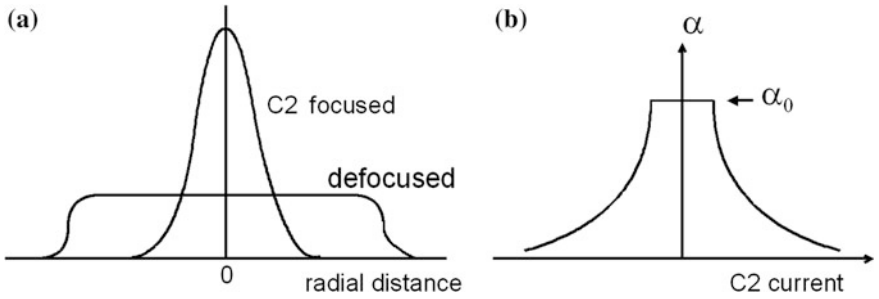


Fig. 3.9 **a** Current density at the specimen as a function of distance from the optic axis, for the illumination fully focused and for defocused (underfocused or overfocused) illumination. **b** Convergence semi-angle of the specimen illumination, as a function of C2-lens excitation

diaphragm is mounted at the end of a support rod that can be moved precisely in the horizontal plane (*x*- and *y*-directions) by turning knobs located outside the microscope column, or by an electric-motor drive. The aperture is correctly aligned (on the optic axis) when varying the C2 current changes the illumination diameter but does *not* cause the magnified disk of illumination to move across the viewing screen. In practice, there are three or four apertures of different diameters ($D \approx 20$–200 μm), arranged along the length of the support rod, so that moving the rod in or out by several mm places a different aperture on the optic axis. Choosing a larger size increases the convergence angle α of the illumination but allows more electrons to reach the specimen, providing a higher intensity in the TEM image.

3.3.2 Condenser Stigmator

The condenser-lens system also contains a stigmator to correct for residual astigmatism of the C1 and C2 lenses. When such astigmatism is present and the *amplitude* control of the stigmator is *set to zero*, the illumination (viewed on the TEM screen, with or without a TEM specimen) expands into an ellipse (rather than a circle) when the C2 lens excitation is increased or decreased from the focused-illumination setting; see Fig. 3.10. To correctly adjust the stigmator, its amplitude control is first set to *maximum* and the orientation control adjusted so that the major axis of the ellipse lies perpendicular to the zero-amplitude direction. The *orientation* is then correct but lens astigmatism has been *overcompensated*. To complete the process, the amplitude setting is reduced until the illumination ellipse becomes a circle, now of smaller diameter than would be possible without astigmatism correction (Fig. 3.10e). In other words, the illumination can be focused more tightly after adjusting the stigmator. To check that the setting is optimum, the C2 current can be varied around the fully focused condition; the illumination should contract or expand but remain circular.

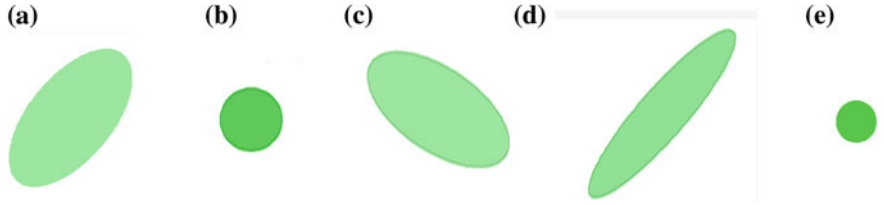


Fig. 3.10 TEM-screen illumination when axial astigmatism is present and the C2 lens is **a** underfocused, **b** fully focused, and **c** overfocused. Also shown: effect of the condenser stigmator with (**d**) its correct orientation (for overfocus condition) but maximum amplitude, and **e** correct orientation and amplitude; note that the focused illumination now has a smaller diameter than with no astigmatism correction

Condenser-lens astigmatism does not directly affect the resolution of a TEM-specimen image. However, it does reduce the maximum intensity (assuming focused illumination) of the TEM-screen image and therefore the ability of the operator to focus on fine details. Therefore the condenser stigmator is routinely adjusted as part of TEM-column alignment.

3.3.3 Illumination Shift and Tilt Controls

The illumination system contains two pairs of coils that apply magnetic fields in the horizontal (x and y) directions, in order to shift the electron beam (incident on the specimen) in the y- and x-directions, respectively. The current in these coils is varied by two illumination-shift controls that are used to *center* the illumination on the TEM screen, correcting for any horizontal drift of the electron gun or slight misalignment of the condenser lenses.

A second pair of coils is used to adjust the *angle* of the incident beam relative to the optic axis. They are located *at* the plane of the C1 image so that their effect on the electron rays does not *shift* the illumination. The currents in these coils are adjusted using (x and y) illumination-tilt controls, which are often adjusted to *align* the illumination parallel to the optic axis (to minimize aberrations of the TEM imaging lenses) but which can also be used to provide *tilted* illumination, as required for *dark-field* images (as discussed in Chap. 4).

3.4 The Specimen Stage

To allow observation in different makes and models of microscope, TEM specimens have a standard circular shape, with a diameter of 3 mm. In some regions at least, the specimen must be thin enough to allow electrons to be transmitted rather than absorbed. The specimen stage is a mechanical device designed to hold the

specimen as stationary as possible, since any drift or vibration will be magnified in the final image, impairing its spatial resolution. But in order to view different regions of the specimen, it may be necessary to move the specimen horizontally over a distance of up to 3 mm.

The design of the stage must also allow the specimen to be inserted without breaking the vacuum of the TEM column. This is achieved by inserting the specimen through an **airlock**, a small chamber into which the specimen is initially placed and which can be evacuated before the specimen enters the TEM column. Not surprisingly, the specimen stage and airlock are mechanically complex and require high-precision manufacture. Two basic stage designs have evolved: side-entry and cartridge-type.

In a **side-entry stage**, the specimen is clamped (for example, by a threaded ring) close to the end of a rod-shaped specimen holder, which is inserted horizontally through the airlock. The airlock-evacuation valve and a high-vacuum valve (at the entrance to the TEM column) are activated by rotation of the specimen holder about its long axis; see Fig. 3.11a.

One advantage of this side-entry design is that it is easy to provide precision motion of the specimen. Translation in the horizontal plane (x- and y-directions) and in the vertical (z) direction is achieved by applying the appropriate movement directly to the specimen rod or to an end-stop that makes contact with the pointed end of the rod. Specimen tilt (rotation to a desired orientation) about the long axis of the rod is easily accomplished by turning the outside end of the specimen holder. Rotation about a perpendicular (horizontal or vertical) axis can be arranged by

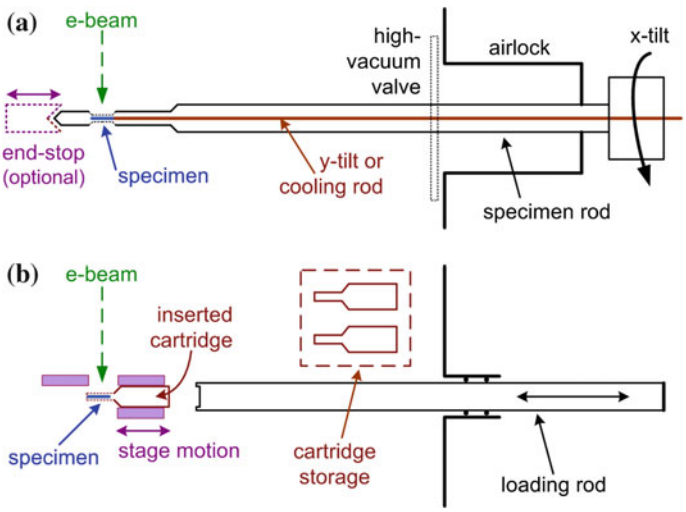


Fig. 3.11 Schematic diagrams of: **a** a side-entry specimen stage, used in the majority of TEMs, **b** a cartridge-type stage, as currently used in STEM instruments from the Nion Company and which allows up to five cartridges to be stored in vacuum and interchanged without air exposure

mounting the specimen on a pivoted ring whose orientation is changed by horizontal movement of a push-rod that runs along the inside of the specimen holder. Precise tilting of the specimen may be required to examine the shape of certain features or to characterize the nature of microscopic defects in a crystalline material.

A further advantage of the side-entry stage is that heating of specimen heating is easily arranged by installing a small heater at the end of the specimen holder, with electrical leads running through the holder to a power supply located outside the TEM. The ability to change the temperature of a specimen allows structural changes in a material (such as phase transitions) to be studied at the microscopic level. Specimen cooling can also be achieved, by incorporating (inside the side-entry holder) a heat-conducting metal rod whose outer end is immersed in liquid nitrogen (at 77 K). If the temperature of a biological-tissue specimen is reduced sufficiently below room temperature, the vapor pressure of ice becomes low enough that the specimen can be maintained in a hydrated state during its examination in the TEM.

The side-entry design also facilitates the use of an **environmental cell** in which the specimen is enclosed between thin windows (typically 50 nm-thick Si_3N_4), allowing gases or liquids to be introduced without degrading the microscope vacuum. Then in situ experiments can be performed, typically involving chemical reactions with the specimen as a catalyst, while chemical and structural changes are observed at a microscopic level, as described in Sect. 7.1. Some side-entry stages include a **Faraday cup**, a small cavity at the end of the specimen rod that can absorb the electron beam and trap most backscattered and secondary electrons (see Chap. 5) so that the beam current can be accurately measured with an external picoammeter.

A disadvantage of the side-entry design is that mechanical vibration, picked up from the TEM column or from acoustical vibrations in the surroundings, is transmitted directly to the specimen. In addition, thermal expansion of the specimen holder may cause drift of the specimen and its TEM image. These problems have been largely overcome by careful design, including choice of materials used to construct the specimen holder. As a result, side-entry holders are now widely used, even for high-resolution imaging.

In a **cartridge stage**, the specimen is clamped to a metal cartridge and inserted into the movable stage by means of a loading arm, which is then detached and retracted. After disengaging the arm, the specimen is insensitive to vibration and temperature change in the TEM environment. In addition, an axially symmetric design is possible, which helps to ensure that thermal expansion occurs radially about the optic axis and becomes vanishingly small on-axis. However, the catrdidge design makes it more difficult to provide tilting, heating, and cooling of the specimen.

Originally the cartridge had a conical collar and was dropped vertically into a conical well of the specimen stage, hence the name **top-entry** stage. For many years this design was preferred for high-resolution imaging because of its superior mechanical stability. But the specimen was held at the bottom end of the cartridge

and only a small fraction of the x-rays generated by the electron beam (and emitted upwards) could be collected, making this design unattractive for analytical microscopy (see Chap. 6).

3.5 TEM Imaging System

The imaging lenses of the TEM are designed to produce a magnified image or an electron-diffraction pattern of the specimen on a viewing screen or camera system. The spatial resolution of the image largely depends on the quality and design of these lenses, especially the first imaging lens: the objective.

3.5.1 Objective Lens

As in the case of a light-optical microscope, the lens closest to the specimen is named the objective. It is a strong lens, with a small focal length, and because of its high excitation current, the objective is cooled with temperature-stabilized water, minimizing any image drift that would result from thermal expansion of the nearby specimen stage. Because focusing power depends on lens excitation, the current for the objective lens must be highly stabilized, using negative feedback within its dc power supply. This power supply must be capable of delivering substantially different lens currents, to retain the same focal length for different electron-accelerating voltages. The power supply also has fine controls that enable the operator to make small adjustments to the objective current, allowing the specimen image to be accurately focused on the viewing screen or camera.

The objective produces a magnified real image of the specimen ($M \approx 50$–100) at a distance v typically 10 cm below the center of the lens. Because of the large magnification, the object distance u is only slightly larger than the focal length, so the specimen is normally within the **pre-field** of the lens: that part of the magnetic field that acts on an electron before it reaches the center of the lens. By analogy with a light microscope, the objective is then referred to as an**immersion lens**.

In a modern *materials-science* TEM, optimized for high-resolution imaging, analytical microscopy, and diffraction analysis of non-biological samples, the specimen is located close to the *center* of the objective lens, where the magnetic field is at a maximum. The objective pre-field then exerts a *strong* focusing effect on the incident illumination and the lens is often called a **condenser-objective**. When the final condenser lens produces a near-parallel beam, the pre-field focuses the electrons into a **nanoprobe** of typical diameter 1–10 nm; see Fig. 3.12a. Such miniscule electron probes are used in analytical electron microscopy to obtain chemical information from very small regions of the specimen, and in

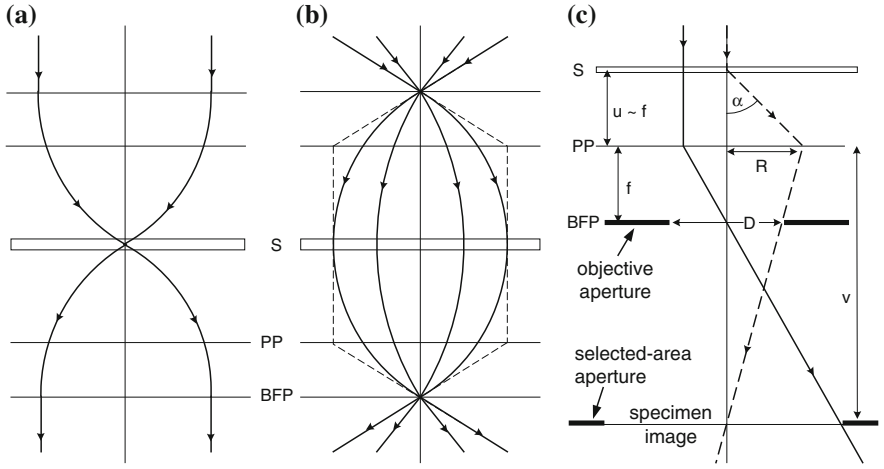


Fig. 3.12 Formation of **a** a small-diameter nanoprobe and **b** parallel illumination at the specimen, with the help of the pre-field of the objective lens. **c** Thin-lens ray diagram for the objective post-field, showing the specimen (S), principal plane (PP) of the objective post-field, and back-focal plane (BFP)

scanning-mode (STEM) operation of the TEM. Alternatively, the condenser system can focus electrons to a crossover at the front-focal plane of the pre-field, giving illumination at the specimen that is approximately parallel, as required for most TEM imaging (Fig. 3.12b). The **post-field** of the objective then acts as the first imaging lens, with a focal length f of around 2 mm. This small focal length optimizes the image resolution by providing small coefficients of spherical and chromatic aberration, as discussed in Chap. 2.

In a *biological* TEM, atomic-scale resolution is often unnecessary and the objective focal length can be somewhat larger. Larger f allows higher image contrast (for a given objective-aperture diameter), often of prime concern because the contrast in tissue-sample images can be very low.

3.5.2 Objective Aperture

An objective diaphragm can be inserted located at the **back-focal plane** (BFP) of the objective-lens post-field, the plane at which a diffraction pattern of the specimen is first produced. Within this plane, *distance* from the optic axis represents the *direction* of travel (angle relative to the optic axis) of an electron that has just left the specimen.

Although in practice the objective behaves as a *thick* lens, we will discuss its properties using a thin-lens ray diagram in which the focusing deflection is

considered to occur at a single plane: theprincipal plane of the lens. The solid ray in Fig. 3.12c shows an electron leaving the specimen away from the optic axis but parallel to it. This electron is deflected at the principal plane and crosses the axis at the back-focal plane, a distance f below the principal plane. Assuming parallel illumination and correctly adjusted tilt controls, such an electron will have *arrived* at the specimen parallel to the axis and must therefore have remained *undeflected* (unscattered) during its passage through the specimen.

The dashed ray in Fig. 3.12c represents an electron that arrives *along* the optic axis and is scattered or diffracted through an angle α by interaction with one or more atoms in the specimen. When it arrives at the principal plane, its radial distance from the axis is $R = u\tan\alpha \approx f\tan\alpha$. After deflection by the objective, this electron crosses the optic axis at the first image plane, a relatively large distance ($v \approx 10$ cm) from the lens. Below the principal plane, the dashed ray is therefore *almost parallel* to the optic axis and its displacement from the axis at the back-focal plane is almost equal to R. By inserting an aperture of diameter D (centered around the optic axis) at the BFP, we can restrict the electrons that pass through the rest of the imaging system to those whose scattering angles are between zero and α, where

$$\alpha \approx \tan\alpha \approx R/f \approx D/(2f) \tag{3.9}$$

Electrons scattered through larger angles are *absorbed* by the diaphragm surrounding the aperture and do not contribute to the final image. By using a small-diameter aperture, we can ensure that most *scattered* electrons are absorbed by the diaphragm. Regions of specimen that scatter electrons strongly will then appear dark (due to fewer electrons) in the final TEM image, which is said to display **scattering contrast** or diffraction contrast.

Besides producing contrast in the TEM image, the objective aperture limits the amount of image blurring that may arise from spherical and chromatic aberration. By restricting the range of scattering angles to values less than α, given by Eq. (3.9), the loss of spatial resolution is limited to an amount $r_s \approx C_s\alpha^3$ due to spherical aberration and $r_c \approx C_c\alpha(\Delta E/E_0)$ due to chromatic aberration. Here, C_s and C_c are aberration coefficients of the objective lens (Chap. 2); ΔE represents the spread in kinetic energy of the electrons emerging from the specimen and E_0 is their kinetic energy before entering the specimen. Because both r_s and r_c increase with aperture size, we might imagine that the best resolution corresponds to the smallest possible aperture diameter.

However, the objective diaphragm gives rise to a diffraction effect that becomes more severe as its diameter decreases, as discussed in Sect. 1.1, which leads to a loss of resolution Δx given by the Rayleigh criterion:

$$\Delta x \approx 0.6\lambda/\alpha \tag{3.10}$$

where we assume that α is small and is measured in radians.

Ignoring *chromatic* aberration for the moment, we can combine the effect of spherical aberration and electron diffraction at the objective diaphragm by adding the two blurring effects together, so that the image resolution Δr (measured in the object plane) is:

$$\Delta r \approx r_s + \Delta x \approx C_s\alpha^3 + 0.6\,\lambda/\alpha \tag{3.11}$$

Since the two terms in Eq. (3.11) have opposite α-dependence, their sum is represented by a curve with a minimum value; see Fig. 3.13a. This value can be found by differentiating Δr with respect to α and setting the total derivative to zero, giving $\alpha* = 0.67(\lambda/C_s)^{1/4}$ as the optimum aperture angle.

A better procedure is to treat the blurring terms r_s and Δx like statistical errors and combine their effect in quadrature:

$$(\Delta r)^2 \approx (r_s)^2 + (\Delta x)^2 \approx (C_s\alpha^3)^2 + (0.6\lambda/\alpha)^2 \tag{3.12}$$

Taking the derivative and setting it to zero then results in:

$$\alpha^* = 0.63\,(\lambda/C_s)^{1/4} \tag{3.13}$$

As an example, if we take C_s = 1.2 mm and E_0 = 200 keV so that λ = 2.5 pm, Eq. (3.13) gives α* = 4.3 mrad, corresponding to an objective aperture of diameter $D \approx 2\alpha^* f \approx 15$ μm if f = 1.7 mm. Using Eq. (3.12), the optimum resolution is

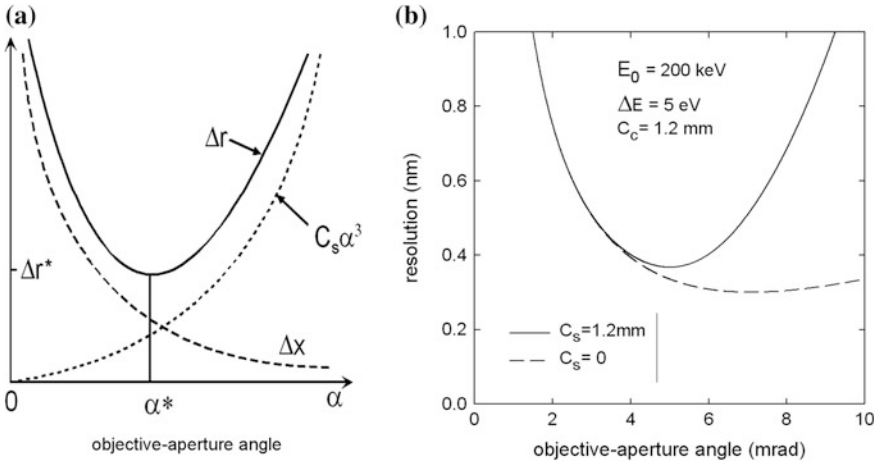


Fig. 3.13 **a** Loss or resolution due to objective-lens spherical aberration and diffraction at the objective aperture. The *solid curve* shows the combined effect. **b** Calculated values for a typical TEM, with spherical aberration (C_s = 1.2 mm) and without (C_s = 0)

$\Delta r^* = 0.34$ nm. Inclusion of *chromatic* aberration in Eq. (3.12) increases Δr^* somewhat; see Fig. 3.13b. If the objective lens has an image-aberration corrector, we can set $C_s \approx 0$ and the resolution is improved, now limited by chromatic aberration and diffraction (dashed curve in Fig. 3.13b).

In fact, our calculations are somewhat pessimistic because they assume that electrons are present in equal numbers at all scattering angles up to the aperture semi-angle α. A more exact treatment is possible, based on wave optics rather than geometrical optics. In practice, a modern 200 kV TEM can achieve a point resolution below 0.2 nm, even without aberration correction, allowing atomic resolution under proper conditions, which include a low vibration level in the microscope room and a low ambient ac magnetic field.

3.5.3 Objective Stigmator

Our resolution estimates assume that the imaging system does not suffer from astigmatism. In practice, axial astigmatism can arise from hysteresis effects in the lens polepieces or electrostatic charging of contamination layers (on the specimen or on an objective diaphragm), and may be different for each specimen. The TEM operator can correct for axial astigmatism of the objective (and other imaging lenses) by using an *objective* stigmator, located just below the objective lens. The correct setting requires adjustment of both amplitude and orientation, as discussed in Sect. 2.6. One way of setting these controls is to insert an amorphous specimen, whose image contains small point-like features. With astigmatism present, there is a "streaking effect" (preferred direction) visible in the image, which changes in orientation by 90° as the objective is adjusted from underfocus to overfocus of the specimen image (similar to Fig. 5.17). The stigmator controls are adjusted to minimize this streaking effect.

3.5.4 Selected-Area Aperture

As indicated in Figs. 3.12c, a diaphragm can be inserted in the plane that contains the first magnified image of the specimen, the *image plane* of the objective lens. This selected-area diffraction (SAD) aperture is used to limit the region of specimen from which an electron-diffraction pattern is recorded. Electrons are transmitted through the aperture only if they fall within its diameter D, corresponding to a diameter of D/M at the specimen plane. In this way, diffraction information is obtained from specimen regions whose diameter is as small as 0.2 μm (taking $D \approx 20$ μm and an objective-lens magnification $M \approx 100$).

As suggested by Fig. 3.12c, *sharp* diffraction spots will be formed at the objective back-focal plane only if the electron beam incident on the specimen is almost parallel. This condition is achieved by defocusing the second condenser lens, giving a low convergence angle but a large irradiated area at the specimen (Fig. 3.9). The purpose of the SAD aperture is therefore to provide diffraction information with good angular resolution, combined with sub-μm spatial resolution.

Even better spatial resolution is possible by using the objective-lens pre-field to form a nanoprobe at the specimen plane (Fig. 3.12a). The incident-convergence angle α is then relatively high (several mrad) and the diffraction spots are broadened into disks in a **convergent-beam diffraction** (CBED) pattern. Analysis of the fine structure of such patterns can yield information about the thickness and structure of a crystalline specimen.

3.5.5 Intermediate Lens

A modern TEM contains several lenses between the objective and the final (projector) lens. At least one of these lenses is referred to as the intermediate and the combined effect of all of them can be described in terms of the action of a *single* intermediate lens, as shown in Fig. 3.14. The intermediate serves two purposes. First of all, by changing its focal length in small steps, its image magnification can be changed, allowing the overall magnification of the TEM to be varied over a large range, typically 10^3–10^6.

Secondly, by making a larger change to the intermediate-lens excitation, an electron-diffraction pattern can be produced on the TEM viewing screen. As depicted in Fig. 3.14, this is achieved by *reducing* the current in the intermediate so that this lens produces a diffraction pattern (solid ray crossing the optic axis) at the projector-lens *object* plane, rather than an image of the specimen (dashed ray crossing the axis).

Because the high magnification of the objective lens *reduces* the angles of electrons relative to the optic axis, intermediate-lens aberrations are not of great concern. Therefore the intermediate can operate as a weak lens (focal length of several cm) without degrading the image or diffraction pattern.

3.5.6 Projector Lens

The purpose of the projector lens is to produce an image or a diffraction pattern whose overall diameter (on the TEM screen or camera phosphor) may be several cm. Because of this requirement, some electrons (such as the solid single-arrow ray in Fig. 3.14) arrive at the screen at a large distance from the optic axis, which introduces the possibility of image *distortion* (Chap. 2). To minimize distortion, the projector operates as a strong lens, with a focal length of a few mm. Ideally, the

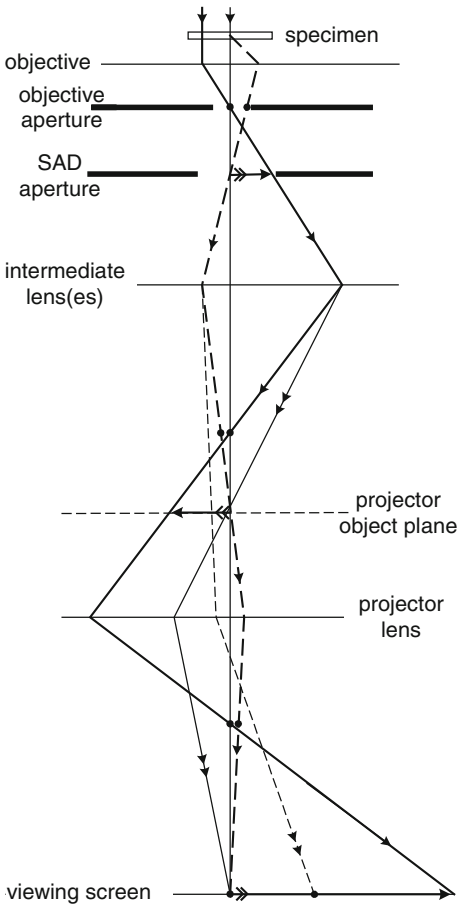


Fig. 3.14 Thin-lens ray diagram of the imaging system of a TEM. As usual, the image rotation has been suppressed so that the electron optics can be represented on a flat plane. Image planes are represented by *horizontal arrows* and diffraction planes by *horizontal dots*. *Double arrowheads* indicate rays that lead to a TEM-screen diffraction pattern (Although illustrative, the diagram is not to scale: the final-image magnification is only 8 here)

final-image diameter is fixed (the image should *fill* the TEM screen) and the projector is kept at a fixed excitation, with a single object distance and magnification $v/u \approx 100$. However, in some TEMs the projector-lens strength can be reduced to give images of relatively low magnification (<1000) at the viewing screen. As in the case of light optics, the final-image magnification is the algebraic product of the magnification factors of each of the imaging lenses.

3.5.7 TEM Screen and Camera System

A phosphor screen is used to convert the electron image to a visible form. It consists of a metal plate coated with a thin layer of powder that fluoresces (emits visible light) under electron bombardment. The traditional phosphor material is zinc sulfide

(ZnS) with small amounts of metallic impurity added but other materials have been developed, with improved sensitivity (electron/photon conversion efficiency). The phosphor emits light in the middle of the spectrum (yellow-green region), where the human eye and camera systems are most sensitive.

Traditionally, the TEM screen is used mainly for *focusing* a TEM image or diffraction pattern, with light-optical binoculars mounted just outside a viewing window to provide additional magnification. This window requires special glass (high lead content) and must be thick enough to absorb the x-rays that are produced when the electrons deposit their energy at the screen.

Nowadays, many microscopes dispense with the viewing screen but nevertheless require a fluorescent screen (also known a scintillator) whose light-optical image is recorded by electronic camera. The camera contains a photosensitive array of several million silicon diodes, each of which provides an electrical signal proportional to the local intensity level. Because the recorded image can be seen almost immediately on a monitor screen, it can be used for focusing and astigmatism correction, as well as for permanent recording. The final image is stored in a digital computer, which allows automatic adjustment of intensity and contrast, as well as various image-processing procedures. One such procedure is to calculate a Fourier transform of the image to yield a **diffractogram** that can reveal artifacts such as astigmatism and image drift. Benefiting from the speed of modern electronics, the diffractogram can be calculated in real time and used to optimize image quality (correction of astigmatism and lens aberrations).

To avoid introducing granularity, a fluorescent screen should be made of single-crystal or a fine-grain polycrystalline material, and for maximum sensitivity it should be thick enough to completely absorb the electrons. However, electrons are scattered sideways as they slow down (see Sect. 5.2) so the spatial resolution of the image is improved by using a thinner screen. Therefore a compromise is necessary, with thicker screens being favored for higher-energy (more penetrating) electrons. Although a reflection screen (phosphor-coated metal plate) is convenient for direct viewing, a thin transmission screen can provide better resolution and is often used for image recording, with fiber-optic coupling of light into the camera.

Direct-recording cameras have been developed that respond to high-energy electrons without the need for a conversion screen. Requiring a specialized sensor and advanced electronics, they are relatively expensive but offer superior image quality (lower noise), especially for low-intensity images. They are used by biologists who require near-atomic resolution from beam-sensitive specimens and by materials scientists conducting in-situ experiments requiring fast recording and low radiation dose.

3.5.8 Depth of Focus and Depth of Field

The image-recording camera is sometimes located several cm below or above the TEM viewing screen. If a specimen image has been brought to exact focus on this

viewing screen, it is strictly speaking *out-of-focus* at any other plane. At a plane that is a height h below or above the true image plane, each point in the image becomes a circle of confusion whose radius is $s = h \tan\beta$, as illustrated in Fig. 3.15a. This radius is equivalent to a blurring in the specimen plane of

$$\Delta s = s/M = (h/M)\tan\beta \tag{3.14}$$

where M is the combined magnification of the entire imaging system and β is the convergence angle at the screen, corresponding to scattering through an angle α in the specimen, as shown in Fig. 3.15a. However, this blurring will be significant only if Δs is comparable to or greater that the resolution Δr of the in-focus TEM image. In other words, the additional image blurring will not be noticeable if $\Delta s \ll \Delta r$.

To compute Δs we first note, from trigonometry of the two large triangles in Fig. 3.15a, that $x = u \tan\alpha = v \tan\beta$, giving $\tan\beta = (u/v)\tan\alpha = (1/M)\tan\alpha$. If an objective diaphragm is in place to remove electrons above $\alpha \approx 5$ mrad, then $\beta \approx \tan\beta \approx \alpha/M = 5 \times 10^{-6}$ rad for $M = 1000$, decreasing to 5×10^{-9} rad for $M = 10^6$.

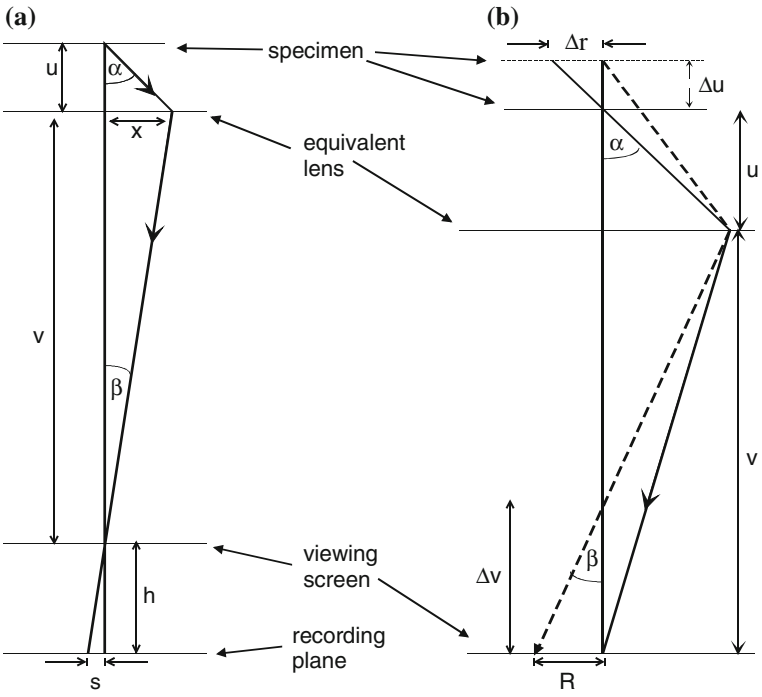


Fig. 3.15 **a** Ray diagram illustrating depth of focus; the TEM imaging lenses have been replaced by a single lens of large magnification M. **b** Ray diagram illustrating depth of field. The radius of blurring (caused by raising the specimen through a distance Δu) is R in the image-recording plane, equivalent to $\Delta r = R/M$ in the specimen plane

These are extremely small angles, corresponding to small values of Δs: for $h = 10$ cm, Eq. (3.14) gives $\Delta s = 0.5$ nm for $M = 1000$ and $\Delta s = 5 \times 10^{-16}$ m for $M = 10^6$. So for typical TEM magnifications, the final image recorded at any plane is equally sharp.

Another way of expressing this result is as follows. Suppose we take $\Delta s = \Delta r$, so that the image resolution is only slightly degraded (by a factor $\approx\sqrt{2}$) by being out-of-focus. The value of h that would produce this Δs, for a given M and α, is known as the **depth of focus** of the image (sometimes called depth of image). Taking $\alpha = 5$ mrad and $\Delta r = 0.2$ nm, our depth of focus is $h = M\,\Delta r/\beta \approx M^2 \Delta r/\alpha = 4$ cm for $M = 1000$, increasing to $h = 40$ km for $M = 10^6$. In practice, a magnification as low as 1000 would not be used to record a high-resolution image, since image detail on a scale $M\Delta r \approx 0.2$ μm would be below the resolution limit of a phosphor or camera. Therefore, practical values of the depth of focus are so large that a magnified image of the specimen is equally sharp at all recording planes.

A related concept is **depth of field**: the distance Δu (along the optic axis) that the *specimen* can be moved without its image (focused at a given plane) becoming noticeably blurred. Replacing the TEM imaging system with a single lens of fixed focal length f and using $1/u + 1/v = 1/f$, the change in image distance is $\Delta v = -(v/u)^2 \Delta u = -M^2\,\Delta u$ (the minus sign denotes a decrease in v, as in Fig. 3.15b). The radius of the disk of confusion in the *image* plane (due to defocus) is $R = \beta|\Delta v| = (a/M)\,(-\Delta v) = \alpha M\,\Delta u$, equivalent to a radius $\Delta r = R/M = \alpha\,\Delta u$ at the *specimen* plane. This same result can be obtained more directly by extrapolating backwards the solid ray in Fig. 3.15b to show that the electrons that arrive at a single on-axis point in the image could have come from anywhere within a circle whose radius (from geometry of the topmost right-angled triangle) is $\Delta r = \alpha\,\Delta u$.

If we take $\alpha = 5$ mrad and set the loss or resolution Δr equal to a TEM resolution of 0.2 nm, we get $\Delta u = (0.2\text{ nm})/(0.005) = 40$ nm as the depth of field. Most TEM specimens are of the order of 100 nm or less in thickness. If the objective lens is focused at the mid-plane of the specimen, features close to the top and bottom surfaces will still be acceptably in focus. In other words, the small value of α ensures that a TEM image is normally a *projection* of all structure present in the specimen. It is therefore difficult or impossible to obtain depth information from a single TEM image. On the other hand, this substantial depth of field avoids the need to focus on different planes within the specimen, as is sometimes necessary in a light-optical microscope where α may be large and the depth of field considerably less than the specimen thickness.

3.6 Scanning Transmission Systems

As indicated in Sect. 1.6, transmission electron microscopy can also be done in a scanning-beam mode, where electrons are focused into a very small probe that is scanned over the specimen while recording a time-dependent signal that contains

the spatial information necessary to form an image. Such images are most often recorded using a TEM that is fitted with scanning coils just above the specimen, taking advantage of the pre-field of the condenser-objective lens to form a sub-nm probe (Fig. 3.12a). Because the image is now formed using the raster-scanning principle (see Sects. 1.6 and 5.1), post-specimen lenses are not essential and the TEM imaging lenses are used only to efficiently transfer electrons into an appropriate detector, typically located below the projector lens. Most STEM images are acquired using a high-angle annular dark-field (HAADF) detector, a ring-shaped semiconductor device similar in geometry and operation to the solid-state detector used to record backscattered electrons in the SEM; see Fig. 5.11b.

When first introduced, a STEM attachment for a TEM (scan coils, STEM detector, display screen, and associated electronics) was difficult to use because the necessary lens currents and deflection-coil settings (for column alignment) differed from those needed for TEM operation. Computer control (including storage of lens and deflector-coil settings) has removed these difficulties and STEM operation is now convenient and offers certain advantages, including compatibility with x-ray and EELS signals. One *disadvantage* of STEM is that atomic resolution is harder to achieve: mechanical vibration of the electron source and noise in the scan-current waveforms both give rise to a blurring of the image. However, these problems can be overcome and dark-field STEM can provide an attractive alternative to phase-contrast TEM; see Fig. 1.17. STEM becomes more competitive when a TEM is fitted with high-brightness (CFEG or Schottky) electron source with a small emitting area.

As outlined in Sect. 1.6, there are also *dedicated* STEMs that operate only in scanning mode. A cold field-emission source is then preferable for high resolution and is sometimes placed at the bottom of the lens column, with electron detectors at the top (gravity changes the electron speed by less than 1 part in 10^7).

3.7 Vacuum System

It is essential to remove most of the air from the inside of a TEM column, allowing the accelerated electrons to obey the principles of electron optics rather than being scattered by gas molecules, and so that the high voltage applied to the gun does not cause an electrical discharge. These requirements suggest a pressure below about 0.1 Pa, or 10^{-3} Torr in older units (1 Torr = 130 Pa = atmospheric pressure/760). However, a LaB_6 source requires a **high vacuum** (pressure below 10^{-4} Pa) to avoid contamination of the electron-emitting surface and a cold field emission gun requires **ultra-high vacuum** (UHV, $\sim 10^{-8}$ Pa) for stable emission; see Table 3.1. These very low pressures cannot be produced with a single vacuum pump; a "rough" vacuum (of the order of 1 Pa) must be obtained first by use of a **roughing pump**.

A mechanical **rotary pump** (RP) is often used for this purpose. It contains a rotating assembly, driven by an electric motor and equipped with internal vanes

(A and B in Fig. 3.16) separated by a coil spring so that the vanes press against the inside cylindrical wall of the pump, forming an airtight seal. The interior is lubricated with a suitable oil (having low vapor pressure) to reduce friction and wear of the sliding surfaces. The rotation axis is offset so that gas drawn from the inlet tube (at A in Fig. 3.16a) expands considerably in volume before being sealed off by the opposite vane (B in Fig. 3.16b). During the remainder of the rotation cycle, the air is compressed (Fig. 3.16c) and driven out of the outlet tube (Fig. 3.16d). Meanwhile, air in the opposite half of the cylinder has started its expansion and will then be compressed to the outlet, giving a continuous pumping action.

An alternative form of roughing pump is the **scroll pump** that contains two internal spirals, rotating with respect to each other to provide a pumping action. It is an example of a "dry" or oil-free pump, sometimes preferred because it reduces the presence of hydrocarbon molecules (that can lead to electron-beam contamination; see later) within the vacuum system.

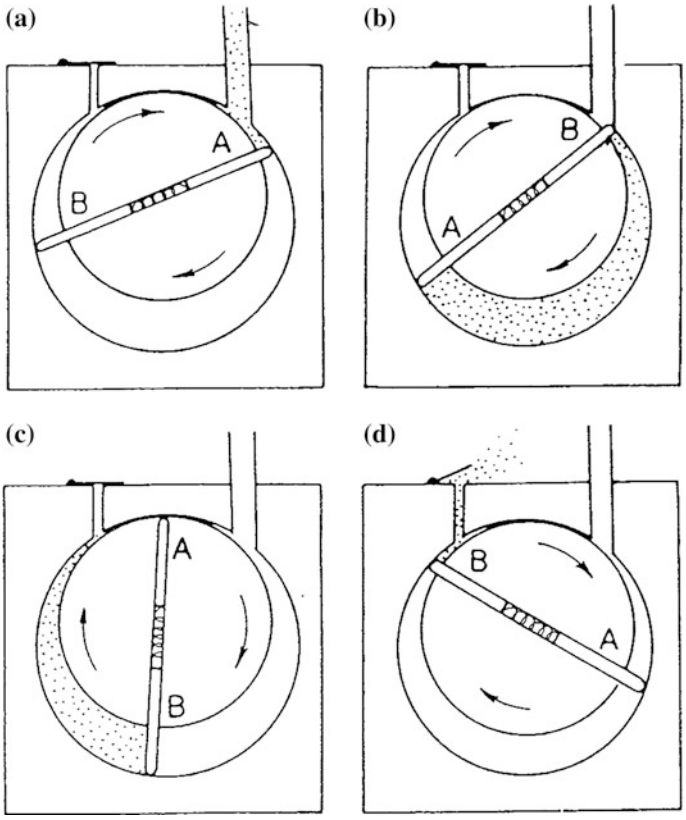


Fig. 3.16 Schematic diagram of a rotary vacuum pump, illustrating one complete cycle of the pumping sequence. Spring-loaded vanes A and B rotate clockwise, each drawing in gas molecules and then compressing them towards the outlet

High vacuum is traditionally produced by a **diffusion pump** (DP), which consists of a vertical metal cylinder containing a small amount of a carbon- or silicon-based oil having very low vapor pressure and a boiling point of several hundred degrees Celsius. At the base of the pump (Fig. 3.17) an electrical heater causes this liquid to boil and turn into a vapor that rises and is then deflected downwards through jets by an internal baffle assembly. During their descent, the oil molecules collide with air molecules, imparting a downward momentum and resulting in a flow of gas from the inlet of the pump. When the oil molecules arrive at the cooled inside wall of the DP, they condense back into a liquid that falls under gravity to the base of the pump, allowing the oil to be continuously recycled. To prevent oxidation of the oil, the pressure at the base of the diffusion pump must be below about 10 Pa before the heater is turned on. A roughing pump is therefore connected to the bottom end of the diffusion pump and acts as a "backing pump" (Fig. 3.17). This roughing pump (or a second one) is also used to remove most of the air from the TEM column before the high-vacuum valve (connecting the DP to the column) is opened.

Nowadays, a **tubomolecular pump** (TMP) often replaces or supplements the diffusion pump. It is essentially a high-speed turbine with multiple blades arranged in layers on the same shaft and driven by an electric motor. Magnetic bearings are often used to reduce vibration. The cryogenic pump (**cryopump**) is another form of oil-free pump; it uses a refrigeration system to cool a surface (often to below 50 K) onto which gas molecules condense. Eventually the surface becomes saturated with condensate and the pump has to be regenerated by warming to room temperature or higher.

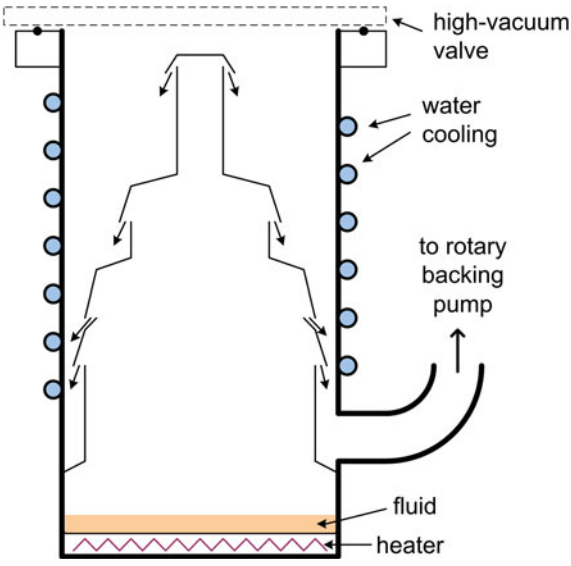


Fig. 3.17 Cross-section through a diffusion pump. The *arrows* show oil vapor leaving jets within the baffle assembly. Water flow within a coiled metal tube keeps the walls cool

To provide ultra-high vacuum, an **ion pump** (IP in Fig. 3.18) is required. By applying a potential difference of several kV between large electrodes, a low-pressure discharge is set up (aided by the presence of a magnetic field) that ionizes gas molecules. Positive ions are attracted to the negative electrode (cathode, made of the reactive metal titanium), where they remove Ti atoms from the surface in the process known as sputtering. The Ti atoms condense on internal surfaces and chemically bind with gas atoms, reducing the gas pressure. The discharge current falls as the pressure drops and can be used as an approximate measure of pressure.

These various pumps must work together in the correct sequence. If a TEM column has been at atmospheric pressure (perhaps for maintenance) most of the air is removed by a roughing pump, which may then function as the *backing pump* for the high-vacuum pump. The pumping sequence is controlled by opening and closing

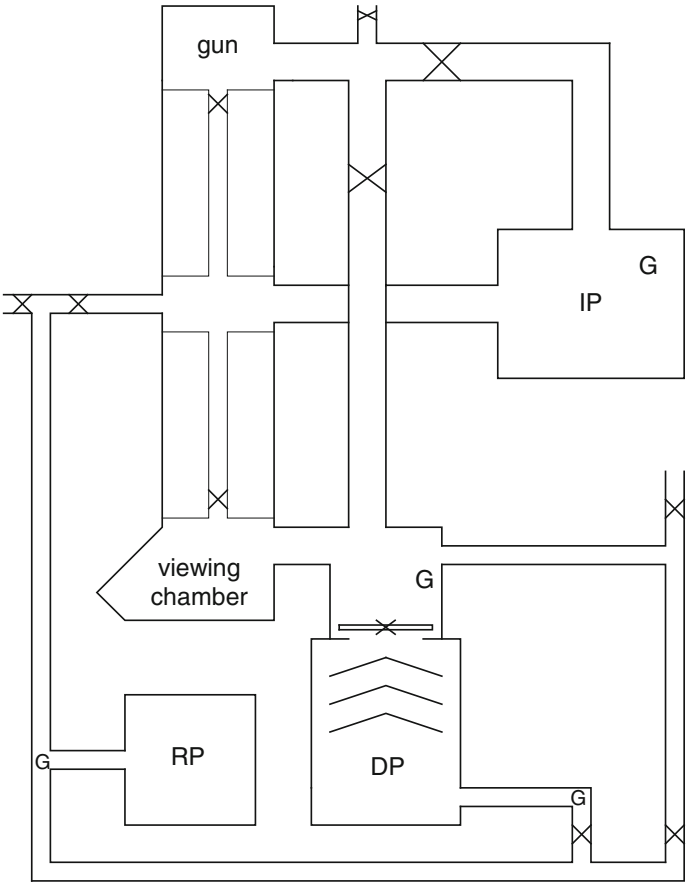


Fig. 3.18 Simplified diagram of the vacuum system of a typical TEM. Vacuum valves are indicated by × and vacuum gauges by G. Sometimes a turbomolecular or cryogenic pump is used in place of the diffusion pump, resulting in a cleaner (oil-free) vacuum system

vacuum valves in response to the local pressure, which is monitored by vacuum gauges located in different parts of the system. In a modern TEM the vacuum system is automated and under the digital control of a dedicated microprocessor.

Liquid nitrogen is often used to improve the vacuum inside a TEM, using the same process as in a cryopump. With a boiling point of 77 K, liquid nitrogen can cool internal surfaces sufficiently to cause water vapor and organic vapors (hydrocarbons) to condense onto them. These vapors are therefore removed locally, so long as the surface remains cold. In a diffusion-pumped vacuum system, liquid nitrogen poured into a "trap" (a contained just above the DP) reduces the "backstreaming" of diffusion-pump vapor into the TEM column.

Organic vapors can also enter the vacuum from components such as the rubber o-ring seals. When any surface is irradiated with electrons, hydrocarbon molecules that arrive at the surface can become chemically bonded together to form a solid *polymer*. This **electron-beam contamination** can build up on the top and bottom surfaces of a TEM specimen, causing additional scattering of the electrons and reducing the contrast of the TEM image. To combat this process, liquid nitrogen is added to a decontaminator trap attached to the side of the TEM column. By conduction through metal rods or braiding, it cools metal plates located just above or below the specimen, condensing water and organic vapors so that the partial pressure of these components is reduced in the immediate vicinity of the specimen.

3.8 Further Reading

Transmission Electron Microscopy (Springer, 2nd edition, 2009) by D.B. Williams and C.B. Carter provides a readable and highly practical guide to understanding the operation of the TEM and covers all the important aspects including analytical electron microscopy. It is aimed at graduate or advanced undergraduate students and uses a conversational style. In common with Williams and Carter, I have maintained a distinction between the terms diaphragm and aperture, although in ordinary parlance "aperture" is most often used for both the circular disk and the hole within it.

There are numerous websites giving practical information about electron microscopy, including the following.

www.amc.anl.gov gives information about the Argonne National Laboratory contributions to microscopy and microanalysis, including software tools written by N.J. Zaluzec.

www.microscopy.com is home to a discussion forum and lists meetings, courses, and other microscopy sites around the world.

www.Rodenburg.org contains advice from John Rodenburg about TEM and STEM alignment, and diffractive imaging.

http://www.emal.engin.umich.edu/courses/CBEDMSE562/ provides an introduction to convergent-beam diffraction (CBED) and its applications.

Chapter 4
TEM Specimens and Images

In the transmitted-light microscope, intensity variation within an image is caused by differences in the *absorption* of photons within different regions of the specimen. In the case of a TEM, however, essentially *all* of the incoming electrons are transmitted through the specimen, provided it is suitably thin. Although not absorbed, these electrons are **scattered** (deflected in their path) by the atoms of the specimen. To understand the formation and interpretation of TEM images, we need to examine the nature of this scattering process.

Many of the concepts involved can be understood by considering just a *single* "target" atom, which we represent in the simplest way as a central nucleus surrounded by orbiting particles, the atomic electrons; see Fig. 4.1. This is the classical planetary model, first established by Ernest Rutherford from analysis of the scattering of alpha particles. In fact, our analysis will largely follow that of Rutherford. Instead of positive alpha particles, we have negative incident electrons but the cause of the scattering remains the same: the electrostatic **Coulomb interaction** between charged particles.

Interaction between the incoming primary electron and an atomic *nucleus* gives rise to **elastic scattering** where (as we shall see) almost no energy is transferred. Interaction between the fast-moving primary electron and atomic *electrons* results in**inelastic scattering**, in which the transmitted electron can lose an appreciable amount of energy. We will discuss these two types of events in turn, dealing first with the **kinematics** of scattering, where we consider the momentum and energy of the fast electron, rather than the forces involved. Since the nature of the force does not matter, electron-scattering kinematics employs the same principles of classical mechanics that are used to treat the collision of larger (macroscopic) objects.

R.F. Egerton, *Physical Principles of Electron Microscopy*,
DOI 10.1007/978-3-319-39877-8_4

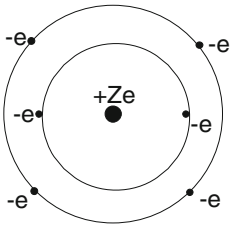


Fig. 4.1 Rutherford model of an atom (in this case, carbon with atomic number $Z = 6$). The electrostatic charge of the central nucleus (mass M) is $+Ze$; the balancing negative charge is provided by Z atomic electrons, each with charge $-e$

4.1 Kinematics of Scattering by an Atomic Nucleus

The nucleus of an atom is truly miniscule (about 3×10^{-15} m in diameter) whereas the whole atom has a diameter of around 3×10^{-10} m. The fraction of space occupied by the nucleus is therefore of the order of $(10^{-5})^3 = 10^{-15}$ and the probability of an electron actually "hitting" the nucleus is almost zero. However, the electrostatic field of the nucleus extends throughout the atom and incoming electrons are deflected (scattered) by this field.

Applying the principle of conservation of energy to the electron-nucleus system (Fig. 4.2) gives:

$$mv_0^2/2 = mv_1^2/2 + MV^2/2 \tag{4.1}$$

where v_0 and v_1 represent the speeds of the electron before and after the "collision", while V is the speed of the nucleus (assumed initially at rest) immediately after the interaction. Because v_0 and v_1 apply to the electron when it is distant from the nucleus, there is no need to include potential energy in Eq. (4.1). For simplicity, we are using the classical (rather than relativistic) formula for kinetic energy, taking m and M as the *rest masses* of the electron and nucleus. As a result, our analysis will not be quantitatively accurate for incident energies E_0 above 50 keV, but this inaccuracy will not affect our general conclusions.

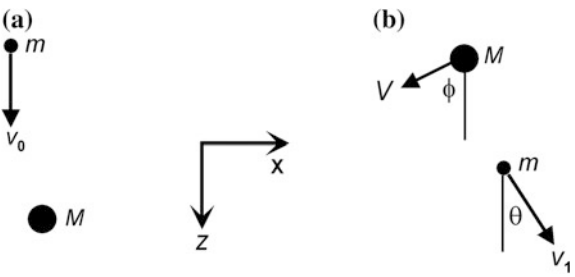


Fig. 4.2 Kinematics of electron-nucleus scattering, showing **a** the speed and direction of the electron (mass m) and nucleus (mass M) before and **b** after the collision

Applying conservation of momentum to velocity components in the z- and x-directions (parallel and perpendicular to the original direction of travel of the electron) gives:

$$mv_0 = mv_1 \cos\theta + MV \cos\phi \tag{4.2}$$

$$0 = mv_1 \sin\theta - MV \sin\phi \tag{4.3}$$

In the TEM we image scattered electrons, not recoiling atomic nuclei, so the unknown nuclear parameters V and ϕ need to be eliminated from Eqs. (4.1)–(4.3). The angle ϕ can removed by using the formula (equivalent to the Pythagoras rule): $\sin^2\phi + \cos^2\phi = 1$. Taking $\sin\phi$ from Eq. (4.3) and $\cos\phi$ from Eq. (4.2) gives:

$$M^2V^2 = M^2V^2\cos^2\phi + M^2V^2\sin^2\phi = m^2v_0^2 + m^2v_1^2 - 2m^2v_0v_1\cos\phi \tag{4.4}$$

where we have used $\sin^2\theta + \cos^2\theta = 1$ to further simplify the right-hand side of Eq. (4.4). Using Eq. (4.4) to substitute for $MV^2/2$ in Eq. (4.1), we eliminate V and obtain:

$$mv_0^2/2 - mv_1^2/2 = MV^2/2 = (1/M)\left[m^2v_0^2 + m^2v_1^2 - 2m^2v_0v_1\cos\theta\right] \tag{4.5}$$

The left-hand side of Eq. (4.5) represents the amount of kinetic energy E lost by the electron (and gained by the nucleus) as a result of the Coulomb interaction, so the *fractional* loss of kinetic energy is:

$$E/E_0 = 2E/\left(mv_0^2\right) = (m/M)\left[1 + v_1^2/v_0^2 - 2\,(v_1/v_0)\cos\theta\right] \tag{4.6}$$

The *largest* value of E/E_0 occurs when $\cos\theta \approx -1$, giving:

$$E/E_0 \approx (m/M)\left[1 + v_1^2/v_0^2 + 2(v_1/v_0)\right] = (m/M)\left[1 + v_1/v_0\right]^2 \tag{4.7}$$

This situation corresponds to a head-on collision, where the electron makes a 180° turn behind the nucleus, as illustrated by the dashed trajectory in Fig. 4.3a.

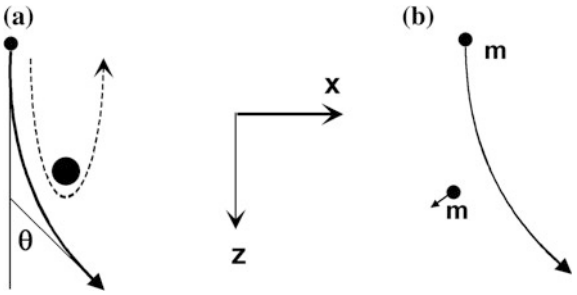


Fig. 4.3 Hyperbolic trajectory of an incident electron during **a** elastic scattering (the case of a large-angle collision is indicated by the dashed curve) and **b** inelastic scattering

Kinematics cannot tell us the value of v_1 but we know that $v_1 < v_0$ from Eq. (4.1), so Eq. (4.7) gives:

$$E/E_0 < (m/M)\,[1+1]^2 = 4m/M = 4m/(uA) \tag{4.8}$$

where u is the atomic mass unit and A is the *atomic mass number* (number of protons + neutrons in the nucleus), often called the *atomic weight* of the atom. Since $m/u \approx 1/1836$, Eq. (4.8) gives $E/E_0 < 0.002/A$, so even for hydrogen ($A = 1$) and 180° scattering, only 0.2 % of the kinetic energy of the electron is transmitted to the nucleus. As we will see, a typical scattering angle is closer to 1° (0.02 rad) and Eq. (4.5) gives $E/E_0 \approx (m/M)\left[1+1-2\left(1-\theta^2/2\right)\right] = (m/M)\theta^2 \approx 10^{-5}/A$ if we take $v_1 \approx v_0$ and $\cos\theta \approx 1-\theta^2/2$. Electron scattering by the electrostatic field of a nucleus is therefore termed *elastic*, meaning that the *scattered electron* retains almost all of its kinetic energy.

Nuclear scattering is also elastic in the more usual sense, as applied to collisions of two macroscopic objects: the *sum* of the *kinetic* energy of the two particles remains unchanged. This last criterion remains true even for large-angle scattering of high-energy electrons, where the energy E acquired by the atomic nucleus may be tens of eV and sufficient to displace it from its usual site in a crystalline material, producing an effect referred to as knock-on or displacement damage.

4.2 Electron-Electron Scattering

If the incident electron passes close to one of the atomic electrons that surround the nucleus, both particles experience a *repulsive* force. If we ignore the presence of the nucleus and the other atomic electrons, we again have a two-body encounter (Fig. 4.3b) similar to the nuclear interaction depicted in Figs. 4.2 and 4.3a. Equations (4.1)–(4.6) should still apply, provided we replace the nuclear mass M by the mass m of an atomic electron (identical to the mass of the incident electron, since we are neglecting relativistic effects). In this case, Eq. (4.6) becomes:

$$E/E_0 = 1 + v_1^2/v_0^2 - 2\,(v_1/v_0)\,\cos\theta \tag{4.9}$$

Because Eq. (4.9) no longer contains the factor (m/M), the energy loss E can be much larger than for an elastic collision and the scattering is described as *inelastic*. In the extreme case of a head-on collision, $v_1 \approx 0$ and $E \approx E_0$: the incident electron loses *all* of its original kinetic energy. But for most inelastic collisions, the scattering angle is small and E is in the range 10–50 eV. A more accurate treatment of inelastic scattering is based on the wave-mechanical theory and includes the *collective* response of many electrons, including those in *nearby* atoms.

In practice, an electron traveling through a solid will experience both repulsive forces (from other electrons) and attractive forces (from atomic nuclei). But in the scattering theory, it turns out to be possible to divide the scattering into elastic and inelastic components that are to a first approximation independent.

4.3 The Dynamics of Scattering

Since the force acting on a primary electron is electrostatic, its magnitude F is described by Coulomb's law. For *elastic* scattering of an electron by the electrostatic field of a single nucleus:

$$F = K(e)(Ze)/r^2 \tag{4.10}$$

where $K = 1/(4\pi\epsilon_0) = 9.0 \times 10^9\,\mathrm{Nm^2C^{-2}}$ is the Coulomb constant. Applying Newtonian mechanics ($F = ma$) to this problem, the electron trajectory can be shown to be a hyperbola, as depicted in Fig. 4.3a.

The **angular distribution** of elastic scattering can be expressed as the probability $P(\theta)$ of an electron being scattered through a given angle θ. But to understand TEM contrast, we are more interested in the probability $P(>\alpha)$ that an electron is scattered through any angle that *exceeds* the semi-angle α of the objective aperture. Such an electron will be absorbed within the diaphragm material that surrounds the aperture, resulting in a reduction in intensity within the TEM image.

We can think of the nucleus as presenting (to each incident electron) a *target* of area σ, known as a**scattering cross section**. Atoms being round, this target takes the form of a *disk* of radius a and area $\sigma = \pi a^2$. For an electron to be scattered through an angle θ that *exceeds* the aperture semi-angle α, it must pass through the target area σ, at a distance *less* than a from the nucleus. But relative to any nucleus, electrons arrive *randomly* with various values of the **impact parameter** b, defined as the distance of closest approach if we *neglect* the hyperbolic curvature of the trajectory, as shown in Fig. 4.4. If $b < a$, the electron is scattered through an angle greater than α and is later absorbed by the objective diaphragm.

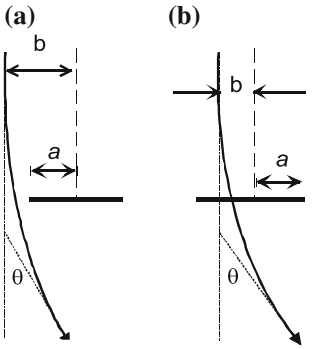


Fig. 4.4 Electron trajectory for elastic scattering through an angle θ which is **a** less than and **b** greater than the objective-aperture semi-angle α. The impact parameter b is defined as the distance of closest approach to the nucleus if the particle were to continue in a straight-line trajectory. It is approximately *equal* to the distance of closest approach when the scattering angle (and the curvature of the trajectory) is small

By *how much* the scattering angle θ exceeds α depends on just how close the electron gets to the center of the atom: small b leads to large θ and vice versa. Algebraic analysis of the force and velocity components, making use of Newton's second law of motion (see Appendix A.2), gives:

$$\theta \approx K\,Z\,e^2/(E_0\,b) \tag{4.11}$$

As expected, Eq. (4.11) specifies an *inverse* relation between b and θ; electrons with smaller impact parameter pass closer to the nucleus and experience a stronger attractive force. Because Eq. (4.11) is a general relationship, it must hold for θ = α, which corresponds to $b = a$. Therefore we can rewrite Eq. (4.11) for this particular case, giving:

$$\alpha \approx KZ\,e^2/(E_0\,a) \tag{4.12}$$

As a result, the cross-section for *elastic* scattering of an electron through any angle greater than α can be written as:

$$\sigma_e = \pi\,a^2 \approx \pi\left[KZe^2/(aE_0)\right]^2 = Z^2e^4/\left(16\pi\,\epsilon_0^2\,E_0^2\,\alpha^2\right) \tag{4.13}$$

Because σ_e has units of m^2, it cannot directly represent scattering probability; we need an additional factor with units of m^{-2} to give the dimensionless number $P_e(>\alpha)$. Also, our TEM specimen contains *many* atoms, each capable of scattering an incoming electron, whereas σ_e is the elastic cross-section for a single atom. Consequently, the total probability of elastic scattering in the specimen is:

$$P_e(>\alpha) = N\,\sigma_e \tag{4.14}$$

where N is the number of atoms *per unit area* of the specimen (viewed in the direction of an approaching electron), sometimes called an **areal density** of atoms.

For a specimen with n atoms *per unit volume*, $N = nt$ where t is the specimen thickness. If the specimen contains only a single element of atomic number A, the atomic density n can be written in terms of the physical density: ρ = (mass/volume) = (atoms per unit volume) (mass/atom) = $n(uA)$, where u is the atomic mass unit (1.66×10^{-27} kg). Therefore:

$$P_e(>\alpha) = [\rho/(uA)]\,t\,\sigma = (\rho\,t)\left(Z^2/A\right)e^4/\left(16\pi\epsilon_0^2\,u\,E_0^2\,\alpha^2\right) \tag{4.15}$$

Equation (4.15) indicates that the number of electrons absorbed at the angle-limiting (objective) diaphragm is proportional to the**mass-thickness** (ρt) of the specimen and *inversely* proportional to the square of the aperture size.

Unfortunately, Eq. (4.15) fails for the important case of small α; our formula shows $P_e(>\alpha)$ increasing towards infinity as the aperture angle is reduced to zero. Since any probability must be less than 1, this prediction indicates that at least one of the assumptions used in our analysis is incorrect. By using Eq. (4.10) to describe

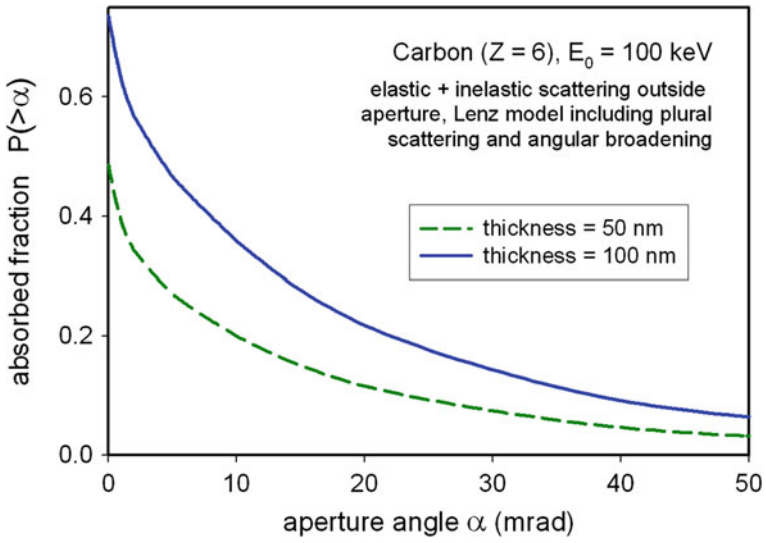


Fig. 4.5 Probability $P(>\alpha)$ of scattering of 100 keV electrons in carbon films of different thickness, as predicted by wave-mechanical theory that includes screening of the nuclear field. The effects of plural elastic and inelastic scattering are also included

the electrostatic force, we have assumed that the electrostatic field of the nucleus extends to infinity (although diminishing with distance). In practice, this field is *terminated* within a neutral atom, due to the presence of the atomic electrons that surround the nucleus. Stated another way, the electrostatic field of each nucleus is **screened** by the atomic electrons, for locations outside the atom. Including such screening results in a scattering probability that rises asymptotically to a *finite* value $P_e(>0)$ as α falls to zero, $P_e(>0)$ being the probability of elastic scattering through *any* angle; see Fig. 4.5.

Even after correction for screening, Eq. (4.15) may still give rise to an unphysical result, since if we increase the specimen thickness t sufficiently, $P_e(>\alpha)$ becomes greater than 1. This situation arises because Eq. (4.15) is a **single-scattering approximation**: we have calculated the fraction of electrons that are scattered only *once* in a specimen. For very thin specimens, this fraction increases in proportion to the specimen thickness, as implied by Eq. (4.15). But in thick specimens, the probability of single scattering must *decrease* with increasing thickness because most electrons become scattered *several* times. If we use Poisson statistics to calculate the probability P_n of n-fold scattering, each probability comes out less than 1, as expected. In fact, the sum ΣP_n over all n (including $n = 0$) is equal to 1 because all electrons are transmitted, except for specimens that are far too thick for transmission microscopy.

The above arguments could be repeated for the case of *inelastic* scattering, taking the electrostatic force as $F = K(e)(e)/r^2$ since the incident electron is now scattered by an atomic electron rather than by the nucleus. The result is an equation

similar to Eq. (4.15) but with the factor Z^2 missing. However, this result would apply to inelastic scattering by only a *single* atomic electron. Considering all Z electrons within the atom, the inelastic-scattering probability must be multiplied by Z, giving:

$$P_{\mathrm{i}}(>\alpha) = (\rho\, t)\,(Z/A)e^4/\left(16\pi\varepsilon_0^2\, u\, E_0^2\, \alpha^2\right) \tag{4.16}$$

Comparison of Eq. (4.16) with Eq. (4.15) shows that the amount of inelastic scattering, *relative* to elastic scattering, is

$$P_i(>\alpha)\,/\,P_{\mathrm{e}}(>\alpha) = 1\,/\,Z \tag{4.17}$$

Equation (4.17) predicts that inelastic scattering makes a relatively small contribution for most elements. However, it is based on Eq. (4.15), which is accurate only for larger aperture angles. For small α, where screening of the nuclear field reduces the amount of elastic scattering, the inelastic/elastic ratio is much higher. More accurate theory (treating the incoming electrons as waves) and experimental measurements of the scattering show that $P_{\mathrm{e}}(>0)$ is typically proportional to $Z^{1.5}$ (rather than Z^2), that $P_{\mathrm{i}}(>0)$ is proportional to $Z^{0.5}$ (rather than Z), and that, considering scattering through all angles,

$$P_{\mathrm{i}}(>0)/P_{\mathrm{e}}(>0) \approx 20\,/\,Z \tag{4.18}$$

Consequently, inelastic scattering does make a significant contribution to the total scattering (through any angle) in the case of light (low-Z) elements.

Although our classical-physics analysis turns out to be rather inaccurate in predicting *absolute* scattering probabilities, the thickness and material parameters involved in Eq. (4.15) and Eq. (4.16) do provide a reasonable explanation for the contrast (variation in intensity level) in TEM images obtained from non-crystalline (amorphous) specimens such as glasses, certain kinds of polymers, amorphous thin films, and most types of biological material.

4.4 Scattering Contrast from Amorphous Specimens

Most TEM images are viewed and recorded with an objective aperture (diameter D) inserted and centered about the optic axis of the TEM objective lens (focal length f). As shown by Eq. (3.9), this aperture absorbs electrons that are scattered through an angle greater than $\alpha \approx 0.5D/f$. However, any part of the *field* of view that contains *no* specimen (such as a hole or a region beyond the specimen edge) is formed from electrons that remain unscattered, so that part appears *bright* relative to the specimen. As a result, this central-aperture image is referred to as a **bright-field** image.

Biological tissue, at least in its dry state, is mainly carbon and so for this common type of specimen we can take $Z = 6$ and $A = 12$ but $\rho \approx 1$ g/cm^2 = 1000 kg/m^3. For

a biological TEM, typical parameters are: E_0 = 100 keV = 1.6×10^{-14} J and $\alpha \approx 0.5D/f$ = 10 mrad = 0.01 rad, taking an objective-lens focal length f = 2 mm and objective-aperture diameter D = 40 μm. With these values, Eq. (4.15) gives $P_e(>\alpha) \approx 0.47$ for a specimen thickness of t = 40 nm. The same parameters inserted into Eq. (4.16) give $P_i(>\alpha) \approx 0.08$ and the total fraction of electrons absorbed by the objective diaphragm is: $P(>10 \text{ mrad}) \approx 0.47 + 0.08 = 0.55$.

We therefore predict that over half of the transmitted electrons are intercepted by a typical-size objective aperture, even for a very thin (40 nm) specimen. This fraction might be even larger for a thicker specimen, but our single-scattering approximation would not be valid. In practice, the specimen thickness should be less than about 200 nm, assuming an accelerating potential of 100 kV and a specimen consisting mainly of low-Z elements. If the specimen is appreciably thicker, only a small fraction of the transmitted electrons pass through the objective aperture and the bright-field image appears very dim on the TEM screen.

Since (A/Z) is approximately the same (≈2) for all elements, Eq. (4.16) indicates that $P_i(>\alpha)$ increases only slowly with increasing atomic number, due to the density term ρ which tends to increase with increasing Z. But using the same argument, Eq. (4.15) implies that $P_e(>\alpha)$ is approximately proportional to ρZ. Therefore specimens that contain mainly heavy elements scatter electrons more strongly and would have to be even thinner, placing unrealistic demands on the specimen preparation. Such specimens are usually examined in a "materials science" TEM that employs an accelerating voltage of 200 kV or higher, taking advantage of the reduction in σ_e and P_e with increasing E_0; see Eqs. (4.13) and (4.15).

For imaging non-crystalline specimens, the main purpose of the objective aperture is to provide **scattering contrast** in the TEM image of a specimen whose composition or thickness varies between different regions. Thicker regions of the sample scatter a larger fraction of the incident electrons, many of which are absorbed by the objective diaphragm, so the *corresponding* regions in the image appear *dark*, giving rise to **thickness contrast** in the image. Regions of higher atomic number also appear dark relative to their surroundings, due mainly to an increase in the amount of the *elastic* scattering described by Eq. (4.15), giving **atomic-number contrast** (Z-contrast). Taken together, these two effects are often described as **mass-thickness contrast**. They provide the information content of TEM images of amorphous materials, in which the atoms are arranged more-or-less randomly and not in a regular array, as in a crystal.

Stained Biological Tissue

TEM specimens of biological (animal or plant) tissue are made by cutting very thin slices (sections) from a small block of *embedded* tissue (held together by epoxy glue) using an instrument called an **ultramicrotome** that employs a glass or diamond knife as the cutting blade. To prevent the sections from curling up, they are floated onto a water surface, which supports them evenly by surface tension. A fine-mesh copper grid (3 mm diameter, held at its edge by tweezers) is then introduced below the water surface and slowly raised, leaving the tissue section

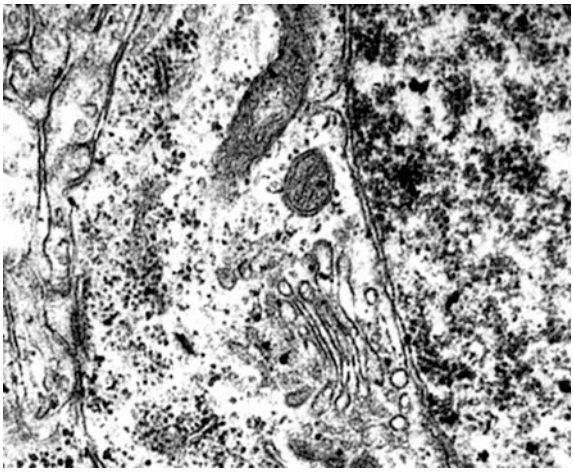


Fig. 4.6 Small area (about 4 μm × 3 μm) of visual cortex tissue stained with uranyl acetate and lead citrate. Cell membranes and components (organelles) within a cell appear dark due to elastic scattering and subsequent absorption of scattered electrons at the objective aperture

supported by the grid. After drying in air, the tissue remains attached to the grid by local mechanical and chemical forces.

Tissue sections prepared in this way are fairly uniform in thickness, so almost no contrast arises from the thickness term in Eq. (4.15). Their atomic number also remains approximately constant ($Z \approx 6$ for dry tissue), therefore the overall contrast is very low and the specimen appears featureless in the TEM. To produce scattering contrast, the sample is chemically treated by a process called **staining**. Before or after slicing, the tissue is immersed in a solution that contains a heavy (high-Z) metal. The solution is absorbed *non-uniformly* by the tissue; a **positive** stain, such as lead citrate or uranyl acetate, tends to migrate to structural features (organelles) within each cell.

As illustrated in Fig. 4.6, these regions appear *dark* in the TEM image because Pb or U atoms strongly scatter the incident electrons and most of the scattered electrons are absorbed by the objective diaphragm. A **negative** stain (such as phosphotungstic acid) tends to *avoid* cellular structures, which in the TEM image appear bright relative to their surroundings, since they contain fewer tungsten atoms.

Surface replicas

Without making a very thin specimen, it is possible to obtain TEM images that reflect a material's surface features. For example, grain boundaries separating small crystalline regions of a polycrystalline metal may form a groove at the external surface, as discussed in Chap. 1. These grooves are made more prominent by treating the material with a chemical etch that preferentially attacks grain-boundary regions where the atoms are bonded to fewer neighbors. The surface is then coated with a very thin layer of a plastic (polymer) or amorphous carbon, which fills the grooves; see Fig. 4.7a. This surface film is stripped off (by dissolving the metal, for example), mounted on a 3 mm-diameter grid, and examined in the TEM. The image

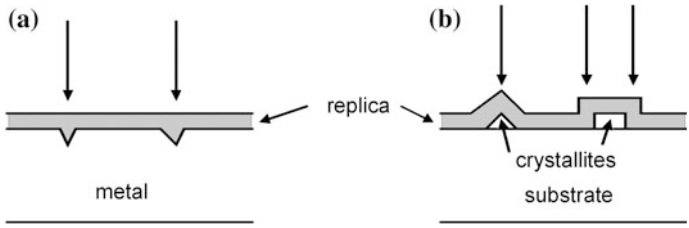


Fig. 4.7 Plastic or carbon replica coating the surface of **a** a metal containing grain-boundary grooves and **b** a flat substrate containing raised crystallites. *Vertical arrows* indicate electrons that travel through a greater thickness of the replica material, and therefore have a greater probability of being scattered and intercepted by an objective diaphragm

of such a **surface replica** appears dark in the region of the surface grooves, due to the increase in thickness and the additional electron scattering.

Alternatively, the original surface might contain raised features. For example, when a solid is heated and evaporates (or sublimes) in vacuum onto a flat substrate, the thin-film coating initially consists of isolated crystallites (very small crystals) attached to the substrate. If the surface is coated with carbon (also by vacuum sublimation) and the surface replica is subsequently detached, it displays thickness contrast when viewed in a TEM.

In many cases, **thickness-gradient contrast** is obtained: through diffusion, the carbon acquires the same thickness (measured perpendicular to the local surface) but its *projected* thickness (in the direction of the electron beam) is greater at the edges of protruding features (Fig. 4.7b). These features therefore appear dark *in outline* in the scattering-contrast image of the replica; see Fig. 4.8.

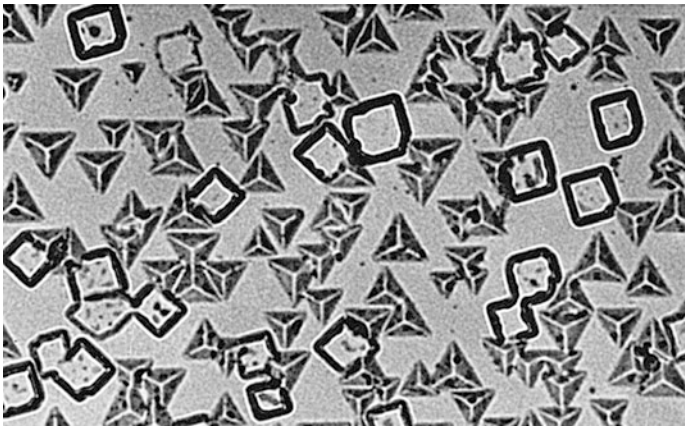


Fig. 4.8 Bright-field TEM image of carbon replica of lead selenide (PbSe) crystallites (each about 0.1 μm across) deposited in vacuum onto a flat mica substrate

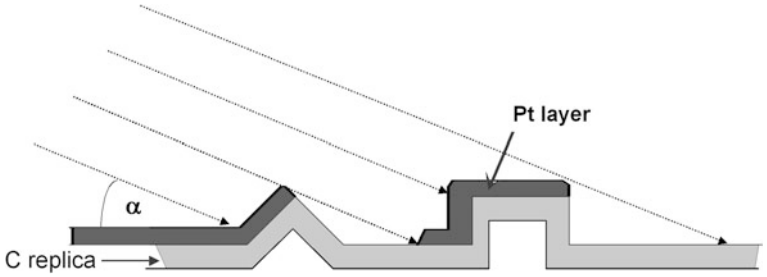


Fig. 4.9 Shadowing of a surface replica with platinum atoms deposited at an angle α relative to the surface

Because most surface replicas consist mainly of carbon, which has a relatively low scattering cross-section, they tend to give low image contrast in the TEM. Therefore the contrast is often increased using a process known as **shadowing**. A heavy metal such as platinum is evaporated (in vacuum) at an oblique angle of incidence onto the replica, as shown in Fig. 4.9. After landing on the replica, platinum atoms are (to a first approximation) immobile, therefore raised features present in the replica cast "shadows" within which platinum is *absent*. When viewed in the TEM, the shadowed replica shows strong atomic-number contrast. Relative to their surroundings, the shadowed areas appear *bright* on the TEM screen, as seen in Fig. 4.10. The contrast can easily be reversed in an electronically recorded image, resulting in a realistic three-dimensional appearance, similar to that of oblique illumination of a rough surface by light.

Besides increasing contrast, shadowing allows the height h of a protruding surface feature to be estimated from its shadow length L, given (see Fig. 4.9) by:

$$h = L \tan \alpha \tag{4.19}$$

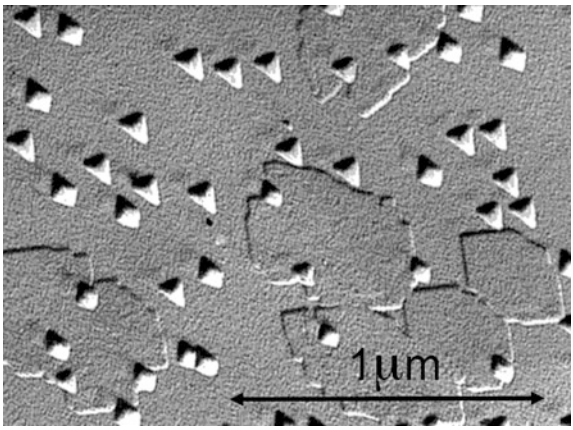


Fig. 4.10 Bright-field image ($M \approx 40.00$) of a platinum-shadowed carbon replica of PbSe crystallites, as it appears on the TEM screen. Compare the contrast with that of Fig. 4.8

Therefore h can be measured if the shadowing is carried out at a known angle of incidence α. Shadowing has also been used to ensure the visibility of small objects such as virus particles or DNA molecules, mounted on a thin-carbon support film.

4.5 Diffraction Contrast from Polycrystalline Specimens

Many inorganic materials, such as metals and ceramics (metal oxides), are polycrystalline: they contain small crystals (**crystallites**, also known as **grains**) within which the atoms are arranged regularly in a **lattice** of rows and columns. However, the direction of atomic alignment varies from one crystallite to the next; the crystallites are separated by grain boundaries where this change in orientation occurs rather abruptly (within a few atoms).

TEM specimens of a polycrystalline material can be prepared by mechanical, chemical, or electrochemical means; see Sect. 4.10. Individual grains then appear with different intensities in the TEM image, implying that they scatter electrons (beyond the objective aperture) to a different extent. Since this intensity variation occurs even for specimens that are uniform in thickness and composition, there must be another factor, besides those appearing in Eq. (4.15), that determines the angular distribution of electron scattering within a *crystalline* material. This additional factor is the *orientation* of the atomic rows and columns relative to the incident electron beam. Because the atoms in a crystal can also be thought as arranged in an orderly fashion on equally-spaced **atomic planes**, we can specify the orientation of the beam relative to these planes. To understand why this orientation matters, we need to abandon our particle description of the incident electrons and consider them as de Broglie (matter) waves.

A useful comparison is with x-rays, which are diffracted by the atoms in a crystal. In fact, interatomic spacings are usually *measured* by recording the diffraction of *hard* x-rays, whose wavelength is comparable to the atomic spacing. The simplest way of understanding x-ray diffraction is in terms of **Bragg reflection** from atomic planes. Reflection implies that the angles of incidence and reflection are equal, as with light reflected from a mirror. But whereas a mirror reflects light incident in any direction, Bragg reflection occurs only when the angle of incidence (here measured between the incident direction and the planes) is equal to a Bragg angle θ_B that satisfies **Bragg's law**:

$$n\lambda = 2d \sin\theta_B \tag{4.20}$$

Here, λ is the x-ray wavelength and d is the spacing between atomic planes, measured in a direction perpendicular to the planes; n is an integer that represents the *order* of reflection, as in the case of light diffracted from an optical diffraction grating. But whereas diffraction from a grating takes place at its surface, x-rays penetrate through many planes of atoms and diffraction occurs within a certain *volume* of the crystal, which acts as a kind of *three-dimensional* diffraction grating.

In accord with this concept, the diffraction condition, Eq. (4.20), involves a spacing d measured *between* diffracting planes rather than *within* a surface plane (as in the case of a diffraction grating).

The fast-moving electrons used in a *transmission* electron microscope also penetrate through many planes of atoms and are diffracted within crystalline regions of a solid, just like x-rays. However, their wavelength (3.7 pm for $E_0 \approx 100$ keV) is far below a typical atomic-plane spacing (≈0.3 nm) so the Bragg angles are *small*, as required by Eq. (4.20) when $\lambda << d$. The integer n in Eq. (4.20) is usually taken as one, the n'th order diffraction from planes of spacing d being described as first-order diffraction from planes of spacing d/n. Using the small-angle approximation, Eq. (4.20) can therefore be rewritten as:

$$\lambda \approx 2\theta_B d = \theta\, d \tag{4.21}$$

where $\theta = 2\theta_B$ is the angle of scattering (deflection angle) of the electron resulting from the diffraction process; see Fig. 4.11a.

For a few particular orientations of a crystallite relative to the incident beam, Eq. (4.21) will be satisfied and a crystallite will strongly diffract the incident electrons. Provided the corresponding deflection angle θ exceeds the semi-angle α of the objective aperture, the diffracted electrons will be absorbed by the objective diaphragm and the crystallite will appear dark in the TEM image. Crystallites whose atomic-plane orientations do *not* satisfy Eq. (4.21) will appear bright, since most electrons passing through them will remain *undiffracted* (undeviated) and will pass through the objective aperture. The grain structure of a polycrystalline material can therefore be seen in a TEM image as an intensity variation between different

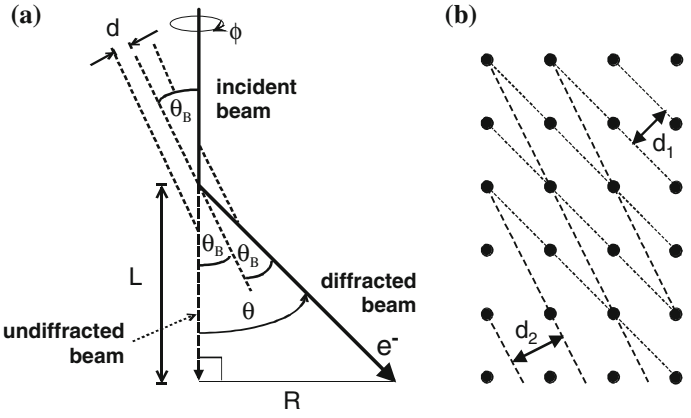


Fig. 4.11 **a** Geometry of x-ray or fast-electron diffraction from atomic planes, which are shown as parallel *dashed lines*; R is the radius of the diffraction ring at the recording plane (not to scale). **b** Schematic cross-section through a cubic lattice with an atom at each lattice point. The *diagonal lines* represent two sets of atomic planes (actually, their intersection with the plane of the diagram) whose spacings are d_1 and d_2

regions; see Fig. 4.12a or Fig. 4.21. A TEM image is said to display **diffraction contrast** in any situation in which the intensity variations arise from differences in diffracting conditions between different regions of the specimen.

4.6 Dark-Field Images

Instead of selecting the undiffracted beam to form an image, we could horizontally displace the objective aperture so that it admits diffracted electrons. Strongly diffracting regions of specimen would then appear *bright* relative to their surroundings, resulting in a **dark-field image** where any part of the field of view that contains *no* specimen appears dark.

Figure 4.12b shows a dark-field image and illustrates the fact that its contrast is *inverted* relative to a bright-field image of the same region of specimen (see also Fig. 4.18b, c). If it were possible to collect *all* of the scattered electrons to form the dark-field image, the two images would be exactly **complementary**: the sum of their intensities would be constant (zero contrast) because every electron has been recorded and practically none are absorbed in a thin specimen.

Unless stated otherwise, TEM micrographs are usually *bright-field* images created with an objective aperture centered about the optic axis. However, one advantage of a dark-field image is that, if the displacement of the objective aperture has been calibrated, the atomic-plane spacing (d) of a crystalline precipitate can be calculated and sometimes used to identify the material.

With the procedure just described, the image has inferior resolution compared to a bright-field image because electrons forming the image pass though the imaging lenses at larger angles relative to the optic axis, giving rise to increased spherical and chromatic aberration. However, such loss of resolution can be avoided by keeping the objective aperture on-axis and re-orienting the electron beam arriving at the specimen, using the deflection-tilt coils that are built into the illumination system of a typical TEM.

4.7 Electron-Diffraction Patterns

Diffraction is the *elastic* scattering of electrons (deflection by the Coulomb field of atomic nuclei) in a *crystalline* material. The regularity of the spacing of the atomic nuclei results in a *redistribution* of the angular dependence of scattering probability, relative to that obtained from a single atom (discussed in Sect. 4.3). Instead of a *continuous* angular distribution, as depicted in Fig. 4.5, the scattering is concentrated into sharp peaks, known as Bragg peaks; see Fig. 4.12d. These peaks correspond to scattering angles that are *twice* the Bragg angle θ_B for atomic planes of a particular orientation: the sum of incident and reflection angles, as illustrated in Fig. 4.11a.

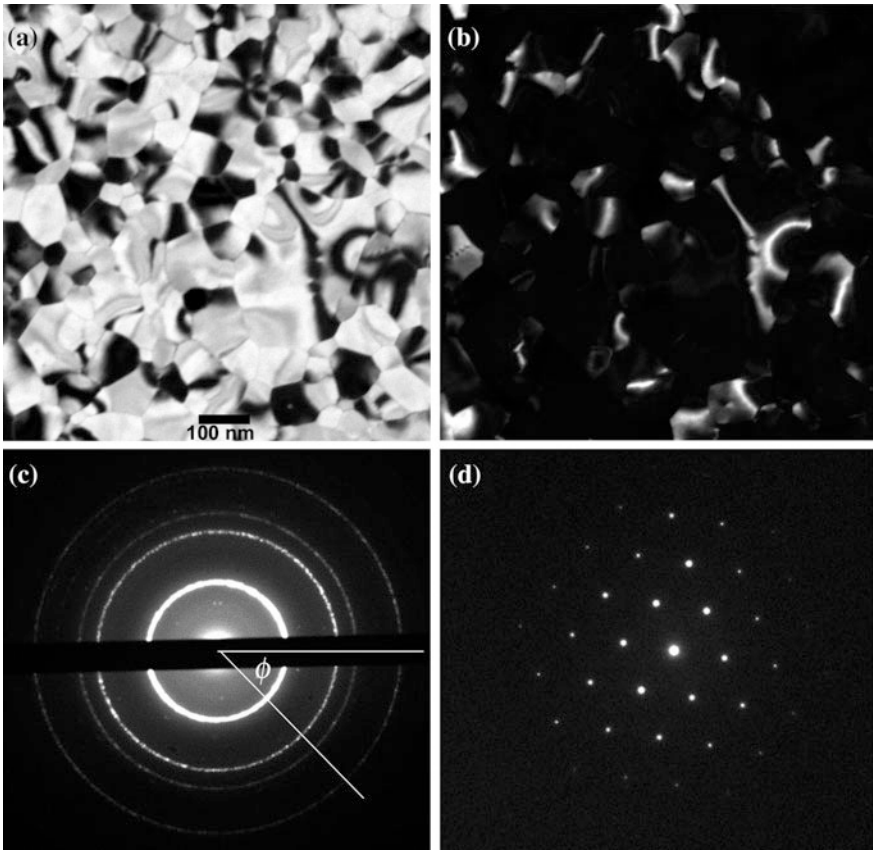


Fig. 4.12 **a** Bright-field TEM image of a polycrystalline thin film of bismuth, including bend contours that appear dark. **b** Dark-field image of the same area; bend contours appear bright. **c** Ring diffraction pattern obtained from many crystallites; for recording purposes, the central (0,0,0) diffraction spot has been masked by a wire. **d** Spot diffraction pattern recorded from a single Bi crystallite, whose trigonal crystal axis was parallel to the incident beam. Courtesy of Marek Malac, National Institute of Nanotechnology, Canada

This angular distribution of scattering can be displayed on the TEM screen by weakening the intermediate lens, so that the intermediate and projector lenses magnify the small diffraction pattern formed at the back-focal plane of the objective lens. Figure 4.12c is typical of the pattern obtained from a polycrystalline material. It consists of concentric rings, centered on a bright central spot that represents the *undiffracted* electrons. Each ring corresponds to atomic planes of different orientation and different interplanar spacing d, as illustrated in Fig. 4.11b.

Close examination of the rings reveals that they consist of a large number of spots, each arising from Bragg reflection from an individual crystallite.

Equation (4.21) predicts only the *radial* angle ($\theta = 2\theta_B$) that determines the distance of a spot from the center of the diffraction pattern. The azimuthal angle ϕ of a spot is determined by the azimuthal orientation of a crystallite about the incident-beam direction (the optic axis). Since the grains in a polycrystalline specimen are randomly oriented, diffraction spots occur at all azimuthal angles and give the appearance of continuous rings if many grains lie within the path of the electron beam (grain size $<<$ beam diameter at the specimen).

The radius R of a given diffraction ring depends not only on the scattering angle θ but also on how much the TEM magnifies the initial diffraction pattern formed at the objective back-focal plane. If there were no imaging lenses and a diffraction pattern was recorded at a distance L from the specimen, simple geometry (Fig. 4.11a) gives:

$$R = L \tan\theta \approx L\theta \tag{4.22}$$

In the TEM, however, the relationship between R and θ depends on the excitation of the imaging lenses. We can still use Eq. (4.22) to represent this relationship but L no longer represents a real physical distance. Instead, L is a measure of the *scale* of the diffraction pattern and is known as the **camera length**. Its value can be changed (typically over the range 10–150 cm) by varying the excitation current of an intermediate lens. In a modern TEM, the camera length is displayed on the control panel whenever diffraction mode is selected.

Knowing L and (from the accelerating voltage V_0) the electron wavelength λ, the d-spacing of each set of atomic planes that contribute to the diffraction pattern can be calculated using Eqs. (4.21) and (4.22). The nearest-neighbor spacing of atoms in the specimen is, in general, not equal to d; it depends also on the arrangement of the atoms within the **unit cell** (the smallest repeat unit) of the crystal. Many metals, such as copper and aluminum, have a **face-centered cubic** (fcc) unit cell, depicted in Fig. 4.13, for which:

$$d = a/\left(h^2 + k^2 + l^2\right)^{1/2} \tag{4.23}$$

Here, a is the **lattice parameter** (dimension of the unit cell) and h, k, and l are integers (known as **Miller indices**) that describe the orientation of planes of spacing d, relative to the x, y, and z crystal axes, which are parallel to the edges of the unit cell. From Eq. (4.23), larger Miller indices correspond to smaller d and, according to Eqs. (4.21) and (4.22), to *larger* θ_B and R. So *higher-order* diffraction spots lie further from the center of the diffraction pattern.

For a material whose lattice parameter is known (for example, from x-ray diffraction), the expected d-spacings can be calculated, using the additional condition that (for the fcc structure) h, k, and l must be *all odd or all even*, otherwise the electron interference is destructive rather than constructive. Therefore the first few rings in the diffraction pattern of a polycrystalline fcc metal correspond to

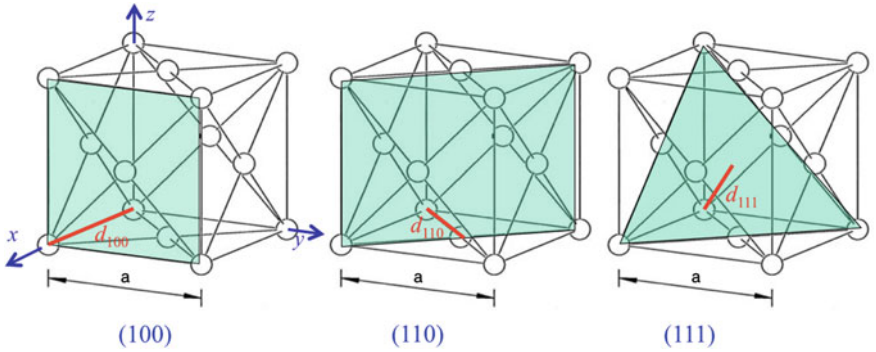


Fig. 4.13 Unit cell for a face-centered cubic crystal: a cube of side a with atoms located at each corner and at the center of each face. The (100), (110), and (111) places are shown shaded in green. Red lines show the perpendicular distance d_{hkl} of each plane from the origin of the unit cell, which is also the d-spacing of the plane since a plane of the same type passes through the origin

$$(hkl) = (1\,1\,1),\ (2\,0\,0),\ (2\,2\,0),\ (3\,1\,1),\ (2\,2\,2), \ldots \tag{4.24a}$$

Then Eq. (4.23) gives $(a/d)^2$ = 3, 4, 8, 11, 12 … and knowing a, the d-spacing for each ring can be calculated and its electron-scattering angle θ obtained from Eq. (4.21). A value of L can be deduced for each ring using Eq. (4.22) and these individual values averaged to give a more accurate value of the camera length. Recording the diffraction pattern of a material with *known* structure and lattice parameter therefore provides a way of *calibrating* the camera length, for a given intermediate-lens setting.

Knowing L, the use of Eq. (4.22) and then Eq. (4.21) can provide a list of the d-spacings represented in the electron-diffraction pattern recorded from an *unknown* material. In favorable cases, these interplanar spacings can then be matched to the d-values calculated for different crystal structures, taking h, k, and l from rules similar to Eq. (4.24) but different for each structure, and with a as an adjustable parameter. For example, another common crystal structure (that of transition metals such as Cr, Ti, and V) is the **body-centered cubic** (bcc) structure, whose unit cell is a cube with eight corner atoms and a single atom located at the geometrical center of the cube. Here, constructive interference (diffraction) occurs only if $h + k + l$ is an even number, leading to the selection rule:

$$(h\,k\,l) = (1\,1\,0),\ (2\,0\,0),\ (2\,1\,1),\ (2\,2\,0),\ (3\,1\,0), \ldots \tag{4.24b}$$

which gives $(a/d)^2$ = 2, 4, 6, 8, 10 … For small θ, Eqs. (4.21)–(4.23) predict a/d proportional to R, so a visual comparison with measured ring radii can distinguish between fcc and bcc materials. If L and λ are known, the lattice parameter a can be calculated and (by comparison with tabulated values) can sometimes be used to identify the diffracting material.

Such diffraction analysis can be extended to non-cubic materials but the procedure is more cumbersome because the x, y, z axes of the unit cell may not be orthogonal and Eq. (4.23) must be replaced by a more complicated formula. In any event, electron diffraction in the TEM is of limited accuracy: the value of L is sensitive to small differences in specimen height (relative to the objective-lens polepieces) and to magnetic hysteresis of the TEM lenses, such that a given lens current does not correspond to a unique magnetic field or focusing power. As a result, d-spacings can only be measured to about 1 %, compared with a few parts per million in the case of x-ray diffraction, where no lenses are used. However, x-ray diffraction requires a macroscopic specimen, with a volume approaching 1 mm^3. In contrast, electron-diffraction data can be obtained from microscopic volumes (<1 μm^3) by use of a thin specimen and an area-selecting aperture or by focusing the incident beam into a very small probe. Electron diffraction can be used to identify sub-μm precipitates in metal alloys, for example.

A few materials such as silicon can be fabricated as **single-crystal** material, without grain boundaries. Others have a crystallite size which is sufficiently large that a focused electron beam passes through only a single grain. In these cases, there is *no* azimuthal randomization of the diffraction spots: the diffraction pattern is a **spot pattern**, as illustrated in Fig. 4.12d. Equation (4.22) still applies to each spot, R being its distance from the central (0 0 0) spot. In this case, the *symmetry* of the diffraction spots is related to the symmetry of arrangement of atoms in the crystal. For example, a cubic crystal produces a *square* array of diffracted spots, *provided* the incident beam travels down one of the crystal axes (parallel to the edges of the unit cell). Using a double-tilt specimen holder, a specimen can often be oriented to give a recognizable diffraction pattern, which may help to identify the phase and/or chemical composition of small crystalline regions within it.

4.8 Diffraction Contrast from a Single Crystal

Close examination of the TEM image of a polycrystalline specimen shows there can be a variation of electron intensity *within* each crystallite. This diffraction contrast arises either from atomic-scale defects within the crystal *or* from the crystalline nature of the material itself, combined with the wave nature of the transmitted electrons.

Defects in a crystal, where the atoms are displaced from the their regular lattice-site positions, can take several different forms. A **dislocation** is an example of a *line* defect. It represents either the termination of an atomic plane within the crystal (edge dislocation) or the axis of a "spiral staircase" where the atomic planes are displaced in the direction of the dislocation line (screw dislocation). The structure of an edge dislocation is depicted in Fig. 4.14, which shows how planes of atoms bend in the vicinity of the dislocation core to accommodate the extra "half-plane" of atoms. Since the Bragg angles for electron diffraction are small, this bending may cause the angle between an atomic plane and the incident beam to become approximately equal to the Bragg angle θ_B at some locations within this

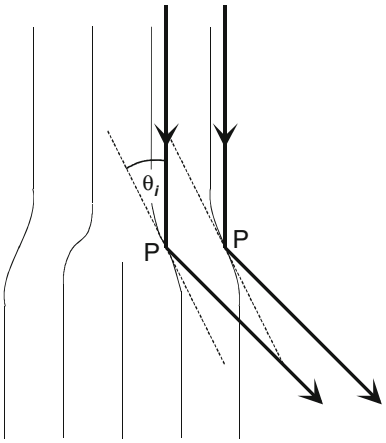


Fig. 4.14 Bending of the atomic planes around an edge dislocation, whose direction runs perpendicular to the plane of the diagram. At points P, the angle of incidence θ_i is equal to the Bragg angle of the incident electrons, which are therefore strongly diffracted

strained region. At such locations, electrons are strongly diffracted and most of these scattered electrons are absorbed at the TEM objective aperture. Consequently, the dislocation appears dark in the TEM image, as shown in Fig. 4.15.

Because they are mobile, dislocations play an important role in the deformation of metals under applied mechanical stress. Although they had been predicted to explain the discrepancy between the measured and theoretical strengths of metals,

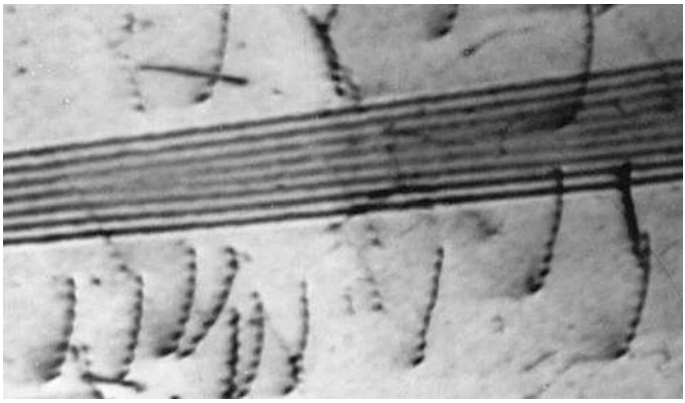


Fig. 4.15 Dislocations and a stacking fault in cobalt metal. The dislocations appear as curved dark lines running through the crystal, whereas the stacking fault is represented by a series of almost-parallel interference fringes. (The specimen thickness is slightly less on the left of the picture, resulting in a reduction in the projected width of the stacking fault.) Courtesy of Prof. S. Sheinin, University of Alberta

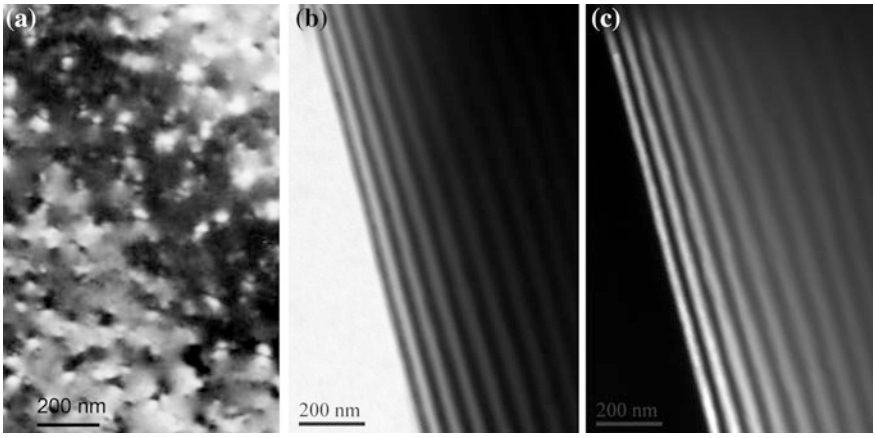


Fig. 4.16 **a** Diffraction-contrast image of point-defect clusters produced in graphite by irradiation with 200 keV electrons at an elevated temperature. A bend contour, visible as a broad dark band, runs diagonally across the image. **b** Bright-field and **c** dark-field images of thickness fringes running parallel to the edge of a wedge-shaped silicon specimen (wedge angle = 55°), recorded with 300 keV electrons (courtesy of Dr. M Malac and NINT)

TEM images provided the first direct evidence for the existence and morphology (shape) of dislocations; see Fig. 1.11.

Crystals can also contain *point* defects, the simplest example being an **atomic vacancy**: a missing atom. Lattice planes become distorted in the immediate vicinity of the vacancy because atoms are pulled towards its center. The bending is small, so single vacancies are not usually visible in a TEM image. However, the local strain produced by a *cluster* of vacancies (a small void) gives rise to a characteristic bright/dark diffraction contrast; see Fig. 4.16a.

One example of a *planar* defect is the **stacking fault**, a plane within the crystal structure where atomic planes are abruptly displaced by a fraction of the interatomic spacing. In a TEM image, the stacking fault gives rise to a series of narrowly spaced fringes (see Fig. 4.15) whose explanation requires a Bloch-wave treatment of the electrons, as outlined below.

Other features in diffraction-contrast images originate from the fact that the crystal can become slightly bent as a result of internal strain. This strain may arise from the crystal-growth process or from subsequent mechanical deformation. As a result, the orientation of the lattice planes gradually changes across a TEM specimen and at some locations the Bragg equation is satisfied, resulting in strong diffraction and occurrence of a **bend contour** as a dark band in the bright-field image, as in Fig. 4.16a. Unlike our previous examples, the bend contour does *not* represent a *localized* feature present in the specimen. If the specimen or the TEM illumination is tilted slightly, relative to the original condition, the Bragg condition is satisfied at some *different* location within the crystal and the bend contour *shifts* to a new position in the image.

Thickness fringes occur in a crystalline specimen whose thickness is non-uniform. Typically they take the form of dark and bright fringes running parallel to the edge of a specimen, where its thickness falls to zero; see Fig. 4.16b, c. To understand the reason for these fringes, imagine the specimen oriented so that a particular set of lattice planes (indices *h k l*)) satisfies the Bragg condition. As electrons penetrate further into the specimen and become diffracted, the intensity of the (*h k l*) diffracted beam increases. However, these diffracted electrons are traveling in exactly the right direction to be Bragg-reflected back into the original (0 0 0) direction of incidence (central spot in Fig. 4.12d). So beyond a certain distance into the crystal, the diffracted-beam intensity decreases and the (0 0 0) intensity starts to increase. At a certain depth (the **extinction distance**) the (*h k l*) intensity becomes almost zero and the (0 0 0) intensity goes through a maximum value. If the specimen is thicker than the extinction depth, the above process is repeated, resulting in an oscillating intensity as in Fig. 4.16b.

In the case of a wedge-shaped (variable-thickness) crystal, the angular distribution of electrons leaving its exit surface corresponds to the diffraction condition prevailing at a depth equal to its thickness. In terms of the diffraction pattern, this means that the intensities of the (0 0 0) and (*h k l*) components *oscillate* as a function of crystal thickness. For a TEM with an *on-axis* objective aperture, the *bright-field* image will show alternate bright and dark thickness fringes. With an off-axis objective aperture, similar but *displaced* fringes appear in the dark-field image (Fig. 4.16c); their intensity distribution is complementary to that of the bright-field image.

A mathematical analysis of this intensity oscillation is based on the fact that the electrons traveling through a crystal can be represented by **Bloch waves** whose wave-function amplitude ψ is modified as a result of the regularly repeating electrostatic potential of the atomic nuclei. If there are *n* Bragg beams in the electron-diffraction pattern, there are *n* Bloch waves with slightly differing wavevectors, and thickness fringes represent an interference (beating effect) between these different Bloch waves. An electron-wave description of the outer-shell (conduction) electrons in a crystalline solid makes use of this same concept; the degree to which these electrons are diffracted determines whether solid is an electrical conductor or an insulator, as discussed in textbooks dealing with solid-state physics.

4.9 Phase Contrast in the TEM

Small-scale features seen in some TEM images depend on the *phase* of the electron waves at the exit plane of the specimen. Although this phase cannot be measured directly, it gives rise to interference between electron waves that have passed through different parts of the specimen. Such electrons are brought together when a TEM image is *defocused* by changing the objective-lens current slightly. Unlike the

case of diffraction-contrast images, a large-diameter objective aperture (or no aperture) is used to enable several diffracted beams to contribute to the image.

A simple example of phase contrast is the appearance (in a defocused image) of a dark **Fresnel fringe** just inside a small hole in the specimen. The separation of this fringe from the edge of the hole increases with the amount of defocus. In fact, its separation from the edge will change if the objective focusing power varies for electrons traveling along different azimuthal directions relative to the optic axis, i.e., if objective astigmatism is present. Therefore the Fresnel fringe inside a circular hole is sometimes used as a guide to adjusting the objective-lens stigmator.

Phase contrast also occurs as a result of the regular spacing of atoms in a crystal, particularly if the specimen is oriented so that columns of atoms are parallel to the incident-electron beam. The contrast observable in high-magnification ($M > 500{,}000$) images is of this type and allows the lattice of atoms (projected onto a plane perpendicular to the incident direction) to be seen directly in the image. The electron-wave explanation is that electrons traveling down the center of an atomic column have their phase *retarded* relative to those that travel in between the columns. However, the basic effect can also be understood by treating the electrons as particles, as follows.

In Fig. 4.17, we represent a crystalline TEM specimen by a single layer of atoms (of equal spacing d) irradiated by a parallel electron beam with a uniform current density. The atomic nuclei scatter electrons elastically through angles that depend *inversely* on the impact parameter b of each electron, as indicated by Eq. (4.11). If the objective lens is focused on the atomic plane and all electrons are collected to form the image, the uniform illumination will ensure that the image contrast is zero. But if the objective current is increased slightly (**overfocus** condition), the TEM will image a plane *below* the specimen, where the electrons are more numerous at positions vertically below the atom positions, causing the atoms to appear *bright* in the final image. On the other hand, if the objective lens is weakened so as to image a virtual-object plane just *above* the specimen, the electrons appear to come from positions halfway between the atoms, as shown by the dashed lines that represent extrapolations of the electron trajectories. In this **underfocus** condition, the image will show atoms that appear *dark* relative to their surroundings.

The *optimum* amount of defocus Δz can be estimated by considering a close collision with impact parameter $b << d$. Geometry of the right-angled triangle in Fig. 4.17 gives $\Delta z = (d/2 - b)/\tan\theta \approx (d/2)/(\tan\theta) \approx d/(2\theta)$ if the scattering angle θ is small. But for atoms of uniform spacing d, the scattering angle should satisfy Bragg's law, therefore Eq. (4.21) implies $\theta \approx \lambda/d$ so that:

$$\Delta z \approx d/(2\theta) \approx d^2/(2\lambda) \tag{4.25}$$

For $E_0 = 200$ keV, $\lambda = 2.5$ pm and taking $d = 0.3$ nm we obtain $\Delta z \approx 18$ nm, so the amount of defocus required for atomic resolution is small. Wave-optical theory gives an expression identical to Eq. (4.25) for an objective lens with no spherical aberration. But if the TEM has no aberration corrector, spherical aberration is important and should be included in the theory. In general, the interpretation of

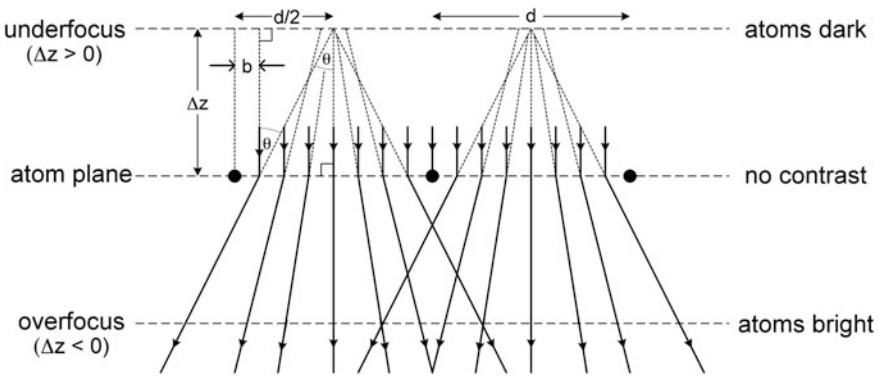


Fig. 4.17 Equally spaced atomic nuclei that are elastically scattering the electrons within a broad incident beam, which provides a continuous range of impact parameter *b*. The electron current density appears non-uniform at planes above and below the atom plane

phase-contrast images requires that the spherical-aberration coefficient of the objective and the amount of defocus are known. In fact, a detailed interpretation of atomic-resolution **lattice images** often requires recording a **through-focus series** of images with positive and negative Δz.

High-magnification phase-contrast images are particularly valuable for examining the atomic structure of *interfaces* within a solid, such as the grain boundaries within a polycrystalline material or the interface between a deposited film and its substrate; see Fig. 4.18. Sometimes the lattice fringes (evidence of ordered planes of atoms) extend right up to the boundary, indicating an abrupt interface. In other

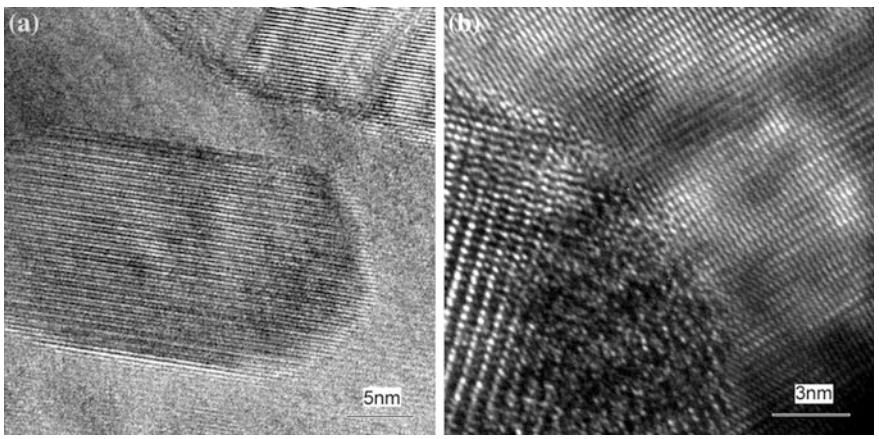


Fig. 4.18 **a** Lattice image of ZnS crystals embedded within a sapphire matrix. **b** Detail of a grain boundary between a PbS crystal and its sapphire matrix. Courtesy of Al Meldrum, University of Alberta

cases, they disappear a short distance from the interface, as the material becomes amorphous. Such amorphous layers can have a profound effect on the mechanical or electrical properties of a material.

A further type of phase-contrast image occurs in specimens that are strongly magnetic (ferromagnetic), such as iron, cobalt, and alloys containing these metals. Again, phase contrast is produced only when the specimen is defocused. The images are known as **Lorentz images**; they are capable of revealing the magnetic-domain structure of the specimen, which is not necessarily related to its crystallographic structure. Because the objective lens of a modern TEM generates a high field that is likely to change the domain structure, Lorentz microscopy is usually performed with the objective turned off or considerably weakened. With only the intermediate and projector lenses operating, the image magnification (and resolution) is relatively low.

Although related to the phase of the electron exit wave, these Lorentz images can again be understood by considering the electron as a particle that undergoes angular deflection when it travels through the field within each magnetic domain; see Fig. 4.19a. If the magnetic field has a strength B in the y-direction, an incident electron traveling with speed v in the z-direction experiences a force $F_x = evB$ and an acceleration $a_x = F_x/m$ in the x-direction. After traveling through a sample of thickness t in a time $T = t/v$, the electron (relativistic mass m) has acquired a lateral velocity $v_x = a_x T = (evB/m)(t/v)$.

From the velocity triangle of Fig. 4.19a, the angle of deflection of the electron (due to the magnetic field inside the specimen) is:

$$\theta \approx \tan\theta = v_x/v_z \approx v_x/v = (e/m)\,(Bt/v) \tag{4.26}$$

Taking $B = 1.5$ T, $t = 100$ nm, $v = 2.1 \times 10^8$ m/s, and $m = 1.3 \times 10^{-30}$ kg (for $E_0 = 200$ keV), Eq. (4.26) gives $\theta \approx 0.1$ mrad. The deflection is therefore small, but can be detected by focusing on a plane sufficiently far above or below the specimen (see Fig. 4.19a) so that the magnetic domains become visible, as in Fig. 4.19b.

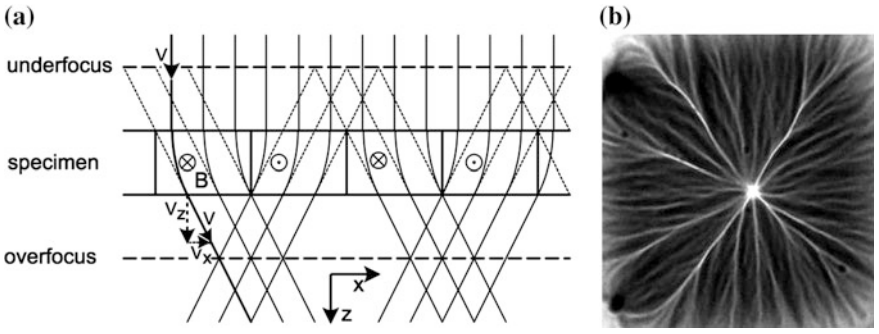


Fig. 4.19 **a** Magnetic deflection of electrons within adjacent domains of a ferromagnetic specimen. **b** Lorentz image showing magnetic-domain boundaries in a thin film of cobalt

With the objective lens (focal length *f*) turned on, this deflection causes a *y*-displacement of each diffraction spot equal to ($\pm\ \theta\ f$) at the objective-aperture plane, due to electrons that have passed through oppositely polarized domains. This can be sufficient to produce an observable splitting of the spots in the electron-diffraction pattern recorded from a crystalline specimen. If the TEM objective aperture or illumination tilt is adjusted so as to admit only *one* spot from a pair, a **Foucault image** is produced in which adjacent magnetic domains appear with different intensity. Its advantage (over a Fresnel image) is that the image appears in focus and can have good spatial resolution, allowing crystalline defects such as grain boundaries to be observed in relation to the magnetic domains.

4.10 STEM Images

Our discussion of diffraction-contrast effects in TEM images assumed a parallel incident beam, which is a reasonable approximation for defocused TEM illumination. As seen in Fig. 3.9b, the convergence angle α of the beam increases when the illumination is focused into a small diameter *d*, in accordance with the brightness theorem, Eq. (3.4), according to which the product αd remains constant. Consequently, the small probe used in high-resolution STEM imaging is relatively convergent (α typically 10 mrad or more) and many of the diffraction-contrast features seen in TEM images of crystalline specimens appear with much reduced contrast in STEM images. The reason is obvious in the case of bend contours, which shift as the angle of incidence is changed and are therefore broadened in the case of convergent illumination; a generally beneficial effect since their absence allows real specimen features to be seen more clearly.

The so-called **reciprocity theorem**, based on the fact that electron (or photon) ray paths are reversible in direction, states that a STEM image looks like a TEM image if the geometries of the electron source and detector are interchanged. Accordingly, the large STEM incident convergence corresponds to collecting TEM electrons over a large angular range (implying low diffraction contrast) *provided* a small on-axis detector is used. However, the reciprocity theorem describes image contrast, not intensity, and a small STEM detector collects only a fraction of the transmitted electrons. Consequently, bright-field STEM is inefficient, a particular disadvantage for specimens that damage in the electron beam.

Conversely, the HAADF detector used in STEM is *more* efficient than its TEM equivalent (an annular condenser aperture), although the collection efficiency remains low because relatively few electrons are elastically scattered through large angles. As a result, STEM has been little used by biologists, who often deal with radiation-sensitive specimens. To achieve high resolution, they typically require phase contrast, achieved using a large specimen defocus or else a device called a **phase plate** inserted in the objective-aperture plane. Phase contrast is possible in STEM and schemes for its efficient exploitation are regularly discussed; see, for example, Ophus et al. (2015).

One advantage of STEM is that the energy spread arising from inelastic scattering of electrons does not lead to any loss of resolution due to chromatic aberration. STEM resolution depends on the probe size, which is increased by chromatic aberration of the probe-forming lens but not greatly, since the energy spread in the incident beam is only that of the electron source, usually a fraction of an eV; see Table 3.1. As a result, STEM can provide useful images of relatively thick specimens, with much less chromatic broadening than in stationary-beam TEM mode.

Because the current in a small probe is often a small fraction of the TEM incident-beam current, STEM images typically take longer to record, a disadvantage if the resolution is limited by specimen drift or electrostatic charging. On the other hand, STEM mode is preferred for analytical electron microscopy, since it allows images to be produced from the x-rays emitted as a result of electron impact (Chap. 5).

As a further consequence of the convergent-beam illumination, a diffraction pattern produced in the STEM mode will consist of disks (which may even overlap) rather than sharp diffraction spots. Although this makes standard analysis of the diffraction pattern (Sect. 4.7) inaccurate, the convergent-beam electron-diffraction (**CBED**) pattern can contain detailed information about the crystal structure, which can be extracted by analyzing the fine structure within the diffraction disks.

4.11 TEM Specimen Preparation

As discussed in Sect. 4.4, electrons are strongly scattered within a solid, due to the large forces acting on them when they pass through the electrostatic field within each atom. As a result, TEM or STEM specimens should be very thin: usually in the range 10 nm –1 μm. Specimen preparation involves ensuring that the thickness of at least *some* regions of the specimen is within this thickness range. For some materials, specimen preparation can represent a large part of the work involved in transmission electron microscopy.

Common methods of specimen preparation are summarized in Fig. 4.20. Usually the material of interest must be *reduced* in thickness and the initial reduction involves a **mechanical** method. When the material is in the form of a block or rod, a thin (<1 mm) slice can be made by sawing or cutting. For hard materials such as quartz or silicon, a "diamond wheel" (a fast-rotating disk whose edge is impregnated with diamond particles) can provide a relatively clean cut.

At this stage it is convenient to cut a 3 mm-diameter disk out of the slice. For reasonably soft metals, (e.g., aluminum, copper) a mechanical punch can be used. For harder materials, an **ultrasonic drill** is necessary; it consists of a thin-walled tube (internal diameter = 3 mm) attached to a *piezoelectric* crystal that changes slightly in length when a voltage is applied between electrodes on its surfaces; see Fig. 4.20a. The tube is lowered onto the slice, which has been bonded to a solid support using by heat-setting wax and its top surface pre-coated with a slurry consisting of diamond or SiC powder mixed with water. Upon applying an

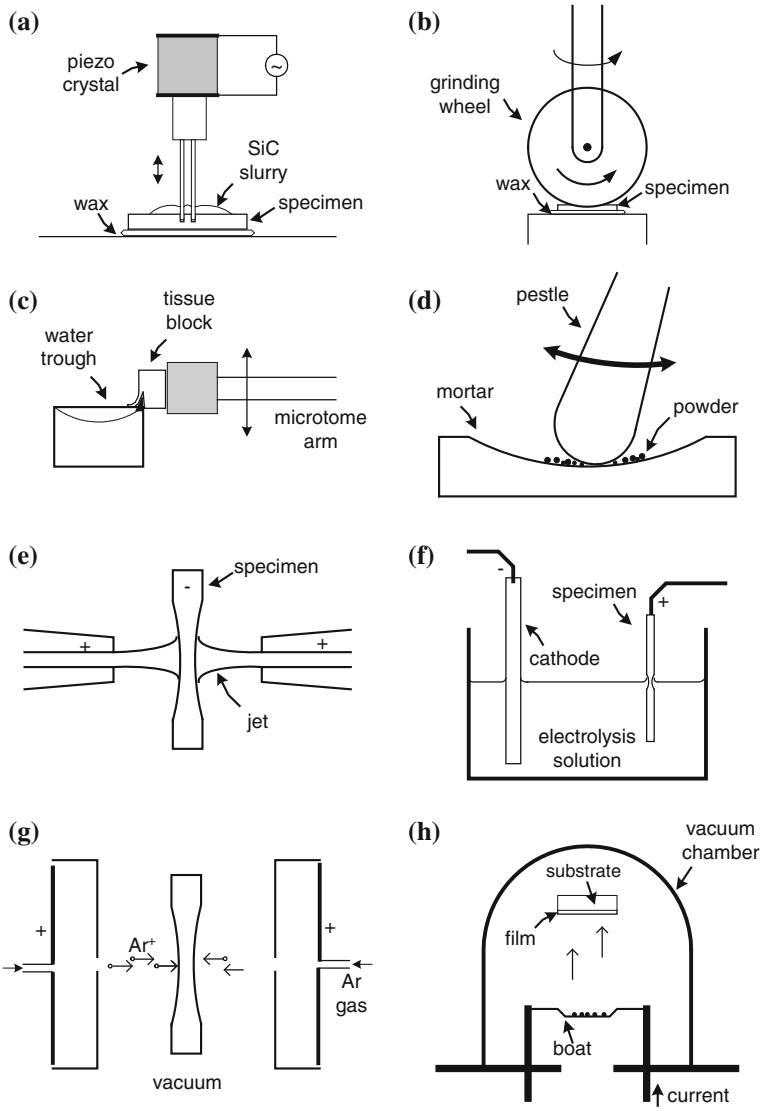


Fig. 4.20 Procedures used in making TEM specimens: **a** disk cutting by ultrasonic drill, **b** dimple grinding, **c** ultramicrotomy, **d** grinding to a powder, **e** chemical jet thinning, **f** electrochemical thinning, **g** ion-beam thinning, and **h** vacuum deposition of a thin film. Methods **a**–**d** are mechanical, **e** and **f** are chemical, **g** and **h** are vacuum techniques

alternating voltage, the tube oscillates vertically thousands of times per second, driving SiC particles against the sample and cutting an annular groove in the slice. When the slice has been cut through completely, the wax is melted and the 3 mm disk is retrieved with tweezers.

The disk is then thinned further by polishing with abrasive paper (coated with diamond or silicon carbide particles) or by using a **dimple grinder**. In the latter case (Fig. 4.20b) a metal wheel rotates rapidly against the surface (covered with a SiC slurry or diamond paste) while the specimen disk is rotated slowly about a vertical axis. The result is a dimpled specimen whose thickness is 10–50 μm at the center but greater (100–400 μm) at the outside, which provides the mechanical strength needed for easy handling.

In the case of biological tissue, a common procedure is to use an **ultramicrotome** to directly cut slices ≈100 nm or more in thickness. The tissue block is lowered onto a glass or diamond knife that *cleaves* the material apart (Fig. 4.20c). The ultramicrotome can also be used to cut thin slices of soft metals, such as aluminum.

Some inorganic materials (e.g., graphite, mica) have a layered structure, with weak forces between sheets of atoms, and are readily cleaved apart by inserting the tip of a knife blade. Repeated **cleavage**, achieved by attaching adhesive tape to each surface and pulling apart, can reduce the thickness to well below 1 μm. After dissolving the adhesive, thin flakes are mounted on 3 mm TEM grids. Other materials, such as silicon, can sometimes be persuaded to cleave along a plane cut at a small angle relative to a natural (crystallographic) cleavage plane. If so, a second cleavage along a crystallographic plane results in a wedge of material whose thin end is transparent to electrons. Mechanical cleavage is also involved when a material is ground into a fine powder, using a pestle and mortar, for example (Fig. 4.20d). The result is a fine powder, whose particles or flakes may be thin enough for electron transmission. They are dispersed onto a 3 mm TEM grid covered with a thin-carbon film, sometimes containing small holes so that some particles are supported only at their edges and can be imaged without any carbon background.

All of the above methods involve the application of a *mechanical* force to achieve thinning. Unfortunately, for all but the most well-behaved materials (e.g., biological tissue, layer materials, silicon), the specimen becomes extremely fragile and cannot be mechanically thinned below about 1 μm. In addition, the mechanical forces involved may leave a damaged surface layer, containing a high density of defects such as dislocations. This damage is undesirable if the TEM is being used to study the *original* defect structure of the specimen. Therefore some non-mechanical method is commonly used for the final thinning.

One such method is **chemical thinning**, in which a chemical solution dissolves the original surface and reduces the specimen thickness to a value suitable for TEM imaging. In the simplest case, a thin piece of material is floated onto the surface of a chemical solution that attacks its lower surface; the sample is retrieved (e.g., by picking up by a TEM grid held in tweezers) before it dissolves completely. More commonly, a jet of chemical solution is directed at one or both surfaces of a thin disk. As soon as a small hole forms in the center (detected by the transmission of a light beam), the polishing solution is replaced by rinse water. If the procedure is successful, regions of the specimen surrounding the hole are thin enough for TEM examination.

Alternatively, **electrochemical thinning** is carried out with a direct current flowing between the specimen (at a negative potential) and a positive electrode, immersed in a chemical solution. In the original window-frame method (Fig. 4.20f), the specimen is in the form of a thin sheet whose four edges are previously painted with protective lacquer to prevent erosion at the edge. When partially immersed in the electrolytic solution, thinning is most rapid at the liquid/air interface, which perforates first. Small pieces of foil are cut adjacent to the perforated edge and mounted on a TEM grid. Nowadays, electrochemical thinning is usually done using the jet-thinning geometry (Fig. 4.20e), by applying a dc voltage between the specimen and jet electrodes. When thinning metals, glycerine is sometimes added to the solution to make the liquid more *viscous*, helping to give the thinned specimen a *polished* (microscopically smooth) surface.

Increasingly, **ion-beam thinning** is used for the final thinning, particularly of materials that are chemically inert. Following mechanical thinning (if necessary), a 3 mm-diameter thin disk of the material is placed in a vacuum system, where it is bombarded by argon ions produced by a gas discharge within an ion gun. These ions transfer energy to surface atoms and remove the material by the process of sputtering; see Fig. 4.20g. A focused-ion-beam (**FIB**) machine uses a beam of gallium or rare-gas ions, which are focused by electrostatic lenses into a small (e.g., 10 nm-diameter) probe that cuts thin slices from a bulk material. This instrument also provides an electron or ion-beam image of the surface of the specimen, allowing the ion beam to be positioned and scanned along a line so as to cut a slice of material at a precise location; see Fig. 4.21.

Instead of thinning a bulk material, TEM specimens are sometimes *built up* in thickness by **thin-film deposition**. Typically, the material is placed in a tungsten "boat" which is electrically heated in vacuum, causing the material to evaporate and then condense onto a substrate; see Fig. 4.20h. If the substrate is soluble (for example, alkali halides dissolve in water), the film is floated onto the liquid surface and captured on a TEM grid, similar to thin sections of biological tissue. Vacuum-deposited thin films are used in the electronics and optical industries but are usually on insoluble substrates such as silicon or glass. For TEM examination, the substrate must be thinned down, for example by dimple grinding or jet thinning from its back surface.

The procedures outlined in Fig. 4.20 yield **plan-view** images in which the sample is viewed perpendicular to its original surface. For some purposes, a **cross-sectional** image is needed, for example to show the growth mechanism of a film. The required specimen is made by using a diamond wheel to cut the film (on its substrate) into several thin strips, which are then turned through 90° about their long axis and glued together by epoxy cement. Dimple grinding followed by ion milling then produces a specimen that is thin at the center, resulting in a TEM image of the substrate and film seen in cross section, as in Fig. 4.21.

Cross-sectional specimens are invaluable for analyzing problems that arise in the manufacture of integrated circuits on a silicon chip. Because device sizes have become so small, it becomes necessary to make use of the high spatial resolution of the TEM for this kind of **failure analysis**. The problem of producing a cross-section

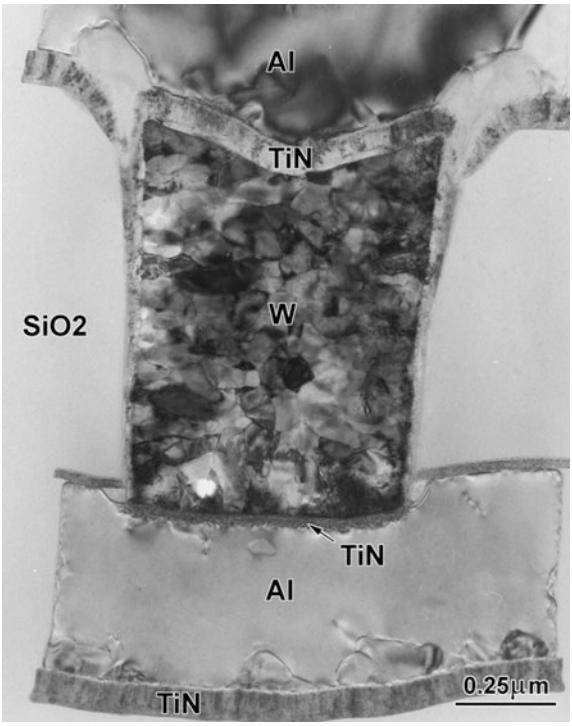


Fig. 4.21 Cross-sectional TEM image showing a polycrystalline tungsten "via" that makes electrical connection between adjacent layers of aluminum in a multilayer integrated circuit. From such images, the thickness and grain structure of each layer can be determined. Reprinted from H. Zhang: *Transmission electron microscopy for the semiconductor industry*, Micron **33** (2002) 517-525, Copyright (2002), with permission from Elsevier

containing a specific component is solved by using a FIB machine, which produces a scanned image of the chip, allowing the ion beam to be precisely positioned and then scanned in a line to cut into the silicon on either side of the component. This leaves a slice only 100 nm in thickness, which can be lifted out and viewed in cross-section in the TEM, as illustrated in Fig. 4.21.

4.12 Further Reading

Transmission Electron Microscopy (5th Edition, Springer, 2007) by L. Reimer and H. Kohl gives a detailed and reliable treatment of the physics of electron scattering. *High-Resolution Electron Microscopy* (4th Edition, Oxford University Press, 2013) by J.C.H. Spence gives a comprehensive account of phase-contrast imaging. For an introduction to phase contrast using the STEM, see the article by Ophus et al. in Nature Communications (2016), DOI:10.1038/ncomms19719.

The contrast features seen in TEM images are also well described by D.B. Williams and C.B. Carter in their *Transmission Electron Microscopy* (2nd Edition, Springer, 2009), which also includes practical advice about use of the TEM and image interpretation, together with a short chapter on specimen preparation.

Introduction to Focused Ion Beams: Instrumentation, Theory, Techniques and Practice by L.A. Giannuzzi and F.A. Stevie (Springer Press, 2004, ISBN 978-0-387-23116-7) reveals some of the tricks behind FIB methods, which have made specimen preparation much more of a precise art.

Chapter 5
The Scanning Electron Microscope

As outlined in Chap. 1, the scanning electron microscope (SEM) was invented soon after the TEM but took longer to be developed into a practical instrument for scientific research. Like the case of the TEM, its spatial resolution improved after magnetic lenses were substituted for electrostatic ones and after a stigmator was added to the lens column. Today, scanning electron microscopes outnumber transmission electron microscopes and they are used in many fields, including medical and materials research, the semiconductor industry, and forensic-science laboratories.

Figure 1.15 (p. 17) shows one example of a commercial high-resolution SEM. Although smaller and generally less expensive than a TEM, the SEM incorporates an electron-optical column that operates according to the principles already discussed in Chaps. 2 and 3. Accordingly, our description will be shorter than for the TEM because we can make use of many of the concepts introduced in those earlier chapters.

5.1 Operating Principle of the SEM

The electron source used in the SEM can be a tungsten filament, a LaB_6 or Schottky emitter, or a tungsten field-emission tip. Because the maximum accelerating voltage (typically 30 kV) is lower than for a TEM, the electron gun is smaller, requiring less insulation. Axially symmetric magnetic lenses are used but they too are smaller than the lenses employed in the TEM; for electrons of lower kinetic energy, the polepieces need not generate such a strong magnetic field. There are also *fewer* lenses; image formation utilizes the scanning principle already outlined in Chap. 1, so *imaging lenses* are not necessary.

Above the specimen, there are typically two or three lenses, acting rather like the condenser lenses of a TEM. But whereas the TEM (if operating in its conventional imaging mode) produces a beam of diameter ≈1 μm or more at the specimen, the incident beam in the SEM (also known as an **electron probe**) needs to be as small

R.F. Egerton, *Physical Principles of Electron Microscopy*,
DOI 10.1007/978-3-319-39877-8_5

as possible: a diameter of 10 nm is typical and 1 nm is possible with a field-emission source. The final lens that forms this very small probe is named the **objective** and its performance (including aberrations) largely determines the spatial resolution of the SEM, as does the objective of a TEM or a light-optical microscope. In fact, the SEM-image resolution can never be better than the incident-probe diameter, as a consequence of the method used to obtain the image.

Whereas the conventional TEM uses a stationary incident beam, the electron probe of an SEM is scanned horizontally across the specimen in two perpendicular (x and y) directions. The x-scan is relatively fast and is generated by a sawtooth-wave generator operating at a **line** frequency f_x; see Fig. 5.2a. This generator supplies current to two scan coils, connected in series and located on either side of the optic axis, just above the objective lens. These coils generate a magnetic field in the y-direction, creating a force on an electron (traveling in the z-direction) that deflects it in the x-direction; see Fig. 5.1.

The y-scan is much slower (Fig. 5.2b) and is generated by a second sawtooth-wave generator running at a **frame** frequency $f_y = f_x/n$ where n is an integer (the number of lines per frame). The entire procedure is known as **raster scanning** and causes the beam to sequentially cover a rectangular area on the specimen (Fig. 5.2d).

During its x-deflection signal, the electron probe moves in a straight line, from A to B in Fig. 5.2d, forming a single **line scan**. After reaching B, the beam is deflected back along the x-axis as quickly as possible (the **flyback** portion of the x-waveform). Because the y-scan generator has increased its output during the line-scan period, it returns *not* to A but to point C, displaced in the y-direction. A second line scan takes the probe to point D, at which point it flies back to E and the process is repeated until n lines have been scanned and the beam arrives at point Z. This entire sequence constitutes a single **frame** of the raster scan. From point Z, the probe quickly returns to A, due to rapid flyback of *both* the line and frame

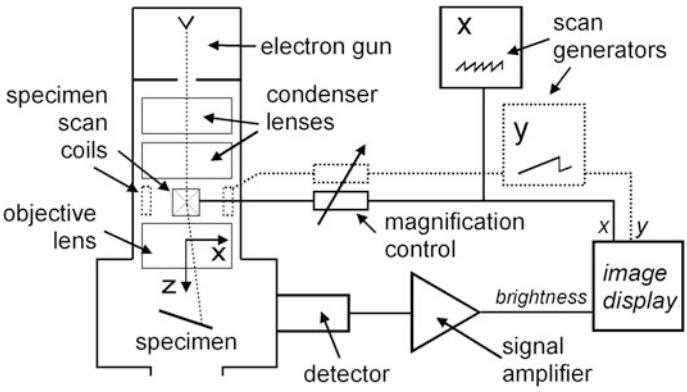


Fig. 5.1 Schematic diagram of a scanning electron microscope. The same x- and y-scan waveforms are applied to the SEM column and to the display device. Signal from a detector modulates the brightness of the display

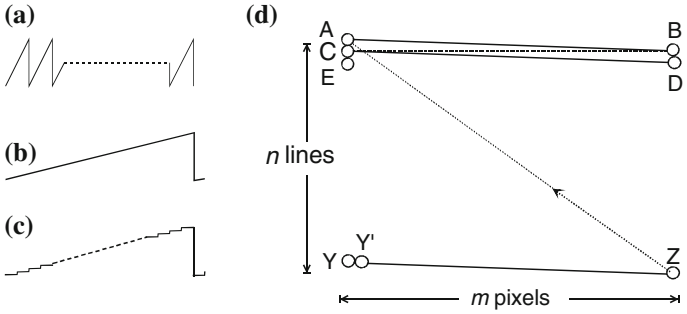


Fig. 5.2 **a** Line-scan waveform (scan current versus time). **b** Analog frame-scan waveform and **c** its digital equivalent. **d** Elements of a single-frame raster scan: AB and YZ are the first and last line scans in the frame, Y and Y′ represent adjacent pixels

generators, and the next frame is executed. This process can run continuously for many frames, as in television and video technology.

The outputs of the two scan generators are also applied to the display device on which the SEM image appears, which was originally a cathode-ray tube (CRT). This CRT contained an electron beam that was scanned exactly in *synchronism* with the beam in the SEM, so for every point on the specimen (within the raster-scanned area) there was an *equivalent* point on its display screen, thereby satisfying Maxwell's first rule of imaging. In order to create *contrast* in the image, a voltage signal was applied to the electron gun of the CRT, to vary the brightness of the scanned spot. This voltage was derived from a detector that responded to *some change in the specimen* induced by the SEM incident probe.

In a modern SEM, the scan signals are generated digitally, by computer-controlled circuitry, and the *x*- and *y*-scan waveforms are *staircase* functions with *m* and *n* levels, respectively; see Fig. 5.2c. This procedure divides the image into a total of *mn* picture elements (**pixels**) and the SEM probe remains stationary for a certain dwell time before *jumping* to the next pixel. One advantage of digital scanning is that the SEM computer "knows" the (*x, y*) address of each pixel and can record the appropriate image-intensity value (as a digitized number) in the appropriate computer-memory location. The digital image, in the form of position and intensity information, is stored in computer memory and more permanently on a magnetic disk or other storage device.

The modern SEM uses a flat-panel display screen in which there is no internal electron beam. Instead, computer-generated voltages are used to sequentially define the *x*- and *y*-coordinates of a screen pixel and the SEM detector signal is applied electronically to that pixel, to change its brightness. In other respects, the raster-scanning principle is the same as for a CRT display.

Image magnification in the SEM is achieved by making the *x*- and *y*-scan distances *on the specimen* a small fraction of the size of the displayed image, since by definition the magnification factor *M* is given by:

$$M = (\text{scan distance in the image})/(\text{scan distance on the specimen}) \quad (5.1)$$

It is convenient to keep the image at a *fixed* size, just filling the display screen, so increasing the magnification involves *reducing* the *x*- and *y*-scan currents, each in the same proportion (to avoid rectangular distortion). Consequently, the SEM is actually working at its hardest (in terms of current drawn from the scan generator) when operating at *low* magnification.

The scanning is sometimes done at video rate (around 50 or 60 frames/second) to generate a rapidly refreshed image that is useful for focusing the specimen or for viewing it at low magnification. At higher magnification, or when making a permanent record of an image, slow scanning (several seconds per frame) is possible; the additional recording time results in a higher-quality image containing less electronic noise.

The signal that modulates (alters) the image brightness can be derived from any property of the specimen that changes in response to electron bombardment. Most commonly, the emission of **secondary electrons** (atomic electrons ejected from the specimen as a result of inelastic scattering) is utilized. Alternatively, a signal derived from **backscattered electrons** (incident electrons elastically scattered through more than 90°) is used. In order to understand these (and other) possibilities, we need to consider what happens when an electron beam enters a *thick* (often called bulk) specimen.

5.2 Penetration of Electrons into a Solid

When accelerated primary electrons enter a solid, they are scattered both *elastically* (by electrostatic interaction with atomic nuclei) and *inelastically* (by interaction with atomic electrons), as discussed in Chap. 4. Most of this interaction involves "forward" scattering, which implies deflection angles of less than 90°. But a small fraction of the primary electrons are elastically backscattered ($\theta > 90°$) with only small fractional loss of energy. Due to their high kinetic energy, these backscattered electrons have a reasonable probability of leaving the specimen and re-entering the surrounding vacuum, in which case they can be collected as a backscattered-electron (**BSE**) signal.

Inelastic scattering involves relatively small scattering angles and therefore contributes little to the backscattered signal. However, it reduces the kinetic energy of the primary electrons until they are eventually brought to rest and absorbed into the solid; in a metal specimen they would become conduction electrons. The depth (below the surface) at which this occurs is called the **penetration depth** or the **electron range**. The volume of sample containing the scattered electrons is called the **interaction volume**, and is often represented as pear-shaped in cross-section (Fig. 5.3) since scattering causes the beam to spread laterally as the electrons penetrate the solid and gradually lose energy.

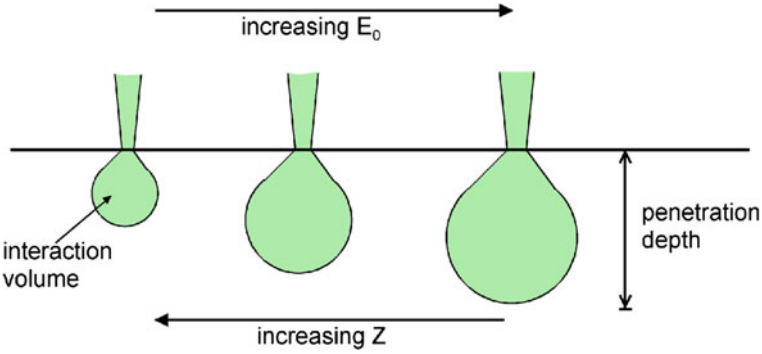


Fig. 5.3 Schematic dependence of the interaction volume and penetration depth, as a function of the kinetic energy E_0 and atomic number Z of the primary electrons

The electron range R for electrons of incident energy E_0 is given by the following approximate formula (Reimer 1998; see Sect. 5.10):

$$\rho R \approx a(E_0)^n \tag{5.2}$$

where $n \approx 1.35$ and ρ is the density of the specimen. If E_0 is given *in* keV in Eq. (5.2), $a \approx 10$ μg/cm^2. Expressing the range as a **mass-thickness** ρR makes the coefficient a roughly independent of atomic number Z but implies that the distance R itself decreases with Z, since the densities of solids tend to increase with atomic number. For carbon ($Z = 6$), $\rho \approx 2$ g/cm^3 and $R \approx 1$ μm for a 10 keV electron. For gold ($Z = 79$) however, $\rho \approx 20$ g/cm^3 and $R \approx 0.2$ μm at $E_0 = 10$ keV. This strong Z-dependence arises mainly because backscattering depletes the number of electrons moving forward into the solid and the probability of such high-angle elastic scattering is proportional to Z^2, as seen from Eq. (4.15). The interaction volume is therefore considerably smaller for materials of high atomic number (Fig. 5.3).

According to Eq. (5.2), the range decreases substantially with decreasing incident energy, not surprising since the probability of inelastic scattering is inversely proportional to E_0 and lower-energy electrons require fewer inelastic collisions to bring them to rest. A 1 keV electron penetrates only about 50 nm into carbon and less than 10 nm into gold. Therefore the interaction volume becomes very small at low incident energy; see Fig. 5.3.

Penetration depth and interaction volume are macroscopic quantities, averaged over a large number of electrons, whereas the behavior of an *individual* electron is highly variable. In other words, scattering is a statistical process. It can be simulated in a computer by running a Monte-Carlo program that contains a random-number generator and information about the angular distributions of elastic and inelastic scattering. Figure 5.4 shows the trajectories of 25 primary electrons entering both aluminum and gold. Sudden changes in direction represent the elastic or inelastic

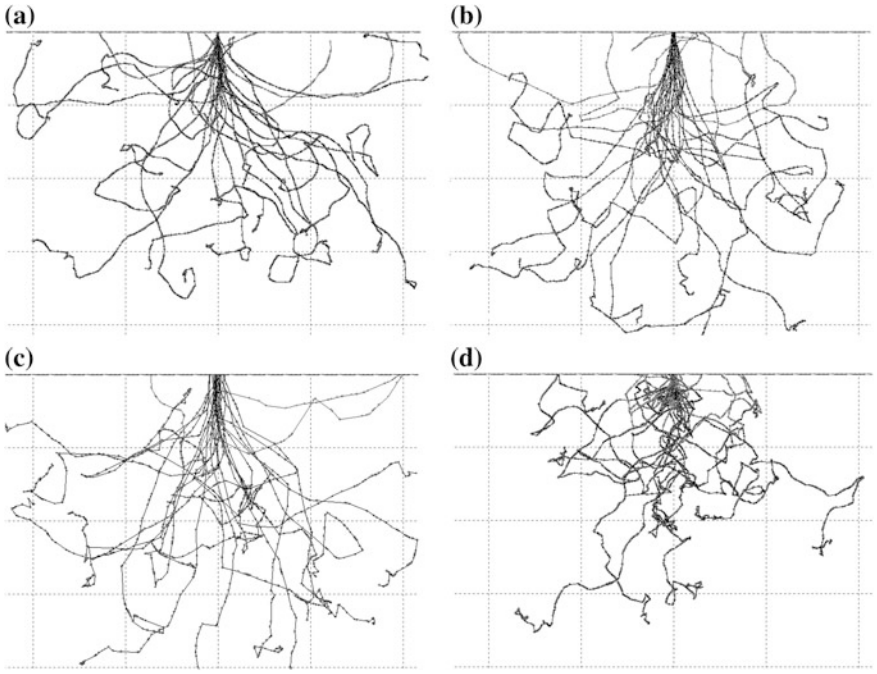


Fig. 5.4 Penetration of **a** 30 keV, **b** 10 keV, and **c** 3 keV electrons into aluminum ($Z = 13$) and **d** 30 keV electrons into gold ($Z = 79$). Note that the dimensional scales are different: the maximum penetration is about 6.4, 0.8, 0.12, and 1.2 μm in (**a**), (**b**), (**c**), and (**d**), respectively. These Monte Carlo simulations were all carried out using the CASINO program (written by R. Gauvin) with 25 primary electrons and an incident-beam diameter of 10 nm

scattering. Although the behavior of each electron is different, the very dissimilar length scales needed to represent the different values of E_0 and Z in Fig. 5.4 are a further illustration of the overall trends represented by Eq. (5.2).

5.3 Secondary-Electron Images

From the principle of conservation of energy, we know that any energy *lost* by a primary electron must appear as a *gain* in energy of the atomic electrons that are responsible for the inelastic scattering. If these are outer-shell (valence or conduction) electrons, which are weakly bound (electrostatically) to atomic nuclei, only a small part of this acquired energy will be used up (as potential energy) to release them from the confines of a particular atom. The rest will be retained as kinetic energy, allowing the ejected electrons to travel through the solid as secondary electrons (abbreviated to **SEs** or **secondaries**). As moving charged particles, the secondaries themselves will interact with other atomic electrons and be scattered

inelastically, gradually losing their kinetic energy. In fact, most SEs start with a kinetic energy of less than 100 eV and travel an average distance of only a few nm, since the probability of inelastic scattering depends *inversely* on kinetic energy as in Eq. (4.16).

As a result, most secondaries are brought to rest *within* the interaction volume of the primary electrons. But those created close to the surface may escape into the vacuum, especially if they are initially traveling *towards* the surface. On average, the *escaping* secondaries are generated only within a very small depth (< 2 nm) below the surface, called the **escape depth**. Since the SE signal used in the SEM is derived from secondaries that escape into the vacuum, the SE image is mainly a property of the *surface structure* (topography) of the specimen rather than any underlying structure; the SEM image is said to display **topographical contrast**.

The average number of *escaping* secondaries *per primary electron* is called the **secondary-electron yield** δ, and is typically in the range from 0.1 to 10; the exact value depends on the chemical composition of the specimen (close to the surface) and on the primary-electron energy E_0. For a given specimen, δ decreases with increasing E_0 because higher-energy primaries undergo less inelastic scattering (per unit distance traveled) and so there will be fewer secondaries generated *within the escape depth*.

As Fig. 5.5 indicates, the secondary-electron yield also depends on the angle between the incoming primary electron and the surface. The yield is lowest for normal (perpendicular) incidence and increases with increasing angle between the primary beam and the surface-normal. The reason is illustrated in Fig. 5.6a, which shows a focused nearly-parallel beam of primary electrons (diameter *d)* incident at two locations on a specimen, where the surface is normal (at A) and inclined (at B) to the incident beam. The region from which secondary electrons can escape is such that all points within it lie within the escape depth λ of the surface. For normal

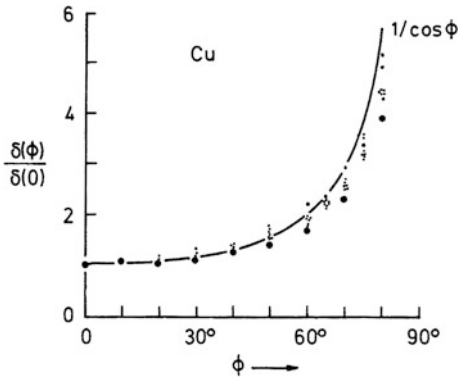


Fig. 5.5 Dependence of SE yield on the angle of tilt φ of the specimen, measured between a plane perpendicular to its surface and the primary-electron beam (φ = 0 corresponds to normal incidence). Data points represent experimental measurements and Monte Carlo predictions for copper; the curve represents a 1/cosφ function. From Reimer (1998), courtesy of Springer-Verlag

incidence, this escape region is a cylinder of radius $d/2$, height λ, and volume $V(0) = (\pi/4)\, d^2\lambda$. For the inclined surface, the escape region is a slanted cylinder with height λ (perpendicular to the surface) and base of cross-sectional area $(\pi/4)(d^2/\cos\phi)$, giving escape volume $V(\phi) = \pi(d/2)^2\,(\lambda/\cos\phi) = V(0)\,/\cos\phi$. Since the SE yield is proportional to the number of SEs generated within the escape region, δ is proportional to the escape volume, resulting in:

$$\delta(\phi) = \delta(0)/\cos\phi \tag{5.3}$$

Measurements of $\delta(\phi)$ support this inverse-cosine formula; see Fig. 5.5.

In qualitative terms, non-normal irradiation of a surface generates more SEs that lie within a *perpendicular* distance λ of the surface and can therefore escape into the vacuum. For a surface with topographical (height) variations (Fig. 5.6b), this orientation dependence of the yield δ results in protruding or recessed features appearing *bright in outline* in the SE image (Figs. 5.6c and 5.7a), somewhat similar to thickness-gradient contrast from a replica TEM specimen.

In practice, there is often some asymmetry if the SE detector is located to one side of the column (Fig. 5.1) rather than directly above. Surface features that are tilted *towards* the detector appear especially bright because electrons emitted from

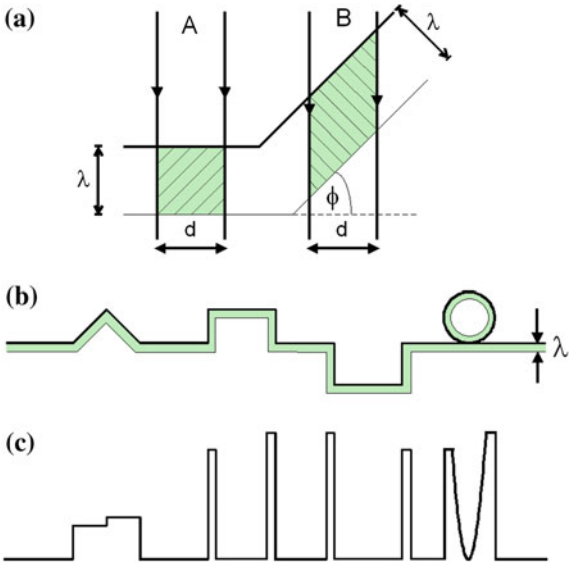


Fig. 5.6 **a** SEM incident beam normal to a specimen surface (at *A*) and inclined to the surface (at *B*). The volume from which secondaries can escape is proportional to the shaded cross-sectional area, which is λd for case A and $\lambda d/\cos\phi$ for a tilted surface (case B). **b** Cross-sectional diagram of a specimen surface that contains triangular and square protrusions, a square-shaped trough or well, and a spherical particle; λ is the secondary-electron escape depth. **c** Corresponding secondary-electron signal (from a line-scan along the surface), assuming an SE detector that is located to the right of the specimen

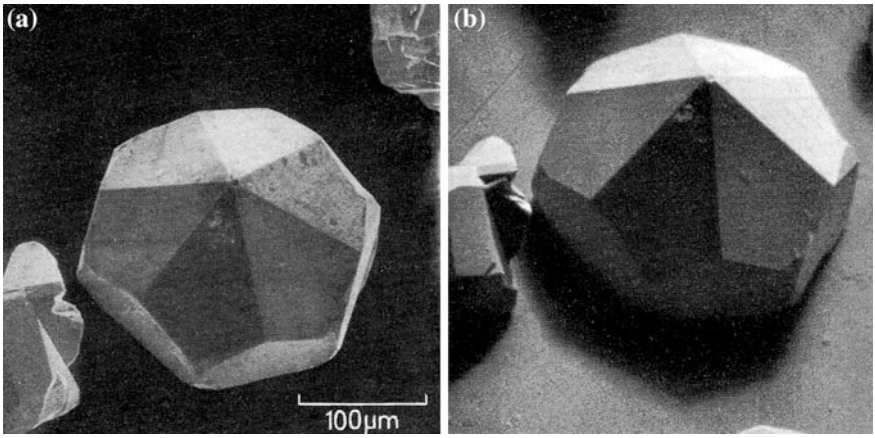


Fig. 5.7 **a** Secondary-electron image of a small crystal; the side-mounted Everhart-Thornley detector is located towards the top of the image. **b** Backscattered-electron image recorded by the same side-mounted detector, which also shows topographical contrast and shadowing effects. Such contrast is much weaker for a BSE detector mounted directly above the specimen. From Reimer (1998), courtesy of Springer-Verlag

those regions have a greater probability of reaching the detector; see Fig. 5.7a. This fact can be used to *distinguish* raised features and depressions in the surface of the specimen, as illustrated in Fig. 5.6c. In general, the SE image has a three-dimensional appearance, similar to that of a rough surface obliquely illuminated by light, which makes the topographical contrast relatively easy to interpret; see also Fig. 5.9a.

Taking advantage of the orientation dependence of δ, the *whole sample* is sometimes tilted away from a horizontal plane and towards the detector, as shown in Fig. 5.1. This increases the overall SE signal, averaged over all regions of the sample, while preserving the topographical contrast due to *differences* in surface orientation.

To further increase the SE signal, a positively biased electrode is used to attract secondary electrons away from the specimen. This electrode could be a simple metal plate, which would absorb the electrons and generate a small current that could be amplified and used to generate an SE image. However, the resulting SE signal would be very weak and noisy. The amount of electronic noise could be reduced by limiting the frequency response (bandwidth) of the amplifier, but the amplified signal would not follow fast changes in emission that occur when the electron probe is scanned rapidly over a non-uniform specimen.

A more satisfactory signal is obtained from an Everhart-Thornley detector, named after the scientists who first applied this design to the SEM. Secondary electrons are first attracted towards a wire-mesh electrode biased positively by a few hundred volts; see Fig. 5.8. Most of the electrons pass *through* this grid and are *accelerated* further towards a **scintillator**, at a positive bias V_s of several *thousand*

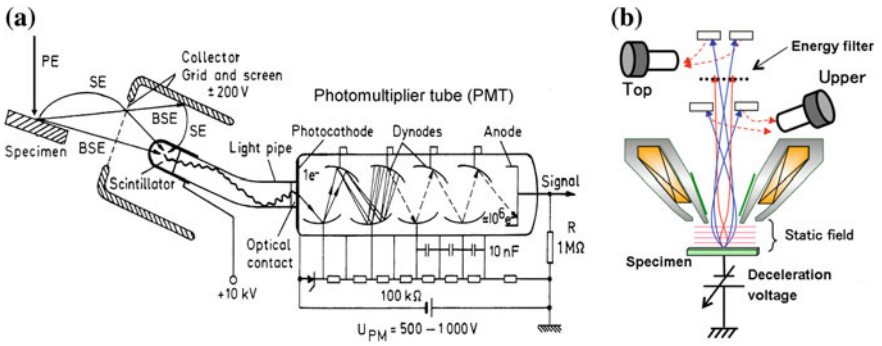


Fig. 5.8 **a** A typical scintillator/PMT (Everhart-Thornley) detector, used for SEM secondary-electron imaging, from Reimer (1998), courtesy of Springer-Verlag. **b** One type of through-the-lens detector, combined with specimen bias to achieve low landing energy. The "Top" and "Upper" detectors can record both secondary electrons *and* backscattered electrons (by conversion to secondaries at an aperture), depending on the electron energies involved and the bias voltage applied to the energy-filter grid. Courtesy of Hitachi High Technology

volts. The scintillator may be a layer of phosphor (similar to the coating on a TEM screen) on the end of a glass rod, or a light-emitting plastic or a garnet (oxide) material, each made conducting by a thin metallic surface coating. The scintillator has the property (**cathodoluminescence**) of emitting visible-light photons when bombarded with charged particles such as electrons. The number of photons generated by each electron depends on its kinetic energy E_k (= eV_s) and is of the order of 100 for $V_s \approx 10$ kV. The scintillator material has a high refractive index, so advantage can be taken of total internal reflection to guide the photons through a light pipe (a solid plastic or glass rod passing through a sealed port in the specimen chamber) to a **photomultiplier tube** (PMT) located outside the vacuum.

The PMT is a highly sensitive detector of visible (or ultraviolet) photons and consists of a sealed glass tube containing a hard (good-quality) vacuum. The light-entrance surface is coated internally with a thin layer of a material with a low work function, which acts as a **photocathode**. When photons are absorbed within the photocathode, they supply sufficient energy to liberate conduction or valence electrons, which may escape into the PMT vacuum as *photoelectrons*. These low-energy electrons are accelerated towards the first of a series of **dynode** electrodes, each biased positively with respect to the photocathode.

At the first dynode, biased at 100–200 V, the accelerated photoelectrons generate *secondary* electrons within their escape depth, just like primary electrons striking an SEM specimen. The dynodes are coated with a material with high SE yield (δ) so that at least two (sometimes as many as ten) secondary electrons are emitted for each photoelectron. The emitted secondaries are accelerated towards a *second* dynode, biased at least 100 V positive with respect to the first dynode, where each secondary produces at least two *new* secondaries. This process is

repeated at each of the n (typically eight) dynodes, resulting in a current amplification factor of $(\delta)^n$, typically $\approx 10^6$ for $n = 8$ and $\delta \approx 4$.

Because each secondary produced at the SEM specimen generated about 100 photoelectrons, the overall amplification (gain) of the PMT/scintillator combination can be as high as 10^8, depending on how much accelerating voltage is applied to the dynodes. Although this conversion of secondaries into photons, then into photoelectrons, and finally back into secondary electrons appears complicated, it is justified by the fact that the detector provides high amplification with relatively little added noise.

For many years, almost every SEM used the scintillator/PMT detector but other designs have recently become prominent. One reason is the quest for high spatial resolution, which (in the absence of aberration correction) requires an objective lens of small focal length. In a high-resolution SEM, the objective is an immersion lens, with the specimen very close and within its magnetic field, just as in the TEM. Secondary electrons released from the specimen spiral within this magnetic field and have little chance of reaching a side-mounted detector. Instead, the secondaries spiral through the objective field and can be collected by a detector mounted above the objective lens; see Fig. 5.8b. It is even possible to use the dispersive properties of the magnetic field to distinguish between fast and slow secondaries, as shown in Fig. 5.8b.

A second major SEM development has been the use of low primary-electron energies (down to and even below 1 keV), which are advantageous for examining uncoated insulating specimens; see Sect. 5.7. With suitable objective-lens design, a spatial resolution better than 1 nm can be achieved at $E_0 = 1$ keV, especially if the specimen is biased negatively so that primaries are focused at high energy but then decelerated to their final **landing energy** (which can be as low as 10 eV) in the resulting electrostatic field below the objective lens. This same electrostatic field *accelerates* the emitted secondary electrons through the objective-lens field and into a detector; see Fig. 5.8b.

Because the signal from such an in-lens (or through-the-lens) detector contains secondaries emitted equally with positive and negative values of ϕ (Fig. 5.6a), the SE image shows very little directional or shadowing effect, as illustrated in Fig. 5.9b.

In-lens detection is sometimes combined with energy filtering of the emitted secondary electrons. For example, perpendicular electric and magnetic fields (Sect. 7.3) can be used to select higher-energy secondaries, which consist mainly of the SE1 component that provides higher spatial resolution (Sect. 5.6).

5.4 Backscattered-Electron Images

A backscattered electron (BSE) is a primary electron that has been ejected from a solid by scattering through an angle greater than 90°. Such deflection could occur as a result of several collisions, some or all of which might involve a scattering angle

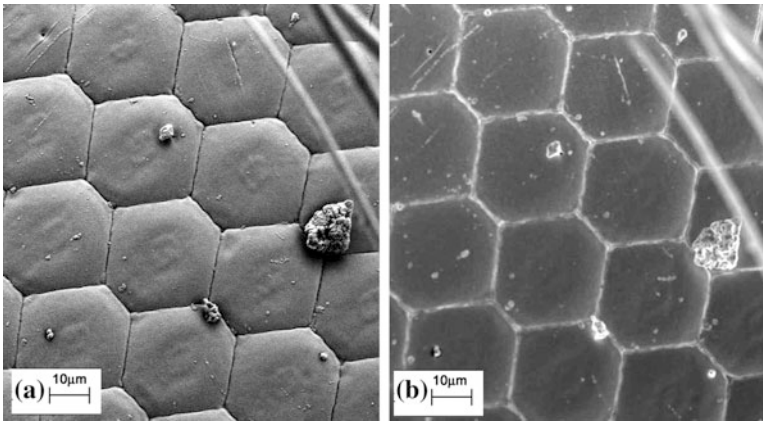


Fig. 5.9 Compound eye of an insect, coated with gold to make the specimen conducting. **a** SE image recorded by a side-mounted detector (located towards the top of the page) and showing a strong directional effect, including dark shadows visible below each dust particle. **b** SE image recorded by an in-lens detector, showing topographical contrast but almost no directional or shadowing effect. Courtesy of Peng Li, University of Alberta

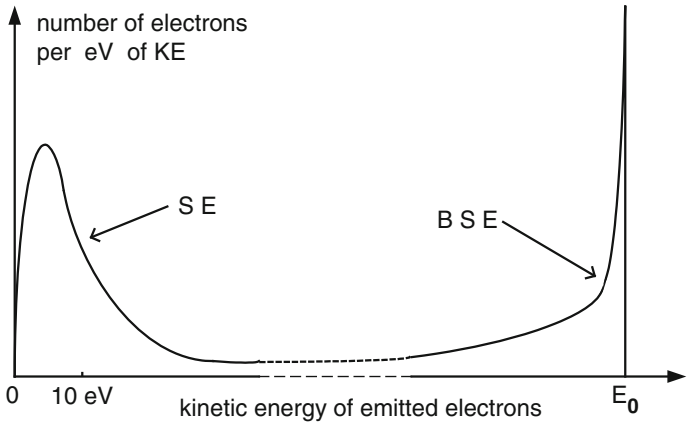


Fig. 5.10 Number of electrons emitted from the SEM specimen as a function of their kinetic energy, illustrating the conventional classification into secondary and backscattered components

of less than 90°; however, a *single* elastic event with $\theta > 90°$ is more likely the cause. Because elastic scattering involves only a *small* energy exchange, most BSEs escape from the sample with energies not too far below the primary-beam energy; see Fig. 5.10. The secondary and backscattered electrons can therefore be distinguished on the basis of their kinetic energy.

Because the cross-section for high-angle elastic scattering is proportional to Z^2, we might expect strong *atomic-number* contrast if backscattered electrons provide

the signal used to modulate the SEM-image intensity. In practice, the **backscattering coefficient** η (the fraction of primary electrons that escape as BSE) *does* increase with atomic number, but almost linearly for low *Z*. In general, BSE images show contrast due to variations in chemical composition of a specimen, whereas SE images reflect mainly its surface topography.

Another difference between the two kinds of images is the *depth* from which the information originates. In the case of a BSE image, the signal comes from a depth of about *half* the penetration depth (if generated at that depth, half the original primary energy is lost during the inward journey and half during the outward outward journey). For primary energies above 3 kV, this depth is some *tens or hundreds* of nm rather than the much smaller SE escape depth (≈1 nm).

Backscattered electrons (BSEs) can be detected by a scintillator/PMT detector if the bias on the first grid is made negative, to *repel* secondary electrons. BSEs are recorded if they impinge directly on the scintillator (causing light emission) or if they strike a surface *beyond* the grid and create secondaries that are then accelerated to the scintillator (see Fig. 5.8). The BSEs travel in almost a straight line, their high energy making them fairly unresponsive to electrostatic fields. Consequently, more of them reach a side-mounted detector if they are emitted from a surface inclined towards it, giving some topographical contrast, as in Fig. 5.7b. However, backscattered electrons are emitted over a *broad* angular range and only those emitted within a small solid angle (defined by the collector diameter) will reach the scintillator, resulting in a weak signal and a noisy image. Therefore a more efficient BSE detector is needed.

In the so-called **Robinson detector**, an annular (ring-shaped) scintillator is mounted immediately below the objective lens and just above the specimen; see Fig. 5.11a. The scintillator subtends a large solid angle and collects a substantial fraction of the backscattered electrons. Light is channeled by internal reflection

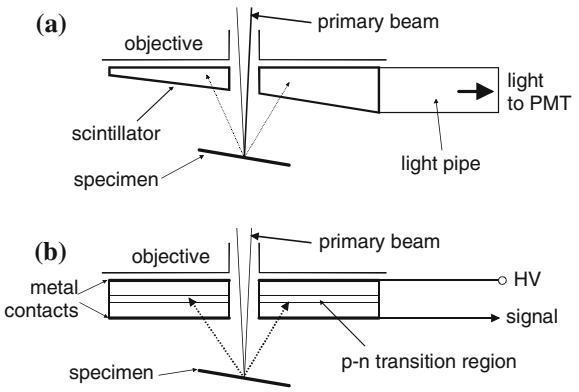


Fig. 5.11 Backscattered-electron detectors installed below the objective lens of an SEM: **a** annular-scintillator/PMT (Robinson) design, and **b** solid-state (semiconductor) detector

through a light pipe and into a PMT, as in the case of the Everhart-Thornley detector.

Another efficient BSE detector is shown in Fig. 5.11b. This **solid-state detector** consists of a large area (several cm^2) silicon diode, mounted just below the objective lens. Impurity atoms (arsenic, phosphorus) are added to the silicon to make it electrically conducting. The diode consists of an n-type layer, in which conduction is by electrons, and a p-type layer in which conduction is by holes (absence of electrons in an otherwise full valence band). At the interface between the p- and n-layers lies a transition region, where current carriers (electrons and holes) are absent because they have diffused across the interface. A voltage applied between the n- and p-regions (via metal surface electrodes) creates a high internal electric field across this high-resistivity region. If a backscattered electron arrives at the detector and penetrates to the transition region, some of its kinetic energy is used to excite electrons from the valence to the conduction band, creating mobile electrons and holes These free carriers move under the influence of the internal field, causing a current pulse to flow between the electrodes and in an external circuit. BSE arrival can therefore be measured by counting current pulses or by measuring the average current, which is proportional to the number of backscattered electrons arriving per second. Because secondary electrons have insufficient energy to reach the transition region, they do *not* contribute to the signal provided by the solid-state detector.

Because the Robinson and solid-state detectors are mounted directly above the specimen, their BSE signal contains very little topographic contrast but does show **Z-contrast** due to differences in local atomic number in the near-surface region of the specimen. The orientation of crystal planes (relative to the incident beam) also affects the electron penetration into the specimen, through diffraction effects. In the BSE image of a polycrystalline specimen, this effect gives rise to **orientation** (or channeling) **contrast** between grains in a polycrystalline specimen. It is also the basis of a technique called **electron-backscattering diffraction** (EBSD), where the angular dependence of the backscattering is recorded using a scintillator/camera system and computer software can produce a map showing the distribution of different crystallographic phases in the specimen.

5.5 Other SEM Imaging Modes

Although SE and BSE images suffice for most SEM applications, other kinds of signal can be used to modulate the image intensity, as we now illustrate with several examples.

A **specimen-current image** is obtained by using a specimen holder that is insulated from ground and connected to the input terminal of a sensitive current amplifier. Conservation of charge implies that the specimen current I_s flowing to ground (through the amplifier) must be equal to the primary-beam current I_p *minus* the rate of loss of electrons from secondary emission and backscattering:

$$I_s = I_p - I_{BSE} - I_{SE} = I_p(1 - \eta - \delta) \tag{5.4}$$

Whereas I_p remains constant, I_{BSE} and I_{SE} vary (due to variations in η and δ) as the probe scans across the specimen. Therefore the specimen-current image contains a mixture of *Z*-contrast and topographical information. To reduce the noise level of the image, the current amplifier must be limited in bandwidth (frequency range), requiring that the specimen be scanned slowly, resulting in a frame time of many seconds.

Electron-beam induced conductivity (**EBIC**) occurs when the primary-electron probe passes near a p-n junction in a semiconductor specimen, such as a silicon integrated circuit (IC) containing diodes and transistors. Additional electrons and holes are created, as in the case of a solid-state detector responding to backscattered electrons, resulting in current flow between two electrodes attached to the specimen surface. If this current is used as the signal that modulates the display device, the junction regions show up bright in the EBIC image. The p-n junctions in ICs are buried below the surface but provided they lie within the penetration depth of the primary electrons, an EBIC signal will be generated. It is even possible to utilize the dependence of penetration depth on primary energy E_0 to image junctions at different depths; see Fig. 5.12.

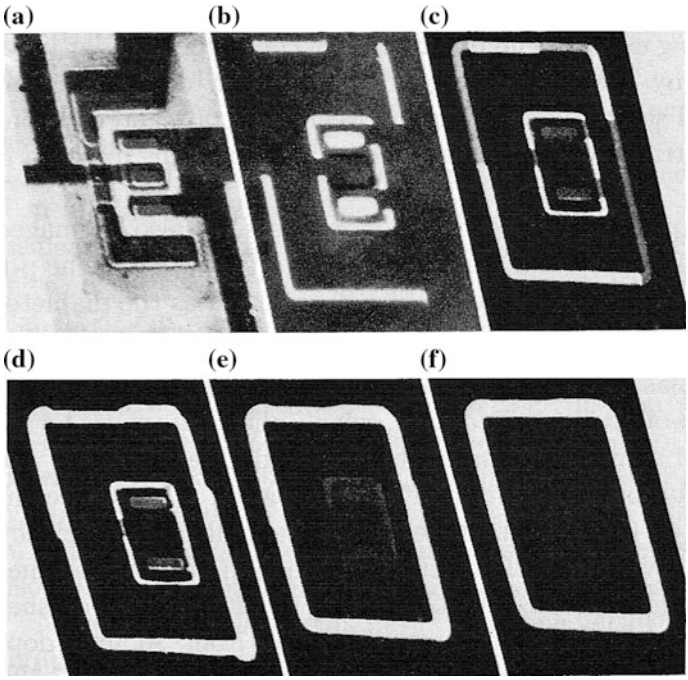


Fig. 5.12 Imaging of perpendicular p-n junctions in a MOS field-effect transistor (MOSFET). **a** SE image, **b–f** EBIC images for increasing primary-electron energy E_0 and therefore increasing penetration depth. Reproduced from Reimer (1998), courtesy of H. Raith and Springer-Verlag

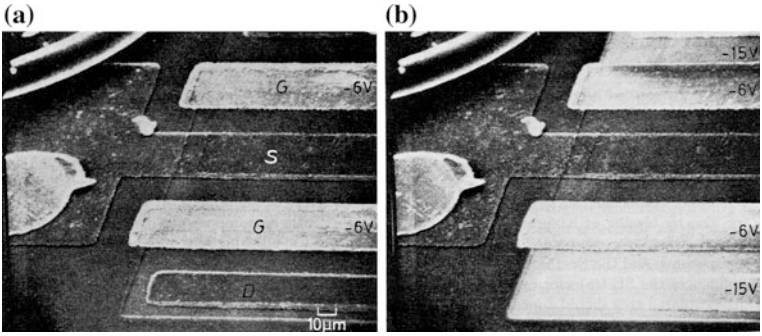


Fig. 5.13 SEM voltage-contrast image of an MOS field-effect transistor with **a** the gate electrode G at −6 V, source S and drain D electrodes grounded; **b** gate at –6 V, source and drain at −15 V. From Reimer (1998), courtesy of Springer-Verlag

Voltage contrast arises when voltages are applied to surface regions of a specimen, usually a semiconductor IC chip. The secondary-electron yield is reduced in regions that are biased positive, as lower-energy secondaries are attracted *back* to the specimen. Conversely, negative regions exhibit a higher SE yield (see Fig. 5.13) because secondaries are repelled and have a higher probability of reaching the detector. The voltage-contrast image is useful for checking whether *supply* voltages applied to an integrated circuit are reaching the appropriate locations. It can also be used to test whether a circuit is operating correctly, with *signal* voltages appearing in the right sequence. Although most ICs (such as microprocessors) operate at too high a frequency for their voltage cycles to be observed directly, this sequence can be slowed down and viewed in a TV-rate SEM image by use of a **stroboscopic** technique. By applying a square-wave current to deflection coils installed in the SEM column, the electron beam can be periodically deflected and intercepted by a suitably placed aperture. If this **chopping** of the beam is performed at a frequency that is slightly different from the operational frequency of the IC, the voltage cycle appears in the SE image at the *beat* frequency (the *difference* between the chopping and IC frequencies), which could be as low as one cycle per second.

As an alternative to collecting electrons to form an SEM image, it is sometimes possible to detect photons emitted from the specimen. As discussed in connection with a scintillator detector, some materials emit visible light in response to bombardment by electrons, the process known as **cathodoluminescence** (CL). In addition to phosphors (see Fig. 5.14), certain semiconductors fall into this category and may emit light uniformly *except* in regions containing crystal defects. In such specimens, CL images have been used to reveal the presence of dislocations, which appear as dark lines.

The CL signal could be detected by a photomultiplier tube, preceded by a color filter so that a limited range of photon wavelengths is recorded. However, spectrometers are available for attachment to an SEM and they provide precise spectral

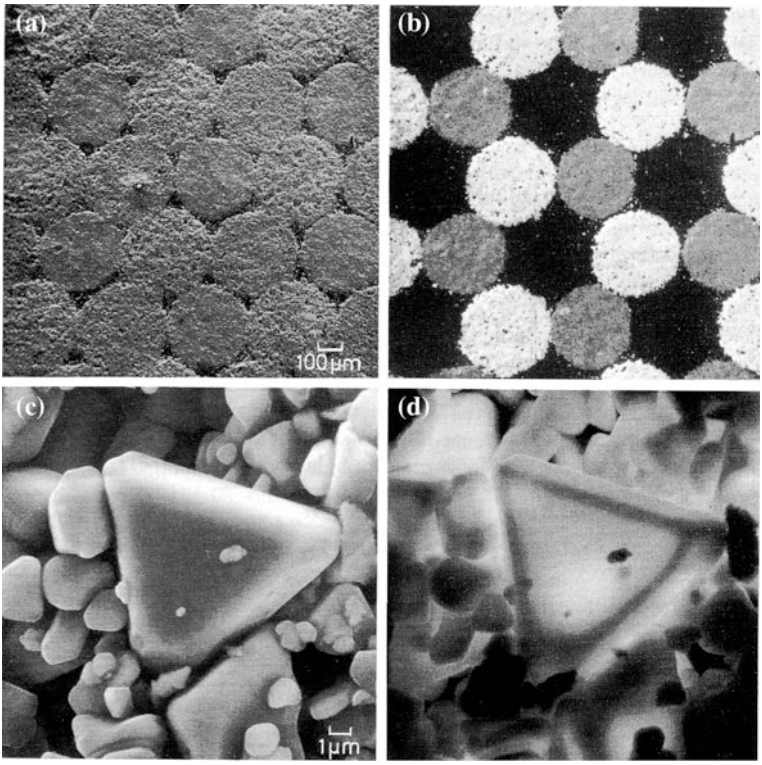


Fig. 5.14 *Red*, *green*, and *blue* phosphor dots in a color-TV screen, imaged using secondary electrons (**a** and **c**) and in CL mode (**b** and **d**) without wavelength filtering. The lower images show the individual grains of light-emitting phosphor imaged at higher magnification. From Reimer (1998), courtesy of J. Hersener, Th. Ricker and Springer-Verlag

information that can lead to the identification of crystalline structures or semiconductor properties. Cathodoluminescence is more efficient at low temperatures, so the specimen is sometimes cooled to below 20 K, using liquid helium as the refrigerant.

Visible light is emitted when an inelastic collision of a primary electron transfers *a few* eV of energy to an outer-shell (valence) electron, which then emits a photon while returning to its lowest-energy state. If the primary electron collides with an *inner-shell* electron, considerably more energy must be transferred to excite the atomic electron to a vacant energy level (an outer orbit or orbital) and then a photon of higher energy (hundreds or thousands of eV) may be emitted as a **characteristic x-ray photon**. The x-ray energy can be measured and used to identify the atomic number of the participating atom, as discussed in Chap. 6. If the characteristic x-ray signal is used to control the scanned-image intensity, the result is an **elemental map** showing the distribution of a particular chemical element within the SEM specimen.

5.6 SEM Operating Conditions

The SEM operator can control several parameters of the SEM, such as the electron-accelerating voltage, the distance of the specimen below the objective lens, known as the **working distance** (WD), and sometimes the diameter of the aperture used in the objective lens to control spherical aberration. The choice of these variables influences the resolution and the quality of the image obtained from the SEM.

The accelerating voltage determines the kinetic energy E_0 of the primary electrons, their penetration depth, and therefore the *information depth* of the BSE image. As we have seen, secondary electrons are generated within a very shallow escape depth below the specimen surface, therefore the SE image might be expected to be independent of the choice of E_0. However, only *part* of the SE signal (the so-called SE1 component) comes from the generation of secondaries by primary electrons close to the surface. Another component (SE2) represents secondaries generated by *backscattered* electrons, as they exit the specimen. A third component (SE3) arises from backscattered electrons that strike an internal surface of the specimen chamber; see Fig. 5.15a. As a result of these SE2 and SE3 components, secondary-electron images can show contrast from structure present well *below* the surface, but within the primary-electron penetration depth. This structure results from changes in backscattering coefficient, for example due to local differences in atomic number. Such effects are more prominent if the primary-electron penetration depth is large, in other words at a higher accelerating voltage, making the sample appear more "transparent" in the SE image; see Fig. 5.16. Conversely, if the accelerating voltage is reduced below 1 kV, the penetration depth can become very

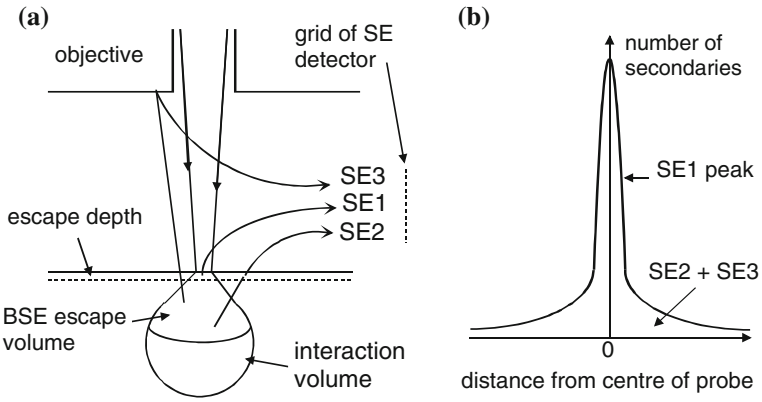


Fig. 5.15 **a** Generation of SE1 and SE2 electrons in a specimen, by primary electrons and by backscattered electrons, respectively. SE3 electrons are generated outside the specimen when a BSE strikes an internal SEM component, in this case the bottom of the objective lens. **b** SE-image resolution function, showing the relative contributions from secondaries generated at different distances from the center of the electron probe

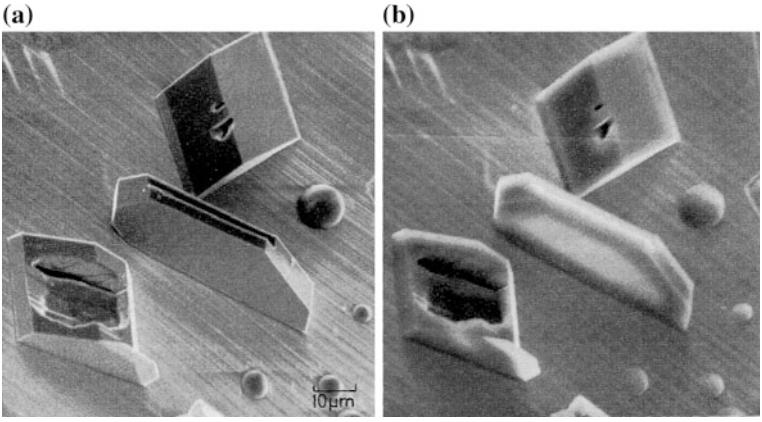


Fig. 5.16 SE images of tridymite crystals and halite spheres on a gold surface, recorded with an SEM accelerating voltage of **a** 10 kV and **b** 30 kV. Note the higher transparency at higher incident energy. From Reimer (1998), courtesy of R. Blaschke and Springer-Verlag

small (less than the secondary-electron escape depth) and only *surface features* are seen in the SE and BSE images.

Since the primary beam spreads laterally as it penetrates the specimen and because backscattering occurs over a broad angular range, some SE2 electrons are generated relatively far from the entrance of the incident probe and reflect the properties of a large portion of the primary-electron interaction volume. The SE3 component also depends on the amount of backscattering from this relatively large volume. In consequence, the spatial resolution of the SE2 and SE3 components is substantially worse than for the SE1 component; the SE2 and SE3 electrons contribute a *tail* or *skirt* to the **image-resolution function**; see Fig. 5.15b. Although this tail reduces the image contrast between closely spaced features, the existence of a *sharp central peak* in the resolution function ensures that *some* high-resolution information remains present in a secondary-electron image.

Because the incident beam spreads out very little within the SE1 escape depth, the width of this central peak (Fig. 5.15b) is approximately equal to the diameter d of the electron probe, which depends on the electron optics of the SEM column. To achieve high *de*magnification of the electron source, the objective lens is *strongly excited*, with a focal length below 2 cm. This in turn implies small C_s and C_c (see Chap. 2), which reduces broadening of the probe by spherical and chromatic aberration. Because of the high demagnification, the image distance of the objective is approximately equal to its focal length (see Sect. 3.5); in other words, the working distance needs to be small to achieve good SEM resolution. The smallest available values of d (below 1 nm) are achieved by employing a *field-emission* source, which provides an effective source diameter of only 10 nm (see Table 3.1), together with a working distance of only a few mm.

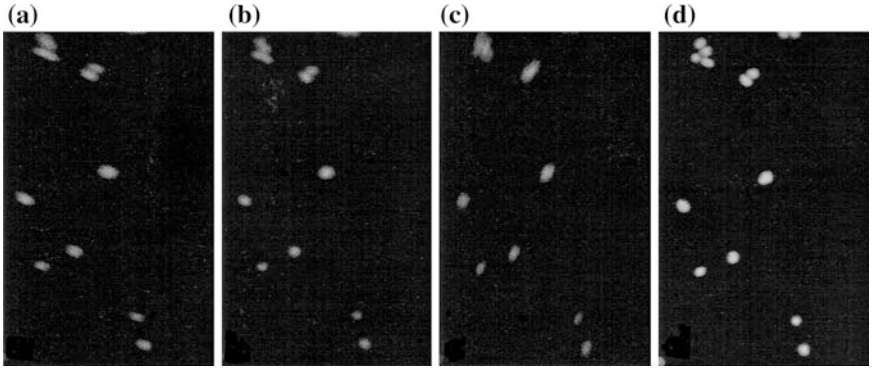


Fig. 5.17 **a–c** Astigmatic SE image showing the appearance of small particles as the objective current is changed slightly; in (**b**) the focus is correct, but the resolution is less than optimum. **d** Correctly focused image after astigmatism correction

Of course, good image resolution is obtained only if the SEM is correctly *focused*, which is done by carefully adjusting the objective-lens current. Small particles on the specimen offer a convenient feature for focusing; the objective lens current is adjusted until their images are as small and sharp as possible. Astigmatism of the SEM lenses can be corrected at the same time; the two stigmator controls are adjusted so that there is no streaking of image features as the image goes through focus (see Fig. 5.17), similar to the commonly used TEM procedure. In the SEM, zero astigmatism corresponds to a round (rather than elliptical) electron probe, equivalent to a radially symmetric resolution function.

As shown in Fig. 5.18, the SEM not only has better resolution than a light microscope but also has a greater **depth of field**. The latter can be defined as the change Δv in specimen height (or working distance) that produces a just-observable loss ($2\Delta r$) in image resolution. As seen from Fig. 5.19b, $\Delta r \approx \alpha/\Delta v$ where α is the convergence semi-angle of the probe. Taking $2\Delta r$ to be equal to the image resolution, so that the latter is degraded by a factor $\approx \sqrt{2}$ by the incorrect focus (quadrature addition, Sect. 1.1) and taking this resolution as equal to the probe diameter d (assuming SE imaging),

$$\Delta v \approx \Delta r/\alpha \approx d/(2\alpha) \tag{5.5}$$

The large SEM depth of field is seen to be a direct result of the relatively small convergence angle α of the electron probe, which in turn is dictated by the need to limit probe broadening due to spherical and chromatic aberration. As shown in Fig. 5.19a,

$$\alpha \approx D/(2v) \tag{5.6}$$

According to Eqs. (5.5) and (5.6), the depth of field Δv can be increased by increasing v or reducing D, although in either case there could be some loss of

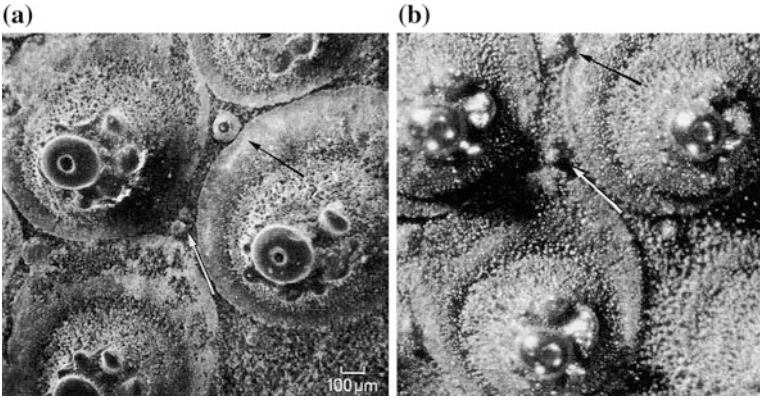


Fig. 5.18 **a** Low-magnification SEM image of a sea-urchin specimen, in which specimen features at different heights are all approximately in focus. **b** Light-microscope image of the same area, in which only one plane is in focus, other features (such as those indicated by *arrows*) appearing blurred. From Reimer (1998), courtesy of Springer-Verlag

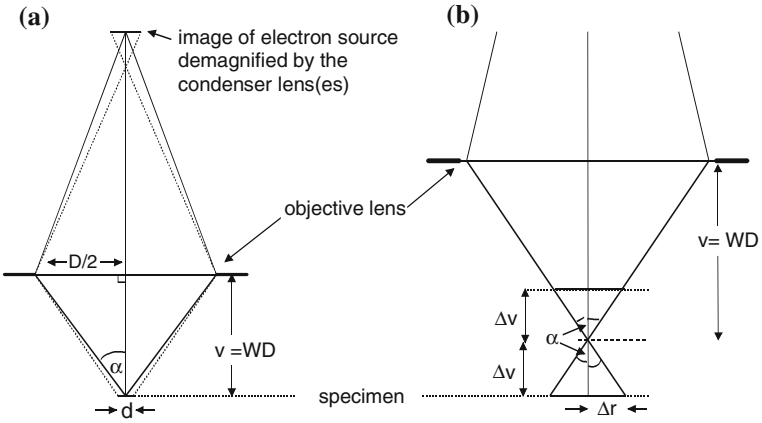


Fig. 5.19 **a** The formation of a focused probe of diameter d by the SEM objective lens. **b** Increase Δr in electron-probe radius for a plane located a distance Δv above or below the plane of focus

resolution because of increased diffraction effects (see page 4). Many SEMs allow the working distance to be changed, by height adjustment of the specimen relative to the lens column. Some microscopes also provide a choice of objective diaphragm, with apertures of more than one diameter.

Because of the large depth of field, the SEM specimen can be tilted away from the horizontal and towards the SE detector (to increase SE yield) without too much loss of resolution away from the center of the image. Even so, this resolution loss can be reduced to zero (in principle) by a technique called **dynamic focusing**. It involves applying the x- and/or y-scan signal (with an appropriate amplitude, which

depends on the angle of tilt) to the objective-current power supply, in order to ensure that the specimen surface remains in focus at all times during the scan. This procedure could be regarded as an adaptation of Maxwell's third rule of focusing (see Chap. 2) to deal with a non-perpendicular image plane. No similar option is available in the fixed-beam TEM, where the post-specimen lenses image all object points simultaneously.

5.7 SEM Specimen Preparation

One major advantage of the SEM (in comparison to a TEM) is the ease of specimen preparation, a result of the fact that the specimen does not have to be made thin. In fact, many conducting specimens require no special preparation before examination in the SEM. On the other hand, specimens of insulating materials do not provide a path to ground for the specimen current I_s and may undergo electrostatic charging when exposed to the electron probe. As seen from Eq. (5.4), this current can be of *either sign*, depending on the values of the backscattering coefficient η and secondary-electron yield δ. Therefore the local charge on the specimen can be positive or negative. Negative charge presents a more serious problem, since it repels the incident electrons and deflects the scanning probe, resulting in image distortion or fluctuations in image intensity.

One solution to the charging problem is to coat the surface of the SEM specimen with a thin film of metal or conducting carbon. This is done in vacuum, using the evaporation or sublimation technique already discussed in Sect. 4.10. Films of thickness 10–20 nm conduct sufficiently to prevent charging of most specimens. Since this thickness is greater than the SE escape depth, the SE signal comes from the *coating* rather than from the specimen material. However, the external contours of a very thin film closely follow those of the specimen, offering the possibility of a faithful topographical image. Gold and chromium are common coating materials. Evaporated carbon is also used; it has a low SE yield but an extremely small grain size so that *granularity* of the coating does not appear (as an artifact) in a high-magnification SE image, masking real specimen features.

Where coating is undesirable or difficult (for example, a specimen with very rough surfaces), specimen charging can often be avoided by careful choice of the SEM accelerating voltage. This option arises because the backscattering coefficient η and secondary-electron yield δ depend on the electron energy E_0. At high primary energy, the penetration depth is large and only a small fraction of the secondary electrons generated in the specimen can escape into the vacuum. In addition, many of the backscattered electrons are generated deep within the specimen and do not have enough energy to escape, so η will be low. A low total yield $(\eta + \delta)$ means that the specimen charges negatively, according to Eq. (5.4). As E_0 is reduced, δ increases and the specimen current I_s required to maintain charge neutrality eventually falls to zero at some incident energy E_2 corresponding to $(\eta + \delta) = 1$. Further reduction in E_0 could result in a positive charge but this would attract secondaries

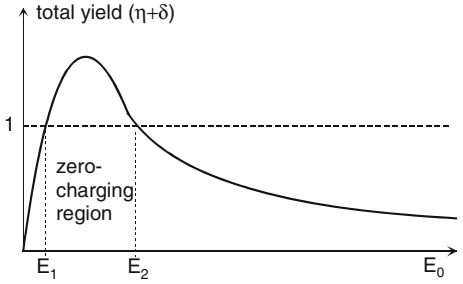


Fig. 5.20 Total electron yield (η + δ) as a function of primary energy, showing the range (E_1 to E_2) over which electrostatic charging of an insulating specimen is not a problem

back to the specimen, neutralizing the charge. So in practice, the specimen becomes slightly positive, with no harmful effect on the image quality.

For a very low incident energy (below some value E_1), the total yield falls below one because the primary electrons do not have enough energy to create secondaries. Since by definition $\eta < 1$, Eq. (5.4) indicates that negative charging will again occur. But in the favorable range $E_1 < E_0 < E_2$ negative charging is absent even for an insulating specimen, as shown in Fig. 5.20. Typically E_2 is in the range 1–10 keV. Although there are tables giving values for common materials (Joy and Joy, 1996; see Sect. 5.10), E_2 is usually found experimentally, by reducing the accelerating voltage until charging artifacts (distortion or pulsating of the image) disappear. E_1 is typically a few hundred volts, below the range of most SEM operation.

The use of a low accelerating voltage is therefore an attractive option for imaging insulating specimens. The main disadvantage of low E_0 is the increased chromatic-aberration broadening of the electron probe, given approximately by $r_c = C_c\alpha(\Delta E/E_0)$. This problem is minimized by using a field-emission source (with low-energy spread: $\Delta E < 0.5$ eV) and by careful design of the objective lens to reduce the chromatic-aberration coefficient C_c. The correction of chromatic aberration is also a practical possibility for the SEM, as well as for TEM (Sect. 2.7).

5.8 The Environmental SEM

An alternative approach to overcoming the specimen-charging problem is to surround the specimen with a gaseous environment rather than high vacuum. In this situation, the primary electrons ionize gas molecules before reaching the specimen. If the specimen charges negatively, positive ions are attracted towards it, largely neutralizing the surface charge.

Of course, there must still be a good vacuum within the SEM column in order to operate a thermionic or field-emission source, to enable high voltage to be used to accelerate the electrons and to allow focusing of electrons without scattering by gas molecules. In an **environmental SEM** (also called a low-vacuum SEM), primary electrons encounter gas molecules only during the last few mm of their journey,

after being focused by the objective lens. A small-diameter aperture in the bore of the objective allows the electrons to pass through but prevents most gas molecules from traveling up the SEM column. Those that do so are removed by continuous pumping. In some designs, a second pressure-differential aperture is placed just below the electron gun, to allow an adequate vacuum to be maintained in the gun, which has its own vacuum pump.

The pressure in the sample chamber can be as high as 5000 Pa (0.05 atmosphere) although a few hundred Pascals are more typical. The gas surrounding the specimen is often water vapor, since this choice allows *wet specimens* to be examined in the SEM without dehydration, provided the specimen-chamber pressure exceeds the saturated vapor pressure (SVP) of water at the temperature of the specimen. At 25 °C, the SVP of water is about 3000 Pa. However, the required pressure can be reduced by a factor of 5 or more by cooling the specimen, using a thermoelectric element incorporated into the specimen stage.

A backscattered-electron image can be obtained in the environmental SEM, using one of the detectors described previously. An Everhart-Thornley SE detector cannot be used because the voltage used to accelerate secondary electrons would cause electrical discharge within the specimen chamber. But if a potential of a few hundred volts is applied to a ring-shaped electrode just below the objective lens, secondary electrons initiate a *controlled* discharge between this electrode and the specimen, resulting in a current that can be amplified and used as the SE signal.

Examples of specimens that have been successfully imaged in the environmental SEM include plant and animal tissue (see Fig. 5.21), textile specimens (which charge easily in a regular SEM), rubber, and ceramics. Oily specimens can be

Fig. 5.21 Finger-like *villi* (providing high surface area for the absorption of nutrients) on the inner wall of the intestine of a mouse, imaged in an environmental SEM. The width of the image is 0.7 mm. Courtesy of ISI/Akashi Beam Technology Corporation

examined without contaminating the entire SEM; hydrocarbon molecules that escape through the differential aperture are quickly removed by the vacuum pumps.

The environmental specimen chamber extends the range of materials that can be examined by SEM and avoids the need for coating the specimen to make it conducting. The main drawback to ionizing gas molecules during the final phase of their journey is that the primary electrons are scattered and deflected from their original path. This effect adds an additional skirt (tail) to the current-density distribution of the electron probe, degrading the image resolution and contrast. Therefore an environmental SEM is usually operated as a high-vacuum SEM (by turning off the gas supply) in the case of conductive specimens that no not have a high vapor pressure.

5.9 Electron-Beam Lithography

Electrons can have a permanent effect on an electron-microscope specimen, known as radiation damage (see Sect. 7.2). Although the landing energies involved in SEM are low enough to make displacement damage unimportant, ionization damage (radiolysis) may occur for insulating specimens. This process can result in shrinkage and warping of the specimen, due to the release of light elements (mass loss), making it difficult to perform high-resolution microscopy SEM on organic materials such as polymers (plastics). Radiolysis also precludes the observation of living tissue at a subcellular level.

However, radiolysis is put to good use in polymer materials known as **resists**, whose purpose is to generate structures that are subsequently transferred to a material of interest in a process referred to as **lithography**. In a **positive** resist, the main radiation effect is bond breakage. As a result, the molecular weight of the polymer decreases and the material becomes more soluble in an organic solvent. An SEM can be modified slightly by connecting its x and y deflection coils to a pattern generator, allowing the electron beam to be scanned in a non-raster manner so that a pattern of radiation damage is produced in the polymer. If the polymer is a thin layer on the surface of a substrate, such as a silicon wafer, subsequent "development" in an organic solvent results in the scanned pattern appearing as a pattern of bare substrate. The *undissolved* areas of polymer form a barrier to chemical or ion-beam etching of the substrate (Sect. 4.10), the polymer acting as a "resist". After etching and removing the remaining resist, the patterned substrate can be used to make useful devices, often in large number. The fabrication of silicon integrated circuits (such as computer chips) makes use of this process, repeated many times to form complex multilayer structures, but using ultraviolet light as the radiation source. Electrons are used for smaller-scale projects, requiring high resolution but where production speed (throughput) is less important. Resist exposure is done using either an SEM or a specialized electron-beam writer.

In a **negative** resist, radiation causes an *increase* in the extent of chemical bonding (by "crosslinking" organic molecules), giving an increase in molecular weight and a

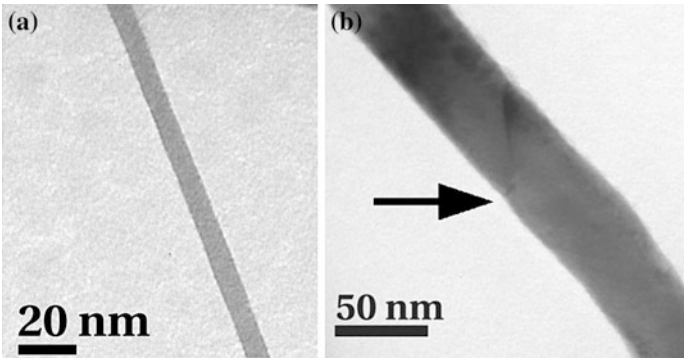


Fig. 5.22 **a** Hydrocarbon-contamination line written onto a thin substrate by scanning a focused probe of 200 keV electrons. **b** A broader line transferred to polycrystalline bismuth by argon-ion etching. The arrow indicates the position of a grain boundary in the bismuth. Courtesy of M. Malac, National Institute of Nanotechnology, Canada

reduction in solubility in exposed areas. An example of this process is electron-beam contamination (Sect. 3.6), in which hydrocarbon molecules adsorbed onto a surface are polymerized into a material of high molecular weight that is impossible to remove by most solvents. Even so, this generally unwanted effect can be put to good use by using the polymerized layer as an ion-beam resist; see Fig. 5.22. As in the case of positive resists, many types of negative resists are commercially available. The choice of the tone (negative or positive) and the chemical nature of the resist are dictated by the application.

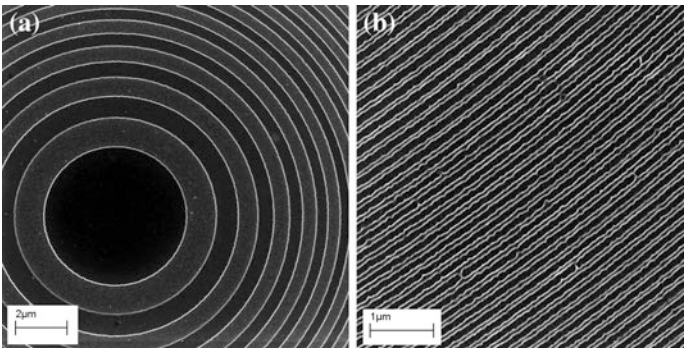


Fig. 5.23 **a** Central region of a zone plate, fabricated by e-beam lithography and imaged with secondary electrons. Bright rings are topographical contrast due to the step between bare Si substrate and PMMA-covered Si, but weak materials contrast is also present: the bare Si appears brighter because of its higher backscattering coefficient, giving a larger SE2 signal. **b** Outer region of the zone plate, where the spatial resolution and accuracy of the pattern represent an engineering challenge. Courtesy of Peng Li and Mirwais Aktary, Applied Nanotools Inc., Edmonton, Canada

For microelectronics and nanotechnology applications, patterns must be generated on a very fine scale and the spatial resolution of the pattern is of prime concern. Electrons can be focused to probes of very small diameter but when they penetrate a thick solid, the beam spreads laterally (Fig. 5.3) and backscattered electrons (which affect a resist coating the surface) cause a loss in resolution, just as in secondary-electron imaging. One solution is to use a thin substrate in which backscattering is minimal, allowing line widths as small as 10 nm; see Fig. 5.22. Another option is to use electrons of low incident energy, where the penetration and lateral spreading of the beam in the substrate are small. In addition to electronics applications, electron-beam lithography has been to fabricate zone plates for focusing x-rays; see Fig. 5.23.

5.10 Further Reading

Scanning Electron Microscopy by L. Reimer (2nd edition. Springer-Verlag, Heidelberg, 1998) provides a valuable account of the physical principles involved in SEM and is still available as Volume 45 of the Springer Series in Optical Sciences: ISBN: 978-3-642-08372-3 (Print) or 978-3-540-38967-5 (Online).

Scanning Electron Microscopy and X-ray Microanalysis (3rd Edition, Springer, 2003; ISBN 978-1-4613-4969-3, ISBN 978-1-4615-0215-9 (eBook), DOI 10.1007/978-1-4615-0215-9) contains many practical details of SEM operation. This book also deals with analytical microscopy in the SEM and contains three chapters on specimen preparation. For information on low-voltage SEM, see D.C. Joy and C.S. Joy, Micron 27 (1996) 247—263.

Monte Carlo Modeling for Electron Microscopy and Microanalysis by D.C. Joy (Oxford, 1995; ISBN 0-19-508874-3) gives examples of Monte Carlo codes and their use in understanding BSE and secondary-electron imaging in the SEM, as well as x-ray microanalysis. The free program CASINO is available from http://www.gel.usherbrooke.ca/casino/What.html.

A very readable introduction to the principles and application of electron-backscattering diffraction and orientation-contrast imaging is given by D.J. Prior and colleagues in American Mineralogist **84** (1999) 1741–1759.

Chapter 6
Analytical Electron Microscopy

The TEM and SEM techniques described in earlier chapters provide valuable information about the external or internal structure of a specimen, but little about its chemical composition. Some of the phenomena involved (diffracted electrons in TEM, BSE in the SEM) depend on the local atomic number Z but not sensitively enough to allow us to distinguish between adjacent elements in the periodic table. For that purpose, we need a signal that is highly Z-specific; for example, an effect that involves the electron-shell structure of an atom. This structure is clearly element-specific because it determines the chemistry of each element in the periodic table.

6.1 The Bohr Model of the Atom

A scientific *model* provides a means of accounting for the properties of an object, preferably using familiar concepts. To understand the electron-shell structure of an atom, we will use the *semi-classical* description of an atom given by Niels Bohr in 1913. Bohr's concept resembles the Rutherford planetary model in that it assumes the atom to consist of a central nucleus (charge = $+Ze$) surrounded by electrons that behave as *particles* (charge $-e$ and mass m). The attractive Coulomb force exerted on an electron (by the nucleus) supplies the centripetal force necessary to keep the electron in a circular orbit of radius r:

$$K(Ze)(e)/r^2 = mv^2/r \tag{6.1}$$

Here $K = 1/(4\pi\varepsilon_0)$ is the Coulomb constant and v is the tangential speed of the electron. The Bohr model *differs* from that of Rutherford by introducing the requirement that an orbit is *allowed* only if it satisfies the condition:

R.F. Egerton, *Physical Principles of Electron Microscopy*,
DOI 10.1007/978-3-319-39877-8_6

$$(mv)r = n(h/2\pi) \tag{6.2}$$

where $h = 6.63 \times 10^{-34}$ Js is the Planck constant. Since the left-hand side of Eq. (6.2) represents the angular momentum of the electron and because n is any *integer*, known as the (principal) **quantum number**, Eq. (6.2) represents the *quantization* of angular momentum. Without this condition, the atom is unstable: the centripetal *acceleration* of the electron would cause it to emit electromagnetic radiation and quickly spiral into the nucleus. A similar problem does not arise in connection with planets in the solar system because they are electrically neutral.

Using Eq. (6.2) to substitute for v in Eq. (6.1), the radius r_n of the orbit of quantum number n is:

$$r_n = n^2(h/2\pi)^2/\left(KmZe^2\right) = n^2 a_0 / Z \tag{6.3}$$

Here $a_0 = 0.053$ nm is the (first) **Bohr radius**, corresponding to the radius of a hydrogen atom ($Z = 1$) in its lowest-energy **ground state** ($n = 1$). We can use Eq. (6.2) to solve for the orbital speed v_n or, more usefully, employ Eq. (6.1) to calculate the total energy E_n of an orbiting electron as the sum of its kinetic and potential energies:

$$\begin{aligned} E_n &= mv^2/2 - K(Ze)(e)/r_n = KZe^2/(2r_n) - KZe^2/r_n = -K\,Z\,e^2/(2r_n) \\ &= -R\left(Z^2/n^2\right) \end{aligned} \tag{6.4}$$

Here $R = Ke^2/(2a_0) = 13.6$ eV is the **Rydberg energy**, the energy needed to remove the electron from a *hydrogen* atom ($Z^2 = 1$) in its ground state ($n = 1$) and therefore equal to the **ionization energy** of hydrogen.

An appealing feature of the Bohr model is that it provides an explanation of the photon-emission and photon-absorption spectra of hydrogen, in terms of electron transitions between allowed orbits or energy levels. When excess energy is imparted to a gas (e.g., by passing an electrical current, as in a low-pressure discharge), each atom can *absorb* only a quantized amount of energy, sufficient to excite its electron to an orbit of higher quantum number. In the **de-excitation** process, the atom *loses* energy and emits a photon of well-defined energy hf given by:

$$hf = -R\,Z^2/n_u^2 - (-RZ^2/n_l^2) = RZ^2(1/n_l^2 - 1/n_u^2) \tag{6.5}$$

where n_u and n_l are the quantum numbers of the *upper* and *lower* energy levels involved in the electron transition; see Fig. 6.1. Equation (6.5) predicts rather accurately the photon energies of the bright lines in the photoemission spectra of hydrogen ($Z = 1$); $n_l = 1$ corresponds to the Lyman series in the ultraviolet region, $n_l = 2$ to the Balmer series in the visible region, etc. Equation (6.5) also gives the energies of the dark (Fraunhofer) lines that result when white light is selectively *absorbed* by hydrogen gas, as when radiation generated in the interior of the sun passes through its outer atmosphere.

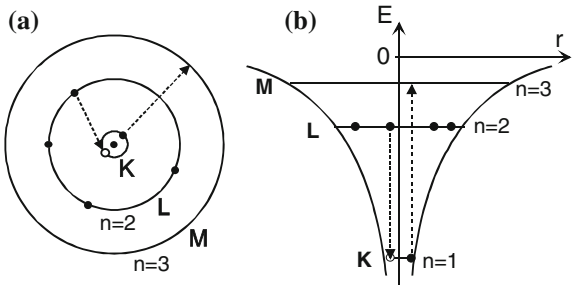


Fig. 6.1 Bohr model of a carbon atom, visualized in terms of **a** electron orbits and **b** the equivalent energy levels. The orbits are also described as electron shells and are designated as K (n = 1), L (n = 2), M (n = 3), etc. The *dashed arrows* illustrate a possible scenario for x-ray emission: a K-shell electron (n_l = 1) is excited to the empty M-shell (n_u = 3) and an L-shell electron fills the K-shell vacancy in a de-excitation process (n_l = 1, n_u = 2)

Unfortunately, Eq. (6.5) is not accurate for elements other than hydrogen, since Eq. (6.1) does not take any account of electrostatic interaction (repulsion) *between* the electrons that orbit the nucleus. To illustrate the importance of this interaction, Table 6.1 lists the ionization energy ($-E_1$) of the lowest-energy (n = 1) electron in several elements, as calculated from Eq. (6.4), and as determined experimentally from spectroscopic measurements.

Many-electron atoms represent a difficult theoretical problem because, in a classical (particle) model, the distance between the different orbiting electrons is always changing. In any event, a more realistic conception of the atom uses **wave mechanics**, treating the atomic electrons as *de Broglie* waves. Analysis then involves solving the Schroedinger wave equation to determine the electron **wave-functions**, represented by **orbitals** (pictured as charge-density clouds) that replace the concept of particle orbits. An exact solution is possible for hydrogen, and results in binding energies that are identical to those predicted by Eq. (6.4). Approximate methods are used for the other elements and in many cases the calculated energy levels are close to those determined from optical spectroscopy.

Other wave-mechanical principles determine the maximum number of electrons in each atomic shell: 2 for the innermost K-shell, 8 for the L-shell, 18 for the M-shell, etc. Since an atom in its *ground* state represents the *minimum-energy*

Table 6.1 K-shell (n = 1) ionization energies for several elements, expressed in eV

Element	Z	$-E_1$ (Bohr)	$-E_1$ (measured)
H	1	13.6	13.6
He	2	54.4	24.6
Li	3	122	54.4
C	6	490	285
Al	13	2298	1560
Cu	29	11,440	8979
Au	79	84,880	80,729

configuration, electrons fill these shells in sequence (with increasing atomic number), starting with the K-shell.

As indicated by Table 6.1, the measured energy levels differ substantially between different elements, resulting in photon energies (always a *difference* in energy between two atomic shells) that can be used to identify each element. Except for H, He, and Li, these photon energies are above 100 eV and lie within the *x-ray* region of the electromagnetic spectrum.

6.2 X-Ray Emission

When a primary electron enters a TEM or SEM specimen, it has a (small) probability of being scattered inelastically by an *inner-shell* (e.g., K-shell) electron, causing the latter to undergo a transition to a higher-energy orbit (or wave-mechanical state) and leaving the atom with an electron vacancy (hole) in its inner shell. However, the scattering atom remains in this *excited* state for only a very brief period of time: within about 10^{-15} s, one of the other atomic electrons fills the inner-shell vacancy by making a *downward* transition from a higher- energy level, as in Fig. 6.1b. In this de-excitation process, energy can be released in the form of a *photon* whose energy (*hf*) is given roughly by Eq. (6.5) but more accurately by the *actual* difference in binding energy between the upper and lower levels.

The energy of this **characteristic x-ray** photon therefore depends on the atomic number Z of the atom involved and on the quantum numbers (n_l, n_u) of the energy levels involved in the electron transition. Characteristic x-rays are traditionally classified according to the following historical scheme. The electron shell in which the original inner-shell vacancy was created, which corresponds to the quantum number n_l, is represented by an upper-case letter; thus K implies that $n_l = 1$, L implies $n_l = 2$, M implies $n_l = 3$, and so on. This Roman symbol is followed by a Greek letter that represents the *change in* quantum number: α denotes $(n_u - n_l) = 1$, β denotes $(n_u - n_l) = 2$, and γ denotes $(n_u - n_l) = 3$. Sometimes a numerical subscript is added to allow for the fact that some energy levels are split into components of slightly different energy, arising from quantum-mechanical effects.

The transition sequence represented in Fig. 6.1b would therefore result in a Kα x-ray being emitted, and in the case of carbon there is no other possibility. With an atom of higher atomic number, containing electrons in its M-shell, an M- to K-shell transition would result in a Kβ x-ray (of greater energy) being emitted. Similarly, a vacancy created (by inelastic scattering of a primary electron) in the L-shell might result in emission of an Lα photon. These possibilities are illustrated in Fig. 6.2, which shows the x-ray emission spectrum recorded from a TEM specimen consisting of a thin film of nickel oxide (NiO) deposited onto a thin-carbon film and supported on a molybdenum (Mo) TEM grid.

Figure 6.2 also demonstrates some general features of x-ray emission spectroscopy. Firstly, each element gives rise to *at least one* characteristic peak and can

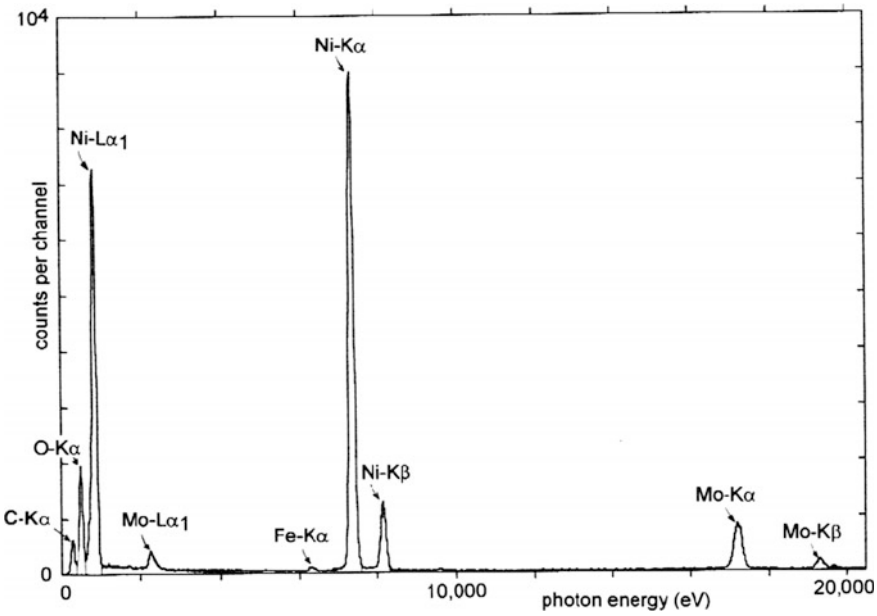


Fig. 6.2 X-ray emission spectrum (number of x-ray photons as a function of photon energy) recorded from a TEM specimen (NiO/C thin film on a Mo grid), showing characteristic peaks due to the elements C, O, Ni, Mo, and Fe. For Ni and Mo, both K- and L-peaks are visible. Mo peaks arise from the grid material; the weak Fe peak comes from scattered electrons hitting the polepieces of the TEM objective lens

be identified from the photon energy associated with this peak. Secondly, medium- and high-Z elements show *several* peaks (K, L, etc.); this complicates the spectrum but can be useful for multi-element specimens where some characteristic peaks may overlap with each other, making the measurement of elemental concentrations difficult if based only on a single peak per element. Thirdly, there are always a few stray electrons outside the focused electron probe (due to spherical aberration, for example), so the x-ray spectrum contains contributions from elements in the nearby environment, such as the TEM support grid or objective-lens polepieces.

High-Z atoms contain a large number of electron shells and can in principle give rise to *many* characteristic peaks. In practice, the number is reduced by the need to satisfy conservation of energy. For example, gold ($Z = 79$) has its K-emission peaks above 77 keV, so in an SEM, where the primary-electron energy is rarely above 30 keV, the primary electrons do not have enough energy to excite K-peaks in the x-ray spectrum.

The characteristic peaks in the x-ray emission spectrum are superimposed on a continuous background that arises from the **bremsstrahlung** process (German for braking radiation, implying deceleration of an electron). If a primary electron passes close to an atomic nucleus, it is elastically scattered and follows a curved (hyperbolic) trajectory, as discussed in Chap. 4. During its deflection, the electron

experiences a Coulomb force and a resulting centripetal acceleration towards the nucleus. According to the principles of electrodynamics, an accelerating charge (equivalent to a varying current) must generate electromagnetic radiation and lose energy. In this case, the acceleration and energy loss depends on the impact parameter of the electron, which is a *continuous* variable, slightly different for each primary electron. Therefore the photons emitted have a broad range of energy and form a background to the characteristic peaks in the x-ray emission spectrum. In Fig. 6.2, this bremsstrahlung background is low but is visible between the characteristic peaks at low photon energies.

Either a TEM or an SEM can be employed to generate an x-ray emission spectrum from a small region of a specimen. The SEM uses a thick (bulk) specimen, into which the electrons may penetrate several micrometers (at an accelerating voltage of 30 kV), so the x-ray intensity is higher than that obtained from the thin specimen used in a TEM. In both kinds of instruments, the volume of specimen emitting x-rays depends on the *diameter* of the primary beam, which can be made very small by focusing the beam into a probe of diameter 10 nm or less. In the case of the TEM, where the sample is thin and lateral spread of the beam (due to elastic scattering) is limited, the analyzed volume can be as small as 10^{-19} cm^2, allowing detection of less than 10^{-19} g of an element. In the SEM, x-rays are emitted from the entire interaction volume, which becomes larger as the incident energy of the electrons is increased (Fig. 5.3).

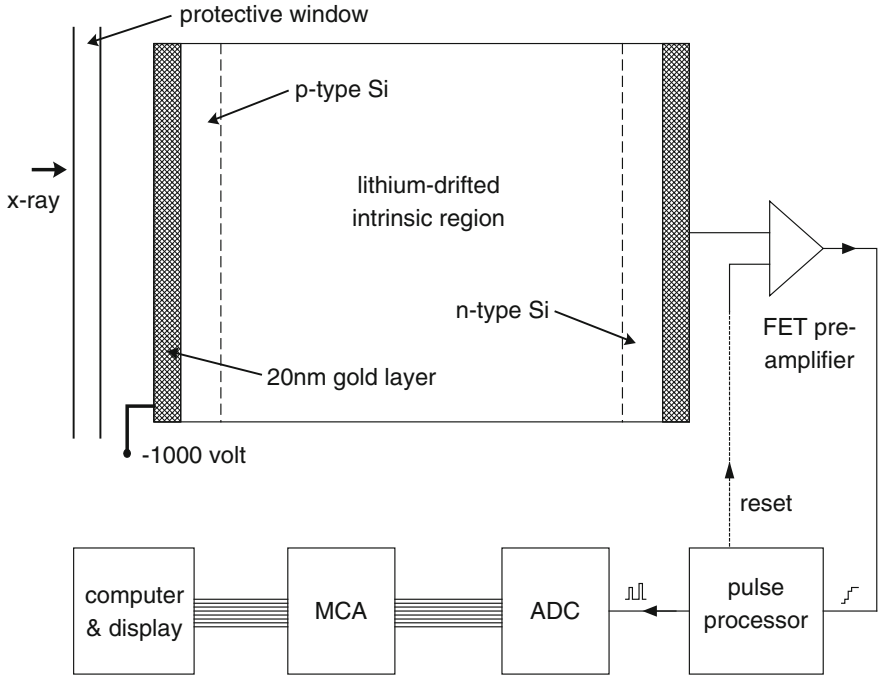


Fig. 6.3 Schematic diagram of a XEDS detector and its signal-processing circuitry

To observe the spectrum of the emitted x-rays, we need some form of *dispersive* device that distinguishes x-ray photons on the basis of their energy ($E = hf$, where f is the frequency of the electromagnetic wave) or their wavelength ($\lambda = c/f = hc/E$, where c is the speed of light in vacuum). Although photon energy and wavelength are closely related, these two options give rise to two distinct forms of spectroscopy, as discussed in Sects. 6.3 and 6.6.

6.3 X-Ray Energy-Dispersive Spectroscopy

In x-ray energy-dispersive spectroscopy(XEDS), the dispersive device is a semiconductor diode, fabricated from a single crystal of silicon (or germanium) and somewhat similar to the BSE detector used in an SEM. If an x-ray photon enters and is absorbed within the transition region between p- and n-doped material, its energy can release a considerable number of outer-shell (valence) electrons from the confinement of a particular atomic nucleus. The process is equivalent to exciting electrons from the valence band to the conduction band (i.e., the creation of electron-hole pairs) and results in electrical conduction by both electrons and holes for a brief period of time. With a reverse-bias voltage applied to the diode, this conduction causes electrical charge to flow through the junction (and around an external circuit), the charge being proportional to the number N of electron-hole pairs generated. Assuming that all of the photon energy (hf) goes into creating electron-hole (e-h) pairs, each pair requiring an *average* energy ΔE, energy conservation implies:

$$N = hf/\Delta E \tag{6.6}$$

For silicon, $\Delta E \approx 4$ eV (just over *twice* the energy gap between valence and conduction bands), therefore a Cu-Kα photon creates about (8000 eV)/(4 eV) = 2000 e–h pairs.

To ensure that essentially *all* of the incoming x-rays are absorbed and generate current pulses in an external circuit, the p-n transition region is made much *wider* than in most semiconductor diodes. In the case of silicon, this can be done by diffusing in the element lithium ($Z = 3$), which annihilates the effect of other electrically active (dopant) impurities and creates a high-resistivity (intrinsic) region several mm in width; see Fig. 6.3a. In recent years, this **Si(Li) detector** has been largely replaced by the **silicon drift detector** (SDD), whose back surface is patterned with ring-shaped electrodes at different voltages. Electric field within the high-resistivity Si accelerates mobile electrons towards a central anode, allowing a higher x-ray count rate.

If the detector were operated at room temperature, *thermal* generation of electron-hole pairs would add considerable electronic noise to the x-ray spectrum. Si(Li) detectors are cooled to about 140 K by conduction along a metal rod that ends in an insulated (dewar) vessel containing liquid nitrogen at 77 K. Silicon-drift

detectors are usually cooled to slightly below room temperature using a thermoelectric cooling element. To prevent water vapor and hydrocarbon molecules (present at low concentration in an SEM or TEM vacuum) from condensing onto the cold surface, a thin protective window is usually placed in front of the detector; see Fig. 6.3. This window was originally a thin (≈8 μm) layer of beryllium, whose low atomic number (Z = 4) allowed most x-ray photons to pass through without absorption. More recently, ultra-thin windows of a low-*Z* material (polymer, diamond, or metal nitride) are used to minimize the absorption of low-energy (<1000 eV) photons and allow the XEDS system to analyze elements of low atomic number via their K-emission peaks.

The current pulses from the detector crystal are fed into an adjacent preamplifier, also cooled to reduce electronic noise. This preamplifier may be a junction field-effect transistor (JFET) that acts as a *charge-integrating* amplifier; for each photon absorbed, its output voltage increases by an amount that is proportional to its input, which is proportional to *N* and to the photon energy, according to Eq. (6.6). The output of the FET is therefore a *staircase* waveform (Fig. 6.4a) with the height of each step proportional to the corresponding photon energy. Before the FET output reaches its saturation level, the voltage is reset to zero by applying an electrical (or optical) trigger signal to the FET. Complementary metal-oxide silicon (CMOS) devices are being used as an alternative to the JFET but they still operate as charge-sensitive preamplifiers, involving periodic reset.

A pulse-processing circuit (Fig. 6.3) converts each voltage step into a signal *pulse* whose height is proportional to the voltage step and therefore proportional to the photon energy; see Fig. 6.4b. An analog-to-digital converter (ADC) then produces a sequence of constant-height pulses from each signal pulse (Fig. 6.4c), their number being proportional to the input-pulse height. This procedure, called

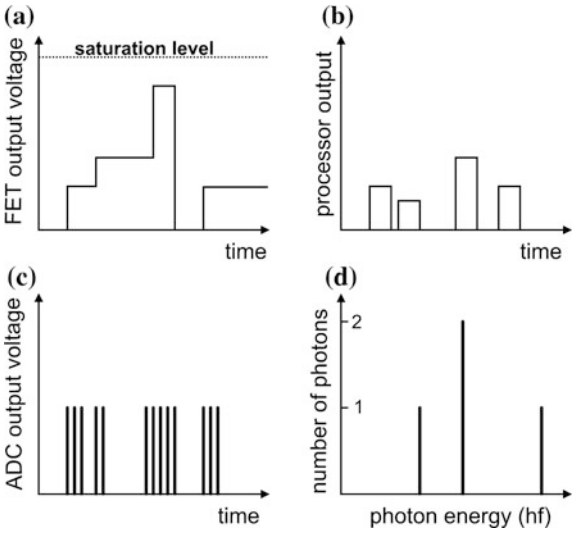


Fig. 6.4 Voltage output signals of **a** the FET preamplifer, **b** the pulse-processor circuit and **c** the analog-to-digital converter (ADC), resulting in **d** the XEDS-spectrum display

pulse-height analysis (PHA), is common in particle-physics instrumentation. Finally, a **multichannel analyzer** (MCA) circuit counts the number (n) of pulses in each ADC-output sequence and increments (by one count) the number stored at the n'th location (address) of computer memory, signifying the recording of a photon whose energy is proportional to n, and therefore to N and to photon energy. At regular time intervals, the computer-memory contents are read out as a spectrum display (Fig. 6.4d) in which the horizontal axis represents photon energy and the vertical scale represents the number of photons counted at that energy.

Typically, the x-ray spectrum contains $2^{12} = 4096$ memory locations (known as channels) and each channel corresponds to a 10 eV range of photon energy, in which case the entire spectrum extends from zero and just over 40 keV, sufficient to include almost all characteristic x-rays. Because the number N of electron-hole pairs produced in the detector is subject to a statistical variation, each characteristic peak has a width of several channels. The peaks therefore appear to have a smooth (Gaussian) profile, centered around the characteristic photon energy but with a width of 100–150 eV, as in Fig. 6.2.

Ideally, each peak in the XEDS spectrum represents an element present *within a known region of the specimen*, defined by the focused probe. In practice, there are often additional peaks due to elements beyond that region or even outside the specimen. Electrons that are backscattered (or in a TEM, forward-scattered through a large angle) strike objects (lens polepieces, parts of the specimen holder) in the immediate vicinity of the specimen and generate x-rays that are characteristic of those objects. Fe and Cu peaks can be produced in this way. In the case of the TEM, a special "analytical" specimen holders is used whose tip (surrounding the specimen) is made from beryllium ($Z = 4$) whose K-emission peak lies below 110 eV and is undetectable by most XEDS systems. Also, the TEM objective aperture is often removed during x-ray spectroscopy, in order to avoid generating backscattered electrons at the aperture diaphragm, which would bombard the specimen from below and produce x-rays far from the focused probe. Even so, spurious peaks sometimes appear, generated from thick regions at the edge of a thinned specimen or by a specimen-support grid. Therefore caution must be used in the interpretation of the x-ray spectrum, to avoid attributing these system peaks to elements in the specimen.

It takes several microseconds for the PHA circuitry to analyze the height of each pulse. Since x-ray photons enter the detector at random times, there is a certain probability of another x-ray photon arriving within this conversion time. To avoid generating a false reading, the PHA circuit ignores such double events, whose occurrence increases as the photon-arrival rate increases. A given recording time therefore consists of two components: **live time**, during which the system is processing data, and **dead time** during which the circuitry is made inactive. The beam current in the TEM or SEM should be kept low enough to ensure that the dead time is less than the live time, otherwise photon detection becomes inefficient; see Fig. 6.5.

The time taken to record a useful XEDS spectrum is dictated by the need to clearly identify the *position* of each significant peak (its characteristic energy) and,

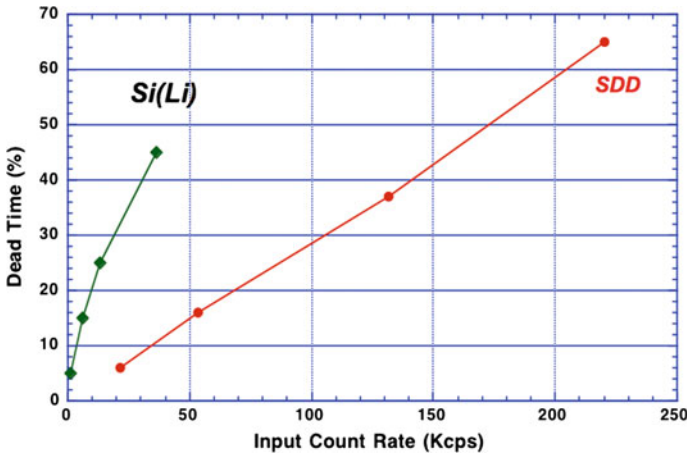


Fig. 6.5 Measurements of dead time as a function of input-photon rate, for SDD and Si(Li) detectors attached to a 300 keV analytical TEM. Courtesy of N.J. Zaluzec, Microscopy and Microanalysis meeting (2003), Microscopy Society of America

for measurement of elemental ratios, the *total number of photon counts* in each peak (the peak integral or area). If the recording time is too short, an insufficient number of x-ray photons will be analyzed and the spectrum will be "noisy". This electronic "shot" noise arises from the fact that the generation of x-rays in the specimen is a statistical process, giving rise to a statistical variation in the number of counts per channel in the XEDS spectrum. According to the rules of (Poisson) statistics, if N randomly arriving photons are recorded, we can expect a *variation* in that number (the standard deviation) equal to $N^{1/2}$ if the experiment is repeated many times. The accuracy of measuring N is therefore $N^{1/2}$ and the fractional accuracy is $N^{1/2}/N = N^{-1/2} = 0.1$ for $N = 100$. As a result, *at least* 100 x-ray counts are needed in a characteristic peak to measure the concentration of the corresponding element to an accuracy of 10 %. This condition may require a recording time of at least 10 s or even several minutes if the x-ray generation rate is low (e.g., with a thin specimen in the TEM).

One important feature of XEDS analysis is that it can provide a *quantitative* estimate of the concentration ratios of elements present in a specimen. The first requirement is to determine the number of characteristic x-ray photons contributing to each peak, by measuring the area (integral) of the peak and subtracting a contribution from the bremsstrahlung background. This background contribution is estimated by measuring the background on either side of the peak, then interpolating or (for greater accuracy) fitting the background to some mathematical function. However, the ratio of two peak integrals is *not* equal to the ratio of the elemental concentrations, since x-rays are not emitted with equal efficiency by different elements. The physics of x-ray production is simplest for the case of a very *thin* specimen, as used in the TEM, so we consider this situation first.

6.4 Quantitative Analysis in the TEM

For XEDS quantification, we need an equation that relates the number N_A of x-ray photons contributing to a characteristic peak of an element A to its concentration n_A (number of atoms of element A per unit volume). We can expect N_A to be proportional to the number N_e of electrons that pass through the sample during the spectrum-recording time, and to the product $n_A t$; for thin specimens, the probability of inner-shell inelastic scattering is just proportional to the specimen thickness t.

However, the product $n_A t$ (the areal density of element A) has dimensions of m^{-2} whereas N_A and N_e are dimensionless numbers. In order to balance the units in our equation, there must be an additional factor with dimensions of m^2. This factor is the **ionization cross-section** σ_A for creating a vacancy in a particular *inner shell* of element A. As discussed in Sect. 4.3, such a cross-section can be interpreted as a *target area* for inelastic scattering of an incident electron by the atom. The value of σ_A depends on the type of inner shell (K, L, etc.) as well as on the atomic number of the element.

Still missing from our equation is a factor ω known as the **fluorescence yield** that allows for the fact that not *every* inner-shell vacancy gives rise to the emission of an x-ray photon. An alternative de-excitation process allows the excited atom to return to its ground state by donating energy to another electron within the atom, which is ejected as an **Auger electron** (named after its discoverer, Pierre Auger). For elements of low atomic number, Auger emission is the more probable outcome and the *x-ray* fluorescence yield ω is considerably less than one ($\omega << 1.0$).

A remaining factor, known as the **collection efficiency** η of the XEDS detector, reflects the fact that x-ray photons are emitted equally in all directions and travel in straight-line paths, so only a fraction η of them reach the detector. Combining all of these factors, the number of x-ray photons contributing to the characteristic peak of an element A is:

$$N_A = (n_A t)\sigma_A \omega_A \eta N_e \tag{6.7}$$

For some other element B present in the sample, we could generate a second equation, identical to Eq. (6.7) but with subscripts B. Dividing Eq. (6.7) by this second equation gives an expression for the *concentration ratio* of the two elements:

$$n_A/n_B = [(\sigma_B \omega_B)/(\sigma_A \omega_A)]\,(N_A/N_B) \tag{6.8a}$$

where we assume that η is the same for both elements. The coefficient in square brackets is known as the ***k*-factor** of element A relative to element B. Although it is possible to *calculate* this *k*-factor from tabulated ionization cross-sections and fluorescence yields, a more accurate value is obtained by *measuring* the peak-intensity ratio (N_A/N_B) in a spectrum recorded from a "standard" sample (such as a binary compound) where the concentration ratio (n_A/n_B) is known. This measured *k*-factor includes any difference in collection efficiency between element

A and element B (due to different absorption of x-rays in the detector window, for example).

The use of standards is common in analytical chemistry. By convention, the concentration ratio in Eq. (6.8a) represents a *weight* ratio of two elements, rather than the *atomic* ratio n_A/n_B, and the *k*-factors reflect this convention. Such *k*-factors have been measured and tabulated for most elements but their precise values depend on the electron energy and detector design, so they are best measured using the *same* TEM/XEDS system as used for microanalysis. By using binary-element standards, the appropriate *k*-factors can be measured and used to determine the ratios of all elements present in a specimen of interest.

One of the problems of this *k*-factor method is that the necessary two-element standards are not always available. Therefore an alternative zeta-factor method has been developed, in which a sensitivity factor ζ (zeta) for *each* required element is measured by recording the x-ray intensity N_A from a pure-element film of thickness t and density ρ, using a recording time T and an electron probe whose current I is also measured, such that: $\zeta_A = (\rho t)(IT/N_A)$. Applied to a specimen of unknown composition, the *relative* concentration of each element A is then:

$$C_A = \zeta_A N_A/(\rho t I T) = \zeta_A N_A/(\Sigma_i \zeta_i N_i) \tag{6.8b}$$

where the denominator involves a sum over all elements present in the specimen. This method is capable of measuring concentrations to an accuracy of 1 %, together with the mass-thickness (ρt) of the TEM specimen.

6.5 Quantitative Analysis in the SEM

In the scanning electron microscope, electrons enter a thick specimen and penetrate a certain distance, which can exceed 1 μm (Sect. 5.2). Quantification is then complicated by the fact that many of the generated x-rays are absorbed before they leave the specimen. The amount of absorption depends on the *chemical composition* of the specimen, which is generally unknown, otherwise there is no need for microanalysis. However, Eq. (6.8) can be used to give an initial estimate of the ratios of *all* of the elements present in the sample, ignoring x-ray absorption. Then an **absorption correction** can be applied for each element, based on x-ray absorption coefficients that have been measured or calculated (as a function of photon energy) and published in graphical or tabulated form. This leads to a second set of elemental ratios, to which revised absorption corrections can be applied, and this iterative process is repeated until the elemental ratios converge to a stable result (within experimental error).

A further complication is x-ray **fluorescence**, a process in which x-ray photons are absorbed but generate photons of lower energy. These lower-energy photons can contribute spurious intensity to a characteristic peak, changing the measured elemental ratio. Again, the effect can be calculated, but only if the chemical

composition of the specimen is known. The solution is therefore to include fluorescence corrections, as well as absorption, in the iterative process described above. Since the ionization cross-sections and yields are *Z*-dependent, the whole procedure is known as **ZAF correction** (making allowance for atomic number, absorption, and fluorescence) and is carried out by running a ZAF program on the computer that handles the acquisition and display of the x-ray emission spectrum.

Because the TEM uses a very thin specimen, absorption is not important *except* for the very low-energy x-rays used to analyze light ($Z < 10$) elements. It can be incorporated into the zeta-factor quantification method by using an iterative scheme that is similar to the ZAF method.

6.6 X-Ray Wavelength-Dispersive Spectroscopy

In x-ray wavelength-dispersive spectroscopy (XWDS), characteristic x-rays are distinguished on the basis of their *wavelength* rather than their photon energy. This process relies on the fact that a crystal behaves as a three-dimensional diffraction grating and reflects (strongly diffracts) x-ray photons if their wavelength λ satisfies the Bragg equation:

$$n\lambda = 2d \sin \theta_i \tag{6.9}$$

As before, n is the *order* of the reflection (usually $n = 1$ is utilized) and θ_i is the angle between the incident x-ray beam and atomic planes (with a particular set of Miller indices) of spacing d in the crystal. Different x-ray wavelengths are selected in turn by continuously changing θ_i so that, with an appropriately located detector, the x-ray intensity is measured as a function of wavelength.

The way in which this is accomplished is shown in Fig. 6.6. A primary-electron beam enters a thick specimen (e.g., in an SEM) and generates x-rays, a small fraction of which travel towards the analyzing crystal. X-rays within a narrow wavelength range are Bragg-reflected, leave the crystal at an angle $2\theta_i$ relative to the incident x-ray beam, and arrive at the detector, usually a gas-flow tube (Fig. 6.6). When an x-ray photon enters the detector through a thin (e.g., beryllium or plastic) window, it is absorbed by a gas molecule within the tube, through the photoelectric effect. This process releases an energetic photoelectron, which ionizes other gas molecules as a result of inelastic scattering. All of the electrons are attracted towards a central wire electrode, connected to a +3 kV supply, causing a current pulse to flow in the power-supply circuit. The pulses are counted to give a signal proportional to the count rate, which represents the x-ray intensity at a particular wavelength.

Because longer-wavelength x-rays are absorbed in air, the detector and the analyzing crystal are held within the microscope vacuum. Gas (often a mixture of argon and methane) is supplied continuously to the detector tube to maintain an internal pressure around 1 atmosphere.

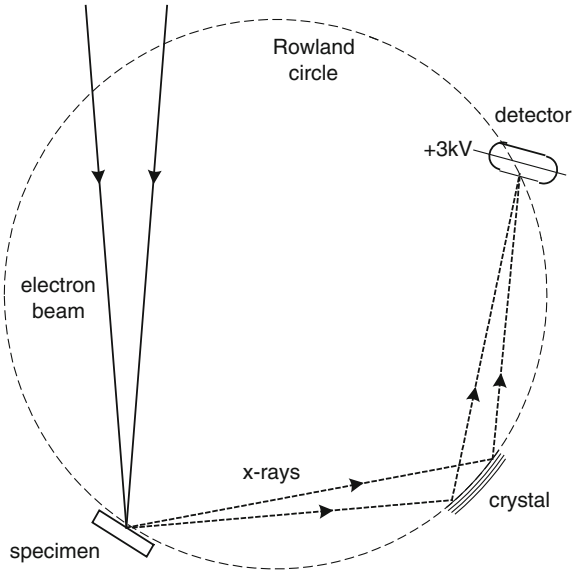


Fig. 6.6 Schematic diagram of an XWDS system. For simplicity, the Bragg-reflecting planes are shown as being parallel to the surface of the analyzing crystal. The crystal and detector move around the *Rowland circle* (*dashed*) in order to record the spectrum

To *decrease* the detected wavelength, the crystal is moved *towards* the specimen, along the arc of a circle (known as a Rowland circle; see Fig. 6.6) such that the x-ray angle of incidence θ_i is *reduced*. At the same time, the detector is moved towards the crystal, also along the Rowland circle, so that the Bragg-reflected beam enters the detector window. Because the deflection angle of an x-ray beam which undergoes Bragg scattering is $2\theta_i$, the detector must be moved at *twice* the angular speed of the crystal to keep the reflected beam at the center of the detector.

The mechanical range of rotation is limited by practical considerations; it is not possible to cover the entire spectral range of interest ($\lambda \approx 0.1$–1 nm) with a single analyzing crystal. Many XWDS systems are therefore equipped with several crystals of different *d*-spacing, such as lithium fluoride, quartz, and organic compounds. To make the angle of incidence θ_i the same for x-rays arriving at different angles, the analyzing crystal is bent (by applying a mechanical force) into a slightly curved shape, which also provides a degree of focusing similar to that of a concave mirror; see Fig. 6.6.

Although XWDS can be done by adding the appropriate mechanical and electronic systems to an SEM, there exists a more specialized instrument, the **electron-probe microanalyzer** (EPMA), which is optimized for such work. It generates an electron probe with a substantial current, as required to record an x-ray spectrum with good statistics (low noise) within a reasonable time period. The EPMA is fitted with several analyzer/detector assemblies, so that more than a single wavelength range can be recorded simultaneously, and sometimes with a XEDS detector also.

6.7 Comparison of XEDS and XWDS Analysis

A major advantage of the XEDS technique is the speed of data acquisition, due largely to the fact that x-rays within a wide energy range can be detected and analyzed almost simultaneously. In contrast, the XWDS system examines only one wavelength at a time and may take many minutes to scan the necessary wavelength range. The XEDS detector can be brought very close (within a few mm) of the specimen, allowing 1 % or even 10 % of the emitted x-rays to be analyzed, whereas the XWDS analyzing crystal needs room to move and subtends a smaller solid angle, so that considerably less than 1 % of the x-ray photons are collected. A commercial XEDS system is also considerably less expensive than a XWDS attachment to an SEM or a separate EPMA machine.

The main advantage of the XWDS system is that it provides *narrow* x-ray peaks (width ≈ 5 eV, rather than >100 eV for XEDS system) that stand out clearly from the background; see Fig. 6.7. This is particularly important when analyzing low-*Z* elements (whose K-peaks are more narrowly spaced) or elements present at low concentration, whose XEDS peaks would have a low signal/background ratio, resulting in a poor quantification accuracy. Consequently, XWDS analysis is the preferred method for measuring *low* concentrations: 200 parts per million (ppm) is fairly routine and 1 ppm is detectable in special cases. At higher concentrations, the narrow XWDS peak allows a measurement accuracy of better than 1 % with the aid of ZAF corrections. In contrast, the accuracy of XEDS analysis is typically around 10 %.

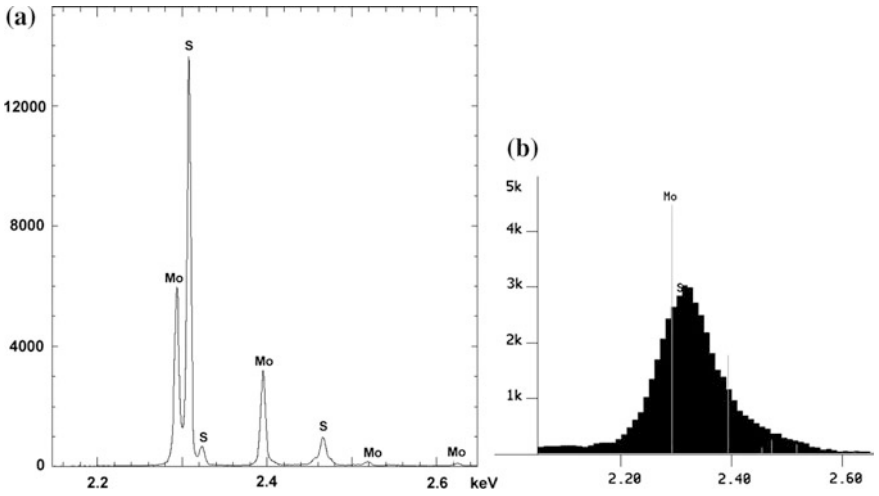


Fig. 6.7 X-ray intensity as a function of photon energy (in the range 2.1–2.7 keV), recorded from a molybdenum disulfide (MoS_2) specimen using **a** XWDS and **b** XEDS. The wavelength-dispersed spectrum displays peaks due to Mo and S, whereas these peaks are not separately resolved in the energy-dispersed spectrum, due to insufficient energy resolution. Courtesy of S. Matveev, Electron Microprobe Laboratory, University of Alberta

6.8 Auger-Electron Spectroscopy

Both XEDS and XWDS spectroscopy encounter problems when applied to the analysis of elements of low atomic number ($Z < 11$) such as B, C, N, O, F. Ultra-thin window XEDS detectors make these elements visible but the characteristic-peak intensities are reduced because of the low x-ray fluorescence yield, which falls continuously with decreasing atomic number; see Fig. 6.8. In addition, the K-emission peaks (which must be used for low-Z elements) occur at energies below 1 keV. In this restricted energy region, the bremsstrahlung background is relatively high and XEDS peaks from different elements tend to overlap. Also, the low-energy x-rays are strongly absorbed, even within a *thin* specimen, requiring a substantial correction for absorption that is not always accurate because the specimen geometry is not completely known.

One solution to the fluorescence-yield problem is to use Auger electrons as the characteristic signal. Because the x-ray yield ω is low, the Auger yield $(1 - \omega)$ is close to one for low-Z elements. Auger electrons can be collected and analyzed using an electrostatic spectrometer, which measures their kinetic energy. However, the Auger electrons of interest have relatively low energy (below 1000 eV) and are absorbed (through inelastic scattering) if they are generated more than one or two nm below the surface of the specimen, just like SE1 electrons. In consequence, the Auger signal measures a chemical composition of the surface of a specimen, which can be different from the composition of underlying material. For example, most specimens accumulate several monolayers of hydrocarbons when exposed to normal air, and some of this material remains on the surface when a specimen is placed in vacuum for electron-beam analysis.

As a result, Auger analysis is only useful for surfaces *prepared* in vacuum, by mechanical cleavage, by Ar-ion sputtering, or by vacuum deposition. Then a sufficiently good vacuum is needed to *maintain* the cleanliness of the surface during measurement; in a typical electron-microscope vacuum of 10^{-4} Pa, about one monolayer of ambient gas molecules arrives at the surface per second and some of these molecules remain bonded to the solid. An Auger microscope uses ultra-high

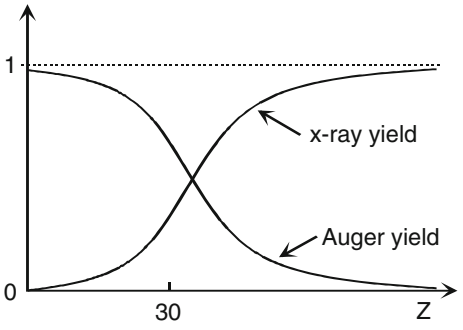


Fig. 6.8 X-ray fluorescence yield (ω), Auger yield and their sum (*dashed line*) for K-electron excitation, as a function of atomic number

vacuum (UHV, $\approx 10^{-8}$ Pa) and can provide surface analysis (without the need for a thin specimen) with a lateral resolution of about 50 nm.

6.9 Electron Energy-Loss Spectroscopy

A third possibility for microanalysis of light (low-*Z*) elements is to perform spectroscopy on the primary electrons that have passed through a thin TEM specimen. Since these electrons are responsible for creating (in an inelastic collision) the inner-shell vacancies that give rise to characteristic x-rays or Auger electrons, they carry atomic-number information in the form of the amount of energy transferred to the specimen. The *number* of electrons with a given *characteristic* energy loss is just equal to the *sum* of the number of x-ray photons and Auger electrons generated within the specimen, so signal strength is not a major problem.

To perform **electron energy-loss spectroscopy** (EELS), it is necessary to detect very small fractional differences in kinetic energy. Primary electrons enter a TEM specimen with a kinetic energy of 100 keV or more and the majority of them emerge with an energy that is lower by an amount between few eV and a few hundred eV. The only form of spectrometer that can achieve sufficient fractional resolution (of the order of one part in 10^5) is a **magnetic prism**, in which a uniform magnetic field ($B \approx 0.01$ Wb) is produced between two parallel faces of an electromagnet. This field exerts a force $F = evB$ on each electron, whose magnitude is constant (the speed v of the electron remains unchanged) and whose direction is always perpendicular to the direction of travel of the electron. This magnetic force supplies the centripetal force necessary for motion in a circle of radius R:

$$evB = F = mv^2/R \tag{6.10}$$

in which m is the relativistic mass of the electron. Because $R = [m/(eB)]v$ depends on the electron speed and kinetic energy, electrons that lose energy in the specimen (through inelastic scattering) have smaller R and are deflected through a slightly larger angle than those that are elastically scattered or remain unscattered. The spectrometer therefore *disperses* the electron beam, similar to the action of a glass prism on a beam of white light.

Unlike a glass prism, the magnetic prism also focuses the electrons, bringing those of the same energy together at the exit of the spectrometer, to form an electron energy-loss spectrum of the specimen. This spectrum can be recorded electronically by a scintillator followed by a CCD camera; see Fig. 6.9. The electrical signal from the camera is fed into a computer, which stores and displays the number of electrons as a function of their energy loss in the specimen.

A typical energy-loss spectrum (Fig. 6.10) contains a **zero-loss peak**, representing electrons that remained *unscattered* or were scattered *elastically* while passing through the specimen. Below 50 eV, one or more peaks represent *inelastic* scattering by *outer-shell* (valence or conduction) electrons in the specimen.

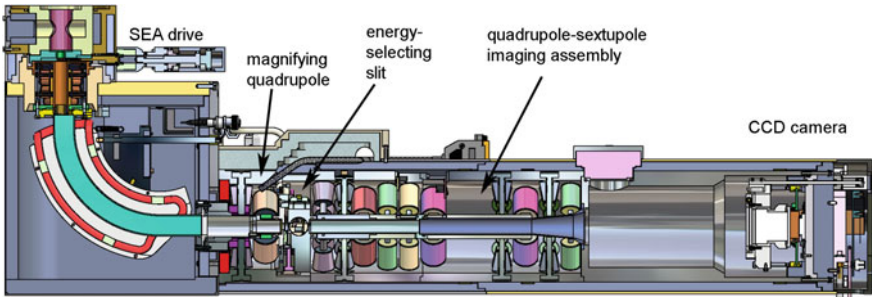


Fig. 6.9 Gatan GIF-Tridiem electron energy-loss spectrometer, which is mounted below the TEM viewing screen. SEA denotes a spectrometer entrance aperture. A uniform magnetic field (perpendicular to the plane of the diagram) bends the electron trajectories through 90° and introduces focusing and dispersion. The dispersion is magnified by quadrupole lenses and a phosphor screen converts the energy-loss spectrum into a photon-intensity distribution that is imaged by a CCD camera. An energy-selecting slit can be introduced at the spectrum plane in order to produce an energy-filtered TEM image or diffraction pattern. Courtesy of Gatan Inc.

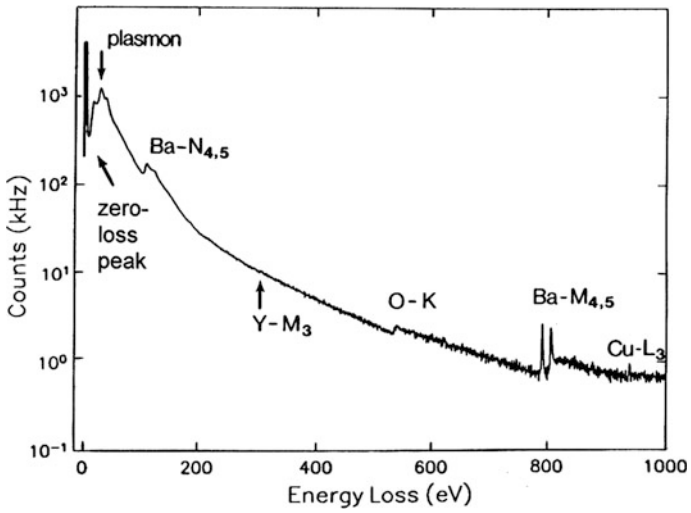


Fig. 6.10 Electron energy-loss spectrum of a high-temperature superconductor, showing N- and M-shell ionization edges of barium, the K-ionization edge of oxygen, and weak ionization edges from copper and yttrium. The plasmon peak represents inelastic scattering from valence electrons. A logarithmic vertical scale is used here to accommodate the large range of electron intensity. Courtesy of D.H. Shin (Ph.D thesis, Cornell University)

This scattering can take the form of a *collective* oscillation (resonance) of *many* outer-shell electrons and is known as a **plasmon** excitation. Inelastic excitation of *inner-shell* electrons causes an abrupt increase in electron intensity (an **ionization edge**) at an energy loss equal to the inner-shell ionization energy. Since this energy

is characteristic of a particular chemical element and is known for every electron shell, the threshold energy of each ionization edge reveals a particular element present within the specimen. Beyond each ionization threshold, the spectral intensity decays more gradually towards the (extrapolated) pre-edge background that arises from electron shells of lower ionization energy. The energy-loss intensity can be integrated (over a region of width typically 50–100 eV) and the background component subtracted to give a signal I_A that is proportional to the concentration of the element A giving rise to the edge. The concentration *ratio* of two different elements (A, B) is given by

$$n_A/n_B = (I_A/ I_B)(\sigma_B/\sigma_A) \tag{6.11}$$

where σ_A and σ_B are ionization cross-sections that can be calculated, knowing the atomic number, type of shell, and the integration range used for each element.

Because the magnetic prism has focusing properties, an electron spectrometer can also be used to form an image from electrons that have undergone a particular energy loss in the specimen. Choosing this energy loss to correspond to the ionization edge of a known element, it is possible to form an **elemental map** that represents the distribution of that element, with a spatial resolution of 1 nm or even atomic dimensions.

Other information is present in the low-loss region of the spectrum, below 50 eV. The amount of inelastic scattering depends on the specimen thickness, so the low-loss intensity relative to the zero-loss peak (Fig. 6.10) can be used to measure local thickness at a known location in a TEM specimen.

The energy-loss spectrum also contains a wealth of information about the atomic arrangement and chemical bonding in the specimen, present in the form of *fine structure* in the low-loss region and the ionization edges. For example, the intensities of the two sharp peaks at the onset of the Ba M-edge in Fig. 6.10 can be used to determine the charge state (valency) of the barium atoms

In most situations, the EELS signal is dominated by **dipole scattering**: the electric field of a primary electron (moving with energy E_0) creates a transient electric dipole within the specimen. For an energy loss E, the inelastic intensity I per unit solid angle Ω is then a Lorentzian function of scattering angle θ:

$$dI/d\Omega \propto 1/(\theta^2 + \theta_E^2) \tag{6.12}$$

where $\theta_E \approx E/(2E_0)$ is a characteristic angle. Even if E is as large as several hundred eV, the angular width of the scattering θ_E is only a few mrad (<0.1°), so only low-angle scattering has to be recorded by the electron spectrometer. Therefore a STEM instrument that produces a dark-field image from high-angle elastic scattering (Sect. 1.6) can simultaneously record an EELS signal that passes though the central hole of the HAADF detector. In fact it is common nowadays to record an

energy-loss spectrum from each pixel of the STEM image, giving **spectrum-image** data that can be stored and subsequently analyzed in detail.

Recent designs of electron monochromator allow an energy resolution of the order of 0.01 eV, highly useful for analyzing EELS fine structure. This energy resolution also permits the detection of vibrational (phonon) modes of energy loss, which show up as peaks whose energy is characteristic of individual chemical bonds, just as in infrared absorption spectroscopy. Because these peaks occur at energy loss below 0.5 eV, the scattering angles given by Eq. (6.12) are very small (microradians) and the transverse momentum of a primary electron after inelastic scattering: $\Delta p \approx m v_0 \ \sin\theta_E \approx (m v_0/2)(E/E_0)$ is also small. From Heisenberg's uncertainty relation: $\Delta p \ \Delta x \approx h$, the uncertainty Δx in the location of the scattering must be large: the inelastic scattering is described as **delocalized**. As a result, vibrational energy losses can be excited by an electron that passes tens of nm outside the edge of a TEM specimen: the so-called **aloof mode** of spectroscopy (see Fig. 6.11). Radiation damage, which results from inelastic scattering with E = 10–100 eV, is more highly localized, so it is possible to collect information from more distant *undamaged* regions if the energy loss being studied is low enough; see Sect. 7.2.

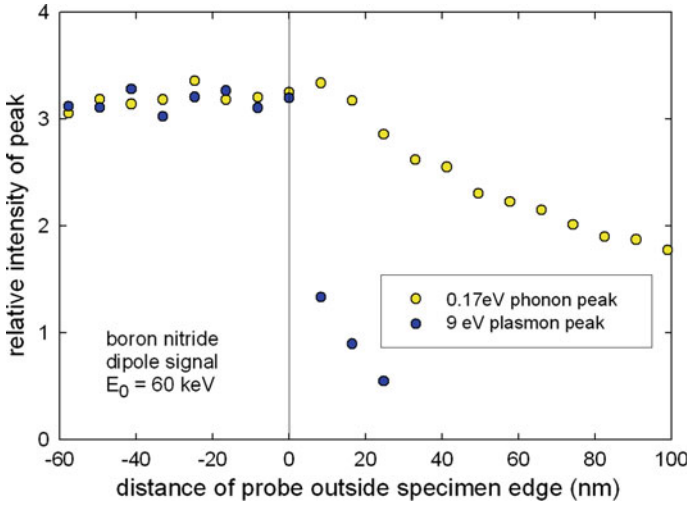


Fig. 6.11 Intensity of the 0.17 eV vibrational-mode peak and the 9 eV plasmon peak, recorded using a Nion monochromated ultraSTEM as a function of the distance of the electron beam from the edge of a BN specimen. Positive distances correspond to aloof-mode spectroscopy, in which the electron beam passes outside the specimen; negative values correspond to the more usual transmission mode, where the beam passes through the specimen

6.10 Further Reading

Scanning Electron Microscopy and X-ray Microanalysis (Kluwer/Plenum, 3rd edition, 2003; DOI 10.1007/978-1-4615-0215-9) deals with practical aspects of energy-dispersive x-ray analysis in the SEM, STEM and TEM. *Transmission Electron Microscopy* by D.B. Williams and C.B. Carter (Springer, 2009) covers both x-ray microanalysis and EELS in Part 4 (Spectrometry).

Scanning Auger Electron Microscopy, ed. M. Prutton and M. El Gomati (Wiley, 2006; ISBN-13: 978-0-470-86677-1) describes the physics, instrumentation and applications involved in Auger-electron imaging.

Electron Energy Loss Spectroscopy by R. Brydson (BIOS, Oxford, 2001) gives an in introduction to EELS, which is treated in more depth by R.F. Egerton in *Electron Energy-Loss Spectroscopy in the Electron Microscope* (3rd edition, 2011, Springer, New York).

Chapter 7
Special Topics

7.1 Environmental TEM

In recent years, ETEM has become a major activity, thanks partly to the availability of special specimen holders fabricated using semiconductor (MEMS) technology. Although analytical electron microscopy allows chemistry to be studied on a local scale, many chemical reactions involve a gas or a liquid, raising problems in relation to the vacuum requirements of a TEM.

One option is to surround the specimen with a low-pressure gas, contained within a restricted volume by means of small apertures above and below the specimen. There are obvious geometrical limitations: if the apertures are too large, gas will leak into sensitive parts of the TEM at an unacceptable rate; if too small, the apertures may restrict the image field of view or the range of scattering angles that can be utilized. The gas pressure and path length between the apertures must be limited to avoid excessive scattering of electrons by gas molecules. In one common arrangement, the aperture diaphragms are attached to the upper and lower pole-pieces of the objective lens, close to the front- and back-focal planes that may contain an electron-beam crossover (see Fig. 3.12b). In practice, there are usually several apertures and the TEM column is differentially pumped, the lowest pressure being in the electron gun.

For quantitative measurements, it is useful to know the gas composition and pressure at the specimen, which in a gas-flow situation may be different from values measured at a remote location. Electron energy-loss spectroscopy, which is sensitive to low-Z elements such as hydrogen and oxygen, has been used for this purpose, at gas pressures up to 0.1 Torr (<10 Pa). From changes in the near-edge fine structure of ionization edges, EELS can also detect variations in the oxidation state of transition-metal or rare-earth catalysts during a chemical reaction. By using an electrically heated specimen holder, such measurements can be made at reaction temperatures up to 1000 °C, as illustrated in Fig. 7.1.

R.F. Egerton, *Physical Principles of Electron Microscopy*,
DOI 10.1007/978-3-319-39877-8_7

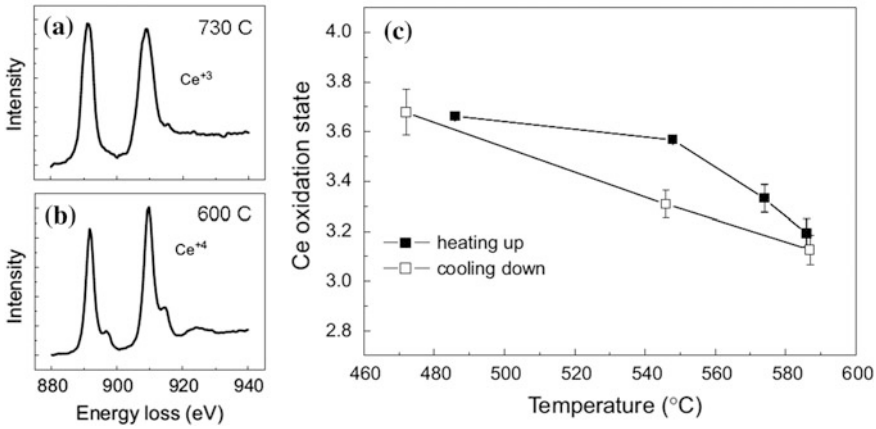


Fig. 7.1 Cerium-based catalyst particles examined in 0.5 Torr of hydrogen at different temperatures, using open-cell ETEM and EELS. *White-line* peaks at the Ce–M_{45} ionization edge were measured at different temperatures, including **a** 730 C and **b** 600 C, resulting in **c** a plot of oxidation state (valency) versus temperature. From the article: in-situ *environmental TEM studies of dynamic changes in cerium-based oxide nanoparticles during redox processes*, by P.A. Crozier, R. Wang and R. Sharma, Ultramicroscopy **108** (2008) 1432–1440, courtesy of Elsevier

As an alternative to the above, a closed-cell design is possible: the specimen is enclosed between two thin (∼50 nm) "windows" made of silicon nitride, its environment protected by o-ring seals; see Fig. 7.2. Graphene (a single layer of graphite) may also be useful as a window material, in view of its remarkable strength in relation to thickness. The window must be very thin in order to minimize additional electron scattering, which reduces the TEM image brightness and contrast.

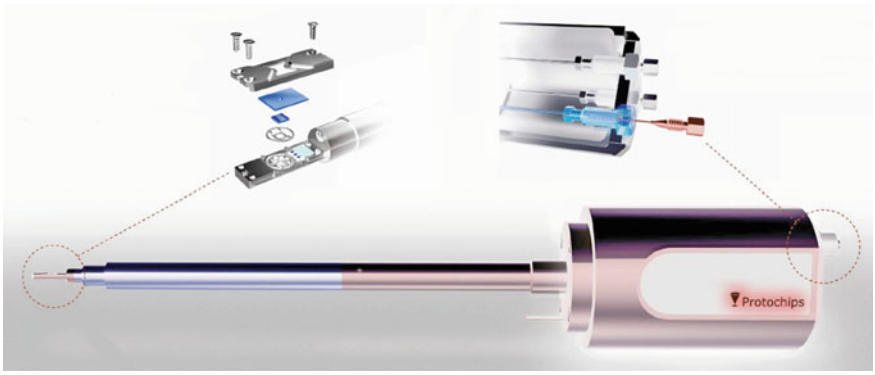


Fig. 7.2 Side-entry closed environmental cell for the TEM, showing the sealing mechanism for vacuum windows and capillary tubing for circulating liquid into the cell. Courtesy of Protochips

Inside this environmental cell, the specimen is surrounded with a gas or a liquid, which can be circulated if necessary using pumps and capillary tubing. In the case of a liquid, its thickness over the field of view must be well below 1 μm, to avoid excessive scattering. An aqueous solution is often used and electrons traveling through it cause radiolysis (see Sect. 7.2) and produce many different chemical species, depending on the pH of the solution and other factors. One useful precaution is therefore to observe the reaction dynamics at different electron-flux densities, to ensure that the behavior seen is similar to that in the absence of an electron beam.

The closed-cell design avoids possible contamination of the TEM and requires no redesign of the TEM vacuum system. It allows gases to be used at pressures up to 1 atmosphere, as well as liquids. The aperture-based open-cell design is restricted to lower-pressure gases but avoids additional scattering from windows and allows standard specimen holders to be used.

Motivations for environmental TEM include the examination of samples in a hydrated state and the observation of chemical growth processes or phases that would not exist under room-temperature high-vacuum conditions. Changing the environment around a specimen may trigger a phase change that can be studied at atomic resolution. In-situ experiments can also include mechanical, thermal, or electrical tests, and exposure to light to simulate photochemical reactions. Although micromanipulation is easier in the SEM, MEMS miniaturization allows many of the same things to be done in the TEM.

7.2 Radiation Damage

High-brightness Schottky and field-emission sources are being used increasingly to provide the small-diameter beams required for analytical TEM and STEM. And besides reducing the electron-probe size, aberration correction of the probe-forming lens has allowed an increase in the probe convergence angle and probe current, resulting in current densities as high as 10^7 A/cm^2. Consequently, beam-induced radiation damage can be a problem, even for specimens previously regarded as beam-stable.

There are several kinds of radiation damage. **Radiolysis** or ionization damage has long been familiar to biologists and polymer chemists using the TEM, where it limits the resolution obtainable from most organic specimens. This damage arises from the inelastic scattering of electrons, in which the transferred energy causes an atomic electron (usually a valence electron) to undergo a transition to a higher-energy state. Although the hole (vacancy) left behind is soon filled by another electron, the excitation process is not always reversible and may result in the breakage or rearrangement of chemical bonds. For example, thermal motion may cause the atom to move while in its excited state. If the original material is crystalline, repeated damage may destroy the crystal structure and will be observed as a fading of the Bragg spots in a diffraction pattern. The electron dose needed to destroy organic crystals can be as low as 10^{-3} C/cm^2. Life processes are even more

sensitive, one reason why living cells cannot be studied in the TEM, even using an environmental cell.

X-rays also produce radiolysis, although recently developed free-electron lasers (XFEL) can deliver 10^{12} x-ray photons within 50 fs, fast enough to provide diffraction data from single macromolecules or viruses before damage can occur. Outrunning radiation damage in the TEM appears unlikely: electrons are charged particles and their Coulomb repulsion limits the available current density. But because radiolysis is influenced by thermal motion, it is slowed down by a factor of 3–10 when a low-temperature specimen holder is used to cool the specimen to 100 K. This option forms the basis for the **cryomicroscopy** of biological specimens, a technique being used to determine the structures of membrane proteins, for example.

Electrically conducting specimens contain a high density of mobile electrons, which prevent radiolysis because any vacant electron state below the Fermi level is filled before atomic motion can occur. However, high-angle *elastic* scattering causes an alternative form of radiation damage: **knock-on displacement** of atoms. As shown in Chap. 4, tens of eV can be supplied direct to the nucleus of a low-Z atom, sufficient to displace atoms from lattice sites in a crystal and resulting in defect clusters (Fig. 4.16) or gradual destruction of the crystalline structure. Such damage can often be prevented by reducing the primary-electron energy E_0 below some threshold, given by Eq. (4.8), so that the energy transfer E is less than the atom-displacement energy. However, atoms at the surface of a specimen are less tightly bound and can be ejected into the vacuum (electron-induced sputtering) unless the incident energy is reduced even further. Because the probability of high-angle elastic scattering is small, knock-on effects are slow and are only observable in the case of small electron probes with high current density.

In the case of a non-conducting *inorganic* material, knock-on displacement might predominate at a high primary energy and low-efficiency radiolysis at a lower energy. However, there are other possibilities. The release of secondary electrons causes an irradiated area to charge positively, deflecting the primary beam and causing image distortion or even electrical breakdown. If the specimen contains mobile ions, these may move under the influence of the internal electric field, which is often highest at the edge of the electron beam. Such effects often depend on dose rate (incident-current density) as well as on the accumulated electron dose. For example, there may be a threshold *current* density, below which damage does not occur or its effects are reversible after removing the beam. Sub-nm probes from a CFEG source have been used to produce small holes in thin films of oxides and to write patterns that represent very dense information storage but the method is too slow to be of practical use.

Some characteristics of these different damage mechanisms are summarized in Table 7.1. Radiation damage can be a serious problem in the TEM (and even SEM) because of the need for high spatial resolution, which involves concentrating electrons into a small region of material. In a beam-sensitive specimen, spatial resolution is not limited by properties of the microscope (electron optics, mechanical or electrical drift) but by the need for a statistically significant number

Table 7.1 Three kinds of radiation damage, together with their effects and remedies

Damage mechanism	Prominent in	Observed effects	Remedy
Knock-on displacement	Conductors, e.g., metals	Amorphization, e-beam sputtering	Reduce primary energy
Radiolysis (ionization)	Poor conductors	Loss of crystallinity, mass loss	Cryomicroscopy
Electrostatic charging	Good insulators	Beam deflection, mass loss (holes may be formed)	Reduce current density, conductive coating

of electrons in the image, diffraction pattern, or analytical signal, which is limited by the radiation dose that the specimen can tolerate without changing its structure. In many organic specimens, this **dose-limited resolution** (DLR) is several nm. To make the DLR a low number, it is necessary to choose an imaging mode that optimizes contrast and signal/noise ratio, besides trying to minimize the damage, so phase-contrast cryoTEM is favored for biological work. For low-loss EELS, it is possible to take advantage of the delocalization of inelastic scattering by working in aloof mode (Fig. 7.3b) or by using a coarse digital raster (Fig. 7.3b).

Damage in the SEM is less problematic because of the less stringent requirement on spatial resolution. However, damage can occur close to the surface, within the penetration depth, and release of light elements (mass loss) in organic specimens can lead to shrinkage and distortion of the structure. Charging is also a problem for insulating specimens, unless the accelerating voltage can be adjusted correctly (Sect. 5.7).

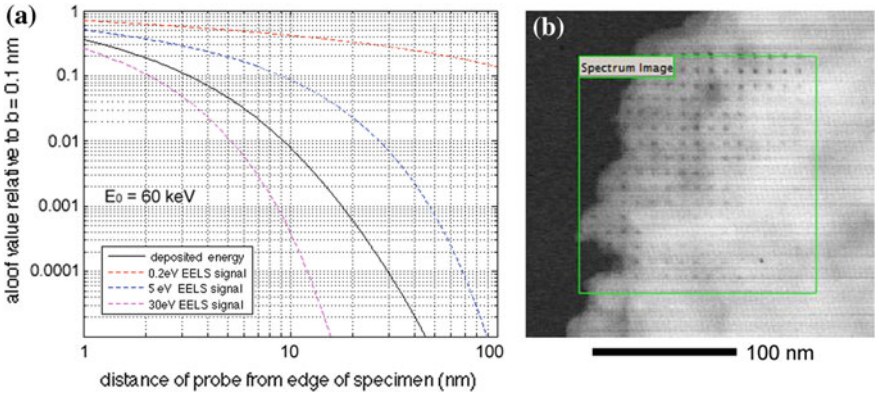


Fig. 7.3 **a** Energy deposition (*solid curve*, proportional to radiation damage) and EELS signal (*dashed curves*, for three values of energy loss) calculated as a function of the distance of an electron probe from the edge of the specimen. **b** Dark-field image of a $Ca(OH)_2$ specimen, showing the effect of a digital raster with a sub-nm electron probe and a step size of 8 nm. Dark spots show partial holes created by radiation damage while vibrational-mode EELS data was recorded from the intervening undamaged material

7.3 Electron Tomography

Tomography implies the assembly of image information in three dimensions. In atom-probe tomography (Sect. 1.9), this is achieved by recording two-dimensional images as a function of time, as atoms are gradually removed from the tip of a specimen. In the TEM, tomography involves recording a series of two-dimensional images, each at a different angle between the electron beam and the specimen. As discussed in Chap. 3, each TEM image represents a projection of the specimen structure in the direction of the electron beam. Computer processing of the tilt series then results in a three-dimensional representation of the specimen. The procedure has much in common with the computer tomography (CT) method used in medical imaging, which employs X-rays rather than electrons.

The information used for TEM tomography can be derived from any signal that changes monotonically (ideally linearly) with specimen thickness. Diffraction contrast is suspect because thickness fringes are periodic in thickness and bend contours represent no real structure. STEM options include HAADF, EELS, and cathodoluminescence signals, which are monotonic provided the specimen is not too thick.

The specimen can be thin slab but with the disadvantage that its thickness in the beam direction increases with the angle of tilt. A cylindrical rod-shaped specimen avoids this complication and allows an unlimited range of tilt angle, thereby avoiding a "missing wedge" problem that leads to a loss of resolution in the incident-beam direction. Typically a single-tilt side-entry stage is used, controlled by a computer that also handles the image acquisition. Before each image readout, the computer checks the specimen focus and position, adjusting x- and y-shifts if necessary to correct for stage drift, although more precise alignment is applied during reconstruction. Computer control is a practical necessity because acquisition of a tilt series can take many minutes or even hours, most of the time being spent in focusing and waiting for stage drift to settle at each orientation.

Reconstruction of the three-dimensional image is based on a computer algorithm. The filtered back-projection (FBP) method first corrects each image for geometric and magnification changes, takes a Fourier transform to generate slices in reciprocal space, assembles the slices, and performs an inverse transform to give the three-dimensional image. Other algorithms are based on matrix algebra and may involve iterative procedures, sometimes giving superior results with less image artifacts.

The final resolution R depends on several factors. Geometrical arguments predict $R \approx D/N$, where D is the diameter of specimen being imaged and N is the number of image slices, but radiation damage may result in a worse resolution for some specimens. In practice, a voxel size down to 1 nm is becoming common, a factor of 10 smaller than synchrotron X-ray tomography and a factor of 100 better than tomography with a bench-top X-ray source. With R = 1 nm and D = 500 nm, each image slice would require $(500)^2$ pixels and the final-image $(500)^3$ voxels: about 1 GB of data with an 8-bit intensity scale. A typical tilt series involves 100 images, recorded at intervals of 1–2° (Fig. 7.4).

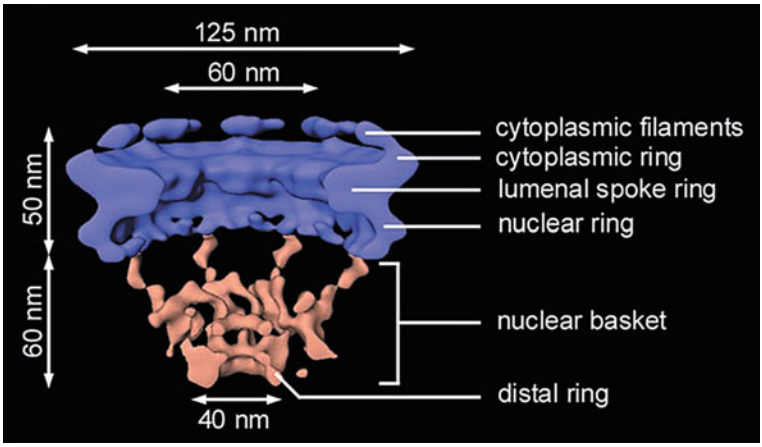


Fig. 7.4 Nuclear pore complex (NPC) from the amoeba *Dictyostelium discoideum*, part of the membrane that controls the exchange of macromolecules between the nucleus and the cytoplasm. This tomographic image was obtained (at about 8 nm resolution) from tilt series of phase-contrast images of frozen-hydrated specimens. Reprinted with license from AAAS, from Beck et al. Science **306** (2004) 1387

7.4 Electron Holography

Although the behavior of electrons can often be understood by treating them as particles, some properties (e.g., distribution of intensity within an electron-diffraction pattern) require a wave description. But when an image or diffraction pattern is viewed or recorded, only the intensity (or amplitude) of the electron wave is retrieved; all phase information is lost, which may include information about the electric and magnetic fields present in the specimen.

In holographic recording, electron or light waves coming from a sample are mixed with a **reference wave**. The resulting interference pattern (a **hologram**) is sensitive to the phase of the wave at the *exit surface* of the specimen. So even though it is *recorded* only as an intensity distribution, the hologram contains both amplitude and phase information, which can be extracted by means of a **reconstruction** procedure. This extraction involves interference between the holographic record and a *reconstruction* wave, not necessarily of the original wavelength or even the same type of radiation.

The *reference* wave must have the same wavelength and a fixed phase relationship with the wave that travels through the specimen. One way of achieving this is to have *both* waves originate from a small (ideally point) source, a distance Δz from the specimen; see Fig. 7.5. This source could be produced by using a strong electron lens L_1 to focus an electron beam into a small probe, as first proposed by Gabor (1948). If the specimen is suitably thin, some electrons are

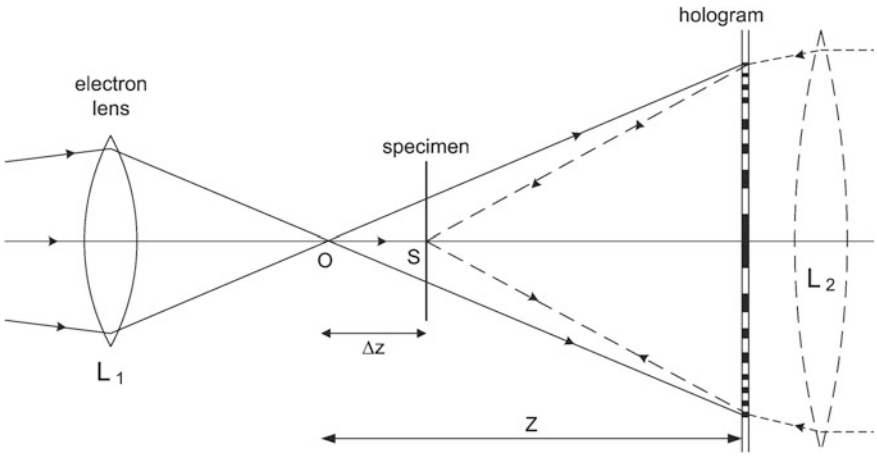


Fig. 7.5 In-line electron hologram formed from spherical waves emanating from a small source O (formed by an electron lens L_1), which interfere with spherical waves emitted from scattering points (such as S) in the specimen. For reconstruction, a light-optical lens L_2 focuses light onto the back (or front) of the hologram, which then focuses the transmitted (or reflected) light onto each point S, forming a magnified image of the specimen

transmitted straight through the specimen and interfere with those that are scattered within it. Interference between transmitted and scattered waves produces a hologram on a nearby screen or recording device.

The hologram has some of the features of a **shadow image**, a projection of the specimen with magnification $M \approx z/\Delta z$. However, each scattering point S in the specimen emits spherical waves (such as those centered on S in Fig. 7.5) that interfere with the spherical waves coming from the point source, so the hologram is also an interference pattern. Bright fringes are formed when there is constructive interference, wavefronts of maximum amplitude coinciding at the hologram plane. In the case of a *three*-dimensional specimen, scattering points S occur at different z-coordinates and their relative displacements (along the z-axis) have an influence on the interference pattern. As a result, the hologram contains *three-dimensional* information, as its name is meant to suggest (holo = entire).

Reconstruction could be accomplished by using visible light of wavelength $M\lambda$, reaching the hologram through the glass lens L_2 in Fig. 7.5. The interference fringes in the hologram act rather like a zone plate or diffraction grating, directing the light into a magnified image of the specimen located at S, as indicated by the dashed rays in Fig. 7.5. Since M can be large (e.g., 10^5), this reconstruction is best done in a separate apparatus. If the defects (aberrations, astigmatism) of the glass lens L_2 match those of the electron lens L_1, the electron-lens defects are *compensated* and the resolution in the reconstructed image is no longer limited by those aberrations.

At the time when Gabor made his proposal, no suitable electron source was available; he demonstrated the holographic principle using visible light. Nowadays, *light-optical* holography has become a practical technique due to the development

of laser sources of highly monochromatic radiation (small spread in wavelength, high *longitudinal coherence*). In *electron* optics, a monochromatic source is one that emits electrons with very similar kinetic energy (low energy spread ΔE). In addition, the electron source should have a very small diameter, giving high *lateral coherence*. Therefore the widespread use of electron holography had to await the commercial availability of the *field-emission* TEM.

There are many alternative holographic schemes, the most popular being **off-axis holography**. In this method, an **electron biprism** is used to combine the **object wave**, representing electrons that pass through the specimen, with a reference wave formed from electrons that pass beyond an edge of the specimen (Fig. 7.6). The biprism consists of a thin conducting wire (<1 μm diameter) maintained at a positive potential (up to a few hundred volts), often inserted at the selected-area-aperture plane of the TEM. The electric field of the biprism bends the electron trajectories (Fig. 7.6) and produces a hologram containing sinusoidal fringes, running parallel to the wire, that are modulated by information about the specimen; see Fig. 7.7. Condenser-lens stigmators are adjusted to provide highly astigmatic illumination, in order to maximize the size of the coherence patch where interfering wavefronts overlap.

Rather than using *optical* reconstruction, it is more convenient to transfer the hologram (from a CCD camera) into a computer, which can simulate the effect of optical processing by running a program that reconstructs the complex image wave.

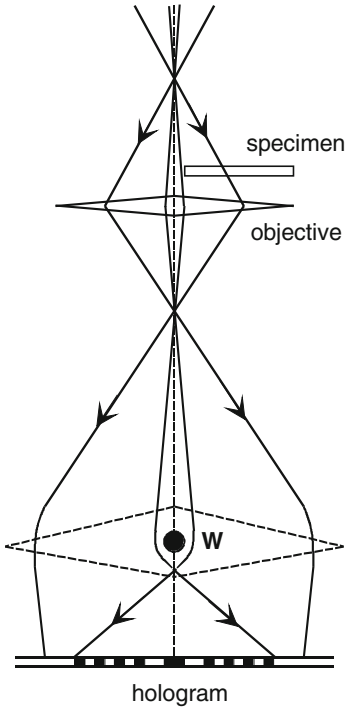


Fig. 7.6 Principle of off-axis electron holography. Electrons that pass through the sample are combined with those of an off-axis beam, due to electrostatic attraction by a positively biased wire W whose axis runs perpendicular to the diagram. W is the electron-optical equivalent of a glass biprism, as indicated by the *dashed lines*

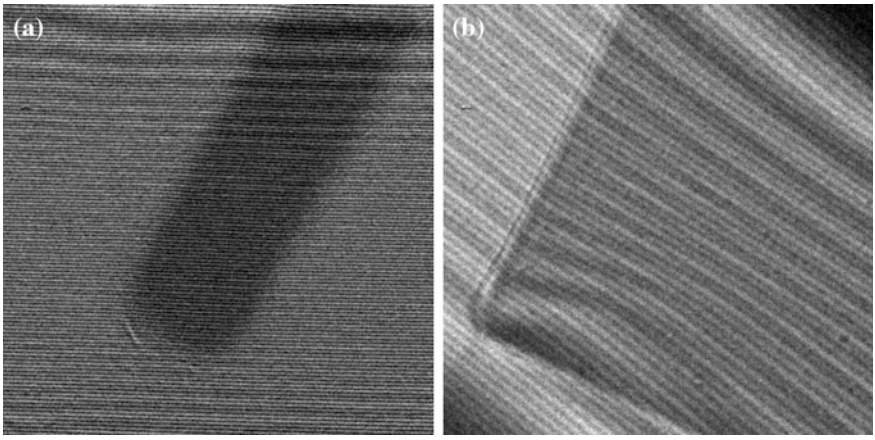


Fig. 7.7 Off-axis low-magnification (objective lens off) electron holograms of a sub-μm permalloy structure supported on a 50 nm-thick SiN membrane. Note the fringe bending corresponding to areas with a strong change in electron phase (for example, near the edges of the permalloy film). Although some of this bending is electrostatic in nature (due to the "inner potential" of the permalloy), most of it results from the internal magnetization (change in magnetic potential). The holograms were recorded using a JEOL 3000F field-emission TEM. Courtesy of M. Malac, M. Schofield and Y. Zhu, Brookhaven National Laboratory

This process involves calculating the two-dimensional Fourier transform of the hologram intensity (its frequency spectrum) and selecting an off-axis portion of the transform, representing modulation of the sinusoidal fringe pattern. Taking an *inverse* Fourier transform of this selected data yields two images: the real and imaginary parts of the processed hologram (representing the cosine and sine harmonics) from which the phase and amplitude of the electron exit wave (i.e., the wave at the exit surface of the specimen) is calculated. The phase image is particularly valuable, since it can reveal changes in magnetic or chemical potential in the specimen.

In ferromagnetic specimens, fringe shifts are mainly due to changes in *magnetic* potential associated with the internal magnetic field. Electron holography therefore supplements the information obtained from Fresnel and Foucault imaging, and is more quantitative.

In non-magnetic specimens, where only electrostatic effects contribute to the shift of the holographic fringes, the phase image can be processed to yield a map of the local (chemical) potential; see Fig. 7.8. Electron holography can therefore be used to examine the distribution of electrically active dopants in semiconductors, with high spatial resolution.

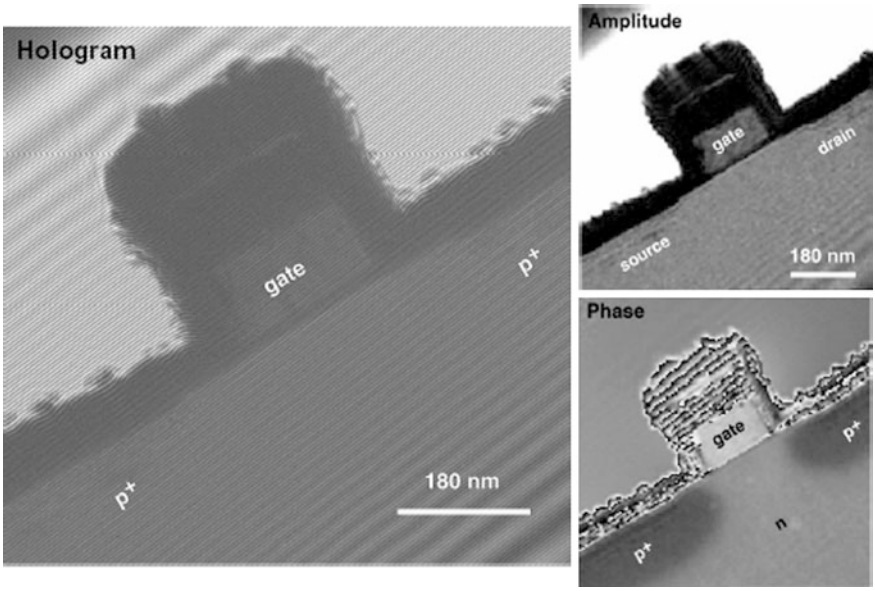


Fig. 7.8 Off-axis electron hologram of a p-MOS transistor, together with the derived amplitude and phase images. In the phase image, the two darker areas represent heavy p^+ doping of the silicon; the abrupt fringes represent phase increments exceeding 2π radian. The cross-sectional specimen was prepared by mechanical polishing followed by Ar-ion milling. Reproduced from: *Mapping of process-induced dopant distributions by electron holography* by W.-D. Rau and A. Orchowski, Microscopy and Microanalysis (February 2004) 1–8, courtesy of Cambridge University Press and the authors

7.5 Time-Resolved Microscopy

Tomography and holography extend microscopy into three spatial dimensions but there is also interest in exploring the fourth (time) dimension. The time dependence of physical or chemical processes can be studied on any microscope fitted with a TV-rate video camera, at time intervals of about 20 ms. Recent TEM-camera advances, which include direct single-electron counting and fast readout, reduce the time resolution down to about 1 ms.

Further improvement in time resolution is possible by using a pulsed beam of electrons. Electrostatic or electromagnetic shutters can create pulses shorter than 1 microsecond and by using a pulsed laser to illuminate a photocathode source, sub-picosecond pulses are possible. In the **pump-probe** technique, laser (pump) pulses illuminate the specimen and excite it to produce the change being studied. Probe pulses applied to the photocathode are delayed by some time T, which can be varied by changing the optical path length. For each value of T, a large number of pulses are used to generate a signal with adequate statistics. Then by gradually varying T, the time dependence of a fast repetitive process can be followed; the

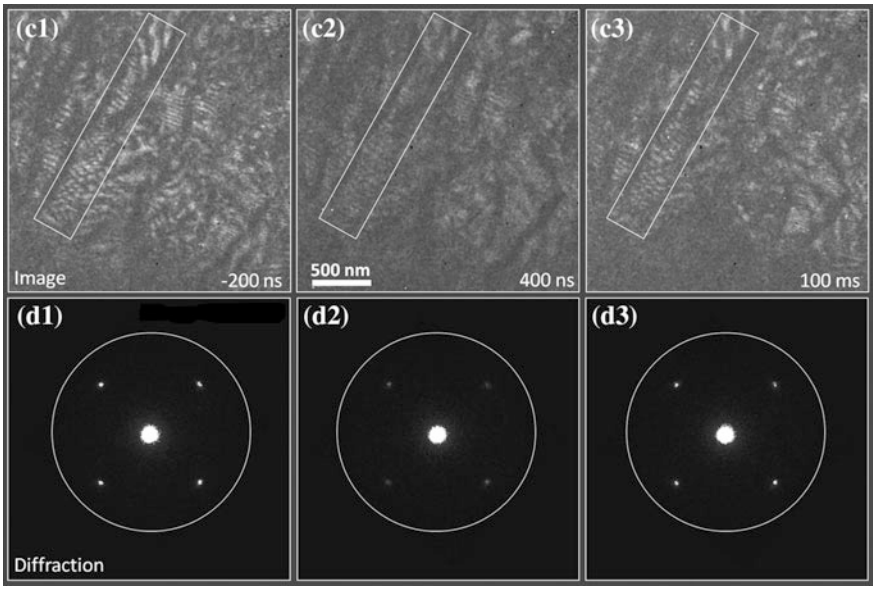


Fig. 7.9 (**c1–c3**) Time-resolved dark-field images and (**d1–d3**) diffraction patterns showing 120-type reflections, recorded (with probe pulse of 10^5 electrons) from poly(ethylene) oxide before (−200 ns), after (400 ns), and long after (100 ms) applying an optical pump pulse. The dark-field contrast and Bragg-spot intensities first decrease due to partial melting of the polymer but then recover due to reordering. From *Macromolecular structural dynamics visualized by pulsed dose control in 4D electron microscopy* by O.-H. Kwon, V. Ortalan and Ahmed H. Zewail, PNAS **108** (2011) 6026–6031, reprinted with permission of National Academy of Sciences

procedure is just a more controllable version of the stroboscopic technique discussed in Sect. 5.5. Since photoemission of electrons requires ultraviolet photons, infrared pump pulses must be converted into probe pulses by frequency doubling or tripling in a nonlinear crystal. Figures 7.9 and 7.10 show an example of pump-probe measurements and the detailed information that can be derived from them.

The pump-probe technique works best for highly repeatable processes, like plasmon excitation and some reversible phase transformations. For non-repeating processes such as mechanical fracture, a **single-shot** technique is necessary, which requires more electrons (typically 10^7–10^{10}) per pulse. But electrons are charged particles and if too many of them are concentrated together in a short pulse, Coulomb repulsion lengthens the pulse, limits the current density, and increases the energy spread. Therefore good time resolution generally implies a loss of spatial resolution. To address this problem, high (relativistic) beam energies, high-brightness sources, and advanced electron optics (including aberration correction) are being explored. At the time of writing, single-shot dynamic TEM (DTEM) can achieve 5–1000 ns time resolution combined with a spatial resolution

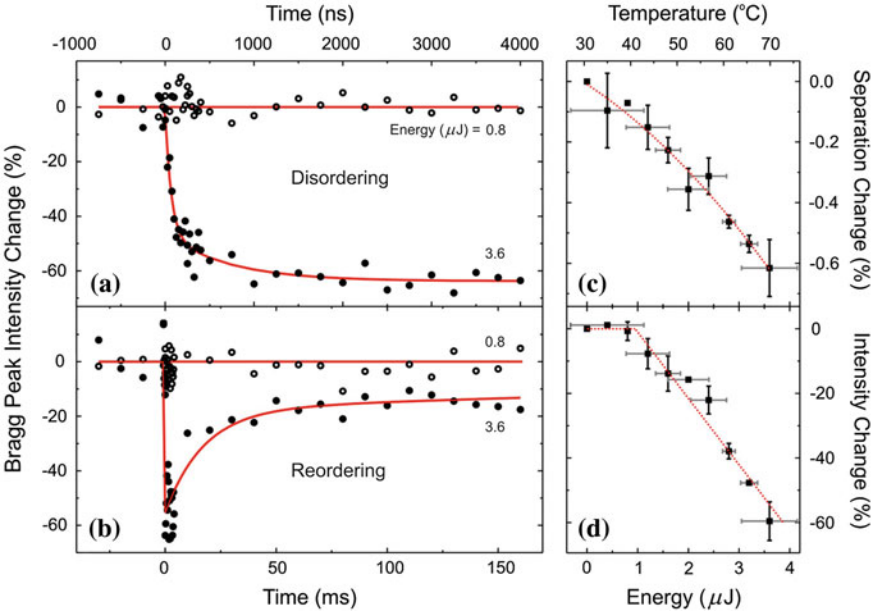


Fig. 7.10 Quantitative information derived from the pump-probe measurements on PEO shown in Fig. 7.9. **a** Fast (ns-scale) decay of Bragg intensity indicating disordering. **b** Much slower recovery of intensity due to atomic reordering. **c** Lattice expansion, seen as a decrease in Bragg-spot separation, as a function of specimen temperature. **d** Decrease in Bragg-spot intensity as a function of pump-pulse energy, showing a threshold at 1 μJ corresponding to a specimen temperature of 40 °C. From *Macromolecular structural dynamics visualized by pulsed dose control in 4D electron microscopy* by O.-H. Kwon, V. Ortalan and Ahmed H. Zewail, PNAS **108** (2011) 6026–6031, reprinted with permission of National Academy of Sciences

of a few nm. Alternatively, single-shot ultrafast diffraction (UED) is capable of sub-ps time resolution with a spatial resolution of the order of 1 μm.

An alternative solution to the Coulomb-repulsion problem is to form pulses containing only one electron. At a pulse-repeat rate of 10 MHz, this is equivalent to an average beam current of 1.6 pA, enough to form a useful image in a few seconds, but time resolution then relies on a pump-probe technique that limits the microscopy to repetitive laser-induced processes.

Materials-science applications of fast microscopy include the study of fracture, plastic deformation, and phase transitions in metal alloys, nm-scale heat conduction, and imaging of reaction fronts in solid-state chemical reactions. Chemical applications include the study of transient processes involved in catalysis and the dynamics of chemical bonding, which would require sub-fs (attosecond) time resolution. Biological applications include examining the shape changes of proteins and macromolecules, over a time scale that may be many microseconds. In conventional cryomicroscopy, molecules are in a crystalline form and not free to move, so liquid-jet methods have been developed for their observation in the TEM.

7.6 Further Reading

Environmental TEM is the subject of a recent compilation: *Controlled Atmosphere Transmission Electron Microscopy*, ed. T.W. Hansen and J.B. Wagner (Springer, 2016; ISBN-10: 3319229877).

Radiation damage in the TEM and SEM is summarized in Micron **35** (2004) 399–409 (doi:10.1016/j.micron.2004.02.003) and Ultramicroscopy **127** (2013) 100–108 (doi:10.1016/j.ultramic.2012.07.006), together with mitigation strategies. *Radiation Damage in Biomolecular Systems* (Springer, 2012, DOI 10.1007/978-94-007-2564-5) provides insight into recent research involving organic systems.

Damage and charging effects in oxides and other inorganic specimens are described by L.W. Hobbs, in *Introduction to Analytical Electron Microscopy*, ed. J. J. Hren, J.I. Goldstein and D.C. Joy (Plenum Press, New York, 1987) pp. 399–445, and by N. Jiang in Rep. Prog. Phys. **79** (2016) 016501, doi:10.1088/0034-4885/79/1/016501.

Applications of X-ray and electron tomography are nicely illustrated by P.A. Midgley, E.P.W.Ward, A.B. Hungria and J.M. Thomas in Chem. Soc. Rev. 36 (2007) 1477–1494 (DOI:10.1039/b701569k). For biological applications, see *Electron Tomography*, ed. J. Frank (2nd edition, Springer, 2010) and for materials science: *Advanced Tomographic Methods in Materials Research and Engineering*, ed. J. Banhart (Oxford, 2008).

Gabor (1948) is the original holography paper: *A new microscopic principle*, in Nature **161** (1948) 777–778. *Introduction to Electron Holography* (Kluwer/Plenum, New York, 1999) contains articles written by experts in the field. Akira Tonomura wrote an extensive (163-page) treatment of electron holography and its applications in *Electron Holography*, Springer Series in Optical Sciences Vol. **70** (1999): ISBN: 978-3-642-08421-8 (Print), 978-3-540-37204-2 (Online). *The Quantum World Unveiled by Electron Waves* (World Scientific, 1998, ISBN 981-02-2510-5) is his non-technical personal account of the general principles involved.

Time-resolved microscopy is given an attractively illustrated overview by A.H. Zewail and J.M Thomas in: *4D Electron Microscopy: Imaging in Space and Time* (Imperial College Press, 2010, 341 pages). For a recent account of ultrafast electron diffraction, see *Femtosecond time-resolved MeV electron diffraction* by Zhu et al., New Journal of Physics 17 (2015) 063004 (doi:10.1088/1367-2630/17/6/063004). Ultrafast electron microscope instrumentation is discussed by B.W. Reed et al. in Microsc. Microanal. **15** (2009) 272–281 (doi:10.1017/S1431927609090394).

Current trends in electron microscopy are reviewed by J.M. Thomas et al., *The rapidly changing face of electron microscopy* in Chemical Physics Letters **631** (2015) 103: http://creativecommons.org/licenses/by-nc-nd/4.0/, and by Y. Zhu and H. Durr, *The future of electron microscopy* in Physics Today 68 (2015) 32 (doi:10.1063/PT.3.2747). Consult http://science.energy.gov/bes/news-and-resources/reports/ for a more detailed discussion.

Appendix
Mathematical Derivations

A.1 The Schottky Effect

The reduction $\Delta\phi$ in the surface-barrier height with applied electric field (Schottky effect) can be calculated by first considering the case where the extraction field is zero. When an electron leaves the conducting cathode, it generates (within the vacuum) its own electric field, which arrives perpendicular to the surface of the conductor and induces an equal and opposite charge $+e$ that is distributed over the cathode surface (Fig. A.1a). The field lines *outside* the cathode are the same as those generated by an electrostatic dipole, where the cathode is replaced by a point charge $+e$ located at an equal distance z *inside* the surface; see Fig. (A.1b). The electron therefore experiences an attractive force given by Coulomb's law:

$$F(z) = K(-e)(+e)/(2z)^2 \tag{A.1}$$

in which $K = 1/(4\pi\varepsilon_0) = 9.0 \times 10^9$ N m^2 C^{-2} is the Coulomb constant. The negative sign of $F(z)$ indicates that the force acts in the $-z$ direction, towards the surface.

With voltage applied to an extraction electrode, an additional field $-E_e$ is created (in the *negative* z-direction) that gives rise to a force $F_e = -e(-E_e) = +eE_e$, the positive sign of F_e denoting that this extraction force acts in the $+z$ direction. The total force is therefore:

$$F = F(z) + F_e = -(K/4)\left(e^2/z^2\right) + eE_e \tag{A.2}$$

R.F. Egerton, *Physical Principles of Electron Microscopy*,
DOI 10.1007/978-3-319-39877-8

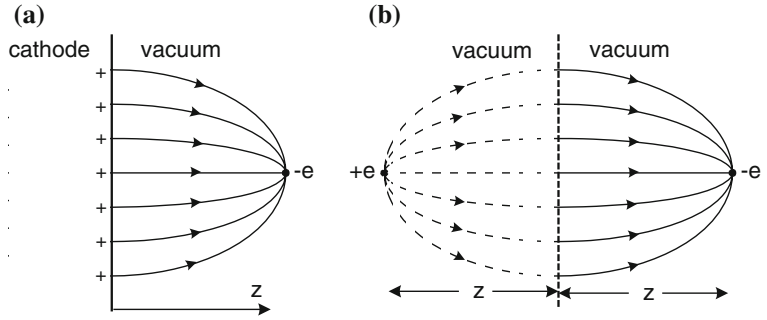


Fig. A.1 **a** Electric field lines and surface-charge distribution produced by an electron located just outside the surface of a flat-surfaced conductor. **b** Equivalent field lines produced by an electrostatic dipole operating in vacuum

This total force can be written as $F = (-e)E(z) = (-e)(-\mathrm{d}V/\mathrm{d}z)$ where $E(z)$ is the total field and V is the corresponding electrostatic potential. The potential can therefore be obtained by integration of Eq. (A.2):

$$V = (1/e)\int F\ \mathrm{d}z = (K/4)\,(e/z) + E_e z \tag{A.3}$$

The potential energy $\phi(z)$ of the electron is therefore:

$$\phi(z) = (-e)(V) = -(K/4)\left(e^2/z\right) - eE_e z \tag{A.4}$$

and is represented by the dashed curve in Fig. 3.4 (p. 60). The top of the potential-energy barrier corresponds to a coordinate z_0 at which the slope $\mathrm{d}\phi/\mathrm{d}z$ and therefore the force F are both zero. Substituting $z = z_0$ and $F = 0$ in Eq. (A.2) results in $z_0^2 = (K/4)(e/E_e)$ and substitution into Eq. (A.4) gives

$$\phi(z_0) = -e^{3/2}K^{1/2}E_e^{1/2} = -\Delta\phi \tag{A.5}$$

From Eq. (3.1), the current density due to thermionic emission is now:

$$J_e = AT^2\exp[-(\phi - \Delta\phi)/kT\ = AT^2\exp(-\phi/kT)\exp(\Delta\phi/kT) \tag{A.6}$$

In other words, the applied field increases J_e by a factor of $\exp(\Delta\phi/kT)$. Taking $E_e = 10^8$ V/m, $\Delta\phi = 6.1 \times 10^{-20}$ J = 0.38 eV and $\exp(\Delta\phi/kT) = 11.5$ for T = 800 K. This accounts for the order-of-magnitude increase in current density J_e (see Table 3.2) for a ZrO-coated Schottky source, compared to a LaB_6 source with the same work function and operating temperature.

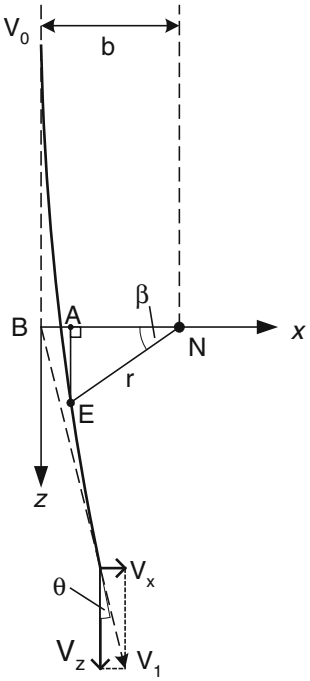


Fig. A.2 Trajectory of an electron E with an impact parameter *b* relative to an unscreened atomic nucleus N, whose electrostatic field causes the electron to be deflected through an angle θ, with a slight change in speed from v_0 to v_1

A.2 Impact Parameter in Rutherford Scattering

Here we retrace Rutherford's derivation of the relation between the impact parameter *b* and the scattering angle θ, replacing his incident alpha particle by an electron and assuming that the scattering angle is small (true for most electrons, if their incident kinetic energy is high). Figure A.2 shows the hyperbolic path of an electron deflected by the electrostatic field of an *unscreened* atomic nucleus. At any instant, the electron is attracted towards the nucleus with an electrostatic force given by Eq. (4.10):

$$F = KZe^2/r^2 \tag{A.7}$$

where $K = 1/(4\pi\varepsilon_0)$ as previously. During the electron trajectory, this force varies in both magnitude and direction; its net effect must be obtained by *integrating* over the path of the electron.

However, only the *x*-component F_x of the force is responsible for angular deflection and this component is:

$$F_x = F\cos\beta = \left(KZe^2/r^2\right)\cos\beta \tag{A.8}$$

The variable β represents the *direction* of the force, relative to the *x*-axis (Fig. A.2). From the geometry of triangle ANE, we have: $r\cos\beta = AN \approx b$ if θ is small, where b is the impact parameter of the initial trajectory. Substituting for r in Eq. (A.8) gives:

$$\left(KZe^2/b^2\right)\cos^3\beta = F_x = m(dv_x/dt) \tag{A.9}$$

where $m = \gamma m_0$ is the *relativistic* mass of the electron and we have applied Newton's second law of motion in the x-direction. The final x-component v_x of electron velocity (resulting from the deflection) is obtained by integrating Eq. (A.9) with respect to time, over the whole trajectory, as follows:

$$\begin{aligned} v_x &= \int (dv_x/dt)\, dt = \int [(dv_x/dt)/(dz/dt)]\, dz \\ &= \int [(dv_x/dt)/v_z]\,(dz/d\beta)\, d\beta \end{aligned} \tag{A.10}$$

In Eq. (A.10), we first replace integration over time by integration over the z-coordinate of the electron, and then by integration over the angle β. Since $z = (EN)\tan\beta \approx b\tan\beta$, basic calculus gives $dz/d\beta \approx b(\sec^2\beta)$. Using this relationship in Eq. (A.10) and making use of Eq. (A.9) to substitute for dv_x/dt, we can write v_x entirely in terms of β as the only variable:

$$v_x = \int [(F_x/m)/v_z]\, b(\sec^2\beta)\, db = \int [KZe^2/(bmv_z)\, \cos^3\beta]\sec^2\beta\, d\beta \tag{A.11}$$

From the vector triangle in Fig. A.2 we have: $v_z = v_1 \cos\theta \approx v_1$ for small θ and also, if θ is small, $v_1 \approx v_0$ because elastic scattering causes only a very small fractional change in kinetic energy of the electron. Since $\sec\beta = 1/\cos\beta$,

$$\begin{aligned} v_x &= \int \left[KZe^2/(bmv_0)\right]\cos\beta\, d\beta = \left[KZe^2/(bmv_0)\right][\sin\beta] \\ &= 2\left[KZe^2/(bmv_0)\right] \end{aligned} \tag{A.12}$$

where we have taken the limits of integration to be $\beta = -\pi/2$ (electron far from the nucleus and approaching it) and $\beta = +\pi/2$ (electron far from the nucleus and receding from it).

Finally, we obtain the scattering angle θ from a parallelogram (or triangle) of velocity vectors in Fig. A.2, using $\theta \approx \tan\theta = v_x/v_z \approx v_x/v_0$ to give:

$$\theta \approx 2KZe^2/\left(bmv_0^2\right) = 2KZe^2/(\gamma m_0 v_0^2\, b) \tag{A.13}$$

Making a non-relativistic approximation for the kinetic energy of the electron ($E_0 \approx mv^2/2$), Eq. (A.13) becomes:

$$\theta \approx K\,Z\,e^2/(E_0\, b) \tag{A.14}$$

A more exact version of Eq. (A.13), relativistic and valid up to large θ, is:

$$\sin\theta\,/(1+\cos\theta) = \tan(\theta/2) = KZe^2/(\gamma m_0 v_0^2 b) \tag{A.15}$$

Index

R.F. Egerton, *Physical Principles of Electron Microscopy*,
DOI 10.1007/978-3-319-39877-8