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 the **micrometer** (1 μm = 10^{-6} m, also called a *micron*) and also the **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 need even smaller units. Later in this book, we will make use of the **picometer** (1 pm = 10^{-12} m).

The diameters of several small objects of scientific or general interest are listed in Table 1-1, together with their approximate dimensions.

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* = 75μm / *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
Uranium atom	0.2 nm = 200 pm	400,000

1.1 Limitations of the Human Eye

Our concepts of the physical world are largely determined by 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, meaning wavelengths 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 the eye's curved front surface, the **cornea**; see Fig. 1-1.

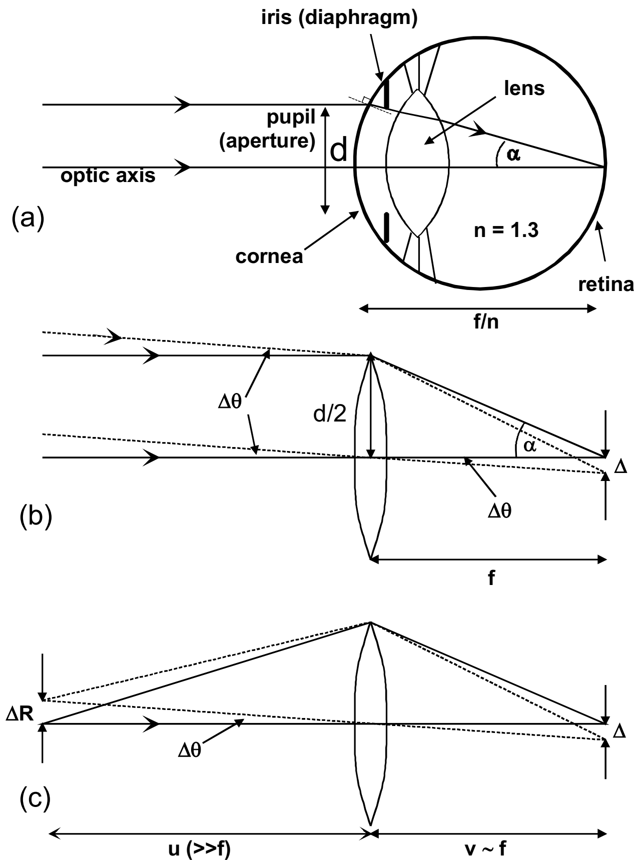

Figure 1-1. (a) A physicist's conception of the human eye, showing two light rays focused to a single point on the retina. (b) Equivalent thin-lens ray diagram for a *distant* object, showing *parallel* light rays arriving from opposite ends (solid and dashed lines) of the object and forming an image (in air) 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 less than f).

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 the local intensity in the image. However, the photochemical processes in the receptor cells work over a limited range of image intensity, therefore the eye controls the amount of light reaching the retina by varying the diameter d (over a range 2 – 8 mm) of the **aperture** of the eye, 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 and still be separately identified from an adjacent and similar object, is determined by three factors: the size of the receptor cells, imperfections in the 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 that is depicted in Fig. 1-2 , a

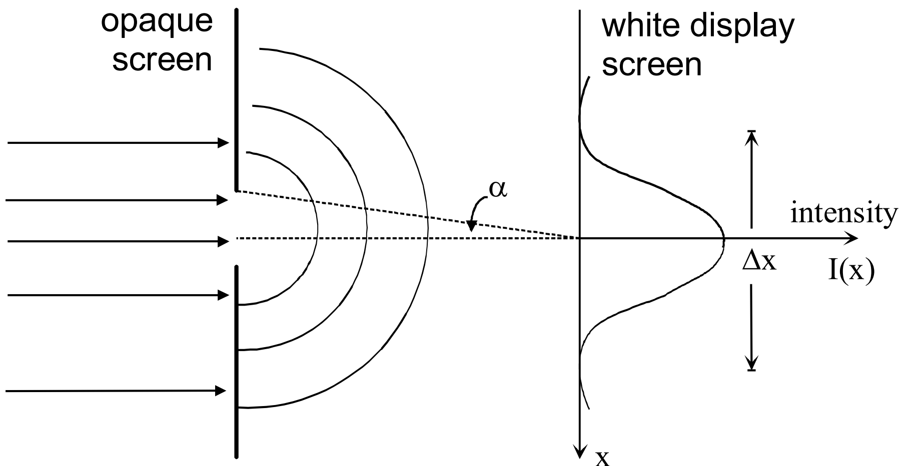

Figure 1-2. Diffraction of light by a slit, or by a circular aperture. Waves spread out from the aperture and fall on a white screen to produce a disk of confusion (Airy disk) whose intensity distribution $I(x)$ is shown by the graph on the right.

parallel beam of light strikes an opaque diaphragm containing a circular aperture whose radius 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 diameter Δx *exceeds* that of the aperture. In fact, for an aperture of small diameter, diffraction effects cause Δx actually to *increase* as the aperture size is reduced, in accordance with the **Rayleigh** criterion:

$$\Delta x \approx 0.6\,\lambda / \sin\alpha \qquad (1.1)$$

where λ is the wavelength of the light being diffracted.

Equation (1.1) can be applied to the eye, with the aid of Fig. 1-1b, which shows an equivalent image formed in air at a distance f from a single focusing lens. For wavelengths in the middle of the visible region of the spectrum, $\lambda \approx 500$ nm and taking $d \approx 4$ mm and $f \approx 2$ cm, the geometry of Fig. 1-1b gives $\tan\alpha \approx (d/2)/f = 0.1$, which implies a small value of α and allows use of the small-angle approximation: $\sin\alpha \approx \tan\alpha$. Equation (1.1) then gives the diameter of the disk of confusion as $\Delta x \approx (0.6)(500\text{ nm})/0.1 = 3$ μm.

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

To a reasonable approximation, these three different 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 overall image blurring Δ is given by:

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

which leads to $\Delta \approx 6$ μm as the blurring of the retinal image. This value corresponds to an *angular* blurring for distant objects (see Fig. 1-1b) of

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

Distant objects (or details within objects) can 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, required a *telescope*, which provided *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. 1c) to a lateral dimension of:

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

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

Because there are many interesting objects *below* this size, including the examples in Table 1-1, an optical device with **magnification factor** M (> 1) is needed to see them; in other words, a **microscope**.

To resolve a small object of diameter D, we need a magnification M^* such that the *magnified* diameter ($M^* D$) at the eye's object plane is greater or equal to the object resolution ΔR (≈ 75 μm) of the eye. In other words:

$$M^* = (\Delta R)/D \qquad (1.5)$$

Values of this minimum magnification are given in the right-hand column of Table 1-1, for objects of various diameter D.

1.2 The Light-Optical Microscope

Light microscopes were developed in the early 1600's, and some of the best observations were made by Anton van Leeuwenhoek, using tiny glass lenses placed very close to the object and to the eye; see Fig. 1-3. By the late 1600's, this Dutch scientist had observed blood cells, bacteria, and structure *within* the cells of animal tissue, all revelations at the time. But this simple one-lens device had to be positioned very accurately, making observation very tiring in practice.

For routine use, it is more convenient to have a **compound microscope**, containing at least two lenses: an **objective** (placed close to the *object* to be magnified) and an **eyepiece** (placed fairly close to the *eye*). By increasing its dimensions or by employing a larger number of lenses, the magnification M of a compound microscope can be increased indefinitely. However, a large value of M does not guarantee that objects of vanishingly small diameter D can be visualized; in addition to satisfying Eq. (1-5), we must ensure that aberrations and diffraction *within the microscope* are sufficiently low.

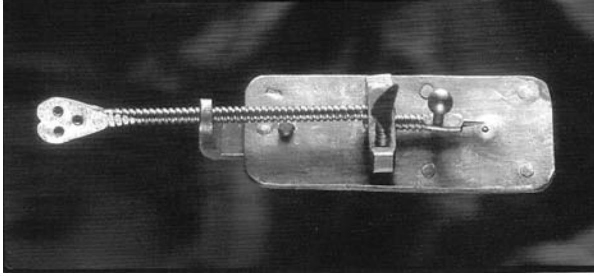

Figure 1-3. One of the single-lens microscopes used by van Leeuwenhoek. The adjustable pointer was used to center the eye on the optic axis of the lens and thereby minimize image aberrations. Courtesy of the FEI Company.

Nowadays, the aberrations of a light-optical instrument can be made unimportant 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 lenses, 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. With a large-aperture lens ($\sin \alpha \approx 1$), 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), this means a best-possible object resolution of about 0.3 μm.

This is a substantial improvement over the resolution (≈ 75 μm) of the unaided eye. But to achieve this resolution, the microscope must magnify the object to a diameter at least equal to ΔR, so that *overall* resolution is determined by microscope diffraction rather than the eye's limitations, requiring a microscope magnification of $M \approx$ (75 μm)/(0.3 μm) = 250. Substantially larger values ("empty magnification") do not significantly improve the sharpness of the magnified image and in fact reduce the **field of view**, the area of the object that can be simultaneously viewed in the image.

Light-optical microscopes are widely used in 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, which creates a real image on the retina of the eye or within an attached camera. Variation in the light intensity (**contrast**) in the image occurs because different parts of the specimen *absorb* light to differing degrees. By using **stains** (light-absorbing chemicals attach themselves preferentially to certain regions of the specimen), the contrast can be increased; the image of a tissue section may then reveal the

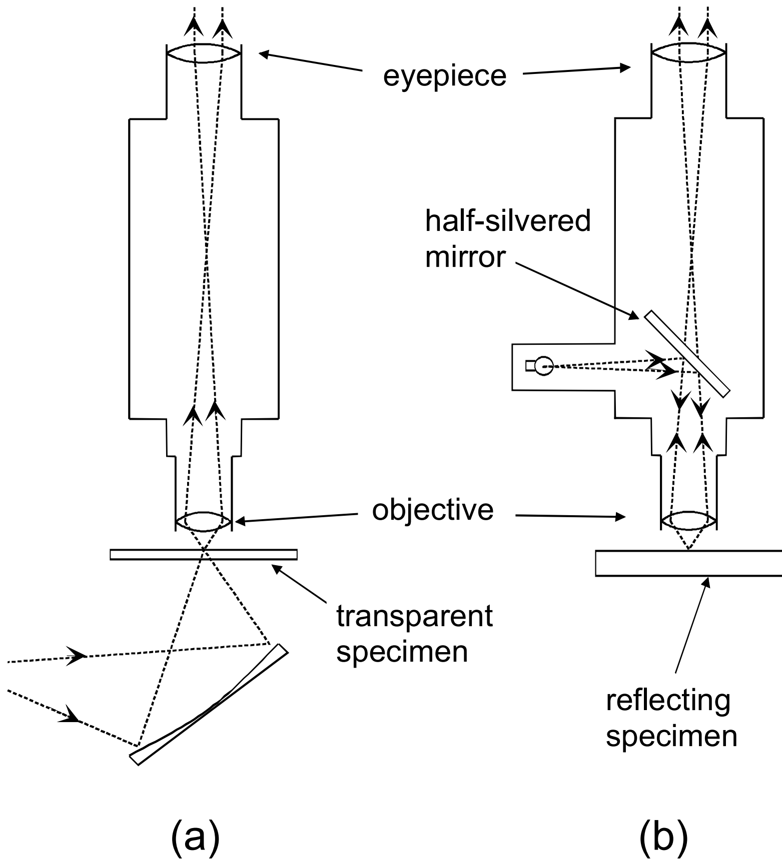

Figure 1-4. Schematic diagrams of (a) a biological microscope, which images light transmitted through the specimen, and (b) a metallurgical microscope, which uses light (often from a built-in illumination source) reflected from the specimen surface.

individual components (organelles) within each biological cell. Because the light travels *through* the specimen, this instrument can also be called a **transmission** light microscope. It is used also by geologists, who are able to 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 cannot easily be made thin enough to be optically transparent. Here, the image is formed by light *reflected* from the surface of the specimen. Because perfectly smooth surfaces 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. In this way,

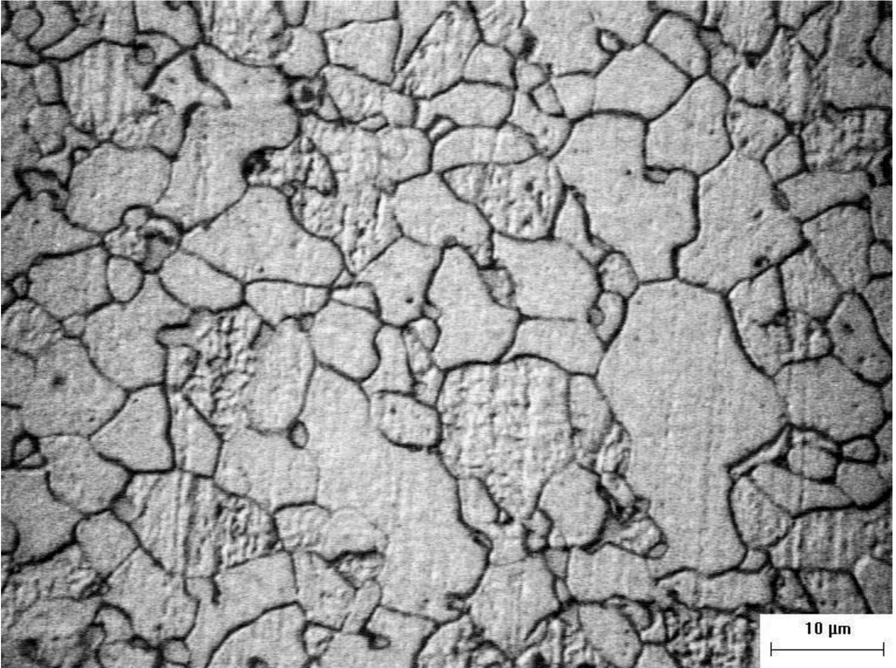

Figure 1-5. Light-microscope image of a polished and etched specimen of X70 pipeline steel, showing dark lines representing the grain boundaries between ferrite (bcc iron) crystallites. Courtesy of Dr. D. Ivey, University of Alberta.

the microscope reveals the microstructure of crystalline materials, such as the different phases present in a metal alloy. Most etches preferentially dissolve the regions between individual crystallites (grains) of the specimen, where the 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 indicated by Eq. (1.1), one possibility for *improving* resolution (which means *reducing* Δx, and therefore Δ and ΔR) is to decrease the wavelength λ of the radiation. The simplest 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 that 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 in resolution comes from using **ultraviolet** (UV) radiation, meaning wavelengths in the range 100 – 300 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 visible light. Because ordinary glass strongly absorbs UV light, 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

Being electromagnetic waves with a wavelength shorter than those of UV light, x-rays offer the possibility of even better spatial resolution. This radiation cannot be focused by convex or concave lenses, as 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 the planes of atoms in a solid, 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 nm to 10 nm. Soft x-rays are diffracted by structures whose periodicity is several nm, such as thin-film multilayers that act as focusing mirrors, or **zone plates**, which are essentially diffraction gratings with circular symmetry (see Fig. 5-23) that focus monochromatic x-rays (those of a single wavelength); as depicted Fig. 1-6.

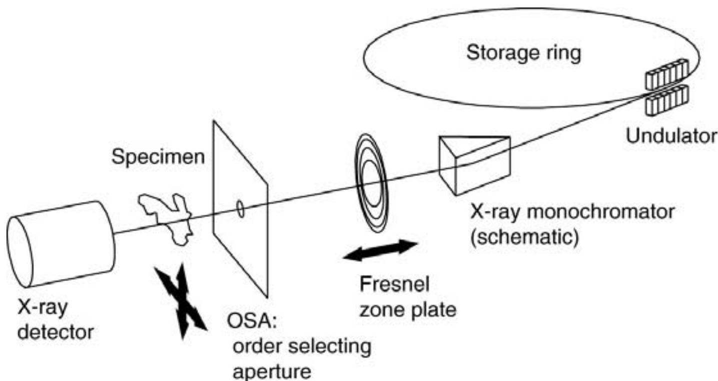

Figure 1-6. Schematic diagram of a scanning transmission x-ray microscope (STXM) attached to a synchrotron radiation source. The monochromator transmits x-rays with a narrow range of wavelength, and these monochromatic rays are focused onto the specimen by means of a Fresnel zone plate. The order-selecting aperture ensures that only a single x-ray beam is focused and scanned across the specimen. From Neuhausler *et al*. (1999), courtesy of Springer-Verlag.

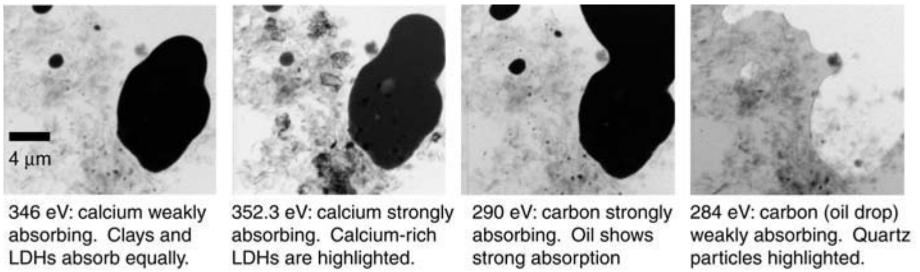

Figure 1-7. Scanning transmission x-ray microscope (STXM) images of a clay-stabilized oil-water emulsion. By changing the photon energy, different components of the emulsion become bright or dark, and can be identified from their known x-ray absorption properties. From Neuhausler *et al.* (1999), courtesy of Springer-Verlag.

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 often recorded over many minutes or hours). This situation prevented the practical realization of an x-ray microscope until the development of an 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 and wobblers can also be inserted into the ring; an array of magnets causes additional deviation of the electron from a straight-line path and produces a strong bremsstrahlung effect, as in Fig. 1-6. Synchrotron x-ray sources are large and expensive (> $100M) but their radiation has a variety of uses; several dozen have been constructed throughout the world during the past 20 years.

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 the microscopy. In this case, x-rays in the wavelength range 2.3 to 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, determined largely by zone-plate focusing, is typically 30 nm.

In contrast, 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 a special "environmental cell," 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. Inspired by Einstein's photon description of electromagnetic radiation, Louis de Broglie proposed that their wavelength is given by

$$\lambda = h/p = h/(mv) \tag{1.5}$$

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. Because this wavelength is comparable to atomic dimensions, such "slow" electrons are strongly diffracted from the regular array of atoms at the surface of a crystal, as first observed by Davisson and Germer (1927).

Raising the accelerating potential to 50 kV, the 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 the solid is crystalline, the electrons are diffracted by atomic planes inside the material, as in the case of 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 G.P. 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 much better than 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 used in cathode-ray tubes, TV display tubes, and computer screens. In fact, the first electron microscopes made use of technology already developed for radar applications of cathode-ray tubes. In a **transmission electron microscope** (TEM), electrons penetrate a *thin* specimen and are then imaged by 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, producing a magnetic field centered along the optic axis. By 1933, Ruska had added a third lens and obtained images of cotton fiber and aluminum foil with a resolution somewhat better than that of the light microscope.

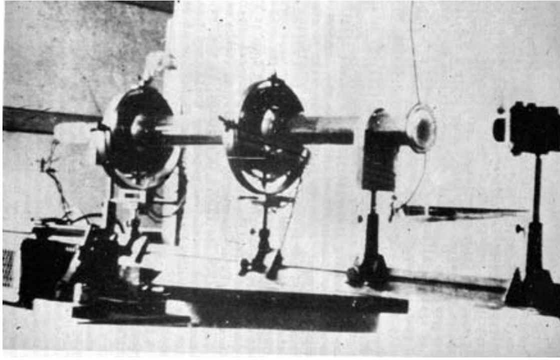

Figure 1-8. Early photograph of a horizontal two-stage electron microscope (Knoll and Ruska, 1932). This material is used by permission of Wiley-VCH, Berlin.

Similar microscopes were built by Marton and co-workers in Brussels, who by 1934 had produced the first images of nuclei within the interior of biological cells. These early TEMs used a horizontal sequence of lenses, as in Fig. 1-8, but such an arrangement was abandoned after it was realized that precise alignment of the lenses along the optic axis is critical to obtaining the best resolution. By stacking the lenses in a vertical column, good alignment can be maintained for a longer time; 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, often constructed out of glass and containing boiling mercury. The electrons were accelerated by applying a high voltage, generated by an electronic oscillator circuit and a step-up transformer. As the transistor had not been invented, the oscillator circuit used vacuum-tube electronics. In fact, vacuum tubes were used in high-voltage circuitry (including television receivers) until the 1980`s because they are less easily damaged by voltage spikes, which occur when there is high-voltage discharge (not uncommon at the time). Vacuum tubes were also used to control and stabilize the dc current applied to the electron lenses.

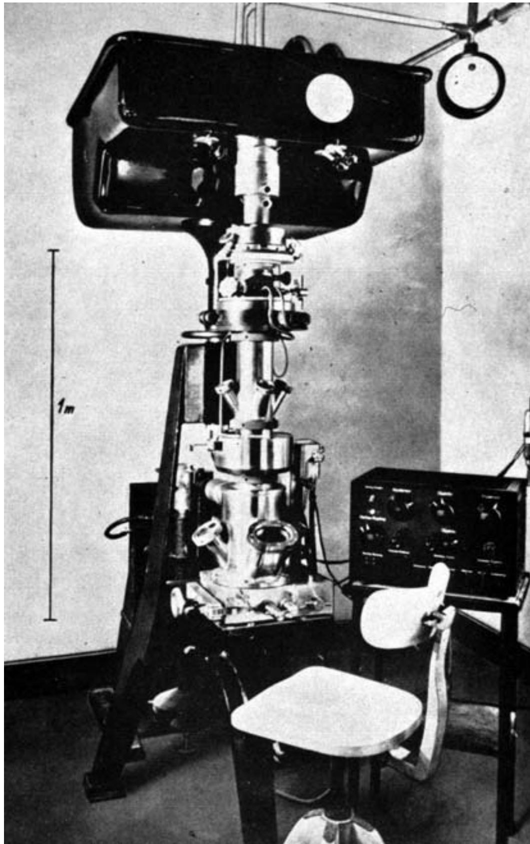

Figure 1-9. First commercial TEM from the Siemens Company, employing three magnetic lenses that were water-cooled and energized by batteries. The objective lens used a focal length down to 2.8 mm at 80 kV, giving an estimated resolution of 10 nm.

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

The further development of the TEM is illustrated by the two JEOL instruments shown in Fig. 1-10. Their model 100B (introduced around 1970) used both vacuum tubes and transistors for control of the lens currents and the high voltage (up to 100 kV) and gave a spatial resolution of 0.3 nm. Model 2010 (introduced 1990) employed integrated circuits and digital control; at 200 kV accelerating voltage, it provided a resolution of 0.2 nm.

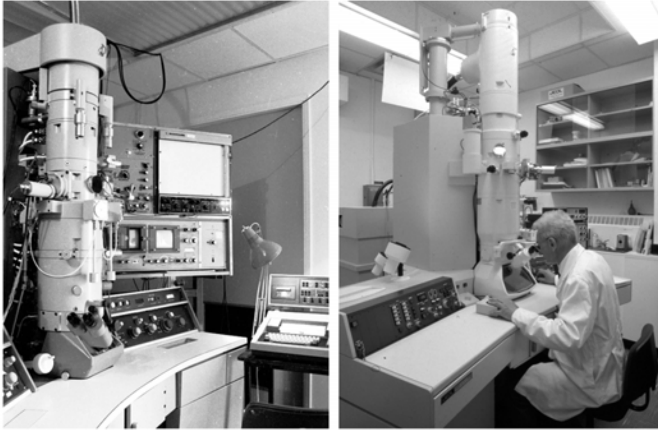

Figure 1-10. JEOL transmission electron microscopes: (a) model 100B and (b) model 2010.

The TEM has proved invaluable for examining the ultrastructure of metals. For example, crystalline defects known as dislocations were first predicted by theorists to account for the fact that metals deform under much lower forces than calculated for perfect crystalline array of atoms. They were first seen directly 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; detail can now be seen within each metal crystallite. With a modern TEM (resolution ≈ 0.2 nm), it is even possible to image individual atomic planes or columns of atoms, as we will discuss in Chapter 4.

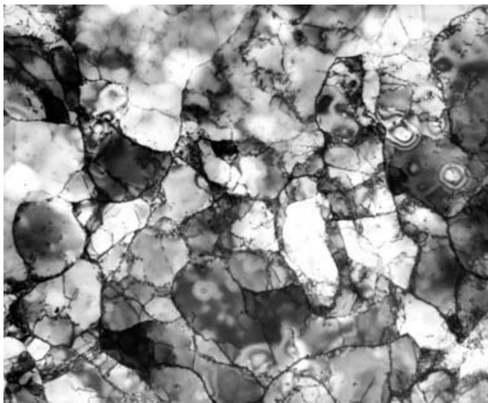

Figure 1-11. TEM diffraction-contrast image ($M \approx 10{,}000$) of polycrystalline aluminum. Individual crystallites (grains) show up with different brightness levels; low-angle boundaries and dislocations are visible as dark lines within each crystallite. Circular fringes (top-right) represent local changes in specimen thickness. Courtsey of M.J. Whelan, Oxford University.

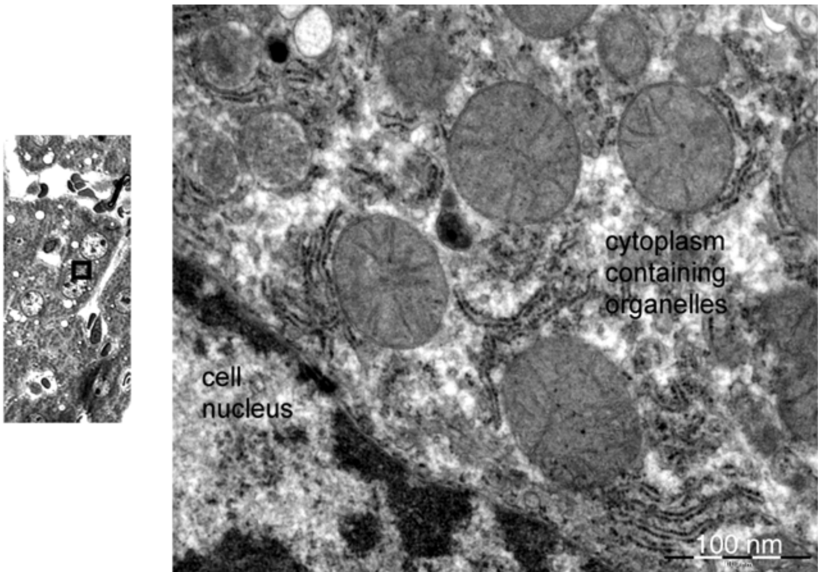

Figure 1-12. TEM image of a stained specimen of mouse-liver tissue, corresponding approximately to the small rectangular area within the light-microscope image on the left. Courtesy of R. Bhatnagar, Biological Sciences Microscopy Unit, University of Alberta.

The TEM has been equally useful 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 TEM image reveals much more structure in the organelles, due to its higher spatial resolution.

Although most modern TEMs use an electron accelerating voltage between 100 kV and 300 kV, a few high-voltage instruments (HVEMs) have been constructed with accelerating voltages as high as 3 MV; see Fig. 1-13. The main motivation was the fact that increasing the electron energy (and therefore momentum) decreases the de Broglie wavelength of electrons and therefore lowers 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 advantageous for looking at thicker specimens, as 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 that it could be used 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 comparable to that at the center of an exploding nuclear weapon. Not surprisingly, therefore, it was found that TEM observation kills living tissue in much less time than needed to record a high-resolution image.

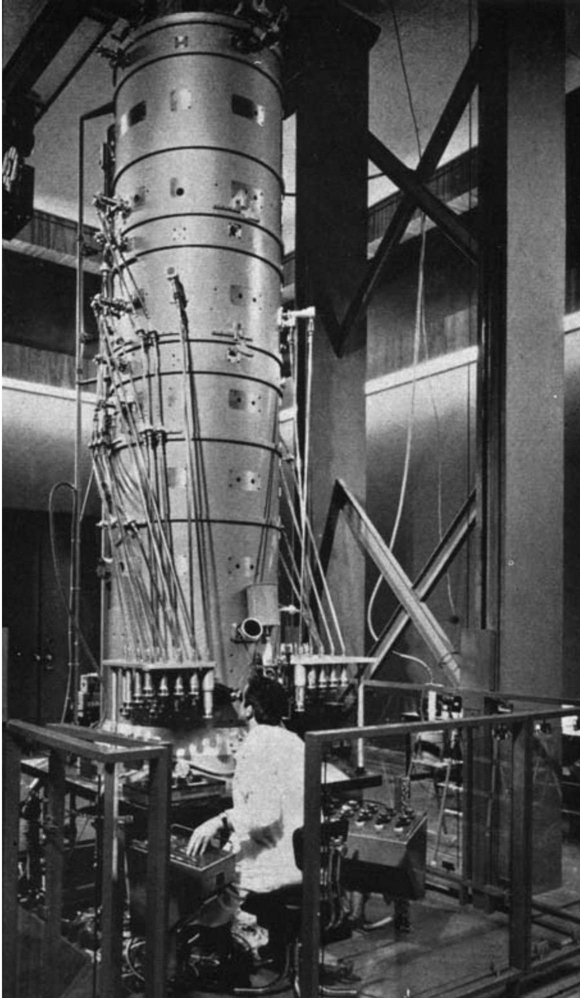

Figure 1-13. A 3 MV HVEM constructed at the C.N.R.S. Laboratories in Toulouse and in operation by 1970. To focus the high-energy electrons, large-diameter lenses were required, and the TEM column became so high that long control rods were needed between the operator and the moving parts (for example, to provide specimen motion). Courtesy of G. Dupouy, personal communication.

1.5 The Scanning Electron Microscope

One limitation of the TEM is that, unless the specimen is made very thin, electrons are strongly scattered within the specimen, or even absorbed rather than transmitted. This constraint has provided the incentive to develop electron microscopes that are capable of examining relatively thick (so-called **bulk**) specimens. In other words, there is need of an electron-beam instrument that is equivalent to the metallurgical light microscope but which offers the advantage of 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 for the incoming (**primary**) electrons to supply energy to the *atomic* electrons that are present in a solid, which can then be released as **secondary electrons**. These electrons are emitted with a range of energies, making it more difficult 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, making use of the fact that electrostatic or magnetic fields, applied at right angles to the beam, can be used to change its direction of travel. By scanning simultaneously in two perpendicular directions, a square or rectangular area of specimen (known as a **raster**) can be covered and an image of this area can be formed by collecting secondary electrons from each point on the specimen.

The same raster-scan signals can be used to deflect the beam generated within a cathode-ray tube (CRT), in exact synchronism with the motion of the electron beam that is focused on the specimen. If the secondary-electron signal is amplified and applied to the electron gun of the CRT (to change the number of electrons reaching the CRT screen), the resulting brightness variation on the phosphor represents a secondary-electron image of the specimen. In raster scanning, the image is generated serially (point by point) rather than simultaneously, as in the TEM or light microscope. A similar principle is used in the production and reception of television signals.

A **scanning electron microscope** (SEM) based on secondary emission of electrons was developed at the RCA Laboratories in New Jersey, under wartime conditions. Some of the early prototypes employed a field-emission electron source (discussed in Chapter 3), whereas later models used a heated-filament source, the electrons being focused onto the specimen by electrostatic lenses. An early version of a FAX machine was employed for image recording; see Fig. 1-14. The spatial resolution was estimated to be 50 nm, nearly a factor of ten better than the light-optical microscope.

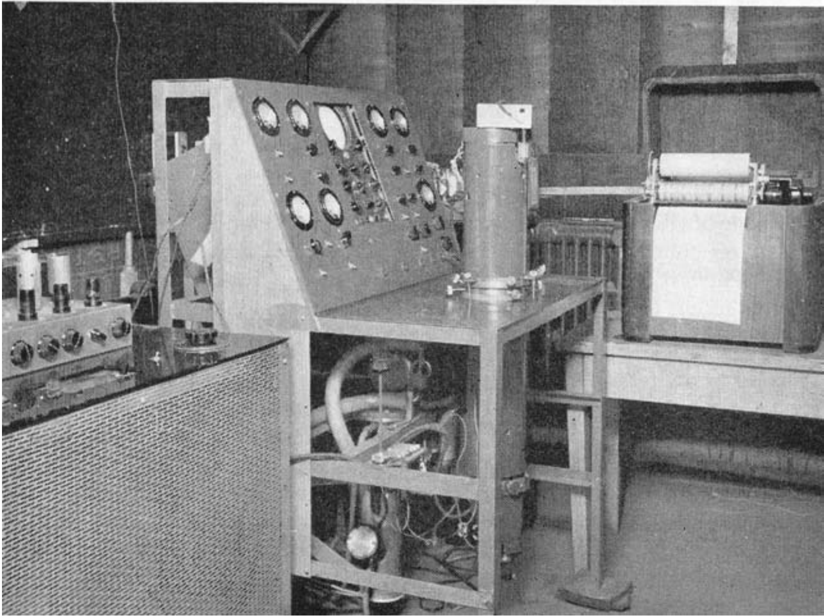

Figure 1-14. Scanning electron microscope at RCA Laboratories (Zwyorkin *et al.*, 1942) using electrostatic lenses and vacuum-tube electronics (as in the amplifier on left of picture). An image was produced on the facsimile machine visible on the right-hand side of the picture. This material is used by permission of John Wiley & Sons, Inc.

Further SEM development occurred after the Second World War, when Charles Oatley and colleagues began a research and construction program in the 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 one example of a modern instrument. Image information is stored in a computer that controls the SEM, and the image appears on the display monitor of the computer.

A modern SEM provides an image resolution typically between 1 nm and 10 nm, not as good as the TEM but much superior to the light microscope. In addition, SEM images have a relatively large **depth of focus**: specimen features that are displaced from the plane of focus appear almost sharply in-focus. As we shall see, this characteristic results from the fact that electrons in the SEM (or the TEM) travel very close to the optic axis, a requirement for obtaining good image resolution.

Figure 1-15. Hitachi-S5200 field-emission scanning electron microscope. This instrument can operate in SEM or STEM mode and provides an image resolution down to 1 nm.

1.6 Scanning Transmission Electron Microscope

It is possible to employ the fine-probe/scanning technique with a thin sample and record, instead of secondary electrons, the electrons that emerge (in a particular direction) from the opposite side of the specimen. The resulting is a **scanning-transmission electron microscope** (STEM). The first STEM was constructed by von Ardenne in 1938 by adding scanning coils to a TEM, and today many TEMs are equipped with scanning attachments, 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 that is often used in the SEM (and TEM) must be replaced by a field-emission source, in which electrons are released from a very sharp tungsten tip under the application 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 scanning mode. The field-emission gun required ultra-high vacuum (UHV), meaning ambient pressures around 10^{-8} Pa. After five years of development, this type of instrument produced the first-ever images of single atoms, visible as bright dots on a dark background (Fig. 1-16).

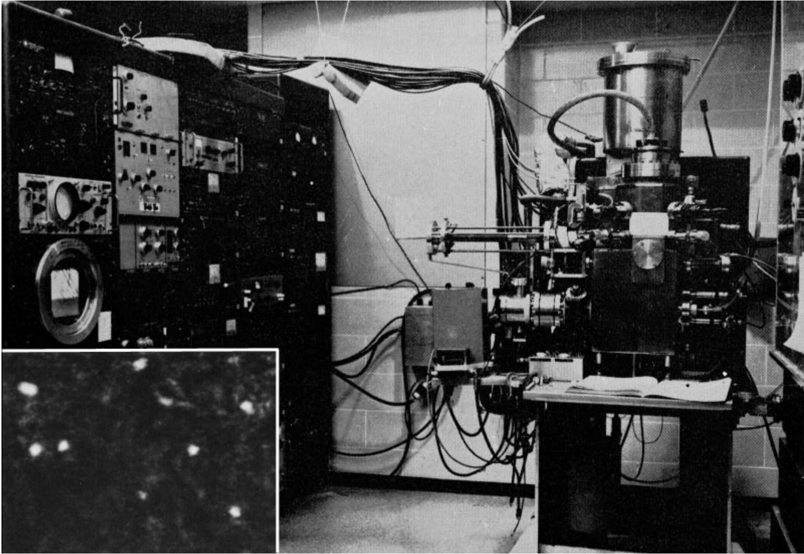

Figure 1-16. Photograph of Chicago STEM and (bottom-left inset) image of mercury atoms on a thin-carbon support film. Courtesy of Dr. Albert Crewe (personal communication).

Atomic-scale resolution is also available in the conventional (fixed-beam) TEM. A crystalline specimen is oriented so that its atomic columns lie parallel to the incident-electron beam, and it is actually *columns* of atoms that are imaged; see Fig. 1-17. It was originally thought that such images might reveal structure within each atom, but such an interpretation is questionable. In fact, the internal structure of the atom can be deduced by analyzing the angular distribution of scattered charged particles (as first done for alpha particles by Ernest Rutherford) without the need to form a direct image.

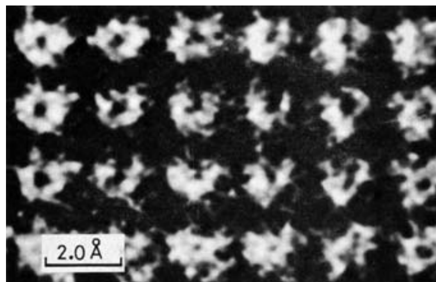

Figure 1-17. Early atomic-resolution TEM image of a gold crystal (Hashimoto *et al.*, 1977), recorded at 65 nm defocus with the incident electrons parallel to the 001 axis. Courtesy Chairperson of the Publication Committee, The Physical Society of Japan, and the authors.

1.7 Analytical Electron Microscopy

All of the images seen up to now provide information about the *structure* of a specimen, in some cases down to the atomic scale. But often there is a need for chemical information, such as the local chemical composition. For this purpose, we require some response from the specimen that is sensitive to the exact atomic number Z of the atoms. As Z increases, the nuclear charge increases, drawing electrons closer to the nucleus and changing their energy. The electrons that are of most use to us are not the outer (valence) electrons but rather the **inner-shell** electrons. Because the latter do not take part in chemical bonding, their energies are unaffected by the surrounding atoms and remain indicative of the nuclear charge and therefore atomic number.

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

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

1.8 Scanning-Probe Microscopes

The raster method of image formation is also employed in a scanning-probe microscope, 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) in which a sharp conducting tip is brought within about 1 nm of the sample and a small potential difference ($\approx$ 1 V) is applied. Provided the tip and specimen are electrically conducting, electrons move between the tip and the specimen by the process of **quantum-mechanical tunneling**. This phenomenon is a direct result of the wavelike characteristics 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. Because the tunneling current increases dramatically with decreasing tip-sample separation, a motorized gear system can be set up to advance the tip towards the sample (in the *z*-direction) until a pre-set tunneling current (e.g. 1 nA) is achieved; see Fig. 1-18a. The tip-sample gap is then about 1 nm in length and fine *z*-adjustments can be made with a piezoelectric drive (a ceramic crystal that changes its length when voltage is applied). If this gap were to decrease, due to thermal expansion or contraction for example, the tunneling current would start to increase, raising the voltage across a series resistance (see Fig. 1-18a). This voltage change is amplified and applied to the piezo *z*-drive so as to *decrease* the gap and return the current to its original value. Such an arrangement is called **negative feedback** because information about the gap length is fed back to the electromechanical system, which acts to keep the gap constant.

To perform scanning microscopy, the tip is raster-scanned across the surface of the specimen in *x*- and *y*-directions, again using piezoelectric drives. If the negative-feedback mechanism remains active, the gap between tip and sample will remain constant, and the tip will move in the *z*-direction in exact synchronism with the 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 modulate the beam in a CRT display device (as in the SEM) or stored in computer memory as a topographical image.

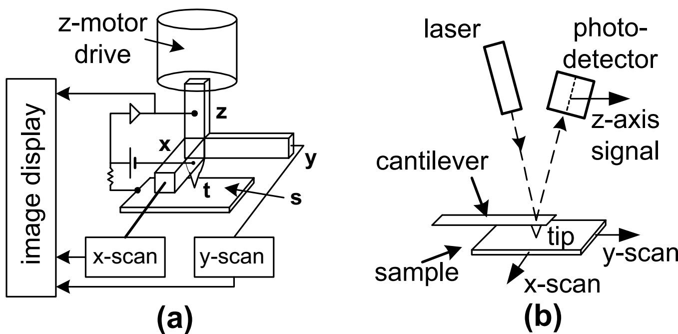

Figure 1-18. (a) Principle of the scanning tunneling microscope (STM); **x** , **y,** and **z** represent piezoelectric drives, **t** is the tunneling tip, and **s** is the specimen. (b) Principle of the scanning force (or atomic force) microscope. In the *x*- and *y*-directions, the tip is stationary and the sample is raster-scanned by piezoelectric drives.

A remarkable feature of the STM is the high spatial resolution that can be achieved: better than 0.01 nm in the *z*-direction, which follows directly 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 when the tip is not atomically sharp (STM tips are made by electrolytic sharpening of a wire or even by using a mechanical wire cutter). The explanation is again in terms of the strong dependence of tunneling current on gap length: most of the electrons tunnel from a single tip atom that is nearest to the specimen, even when other atoms are only slightly further away. As a result, the STM can be used to study the structure of surfaces with single-atom resolution, as illustrated in Fig. 1-19.

Nevertheless, problems can occur due to the existence of "multiple tips", resulting in false features (**artifacts**) in an image. Also, the scan times become long if the scanning is done slowly enough for the tip to move in the *z*-direction. As a result, *atomic-resolution* images are sometimes 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; even for a very smooth surface, the tip would eventually crash into the specimen, damaging the tip and/or specimen.

Figure 1-19. 5 nm × 5 nm area of a hydrogen-passivated Si (111) surface, imaged in an STM with a tip potential of –1.5 V. The subtle hexagonal structure represents H-covered Si atoms, while the two prominent white patches arise from dangling bonds where the H atoms have been removed (these appear non-circular because of the somewhat irregularly shaped tip). Courtesy of Jason Pitters and Bob Wolkow, National Institute of Nanotechnology, Canada.

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

The STM was 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 TEM). It quickly inspired other types of scanning-probe microscope, such as the **atomic force microscope** (AFM) in which a sharp tip (at the end of a cantilever) is brought sufficiently close to the surface of a specimen, so that it essentially touches it 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 nanometers. But in the 1990's, the instrument was refined to give near-atomic resolution.

Initially, the z-motion of the cantilever was detected by locating an STM tip immediately above. Nowadays it is usually achieved 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 (from silicon nitride) in large quantities, using the same kind of photolithography process that yields 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 misleading artifacts such as multiple-tip effects.

The mechanical force is repulsive if the tip is in direct contact with the sample, but at a small distance above, the tip senses an attractive (van der Waals) force. Either regime may be used to provide images. Alternatively, a 4-quadrant photodetector can sense torsional motion (twisting) of the AFM cantilever, which results from a sideways frictional force, giving an image that is essentially a map of the local coefficient of friction. Also, with a modified tip, the magnetic field of a sample can be monitored, allowing the direct imaging of magnetic data-storage media materials for example.

Although is more difficult to obtain atomic resolution than with an STM, the AFM has the advantage that it does not require a conducting sample. In fact, the AFM can operate with its tip and specimen immersed in a liquid such as water, making the instrument valuable for imaging biological specimens. This versatility, combined with its 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 large areas of specimen is very time-consuming; it is less feasible to zoom in and out (by changing magnification) than with an

electron-beam instrument. Also, there is no way of implementing 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. And except in special cases, a scanning-probe image represents only the surface properties of a specimen and not the *internal* structure that is visible using a TEM.

Figure 1-20a shows a typical AFM image, presented in a conventional way in which local changes in image brightness represent variations in surface height (motion of the tip in the z-direction). However, scanning-probe images are often presented as so-called y-modulation images, in which z-motion of the tip is used to deflect the electron beam of the CRT display in the y-direction, perpendicular to its scan direction. This procedure gives a three-dimensional effect, equivalent to viewing the specimen surface at an oblique angle rather than in the 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 but making them more easily visible; see Fig. 1-20b.

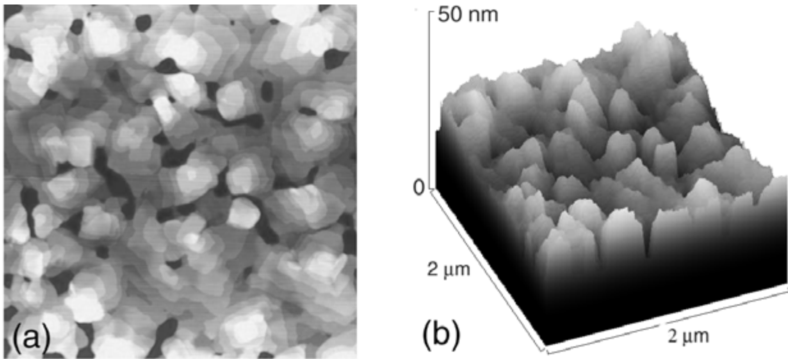

Figure 1-20. AFM images of a vacuum-deposited thin film of the organic semiconductor pentacene: (a) brightness-modulation image, in which abrupt changes in image brightness represent steps (terraces) on the surface, and (b) y-modulation image of the same area. Courtesy of Hui Qian, University of Alberta.

Chapter 2

ELECTRON OPTICS

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

2.1 Properties of an Ideal Image

Clearly, an optical image is closely related 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 relating electric and magnetic fields that underlie all electrostatic and magnetic phenomena, including electromagnetic waves. In a journal article (Maxwell, 1858), remarkable for its clarity and for its frank comments about fellow scientists, he stated the requirements of a perfect image 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 (in practice) when 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 an *equivalent* point in the image. In many forms of microscopy, the connection between these two points is made by some kind of particle (*e.g.,* electron or photon) that leaves the object and ends up at the related image point. It is conveyed from object to image through a focusing device (such as a 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 this rule is obeyed depends on several properties of the lens. For example, if the focusing *strength* is incorrect, the image formed at a particular plane will be **out-of-focus**. Particles leaving a single point in the object then arrive anywhere within a circle *surrounding* the ideal-image point, a so-called **disk of confusion**. But even if the focusing power is appropriate, a real lens can produce a disk of confusion because of lens **aberrations**: particles having different energy, or which take different paths after leaving the object, arrive displaced from the "ideal" image point. The image then appears blurred, with loss of fine detail, just as in the case of an out-of-focus image.

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 define a triangle and their locations in the image represent a triangle that is *similar* in the geometric sense: it contains the same angles. The image triangle may have a different orientation than that of the object triangle; for example it could be **inverted** (relative to the optic axis) without violating Rule 2, as in Fig. 2-1. Also, 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 $M < 1$).

Although the light image formed by a glass lens may appear similar to the object, close inspection often reveals 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 image distortion is equivalent to a variation of the magnification factor 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 M decreasing away from the axis (Fig. 2-2b). As we will see, many electron lenses cause a *rotation* of the image, and if this rotation increases with distance from the axis, the result is **spiral** distortion (Fig. 2-2c).

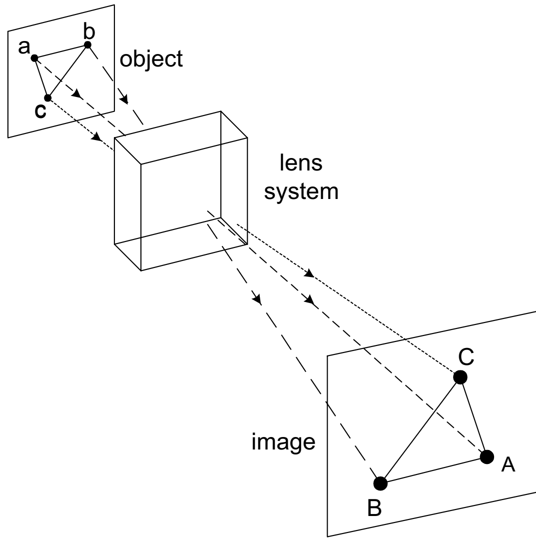

Figure 2-1. A triangle imaged by an ideal lens, with magnification and inversion. Image points A, B, and C are equivalent to the object points a , b, and c, respectively.

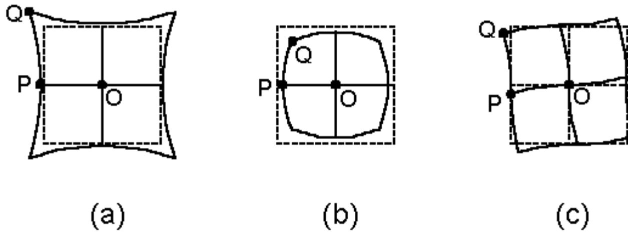

Figure 2-2. (a) Square mesh (dashed lines) imaged with pincushion distortion (solid curves); magnification *M* is higher at point Q than at point P. (b) Image showing barrel distortion, with *M* at Q lower than at P. (c) Image of a square, showing spiral distortion; the counterclockwise rotation is higher at Q than at P.

Rule 3: Images usually exist in two dimensions and occupy a *flat* plane. Even if the corresponding object is three-dimensional, only a single object plane is *precisely* in focus in the image. 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**; the sharpest image would be formed on a surface

that is curved rather than planar. Inexpensive cameras have sometimes compensated for this lens defect by curving the photographic film. Likewise, the Schmidt astronomical telescope installed at Mount Palomar was designed to record a well-focused image of a large section of the sky on 14-inch square glass plates, bent into a section of a sphere using vacuum pressure.

To summarize, focusing aberrations occur when Maxwell's Rule 1 is broken: the image appears blurred because rays coming from a single point in the object are focused into a *disk* rather than a single point in the image plane. Distortion occurs when Rule 2 is broken, due to a change in image magnification (or rotation) with position in the object plane. Curvature of field occurs when Rule 3 is broken, due to a change in focusing power with position in the object plane.

2.2 Imaging in Light Optics

Because electron optics makes use of many of the concepts of light optics, we will quickly review some of the basic optical principles. In light optics, we can employ a glass lens for focusing, 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 the perpendicular to an air/glass interface) is θ_1 in air ($n_1 \approx 1$), the corresponding value θ_2 in glass ($n_2 \approx 1.5$) is smaller by an amount given by Snell's law:

$$n_1 \sin\theta_1 = n_2 \sin_2 \tag{2.1}$$

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 ($\alpha = 0$).

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). 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* toward a single (on-axis) point in the image (Fig. 2-4) and therefore satisfy the first of Maxwell's rules for image formation.

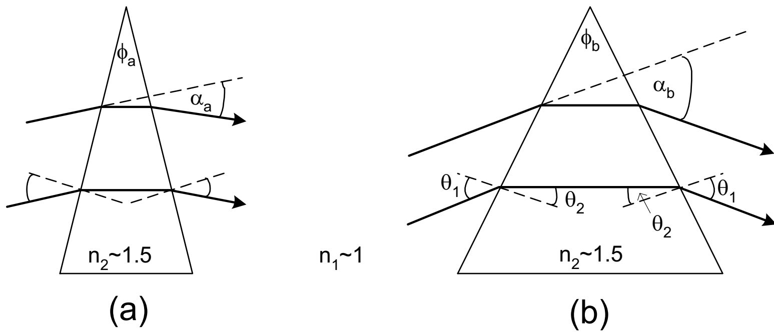

Figure 2-3. Light refracted by a glass prism (a) of small angle ϕ_a and (b) of large angle ϕ_b. Note that the angle of deflection α is independent of the glass thickness but increases with increasing prism angle.

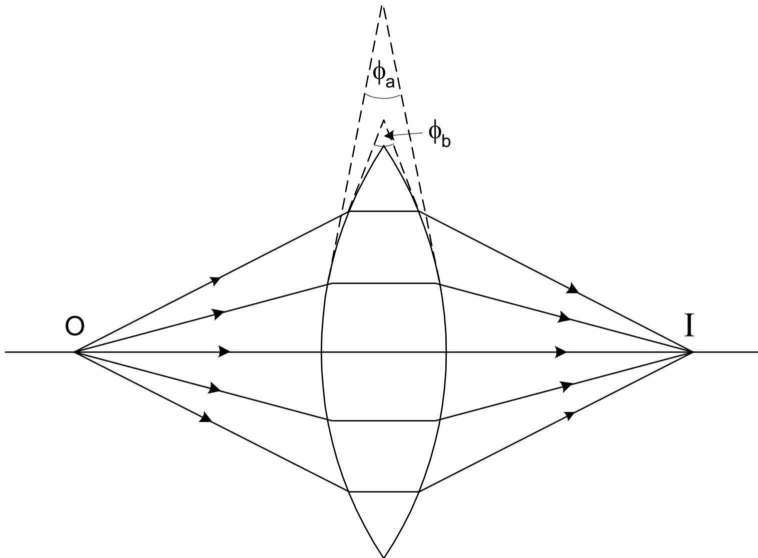

Figure 2-4. A convex lens focusing rays from axial object point O to an axial image point I .

In Fig. 2-4, we have shown 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 may travel at various angles relative to the optic axis. A ray diagram that showed *all* of these rays

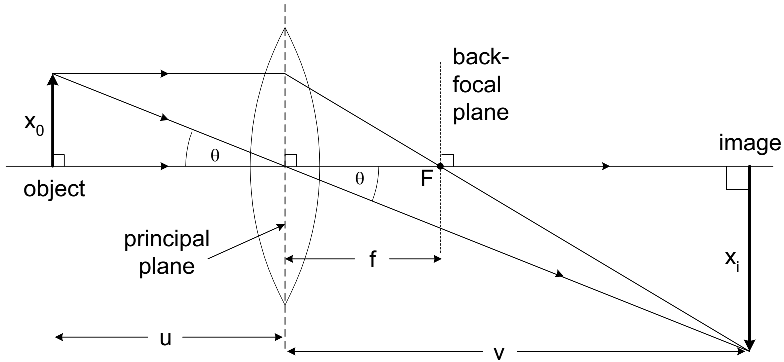

Figure 2-5. A thin-lens ray diagram, in which bending of light rays is imagined to occur at the mid-plane of the lens (dashed vertical line). These special rays define the focal length *f* of the lens and the location of the back-focal plane (dotted vertical line).

would be highly confusing, so in practice it is sufficient to show just a few *special* rays, as in Fig. 2-5.

One special ray is the one that travels *along* the optic axis. Because 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 but happens to pass through the *center of the lens*, remains unchanged in direction. A third example of a special ray is one that leaves the object at distance X_o from the axis and travels *parallel to* the optic axis. The lens bends this ray toward 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, perpendicular to the optic axis and passing through F, is known as the **back-focal plane** of the lens.

In drawing 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 if we can 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 radii of curvature of the lens surfaces are large compared to the focal length, which implies that the lens is *physically* thin, and is therefore 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. Because 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 (not 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 applying Eq. (2.3). The *total* magnification (between second image and original object) is then the *product* of the magnification factors of the two lenses.

If the second lens is placed *within* the image distance of the first, a real image cannot be formed but, 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 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 Chapter 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 that contains a continuous range of wavelengths between its red and violet components. Because the refractive index of glass varies with wavelength (called **dispersion**, as 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 object 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, analogous to visible-light photons. As a result of this analogy, each electron trajectory is often referred to as a *ray* path.

To obtain the equivalent of a convex lens for electrons, we must arrange for the amount of deflection to increase with increasing deviation of the electron ray from the optic axis. For such focusing, we cannot rely on refraction by a material such as glass, as electrons are strongly scattered and absorbed soon after entering a solid. Instead, we take advantage of the fact that the electron has an electrostatic charge and is therefore deflected by an electric field. Alternatively, we can use the fact that the electrons in a beam are moving; the beam is therefore equivalent to an electric current in a wire, and can be deflected by an applied magnetic field.

Electrostatic lenses

The most straightforward example of an electric field is the uniform field produced between two parallel conducting plates. 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, as in a cathode-ray tube, 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 from all points on the electrode and therefore suffers no deflection, whereas an off-axis electron is repelled by the negative charge that lies closest to it and is therefore deflected back toward 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, because 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. 2-5. Note that the electrodes, and therefore the electric fields which give rise to 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.

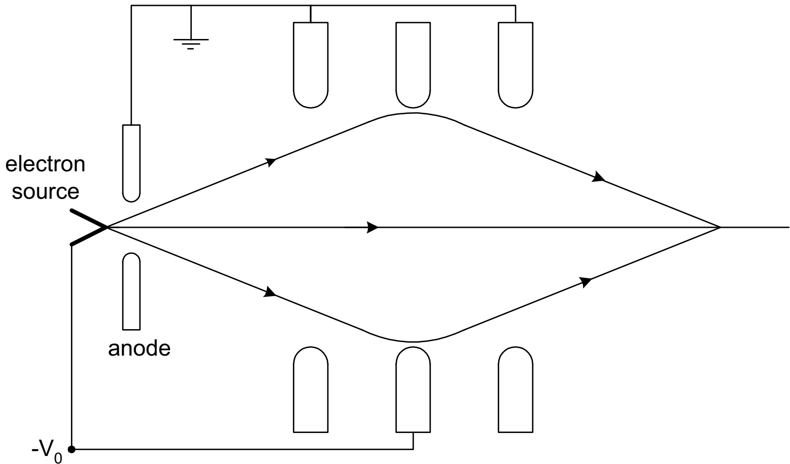

Figure 2-6. Electrons emitted from an electron source, accelerated through a potential difference V_0 toward an anode and then focused by a unipotential electrostatic lens. The electrodes, seen here in cross section, are circular disks containing round holes (apertures) with a common axis, the optic axis of the lens.

Electrostatic lenses have been used in cathode-ray tubes and television picture tubes, to ensure that the electrons emitted from a heated filament are focused back into a small spot on the phosphor-coated inside face of the tube. Although some early electron microscopes used electrostatic lenses, modern electron-beam instruments 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 employs the magnetic field produced by a coil carrying a direct current. As in the electrostatic case, a *uniform* field (applied perpendicular to the beam) would produce overall deflection but no focusing action. To obtain focusing, we need a field with axial symmetry, similar to that of the einzel lens. Such a field is generated by a short coil, as illustrated in Fig. 2-7a. As the electron passes through this non-uniform magnetic field, the force acting on it varies in both magnitude and direction, so it must be represented by a vector quantity $\boldsymbol{F}$. According to 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 $\boldsymbol{B}$ is the magnetic field, representing both the magnitude B of the field (or induction, measured in Tesla) and its direction. The symbol × indicates a

cross-product or **vector product** of ***v*** and ***B*** ; this mathematical operator gives Eq. (2.4) the following two properties.

1. The direction of ***F*** is *perpendicular* to both ***v*** and ***B***. Consequently, ***F*** has *no* component in the direction of motion, implying that the electron speed v (the *magnitude* of the velocity ***v***) remains constant at all times. But because the direction of ***B*** (and possibly ***v***) changes continuously, so does the direction of the magnetic force.

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 ***v*** and ***B*** at the location of the electron. Because B (and possibly v) changes continuously as an electron passes through the field, so does F. Note that for an electron traveling along the coil axis, ***v*** and ***B*** are 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 (***F*** = m d***v***/dt). Such analysis is simplified by considering ***v*** and ***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 from the outset that the magnetic field possesses axial (cylindrical)

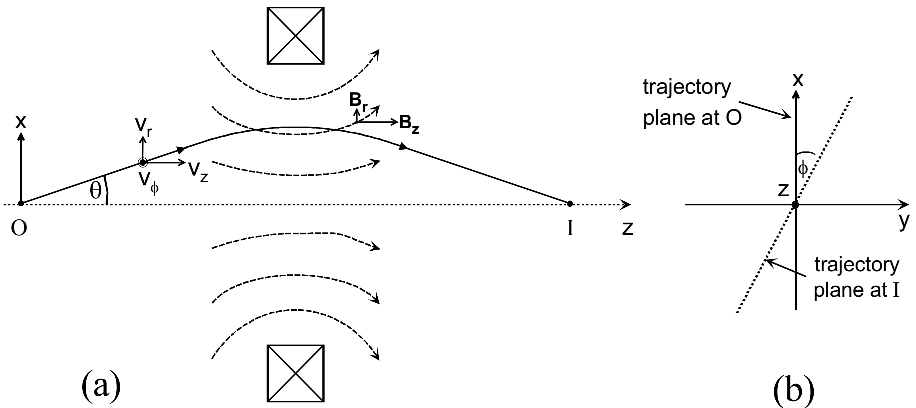

Figure 2-7. (a) Magnetic flux lines (dashed curves) produced by a short coil, seen here in cross section, together with the trajectory of an electron from an axial object point O to the equivalent image point I. (b) View along the z-direction, showing rotation φ of the plane of the electron trajectory, which is also the rotation angle for an extended image produced at I.

symmetry and 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). Therefore, as shown 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:

$$F_\phi = - e\,(v_z B_r) + e\,(B_z v_r) \qquad (2.6a)$$

$$F_r = - e\,(v_\phi B_z) \qquad (2.6b)$$

$$F_z = e\,(v_\phi B_r) \qquad (2.6c)$$

Let us trace the path of an electron that starts from an axial point O and enters the field at an angle θ , defined in Fig. 2-7, relative to the symmetry (z) axis. 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, as 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), also positive, increases. So as a result of both terms in Eq. (2.6a), the electron starts to spiral through the field, acquiring an increasing tangential velocity v_ϕ directed *out* of the plane of Fig. (2-7a). Resulting from this acquired tangential component, a new force F_r starts to act on the electron. According to Eq. (2.6b), this force is negative (toward the z-axis), therefore we have a focusing action: the non-uniform magnetic field acts like a convex lens.

Provided that the radial force F_r toward the axis is large enough, the radial motion of the electron will be *reversed* and the electron will approach the z-axis. Then v_r becomes negative and the *second* term in Eq. (2.6a) becomes negative. And 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; see Fig. 2-7b.

This rotation effect is *not* depicted in Fig (2.7a) or in the 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 effectively suppressing (or ignoring) the rotation effect, we can draw ray diagrams that resemble those of light optics. Even so, it is

important to remember that the trajectory has a rotational component whenever an electron passes through an axially-symmetric magnetic lens.

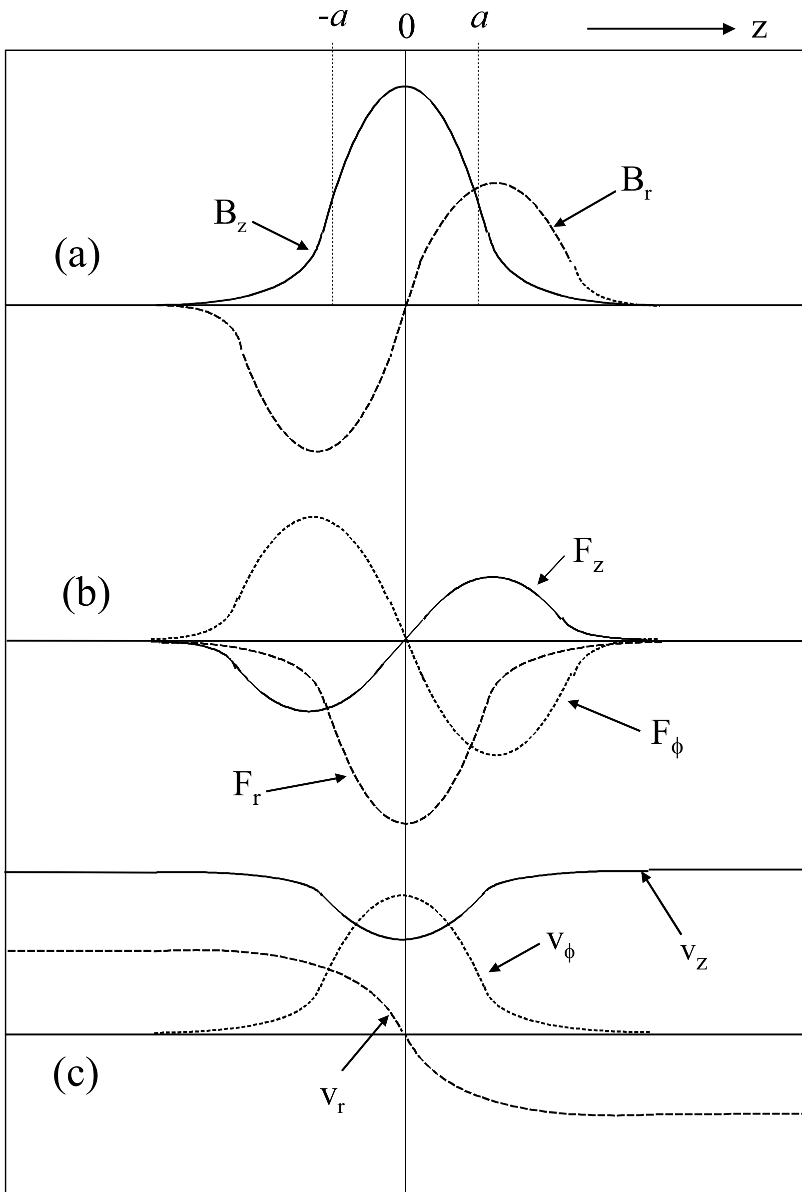

Figure 2-8. Qualitative behavior of the radial (*r*), axial (*z*), and azimuthal (ϕ) components of (a) magnetic field, (b) force on an electron, and (c) the resulting electron velocity, as a function of the *z*-coordinate of an electron going through the electron lens shown in Fig. 2-7a.

As we said earlier, the overall speed v of an electron in a magnetic field remains constant, so the appearance of tangential and radial components of velocity implies that v_z must decrease, as depicted in Fig. 2-8c. This is in accordance with Eq. (2.6c), which predicts the existence of a third force F_z which 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 electron speed is 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 important part in electron focusing. If a long coil (solenoid) were used to generate the field, this component 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. This can be done 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 carries most of the magnetic flux lines. However, the magnetic circuit contains a **gap** filled with nonmagnetic material, so that flux lines appear within the internal **bore** of the lens. The magnetic field experienced by the electron beam can be increased and further concentrated by the use of ferromagnetic (soft iron) **polepieces** of 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 symmetry required for good focusing.

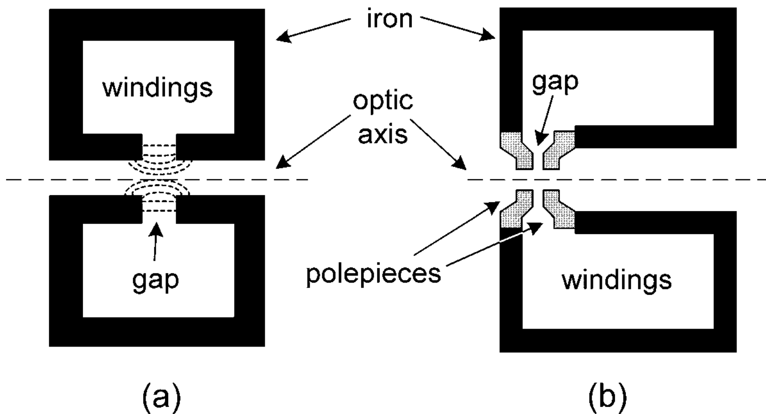

Figure 2-9. (a) Use of ferromagnetic soft iron to concentrate magnetic field within a small volume. (b) Use of ferromagnetic polepieces to further concentrate the field.

A cross section through a typical magnetic lens is shown in Fig. 2-10. Here the optic axis is shown vertical, as is nearly always the case in practice. 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, which would lead to lens **misalignment** (departure of the polepieces from a straight-line configuration). The strong magnetic field (up to about 2 Tesla) in each lens gap is generated by a relatively large coil that contains many turns of wire and typically carries a few amps of direct current. To remove heat generated in the coil (due to its resistance), water flows into and out of each lens. Water cooling ensures that the temperature of the lens column reaches a stable value, not far from room temperature, so that thermal expansion (which could lead to column misalignment) is minimized. Temperature changes are also reduced by controlling the temperature of the cooling water within a refrigeration system that circulates water through the lenses in a closed cycle and removes the heat generated.

Rubber **o-rings** (of circular cross section) provide an airtight seal between the interior of the lens column, which is connected to vacuum pumps, 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 air 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.

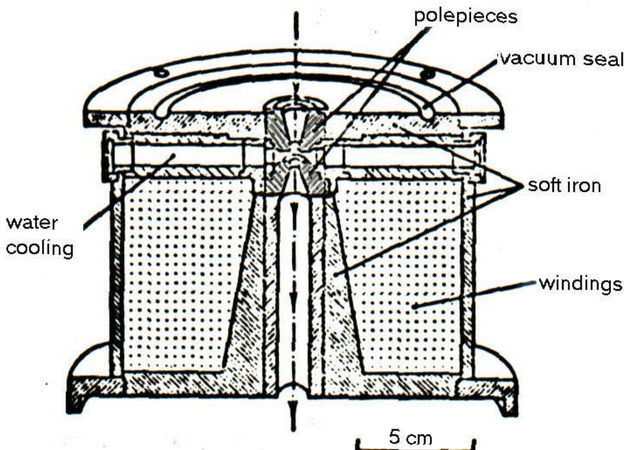

Figure 2-10. Cross section through a magnetic lens whose optic axis (dashed line) is vertical.

2.4 Focusing Properties of a Thin Magnetic Lens

The use of ferromagnetic polepieces results in a focusing field that extends only a few millimeters along the optic axis, so that to a first approximation the lens may be considered thin. Deflection of the electron trajectory can then be considered to take place at a single plane (the principal plane), which allows thin-lens formulas such as Eq. (2.2) and Eq. (2.3) to be used to describe the optical properties of the lens.

The thin-lens approximation also simplifies analysis of the effect of the magnetic forces acting on a charged particle, leading to a simple expression for the **focusing power** (reciprocal of focal length) of a magnetic lens:

$$1/f = [e^2/(8mE_0)] \int B_z^2 \, dz \tag{2.7}$$

As we are considering electrons, the particle charge e is 1.6×10^{-19} C and mass $m = 9.11 \times 10^{-31}$ kg; E_0 represents the kinetic energy of the particles 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 also depends on the lens current and the 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 toward its maximum value B_0 (at the center of the lens) according to a symmetric *bell-shaped* curve described by the **Lorentzian** function:

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

As we can see by substituting $z = a$ in Eq. (2.8), a is the half-width at half maximum (HWHM) of a graph 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. 2-8. 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 valid, the integral in Eq. (2.7) becomes $(\pi/2)aB_0^2$ and the focusing power of the lens is

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

As an example, we can take B_0 = 0.3 Tesla 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. Because 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. In fact, more exact calculations (Reimer, 1997) show 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 must 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 $E_0\,(1 + V_0\,/\,1022\text{ kV})$ 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) enables us to see how the focusing power of a magnetic lens depends on the strength and spatial extent of the magnetic field and on certain properties of particles being imaged (their kinetic energy, charge and mass). Because the kinetic energy E_0 appears in the denominator of Eq. (2.7), focusing power decreases as the accelerating voltage is increased. As might be expected intuitively, faster electrons are deflected less in the magnetic field.

Because 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. However, 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 qualitatively the action of a magnetic field, we saw that the electrons execute a spiral motion, besides being deflected back toward the optic axis. As a result, the plane containing the exit ray is rotated through an angle ϕ relative to the plane containing the incoming electron. Again making a thin-lens approximation ($a \ll f$) and assuming a Lorentzian field distribution, the equations of motion can be solved to give:

$$\phi = [e/(8mE_0)^{1/2}] \int B_z \, dz = [e/(8mE_0)^{1/2}] \pi a B_0 \tag{2.10}$$

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 not unimportant. 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 counterclockwise) if the current through the lens windings were reversed, because B_z (and therefore ϕ) would be reversed in sign. On the other hand, reversing the current does *not* change the focusing power, as 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, for which the focal length f would have to be negative.

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 potential used to accelerate the particles), 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

Some of the differences between electrostatic and magnetic lenses (of axial symmetry) are summarized in Table 2-1 and will now be discussed in turn.

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

The ability of electrostatic lenses to operate with voltages that slowly change (drift) or contain fluctuating components (ripple) is a result of the fact that the voltage supply that is connected to the lens can also be used to accelerate the electrons, as illustrated in Fig. 2-6. If the accelerating voltage V_0 increases, the electrons travel faster; but as it requires a higher lens voltage to focus faster electrons, the two effects cancel (to first order). Electrostatic lenses were therefore preferred in some of the early electron

microscopes, at a time when high-voltage supplies were less stable than is possible today.

On the other hand, 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 are generally preferred for electron microscopy. Magnetic lenses also provide somewhat lower aberrations, for the same focal length, further improving the image resolution. As we will see later in this chapter, lens aberrations are also reduced by making the focal length of the objective lens small, implying a magnetic *immersion* lens with the specimen present *within* the lens field. Such a concept is problematic for an electrostatic objective, where introducing a conducting specimen could greatly modify the electric-field distribution.

Table 2-1. Comparison of electrostatic and electromagnetic lens designs.

Advantages of an electrostatic lens	Advantages of a magnetic lens
No image rotation	Lower lens aberrations
Lightweight, consumes no power	No high-voltage insulation required
Highly stable voltage unnecessary	Can be used as an immersion lens
Easier focusing of ions	

2.6 Defects of Electron Lenses

For a microscope, the most important focusing defects are lens *aberrations*, as they reduce the spatial resolution of the image, even when it has been optimally focused. We will discuss two kinds of **axial aberrations**, those that lead to image blurring even for object points that lie *on the optic axis*. Similar aberrations occur in light optics.

Spherical aberration

The effect of 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. Those that arrive very *close to* the optic axis (*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. When spherical aberration is present, electrons arriving at an appreciable distance x from the axis are focused to a different point F_1 located at 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:

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

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 axially symmetric, the deflection angle α will 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).

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 $x << f$, taking the angle α to be small, and also that $\Delta f << f$, 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 high-voltage electron optics.

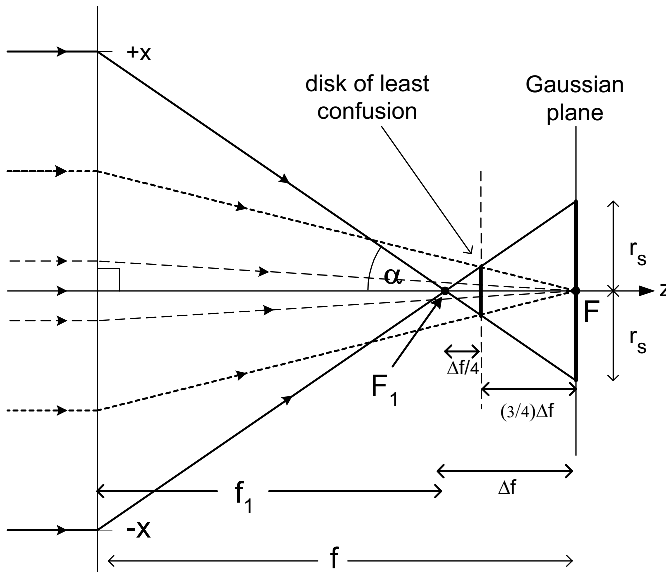

Figure 2-11. Definition of the disk of confusion due to spherical aberration, in terms of the focusing of parallel rays by a thin lens.

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

$$r_s = (\Delta f)\tan\alpha \approx (\Delta f)\,\alpha \quad (2.13)$$

where we once again assume that α is small. As 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:

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

in which we have combined c_2 and f into a single constant C_s, known as 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, with electrons arriving at the lens with all radial displacements (*between* zero and some value x) 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 axial 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 the 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 toward 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.

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 made more applicable 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. Even so, the diagram represents the case of large *de*magnification ($u/f \approx M \ll 1$), not $M \gg 1$ as required for a microscope. To describe the TEM situation, we can 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 (shown 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, the disks overlap substantially, as in Fig. 2-12c. For a further reduction in spacing, the two separate point objects would no longer be distinguishable from the image, and so we take r_s as the spherical-aberration limit to the **point resolution** of a TEM objective lens. This approximates to the Rayleigh criterion (Section 1.1); the current-density distribution in the image consists of two overlapping peaks with about 15% dip between them.

Spherical aberration occurs in TEM lenses *after* the objective but is much less important. 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), while the spherical-aberration blurring depends on the *third power* of this angle, according to Eq. (2.14).

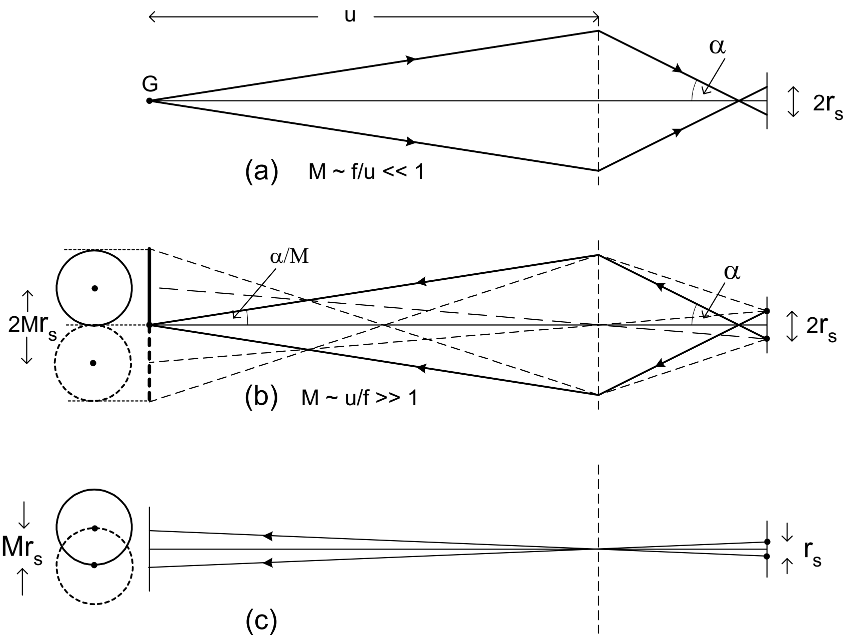

Figure 2-12. (a) Ray diagram similar to Fig. 2-11, with the object distance u large but finite. (b) Equivalent diagram with the rays reversed, showing two image disks of confusion arising from object points whose separation is $2r_s$. (c) Same diagram but with object-point separation reduced to r_s so that the two points are barely resolved in the image (Rayleigh criterion).

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 (Glaeser, 1952). Figure 2.13 shows the calculated C_s and focal length f as a function of the maximum field B_0, for 200kV 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 becomes inaccurate for a strong lens. For the weak lens, $C_s \approx f \approx$ several mm; but for a strong lens ($B_0 = 2$ to 3 T, as used for a TEM objective), C_s falls to about $f/4$. If we take $f = 2$ mm, so that $C_s \approx 0.5$ mm, and require a point resolution $r_s = 1$ nm, the maximum angle of the electrons (relative to the optic axis) must satisfy: $C_s\alpha^3 \approx r_s$, giving $\alpha \approx 10^{-2}$ rad = 10 mrad. This low value justifies our use of small-angle approximations in the preceding analysis.

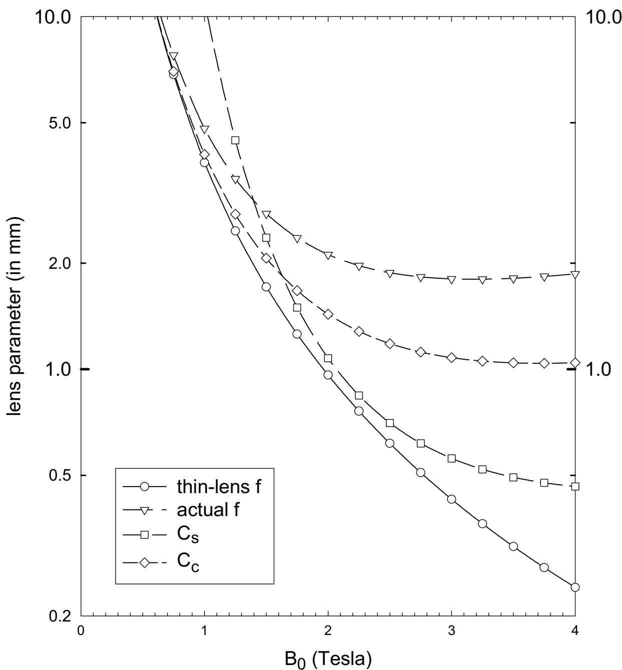

Figure 2-13. Focal length and coefficients of spherical and chromatic aberration for a magnetic lens containing a Lorentzian field with peak field B_0 and half-width $a = 1.8$ mm, focusing 200keV electrons. Values were calculated from Eq. (2.7) and from Glaeser (1952).

The basic physical properties of a magnetic field dictate that the spherical aberration of an axially-symmetric electron lens cannot be eliminated through careful design of the lens polepieces. However, spherical aberration can be *minimized* by using a strong lens (small f). The smallest possible focal length is determined by the maximum field ($B_0 \approx 2.6$ Tesla) obtainable, limited by magnetic saturation of the lens polepieces. The radius r_s of the disk of confusion is also reduced by using an aperture in the lens column to limit the maximum angular deviation α of electrons from the optic axis.

Chromatic aberration

In light optics, chromatic aberration occurs when there is a spread in the wavelength of the light passing through a lens, coupled 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 E_0 , while Eq. (2.7) shows that the focusing power of a magnetic lens depends inversely on the kinetic energy. So if electrons are present with different kinetic energies, they will be focused at a different distances from a lens; for any 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) Different kinetic energies of the electrons emitted from the source. For example, electrons emitted by a heated-filament source have a **thermal spread** ($\approx kT$, 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 still 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 also 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, ignoring 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. Because $1/f$ increases as the electron energy decreases, electrons of energy $E_0 - \Delta E_0$ will have an 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,

$$r_i = \Delta v \tan\beta \approx \beta \, \Delta v \qquad (2.15)$$

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. As previously, we will take the required displacement in the image plane to be *equal* to the disk radius r_i, which will correspond to a displacement *in the object plane* equal to $r_c = r_i/M$, where M is the image magnification given by:

$$M = v/u = \tan\alpha/\tan\beta \approx \alpha / \beta \qquad (2.16)$$

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

$$r_c \approx \beta \, \Delta v/M \approx \alpha \, \Delta v/M^2 \qquad (2.17)$$

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:

$$\Delta v = (v^2/f^2) \, \Delta f \qquad (2.18)$$

For $M >> 1$, 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. (4.7) gives:

$$r_c \approx \alpha \, \Delta f \qquad (2.19)$$

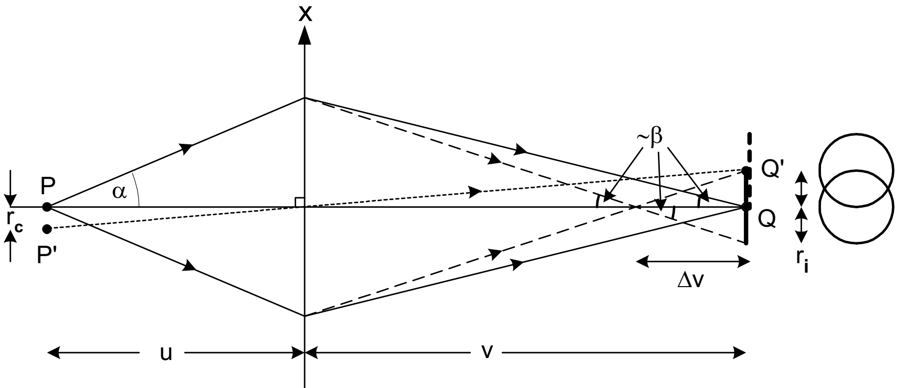

Figure 2-14. Ray diagram illustrating the change in focus and the disk of confusion resulting from chromatic aberration. With two object points, the image disks overlap; the Rayleigh criterion (about 15% reduction in intensity between the current-density maxima) is satisfied when the separation PP' in the object plane is given by Eq. (2.20).

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, and 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:

$$r_c \approx \alpha f\,(\Delta E_0/E_0) \qquad (2.20)$$

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

$$r_c \approx \alpha\ C_c\,(\Delta E_0\,/E_0) \qquad (2.21)$$

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 typically $f/2$ 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) and by using an angle-limiting aperture (restricting α). High electron-accelerating voltage (large E_0) also reduces the chromatic effect.

Axial astigmatism

So far we have assumed complete axial 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 cylindrical 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, and 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 , displaced along the z-axis. In the case of 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.

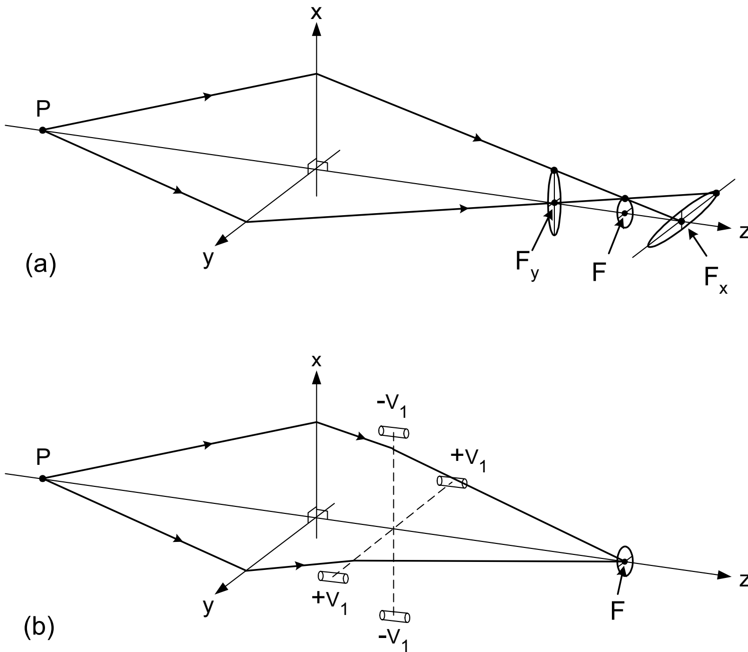

Figure 2-15. (a) Rays leaving an axial image point, focused by a lens (with axial astigmatism) into ellipses centered around F_x and F_y or into a circle of radius R at some intermediate plane. (b) Use of an electrostatic stigmator to correct for the axial astigmatism of an electron lens.

Axial astigmatism also occurs in the human eye, when there is a lack of axial symmetry. It can be *corrected* by wearing lenses whose focal length differs 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 and electrons traveling in that plane are repelled *toward* the axis, resulting in a *positive* focusing power (convex-lens effect). A potential $+V_1$ is applied to the other pair, which therefore *attract* electrons traveling in the *y-z* plane and provide a *negative* focusing power in that plane. By combining the stigmator with an astigmatic lens and choosing V_1 appropriately, the focal length of the 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.

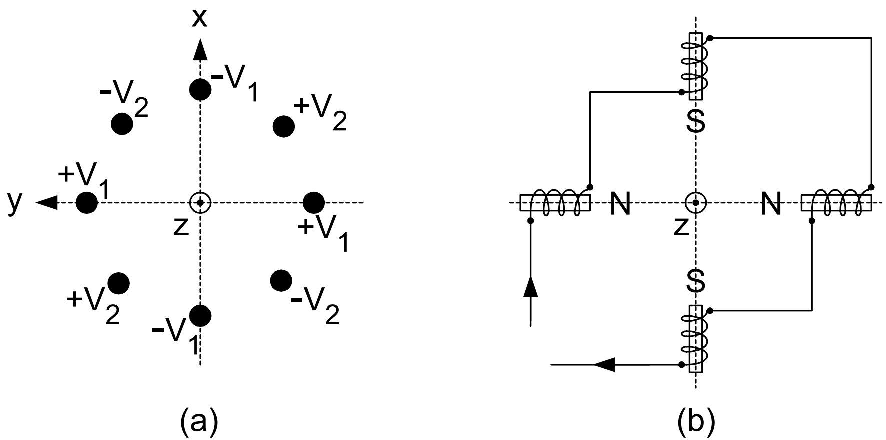

Figure 2-16. (a) Electrostatic stigmator, viewed along the optic axis; (b) magnetic quadrupole, the basis of an electromagnetic stigmator.

In practice, we cannot predict which azimuthal direction corresponds to the smallest or the 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 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 while avoiding the need for a rotating vacuum seal.

A more common form of stigmator consists of a *magnetic* quadrupole: four short solenoid coils with their axes pointing toward 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 north and south magnetic poles face each other, as in Fig. 2-16b. Because the magnetic force on an electron is perpendicular to the magnetic field, the coils that lie along the *x*-axis deflect electrons in the *y*-direction and *vice versa*. In other respects, the magnetic stigmator acts similar to the 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 any mechanical rotation.

The stigmators found in electron-beam columns are *weak* quadrupoles, designed to correct for *small* deviations of 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 is important in the case of particles heavier than the electron.

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 not 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 some importance in a TEM if the instrument is to achieve its highest possible resolution.

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. Distortion changes this ideal relation to:

$$R = M r + C_d\, r^3 \qquad (2.22)$$

where C_d is a constant. If $C_d > 0$, 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 $C_d < 0$, 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, an axially-symmetric electron lens (for which $C_s > 0$) will give $C_d > 0$ 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, because 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 observed when the final (projector) lens is operated at reduced current (therefore large C_s) to achieve a low overall magnification.

Curvature of field is not a serious problem in the TEM or SEM, because 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 plane of viewing is moved along the optic axis) as we will discuss in Chapter 3.