Chapter 2
Energy-Loss Instrumentation

2.1 Energy-Analyzing and Energy-Selecting Systems

Complete characterization of a specimen in terms of its inelastic scattering would involve recording the scattered intensity $J(x, y, z, \theta_x, \theta_y, E)$ as a function of position (coordinates x, y, z) within the specimen and as a function of scattering angle (components θ_x and θ_y) and energy loss E. For an anisotropic crystalline specimen, the procedure would have to be repeated at different specimen orientations. Even if technically feasible, such measurements would involve storing a vast amount of information, so in practice the acquisition of energy-loss data is restricted to the following categories (see Fig. 2.1):

(a) An *energy-loss spectrum* $J(E)$ recorded at a particular point on the specimen or (more precisely) integrated over a circular region defined by an incident electron beam or an area-selecting aperture. Such spectroscopy (also known as energy analysis) can be carried out using a conventional transmission electron microscope (CTEM) or a scanning transmission electron microscope (STEM) producing a stationary probe, either of them fitted with a double-focusing spectrometer such as the magnetic prism (Sections 2.1.1 and 2.2).

(b) A *line spectrum* $J(y, E)$ or $J(\theta_y, E)$, where distance perpendicular to the E-axis represents a single coordinate in the image or diffraction pattern. This mode is obtained by using a spectrometer that focuses only in the direction of dispersion, such as the Wien filter (Section 2.1.3). It can also be implemented by placing a slit close to the entrance of a double-focusing magnetic spectrometer, the slit axis corresponding to the nondispersive direction in the spectrometer image plane; this technique is sometimes called spatially resolved EELS (SREELS).

(c) An *energy-selected image* $J(x, y)$ or *filtered diffraction pattern* $J(\theta_x, \theta_y)$ recorded for a given energy loss E (or small range of energy loss) using CTEM or STEM techniques, as discussed in Section 2.6.

(d) A *spectrum image* $J(x, y, E)$ obtained by acquiring an energy-loss spectrum at each pixel as a STEM probe is rastered over the specimen (Jeanguillaume and Colliex, 1989). Using a conventional TEM fitted with an imaging filter, the same

R.F. Egerton, *Electron Energy-Loss Spectroscopy in the Electron Microscope*,
DOI 10.1007/978-1-4419-9583-4_2, © Springer Science+Business Media, LLC 2011

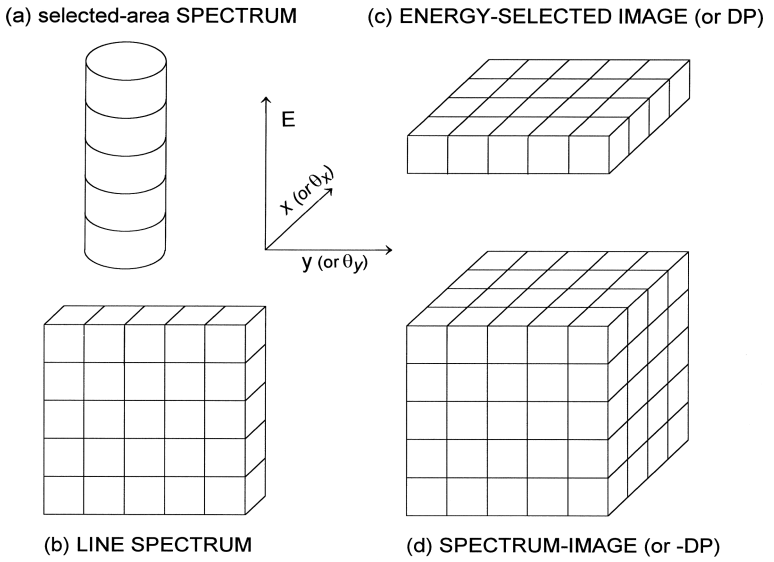

Fig. 2.1 Energy-loss data obtainable from (**a**) a fixed-beam TEM fitted with a double-focusing spectrometer; (**b**) CTEM with line-focus spectrometer or double-focusing spectrometer with an entrance slit; (**c**) CTEM operating with an imaging filter or STEM with a serial-recording spectrometer; (**d**) STEM fitted with a parallel-recording spectrometer or CTEM collecting a series of energy-filtered images

information can be obtained by recording a series of energy-filtered images at successive energy losses, sometimes called an *image spectrum* (Lavergne et al., 1992). This corresponds in Fig. 2.1d to acquiring the information from successive layers, rather than column by column as in the STEM method. By filtering a diffraction pattern or using a rocking beam (in STEM) it is also possible to record the $J(\theta_x, \theta_y, E)$ data cube.

Details of the operation of these energy-loss *systems* are discussed in Sections 2.3, 2.4, 2.5, and 2.6. In the next section, we review the kinds of spectrometer that have been used for TEM-EELS, where an incident energy of the order of 10^5 eV is necessary to avoid excessive scattering in the specimen. Since an energy resolution better than 1 eV is desirable, the choice of spectrometer is limited to those types that offer high resolving power, which rules out techniques such as time-of-flight analysis that are used successfully in other branches of spectroscopy.

2.1.1 The Magnetic Prism Spectrometer

In the magnetic prism spectrometer, electrons traveling at a speed v in the z-direction are directed between the poles of an electromagnet whose magnetic field B is in the

y-direction, perpendicular to the incident beam. Within this field, the electrons travel in a circular orbit whose radius of curvature R is given by

$$R = (\gamma m_0/eB)v \tag{2.1}$$

where $\gamma = 1/(1-v^2/c^2)^{1/2}$ is a relativistic factor and m_0 is the rest mass of an electron. The electron beam emerges from the magnet having been deflected through an angle ϕ; often chosen to be 90° for convenience. As Eq. (2.1) indicates, the precise angular deflection of an electron depends on its velocity v within the magnetic field. Electrons that have lost energy in the specimen have a lower value of v and smaller R, so they leave the magnet with a slightly larger deflection angle (Fig. 2.2a).

Besides introducing bending and dispersion, the magnetic prism also focuses an electron beam. Electrons that originate from a point object O (a distance u from the entrance of the magnet) and deviate from the central trajectory (the optic axis) by some angle γ_x (measured in the radial direction) will be focused into an image I_x a

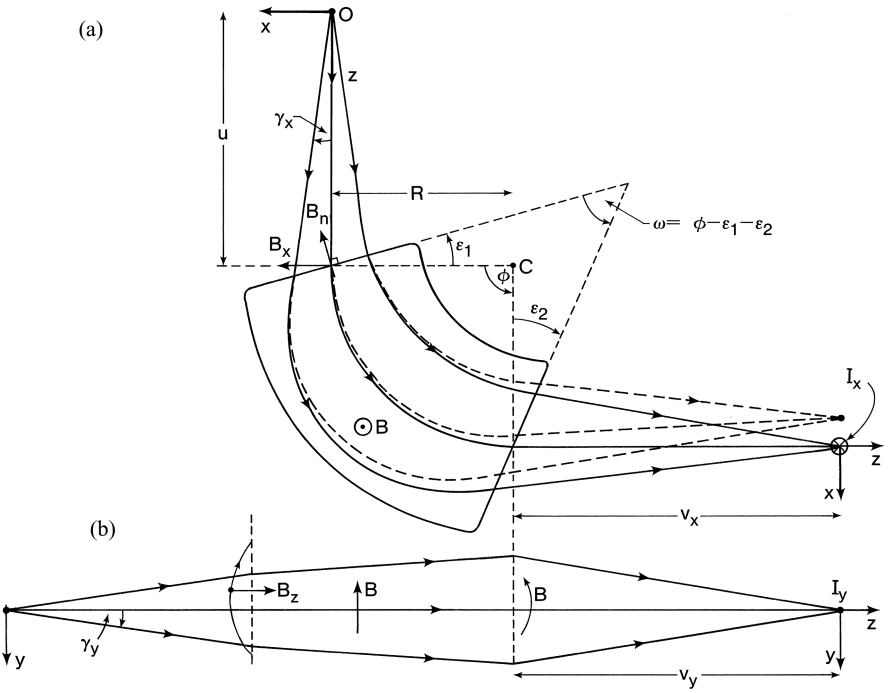

Fig. 2.2 Focusing and dispersive properties of a magnetic prism. The coordinate system rotates with the electron beam, so the x-axis always represents the radial direction and the z-axis is the direction of motion of the central zero-loss trajectory (the optic axis). Radial focusing in the x–z plane (the first principal section) is represented in (**a**); the trajectories of electrons that have lost energy are indicated by *dashed lines* and the normal component B_n of the fringing field is shown for the case $y > 0$. Axial focusing in the y–z plane (a flattened version of the second principal section) is illustrated in (**b**)

distance v_x from the exit of the magnet; see Fig. 2.2a. This focusing action occurs because electrons with positive γ_x travel a longer distance within the magnetic field and therefore undergo a larger angular deflection, so they return toward the optic axis. Conversely, electrons with negative γ_x travel a shorter distance in the field are deflected less and converge toward the same point I_x. To a first approximation, the difference in path length is proportional to γ_x, giving first order focusing in the x–z plane. If the edges of the magnet are perpendicular to the entrance and exit beam ($\varepsilon_1 = \varepsilon_2 = 0$), points O, I_x, and C (the center of curvature) lie in a straight line (Barber's rule); the prism is then properly referred to as a *sector* magnet and focuses *only* in the x-direction. If the entrance and exit faces are tilted through positive angles ε_1 and ε_2 (in the direction shown in Fig. 2.2a), the differences in path length are less and the focusing power in the x–z plane is reduced.

Focusing can also take place in the y–z plane (i.e., in the axial direction, parallel to the magnetic field axis), but this requires a component of magnetic field in the radial (x) direction. Unless a gradient field design is used (Crewe and Scaduto, 1982), such a component is absent within the interior of the magnet, but in the fringing field at the polepiece edges there is a component of field B_n (for $y \neq 0$) that is normal to each polepiece edge (see Fig. 2.2a). Provided the edges are not perpendicular to the optic axis ($\varepsilon_1 \neq 0 \neq \varepsilon_2$), B_n itself has a radial component B_x in the x-direction, in addition to its component B_z along the optic axis. If ε_1 and ε_2 are positive (so that the wedge angle ω is less than the bend angle ϕ), $B_x > 0$ for $y > 0$ and the magnetic forces at both the entrance and exit edges are in the negative y-direction, returning the electron toward a point I_y on the optic axis. Each boundary of the magnet therefore behaves like a convex lens for electrons traveling in the y–z plane (Fig. 2.2b).

In general, the focusing powers in the x- and y-directions are unequal, so that *line foci* I_x and I_y are formed at different distances v_x and v_y from the exit face; in other words, the device exhibits axial astigmatism. For a particular combination of ε_1 and ε_2, however, the focusing powers can be made equal and the spectrometer is said to be *double focusing*. In the absence of aberrations, electrons originating from O would all pass through a single point I, a distance $v_x = v_y = v$ from the exit. A double-focusing spectrometer therefore behaves like a normal lens; if an extended object were placed in the x–y plane at point O, its image would be produced in the x–y plane passing through I. But unlike the case of an axially symmetric lens, this two-dimensional imaging occurs only for a single value of the object distance u. If u is changed, a different combination of ε_1 and ε_2 is required to give double focusing.

Like any optical element, the magnetic prism suffers from aberrations. For example, *aperture* aberrations cause an axial point object to be broadened into an *aberration figure* (Castaing et al., 1967). For the straight-edged prism shown in Fig. 2.2a, these aberrations are predominantly second order; in other words, the dimensions of the aberration figure depend on γ_x^2 and γ_y^2. Fortunately, it is possible to correct these aberrations by means of sextupole elements or by curving the edges of the magnet, as discussed in Section 2.2.

For energy analysis in the electron microscope, a single magnetic prism is the most frequently used type of spectrometer. This popularity arises largely from

the fact that it can be manufactured as a compact, add-on attachment to either a conventional or a scanning transmission microscope without affecting its basic performance and operation. The spectrometer is not connected to the microscope high-voltage supply, so a magnetic prism can be used even for accelerating voltages exceeding 500 keV (Darlington and Sparrow, 1975; Perez et al., 1975). However, good energy resolution demands a magnet current supply of very high stability and requires the high-voltage supply of the microscope to be equally stable. As the dispersive power is rather low, values of around 2 μm/eV being typical for 100-keV electrons, good energy resolution requires finely machined detector slits (for serial acquisition) or post-spectrometer magnifying optics (in the case of a parallel recording detector). On the other hand, the dispersion is fairly linear over a range of 2000 eV, making the magnetic prism well suited to parallel recording of inner-shell losses.

2.1.2 Energy-Filtering Magnetic Prism Systems

As discussed in Section 2.1.1, the edge angles of a magnetic prism can be chosen so that electrons coming from a point object will be imaged to a point on the exit side of the prism, for a given electron energy. In other words, there are points R_1 and R_2 that are fully stigmatic and lie within a real object and image, respectively (see Fig. 2.3a). Because of the dispersive properties of the prism, the plane through R_2 will contain the object intensity *convolved* with the electron energy-loss spectrum of the specimen. Electron optical theory (Castaing et al., 1967; Metherell, 1971) indicates that, for the same prism geometry, there exists a second pair of stigmatic points V_1 and V_2 (Fig. 2.3b) that usually lie within the prism and correspond to *virtual* image points. Electrons that are focused so as to converge on V_1 appear (after deflection by the prism) to emanate from V_2. If an electron lens were used to produce an image of the specimen at the plane passing through V_1, a second lens focused on V_2 could project a real image of the specimen from the electrons that have passed through the prism. An aperture or slit placed at R_2 would transmit only those electrons whose energy loss lies within a certain range, so the final image would be an *energy-filtered* (or energy-selected) image.

Ideally, the image at V_2 should be achromatic (see Fig. 2.3), a condition that can be arranged by suitable choice of the object distance (location of point R_1) and prism geometry. In that case, the prism introduces no additional chromatic aberration, regardless of the width of the energy-selecting slit.

In order to limit the angular divergence of the rays at R_1 (so that spectrometer aberrations do not degrade the energy resolution) while at the same time ensuring a reasonable field of view at the specimen, the prism is ideally located in the *middle* of a CTEM column, between the objective and projector lenses. A single magnetic prism is then at a disadvantage: it bends the electron beam through a large angle, so the mechanical stability of a vertical lens column would be lost. In preference, a multiple deflection system is used, such that the net angle of deflection is zero.

2.1.2.1 Prism–Mirror Filter

A filtering device first developed at the University of Paris (Castaing and Henry, 1962) consists of a uniform field magnetic prism and an electrostatic mirror. Electrons are deflected through 90° by the prism, emerge in a horizontal direction, and are reflected through 180° by the mirror so that they enter the magnetic field a second time. Because their velocity is now reversed, the electrons are deflected downward and emerge from the prism traveling in their original direction, along the vertical axis of the microscope.

In Fig. 2.3, R_2 and V_2 act as real and virtual *objects* for the second magnetic deflection in the lower half of the prism, producing real and virtual images R_3 and V_3, respectively. To achieve this, the mirror must be located such that its apex is at R_2 and its center of curvature is at V_2, electrons being reflected from the mirror back toward V_2. Provided the prism itself is symmetrical ($\varepsilon_1 = \varepsilon_3$ in Fig. 2.3), the virtual image V_3 will be achromatic and at the same distance from the midplane as V_1 (Castaing et al., 1967; Metherell, 1971). In practice, the electrostatic mirror consists of a planar and an annular electrode, both biased some hundreds of volts negative with respect to the gun potential of the microscope. The apex of the mirror depends

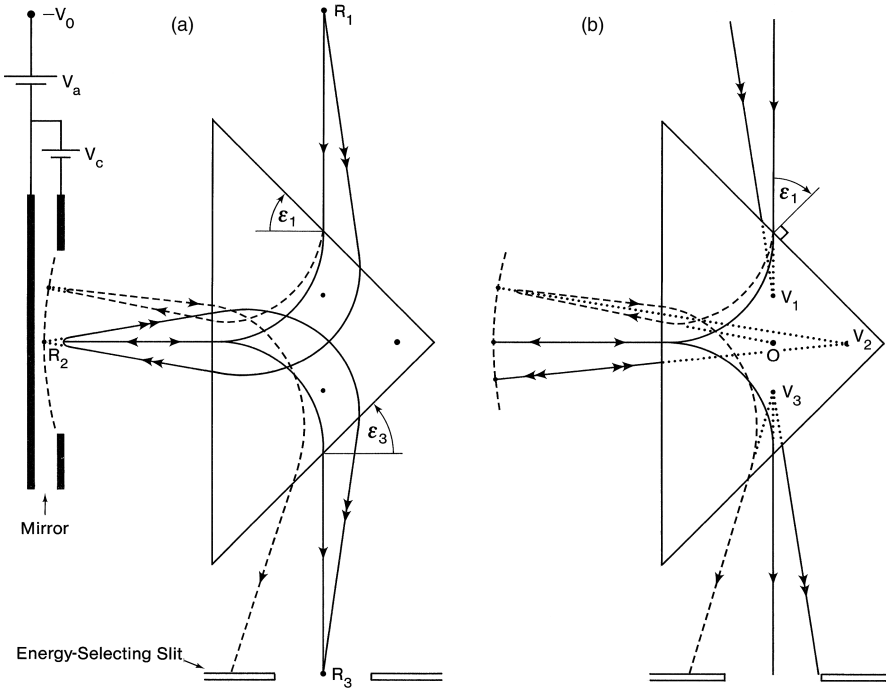

Fig. 2.3 Imaging properties of a magnetic prism, showing (**a**) real image points (R_1, R_2, and R_3) and (**b**) virtual image points (V_1, V_2, and V_3) and an achromatic point O. In this example, electrons are reflected back through the prism by an electrostatic mirror, whose apex and curvature are adjusted by bias voltages V_a and V_c

on the bias applied to both electrodes; the curvature can be adjusted by varying the voltage difference between the two (Henkelman and Ottensmeyer, 1974b).

Although not essential (Castaing, 1975), the position of R_1 can be chosen (for a symmetric prism, $\varepsilon_1 = \varepsilon_3$) such that the point focus at R_3 is located at the same distance as R_1 from the midplane of the system. Since the real image at R_2 is chromatic (as discussed earlier) and since the dispersion is additive during the second passage through the prism, the image at R_3 is also chromatic; if R_1 is a point object, R_3 contains an energy-loss spectrum of the sample and an energy-selecting aperture or slit placed at R_3 will define the range of energy loss contributing to the image V_3. The latter is converted into a real image by the intermediate and projector lenses of the TEM column. Because V_3 is achromatic, the resolution in the final image is (to first order) independent of the width of the energy-selecting slit, which ensures that the range of energy loss can be made sufficiently large to give good image intensity and that (if desired) the energy-selecting aperture can be withdrawn to produce an unfiltered image.

In the usual mode of operation of an energy-selecting TEM, a low-excitation "post-objective" lens forms a magnified image of the specimen at V_1 and a demagnified image of the back-focal plane of the objective lens at R_1. In other words, the object at R_1 is a portion (selected by the objective aperture) of the diffraction pattern of the specimen: the central part, for bright-field imaging. With suitable operation of the lens column, the location of the specimen image and diffraction pattern can be reversed, so that energy-filtered diffraction patterns are obtained (Henry et al., 1969; Egerton et al., 1975; Egle et al., 1984).

In addition, the intermediate lens excitation can be changed so that the intensity distribution at R_3 is projected onto the TEM screen. The energy-loss spectrum can then be recorded in a parallel mode (using a CCD camera) or serially (by scanning the spectrum past an aperture and electron detection system). If the system is slightly misaligned, a line spectrum is produced (Henkelman and Ottensmeyer, 1974a; Egerton et al., 1975) rather than a series of points, more convenient because the lower current density results in less risk of damage in a scintillator or contamination on energy-selecting slits.

The original Castaing–Henry system was improved by curving the prism edges to reduce second-order aberrations (Andrew et al., 1978; Jiang and Ottensmeyer, 1993), allowing a greater angular divergence at R_1 and therefore a larger field of view in the energy-filtered image, for a given energy resolution. Although the mirror potential is tied to the microscope high voltage, the dispersion of the system arises entirely from the magnetic field. Therefore, good energy resolution is dependent upon stability of the high-voltage supply, unlike electrostatic or retarding field analyzers.

2.1.2.2 Omega, Alpha, and MANDOLINE Filters

A more recent approach to energy filtering in a CTEM takes the form of a purely magnetic device known as the omega filter. After passing through the specimen, the objective lens, and a low-excitation post-objective lens, the electrons pass through

a magnetic prism and are deflected through an angle ϕ, typically 90–120°. They then enter a second prism whose magnetic field is in the opposite direction, so the beam is deflected downward. A further two prisms are located symmetrically with respect to the first pair and the complete trajectory takes the form of the Greek letter Ω (Fig. 2.4). The beam emerges from the device along its original axis, allowing vertical alignment of the lens column to be preserved. The dispersion within each magnetic prism is additive and an energy-loss spectrum is formed at a position D_2 that is conjugate with the object point D_1; see Fig. 2.4. For energy-filtered imaging, D_1 and D_2 contain diffraction patterns while planes O_1 and O_3 (located just outside or inside the first and last prisms) contain real or virtual images of the specimen. An energy-selected image of the specimen is produced by using an energy-selecting slit in plane D_2 and a second intermediate lens to image O_3 onto the TEM screen (via the projector lens). Alternatively, the intermediate lens can be focused on D_2 in order

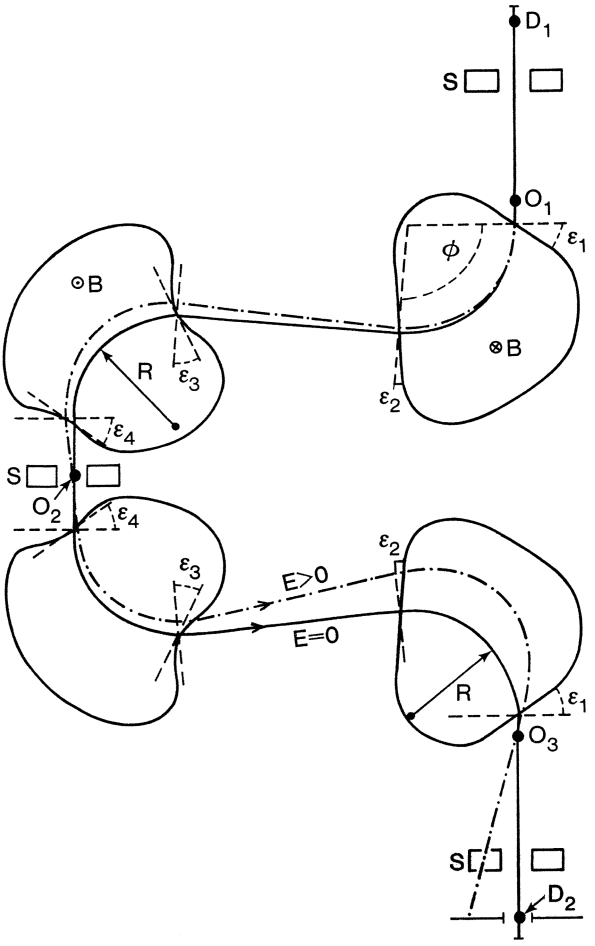

Fig. 2.4 Optics of an aberration-corrected omega filter (Pejas and Rose, 1978). Achromatic images of the specimen are formed at O_1, O_2, and O_3; the plane through D_2 contains an energy-dispersed diffraction pattern. Sextupole lenses (S) placed close to D_1, O_2 and D_2 correct for image-plane tilt

to record the energy-loss spectrum. If the post-objective lens creates an image and diffraction pattern at D_1 and O_1, respectively, an energy-filtered diffraction pattern occurs at O_3.

As a result of the symmetry of the omega filter about its midplane, second-order aperture aberration and second-order distortion vanish if the system is properly aligned (Rose and Plies, 1974; Krahl et al., 1978; Zanchi et al., 1977b). The remaining second-order aberrations can be compensated by curving the polefaces of the second and third prisms, using sextupole coils symmetrically placed about the midplane (Fig. 2.4) and operating the system with line (instead of point) foci between the prisms (Pejas and Rose, 1978; Krahl et al., 1978; Lanio, 1986).

Unlike the prism–mirror system, the omega filter does not require connection to the microscope accelerating voltage. As a result, it has become a preferred choice for an energy-filtering TEM that employs an accelerating voltage above 100 keV (Zanchi et al., 1977a, 1982; Tsuno et al., 1997). Since a magnetic field of the same strength and polarity is used in the second and third prisms, these two can be combined into one (Zanchi et al., 1975), although this design does not allow a sextupole at the midplane.

Another kind of all-magnetic energy filter consists of two magnets whose field is in the same direction but of different strength; the electrons execute a trajectory in the form of the Greek letter α (see later, Fig. 2.6b). An analysis of the first-order imaging properties of the alpha filter is given by Perez et al. (1984).

In the MANDOLINE filter (Uhlemann and Rose, 1994) the first and last prisms of the omega design are combined into one and multipole correction elements are incorporated between all three prisms. This filter provides relatively high dispersion (6 μm/eV at 200 kV) and has been used as a high-transmissivity imaging filter in the Zeiss SESAM monochromated TEM.

2.1.3 The Wien Filter

A dispersive device employing both magnetic and electrostatic fields was reported in 1897 by W. Wien and first used with high-energy electrons by Boersch et al. (1962). The magnetic field (induction B in the y-direction) and electric field (strength E, parallel to the x-axis) are both perpendicular to the entrance beam (the z-axis). The polarities of these fields are such that the magnetic and electrostatic forces on an electron are in opposite directions; their relative strengths obey the relationship $E = v_1 B$ such that an electron moving parallel to the optic axis with speed v_1 and kinetic energy E_1 continues in a straight line, the net force on it being zero. Electrons traveling at some angle to the optic axis or with velocities other than v_1 execute a cycloidal motion (Fig. 2.5) whose rotational component is at the cyclotron frequency: $\omega = eB/\gamma m_0$, where e and m_0 are the electron charge and rest mass and γm_0 is the relativistic mass. Electrons starting from a point ($z = 0$) on the optic axis and initially traveling in the x–z plane return to the z-axis after one complete revolution; in other words, an achromatic focus occurs at $z = 2L$, where $L = (\pi v_1/\omega) = \pi\gamma m_0 E/eB^2$. In addition, an inverted chromatic image of unit

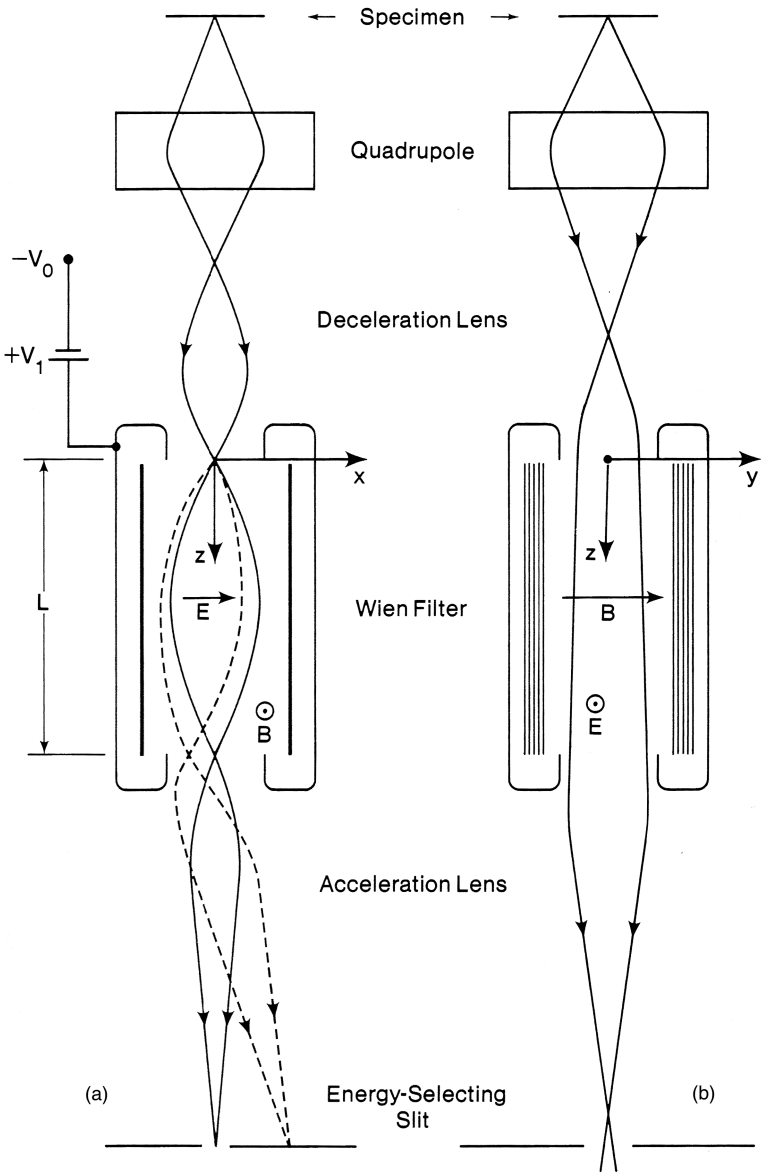

Fig. 2.5 Wien filter spectrometer for a scanning-transmission microscope (Batson, 1985). Electron trajectories are shown (**a**) in the dispersive (x–z) plane and (**b**) in the nondispersive (y–z) plane. *Dashed lines* represent electrons that have lost energy in the specimen. A quadrupole lens has been added to make the system approximately double focusing

magnification ($M_x = -1$) is formed at $z = L$ (i.e., after half a revolution), its energy dispersion being $L/(\pi E_1)$ (Curtis and Silcox, 1971). Velocity components along the y-axis (magnetic field direction) are unaffected by the magnetic and electrostatic fields, so both the chromatic and achromatic images are actually line foci.

The Wien filter is generally used with decelerated electrons. In other words, the filter is operated at a potential $-V_0 + V_1$ which is close to the negative potential $-V_0$ of the electron source. The positive bias V_1 is obtained from a power supply connected to the high-voltage line; its value, typically in the range 100–1000 eV, determines the energy (eV_1) of the electrons that can move in a straight line through the filter. The retarding and accelerating fields at the entrance and exit of the filter act as electrostatic lenses (Fig. 2.5), whose effect must be taken into account in the design of the system.

Although retardation involves the inconvenience of handling high voltages, it provides several advantages. First of all, the dispersion at the chromatic focus is increased by a factor V_0/V_1 for a given length L of the filter; values of 100 μm/eV or more are typical. The electrostatic lens at the exit of the filter can be used to project the spectrum onto the detection plane, with either a decrease or a further increase in the dispersion, depending on the distance of the final image. Second, the required magnitudes and stabilities of B and E are reduced and the mechanical tolerances of the polepieces and electrodes are relaxed. Third, because the electron velocity for straight-line transmission depends on V_1 rather than V_0, fluctuations and drift in V_0 do not affect the energy resolution. This factor is particularly important where high resolution must be combined with long recording times, for example, when recording inner-shell losses using a field-emission STEM (Batson, 1985). A Wien filter used in conjunction with a monochromator (Section 2.1.4) achieved an energy resolution of 5 meV for 30-keV electrons (Geiger et al., 1970), yielding spectra showing vibrational and phonon modes of energy loss (Fig. 1.9). These vibrational modes can be studied by infrared absorption spectroscopy but EELS offers the potential of much better spatial resolution.

Because the system just discussed does not focus in the y-direction, the energy-loss spectrum is produced as a function of distance along a straight line in the entrance plane, this line being defined by an entrance slit. If a diffraction pattern (or a magnified image) of the specimen is projected onto the slit plane, using the lenses of a CTEM, the final image will contain a map of electron intensity as a function of both energy loss and scattering angle (or specimen coordinate). A two-dimensional sensor placed at the final-image plane can therefore record a large amount of information about the specimen (Batson and Silcox, 1983).

The Wien filter can become double focusing if either E or B is made nonuniform, for example, by curving the electric field electrodes, by tilting the magnetic polepieces to create a magnetic field gradient, or by shaping both the electric and magnetic fields to provide a quadrupole action (Andersen, 1967). The device is then suitable for use as an imaging filter in a fixed-beam TEM (Andersen and Kramer, 1972). Aberrations of the filter can be corrected by introducing multipole elements (Andersen, 1967; Martinez and Tsuno, 2008).

2.1.4 Electron Monochromators

Besides being dependent on the spectrometer, the energy resolution of an energy analysis system is limited by energy spread in the electron beam incident on

the specimen. If the electrons are produced by a thermionic source operated at a temperature T_s, the energies of the electrons leaving the cathode follow a Maxwellian distribution, whose full width at half maximum (FWHM) is $\Delta E_s = 2.45(kT_s)$ (Reimer and Kohl, 2008). For a tungsten filament whose emission surface is at a temperature of 2800 K, $\Delta E_s = 0.6$ eV; for a lanthanum hexaboride source at 1700 K, $\Delta E_s = 0.3$ eV. Values are lower for Schottky and field-emission sources.

2.1.4.1 The Boersch Effect

The energy spread ΔE_0 measured in an electron microscope is always larger than ΔE_s, the discrepancy being referred to as the Boersch effect, since Boersch (1954) first investigated the dependence of the measured spread on physical parameters of the electron microscope: cathode temperature, Wehnelt electrode geometry, Wehnelt bias, accelerating voltage, vacuum conditions, and the deployment of magnetic and electrostatic lenses. He found that ΔE_0 increases with the emission current and is further increased when the beam is focused into a crossover. Subsequent experimental work (Martin and Geissler, 1972; Ditchfield and Whelan, 1977; Bell and Swanson, 1979) confirmed these findings.

When electrons are rapidly accelerated to an energy E_0, their energy spread δE remains unaltered, in accordance with the conservation of energy. However, the *axial* velocity spread δv_z is reduced as the axial velocity v_z increases, since (nonrelativistically) $\delta E = \delta(m_0 v_z^2/2)$, *giving* $\delta v_z = \Delta E_s/(m_0 v_z)$. The equivalent axial beam temperature attained is $T_z = (kT_s/E_0)T_s$ (Knauer, 1979) and is very low (<0.1 K) for $E_0 > 10$ keV. If the electrons spend enough time in sufficiently close proximity to one another, so that they interact via Coulomb forces, the difference between the axial and transverse temperatures is reduced, raising δv_z and increasing the measured energy spread. This is known as the "thermal" Boersch effect; the resulting value of ΔE_0 depends on the path length of the electrons and on the current density (Knauer, 1979).

In addition, electrons that are focused into a crossover can suffer "collision broadening" through interaction between their transverse velocity components. The energy broadening depends on the current density at the crossover and on the divergence angle (Crewe, 1978; Knauer, 1979; Rose and Spehr, 1980). The beam current is highest within the electron gun, so appreciable broadening can occur at a gun crossover. A cold field-emission (CFEG) source provides the lowest energy spread (0.3 eV) at low emission currents (<10 nA) but ΔE_0 increases to as much as 1 eV at 100 nA emission, due to Coulomb interaction of electrons just outside the tip (Bell and Swanson, 1979).

2.1.4.2 Types of Monochromator

The Wien filter offers a high dispersion and good energy resolution (a few millielectron volts) when operated with low-velocity electrons (Section 2.1.3). It can therefore be used to produce an incident beam of small energy width if an energy-selecting aperture is placed in an image of its chromatic focus (Boersch et al., 1962,

1964). A second Wien filter (after the specimen) can act as an energy analyzer, making possible energy-loss spectroscopy of vibrational modes (Boersch et al., 1962; Katterwe, 1972; Geiger, 1981). An energy resolution below 6 meV was eventually achieved at 30-keV incident energy.

For analysis of small areas of a TEM specimen, a higher accelerating voltage is useful and a field-emission source maximizes the fraction of electrons allowed through the monochromator (this fraction is approximately the required resolution divided by the energy width of the source). A JEOL-1200EX instrument fitted with two retarding Wien filters and operated at 80 kV achieved an energy resolution of typically 80 meV (Terauchi et al., 1994) or down to 30 meV for energy losses below 5 eV. The microscope accelerating voltage was applied to the analyzing filter to decelerate the electrons, so high-voltage fluctuations were compensated in this design.

An alternative strategy is to place the monochromator within the electron gun, where high dispersion is possible because the electrons have undergone only a limited amount of acceleration. Mook and Kruit (2000) designed a short Wien filter for a high-resolution field-emission STEM (Batson et al., 2000, 2001). Its small length (4 mm) reduced the collision-broadening effect but resulted in a low energy dispersion ($D \approx 4$ μm/eV) for the 800-eV electrons passing through the filter, requiring a very narrow (200 nm) energy selecting to achieve an energy spread in the 50–100 meV range.

For FEI microscopes, Tiemeijer et al. (2001) developed a longer (50 mm) gun monochromator. Electrons are dispersed and then accelerated before reaching the energy-selecting slit, making the latter less sensitive to electrostatic charging. The system has demonstrated a resolution down to 100 meV at low beam current.

The design of Tsuno (2000) is a double Wien filter in which the second half of the filter compensates for energy broadening within the energy-selecting slit. This procedure increases the electron-optical brightness by typically a factor of 3 compared to the single Wien filter design. Since electrons are removed by the energy-selecting slit, the beam current is still reduced, by a factor about equal to the improvement in energy resolution. Tsuno et al. (2005) have reported that a Wien filter is capable of acting as both a monochromator and a lens aberration corrector.

The monochromator design of Rose (1990) is an electrostatic version of the omega filter. An energy-selecting slit is placed at its mid-plane (equivalent to O_2 in Fig. 2.4) and the second half of the filter compensates for energy dispersion within the slit, optimizing the source brightness. This design was commercialized by CEOS GmbH and used in the Zeiss SESAM microscope, where it has demonstrated an energy resolution below 0.1 eV in conjunction with the MANDOLINE energy filter. Advantages of the electrostatic design include low drift (absence of magnetic hysteresis effects) and avoidance of a high-stability current supply running at high potential. Disadvantages are its more complex electron optics and its fixed dispersion: the width of the energy-selecting slit must be changed in order to vary the energy resolution.

A gun monochromator does not compensate for instabilities of the high-voltage supply and the same is true for designs that use the microscope's high voltage to

decelerate the electrons inside the monochromator. Unless this same voltage supply is used to decelerate electrons within the spectrometer, superior design of the high-voltage supply is necessary to achieve 0.1-eV energy resolution (Tiemeijer et al., 2001). Such resolution under practical conditions has also required improved spectrometer design (Brink et al., 2003) and careful attention to minimizing magnetic fields in the microscope environment (Muller and Grazul, 2001). Slow drift can be compensated by storing multiple readouts of the spectrum and shifting them under computer control; even with no monochromator, this technique has yielded a spectral resolution of 0.3 eV (Kimoto and Matsui, 2002) and a precision of 0.1 eV for recording ionization edges (Potapov and Schryvers, 2004).

Krivanek et al. (2009) designed a system employing two purely magnetic non-decelerating filters for the monochromator and analyzer, as shown in Fig. 2.6a. The monochromator is an alpha filter with an energy-selecting slit at its mid-point, the second half of the filter canceling the dispersion of the first half. It uses the same current as the analyzer, so that the latter tracks the energy selected by the monochromator, current drift having no effect on the selected energy. Current absorbed by the two halves of the energy-selecting slit is used as feedback to the high-voltage generator so that the beam remains centered on the slit (Kruit and Shuman, 1985a). Quadrupoles placed before and after the monochromator slit magnify the dispersion from 0.3 μm/eV to about 200 μm/eV, so a 2-μm slit would select 10 meV energy width; see Fig. 2.6b. First- and second-order prism focusing is provided by quadrupole and sextupole lenses, rather than by tilt and curvature of the pole-piece edges. The sextupoles are also designed to correct for chromatic aberration of the probe-forming condenser lenses. For core-loss analysis, monochromation is less essential, so the alpha-filter field could be canceled by diverting the current into separate windings, allowing a straight-line path through the system but preserving the same heat dissipation (to avoid temperature change and mechanical drift). The energy resolution would then be ≈0.3 eV, assuming a cold field-emission source running at 200 kV.

2.1.4.3 Dispersion Compensation

One disadvantage of a conventional monochromator system is that the monochromator reduces the beam current by a large factor if the energy spread of the electron source (including the Boersch effect) greatly exceeds the required energy resolution. Low beam implies longer recording times. A remedy is to eliminate the energy-selecting slit of the monochromator and use a completely symmetrical system of monochromator and analyzer.

For example, if the length of a Wien filter is extended to a value $2L = 2\pi\gamma m_0 E/eB^2$, an *achromatic* focus is formed at the exit plane, as discussed in Section 2.1.3. Because the final image is achromatic, its width is independent of the incident energy spread ΔE_0, a result of the fact that the chromatic aberration of the second half of the system exactly compensates that of the first half. If a specimen is now introduced at the *chromatic* focus ($z = L$), the resulting energy losses provide an energy-loss spectrum at $z = 2L$ but the width of each spectral line remains

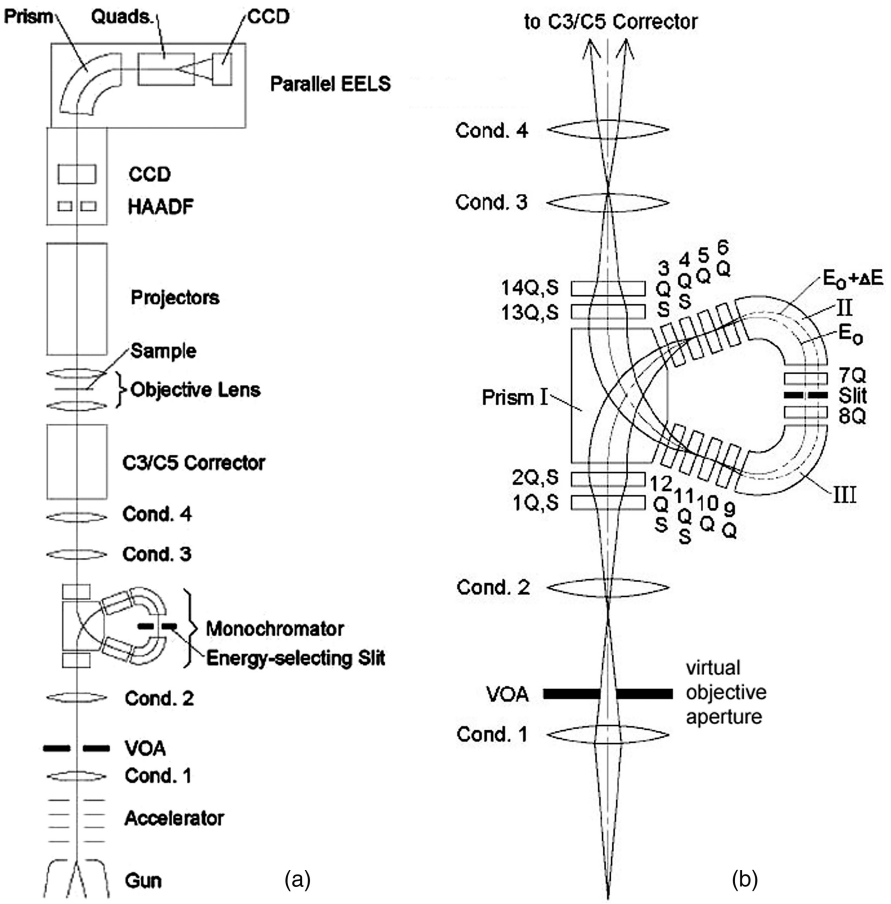

Fig. 2.6 (**a**) Design of a monochromated TEM in which energy drift is minimized by using the same excitation current for both the monochromator and spectrometer and where intensity drift is minimized by feedback from the monochromator slit, so that the beam remains centered at the slit. (**b**) Design of the α-filter monochromator, incorporating three magnetic prisms, quadrupoles (*Q*), and sextupoles (*S*); electrons travel upward from a crossover provided by the condenser 1 lens. From Krivanek et al. (2009), courtesy of the Royal Society, London

independent of the value of ΔE_0. In addition, second-order aperture aberrations of the second half cancel those of the first, so the energy resolution of the system (if perfectly symmetrical) depends only on higher order aberrations and the object size of the electron source. In practice, the two halves of the double Wien filter can be separated by a short distance (to allow room for inserting the specimen) but great care has to be taken to keep the system symmetrical. An experimental prototype based on this principle (Andersen and Le Poole, 1970) achieved an energy resolution of 50 meV (measured without a specimen) using 10-keV transmitted electrons. Scattering in the sample degraded this resolution to about 100 meV but a transmitted

current of up to 0.1 μA was available. The system can be made double focusing at the chromatic image by using two independently excited field coils in each Wien filter (Andersen, 1967), thereby reducing the *y*-spreading of the beam at the specimen. Some spreading in the *x*-direction is unavoidable, because of the energy spread of the electron source and the dispersion produced by the first Wien filter, and would limit the spatial resolution of analysis.

The same principle (known as dispersion compensation or dispersion matching) has been employed in nuclear physics, the target being placed between a pair of magnetic sector spectrometers which bend a high-energy beam of electrons in the same direction (Schaerf and Scrimaglio, 1964). It has been applied to reflection spectroscopy of low-energy (e.g., 5 eV) electrons, using two identical spherical electrostatic sectors (Kevan and Dubois, 1984).

Instead of a monochromator, it would be attractive to have a high-brightness electron source with low energy width. Fransen et al. (1999) examined the field emission properties of individual multiwall carbon nanotubes, mounted on the end of a tungsten wire. In ultrahigh vacuum, the emission was highly stable (less than 10% variation in 50 days) even without "flashing" the tip. Energy widths in the range 0.11–0.2 eV were measured, the source brightness being roughly equivalent to that of a cold tungsten field emitter. Subsequent work has confirmed that high electron-optical brightness may be achievable from a carbon nanotube (De Jonge et al., 2002); a value of $6 \times 10^9 \mathrm{A/cm^2/sr^1}$ at 200 kV, together with good stability, has been reported (Houdellier et al., 2010).

2.2 Optics of a Magnetic Prism Spectrometer

As discussed in Section 2.1.1, a magnetic prism spectrometer produces three effects on a beam of electrons: bending, dispersion, and focusing. Focusing warrants the most attention, since the attainable energy resolution depends on the width of the exit beam at the dispersion plane. Provided the spatial distribution of the magnetic field is known, the behavior of an electron within the spectrometer can be predicted by applying equations of motion, based on Eq. (2.1), to each region of the trajectory. Details are given in Penner (1961), Brown et al. (1964), Brown (1967), and Enge (1967). The aim of this section is to summarize the results of such analysis and to provide an example of the use of a matrix computer program for spectrometer design.

We will adopt the coordinate system and notation of Brown et al. (1964), widely used in nuclear physics. For a negative particle such as the electron, the *y*-axis is antiparallel to the direction of magnetic field. The *z*-axis always represents the direction of motion of an electron traveling along the central trajectory (the optic axis); in other words, the coordinate system rotates about the *y*-axis as the electron proceeds through the magnetic field. The *x*-axis is perpendicular to the *y*- and *z*-axes and points radially outward, away from the center of curvature of the electron trajectories. Using this curvilinear coordinate system, the *y*-axis focusing can be represented on a flat *y*–*z* plane (Fig. 2.2b).

The behavior of electrons at the entrance and exit edges of the magnet is simpler to calculate if the magnetic field is assumed to remain constant up to the polepiece edges, dropping abruptly to zero outside the magnet. This assumption is known as the SCOFF (sharp cutoff fringing field) approximation and is more likely to be realistic if the gap between the polepieces (measured in the y-direction) is very small. In practice, the field strength just inside the magnet is less than in the interior, and a fringing field extends some distance (of the order of the gap length) outside the geometrical boundaries. In the EFF (effective fringing field) model, the z-dependence of field strength adjacent to the magnet boundaries is specified by one or more coefficients, leading to a more accurate prediction of the spectrometer focusing.

An important concept is the *order* of the focusing. Formation of an image is a first-order effect, so first-order theory is used to predict object and image distances, image magnifications, and dispersive power, the latter being first order in energy loss. Second- or higher order analysis is needed to describe image aberration and distortion, together with other properties such as the dispersion-plane tilt.

2.2.1 First-Order Properties

We first consider the "radial" focusing of electrons that originate from a point object O located a distance u from the entrance face of a magnetic prism (Fig. 2.2). For a particular value of u, all electrons that arrive at the *center* of the magnet (after deflection through an angle $\phi/2$) are traveling parallel to the optic axis before being focused by the second half of the prism into a crossover (or image) I_x located a distance v_x from the exit edge. We can regard these particular values of u and v_x as being the focal lengths f_x of the first and second halves of the prism, and their reciprocals as the corresponding focusing powers. In the SCOFF approximation, these focusing powers are given by (Wittry, 1969)

$$1/f_x = [\tan(\phi/2) - \tan\varepsilon]/R \tag{2.2}$$

where $\varepsilon = \varepsilon_1$ for the first half of the prism and $\varepsilon = \varepsilon_2$ for the second half, ε_1 and ε_2 being the tilt angles of the prism edges. Note that a positive value of ε reduces the radial focusing power, leading to longer object and image distances. In this case, the boundaries of the magnet have a divergent focusing action, whereas the effect of the uniform field in the center of the magnet is to gradually return the electrons toward the optic axis, as illustrated in Fig. 2.7.

In contrast, focusing in the axial (y-) direction takes place *only* at the entrance and exit of the magnet. Each boundary can be characterized by a focusing power that is given, in the SCOFF approximation, by

$$1/f_y = \tan(\varepsilon)/R \tag{2.3}$$

Because $1/f_x$ and $1/f_y$ change in opposite directions as ε is varied, the entrance- and exit-face tilts can be chosen so that the net focusing powers in the radial and

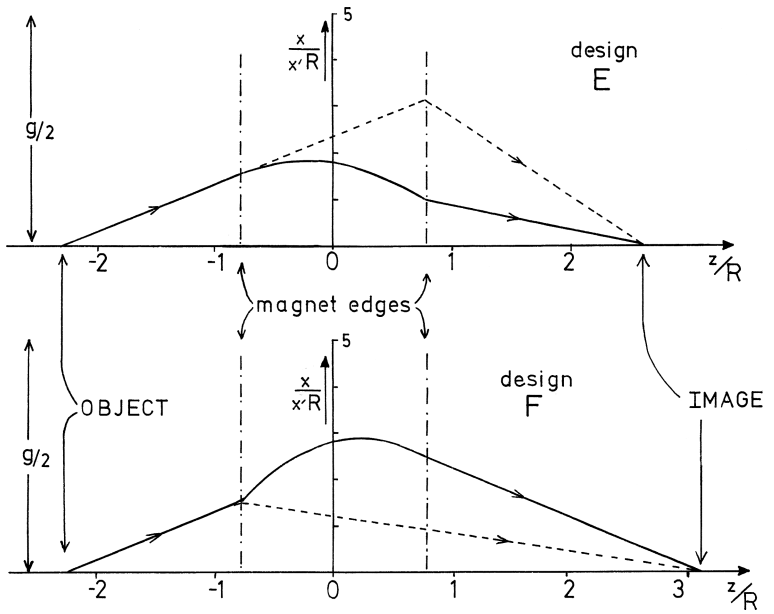

Fig. 2.7 Trajectories of electrons through a magnetic prism spectrometer. *Solid lines* represent the component of motion in the x–z plane (first principal section); *dashed lines* represent motion in the y–z plane (second principal section). The *horizontal axis* indicates distance along the optic axis, relative to the center of the prism. For design E, $\varepsilon_1 = 0$ and $\varepsilon_2 = 45°$; for design F, $\varepsilon_1 = 45°$ and $\varepsilon_1 = 10°$. From Egerton (1980b), copyright Elsevier

axial directions are equal; the prism is then *double focusing*. Although not essential, an approximation to double focusing is generally desirable because it minimizes the width (in the y-direction) of the image at the energy-selecting plane, making the energy resolution of the system less dependent on the precise orientation of the detector about the z-axis.

For a bend angle of 90°, the most symmetrical solution of Eqs. (2.2) and (2.3) corresponds to the double-focusing condition: $u = v_x = v_y = 2R$ and $\tan\varepsilon_1 = \tan\varepsilon_2 = 0.5$ (i.e., $\varepsilon = 26.6°$). In practice, the object distance u may be dictated by external constraints, such as the location of the projector lens crossover in an electron microscope column. The spectrometer will still be double focusing provided the prism angles ε_1 and ε_2 satisfy the relation (valid in the SCOFF approximation)

$$\tan\varepsilon_2 = \frac{1}{2}\left[\frac{1-(\tan\varepsilon_1 + R/u)\tan\phi}{\tan\varepsilon_1 + R/u + \cot\phi} - \frac{\tan\varepsilon_1 - R/u}{1-\phi(\tan\varepsilon_1 - R/u)}\right] \tag{2.4}$$

As ε_1 increases, the required value of ε_2 decreases, as illustrated in Fig. 2.8. The image distance $v_x = v_y = v$ is given by

$$\frac{v}{R} = \left[\frac{\tan\varepsilon_1 - R/u}{1-\phi(\tan\varepsilon_1 - R/u)} + \tan\varepsilon_2\right]^{-1} \tag{2.5}$$

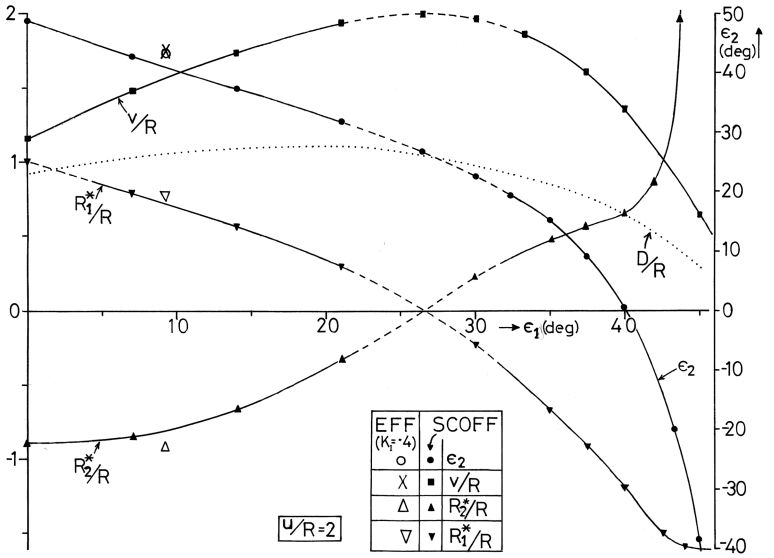

Fig. 2.8 Double-focusing parameters of a magnetic prism, for a fixed object distance ($u = 2R$) and bend angle $\phi = 90°$. The curves were calculated using the SCOFF approximation; *dashed lines* indicate the region in which correction of second-order aberrations requires excessive edge curvatures, as determined by Eq. (2.18). One set of points is given for an extended fringing field (EFF) with $K_1 = 0.4$. From Egerton (1980b), copyright Elsevier

A large difference between ε_1 and ε_2 leads to stronger focusing, reflected in a shorter image distance (Fig. 2.8). The dispersive power $D = dx/dE_0$ at the image plane is (Livingood, 1969)

$$D = \left(\frac{R}{2\gamma T}\right) \frac{\sin\phi + (1-\cos\phi)(\tan\varepsilon_1 + R/u)}{\sin\phi[1-\tan\varepsilon_2(\tan\varepsilon_1 + R/u)] - \cos\phi(\tan\varepsilon_1 + \tan\varepsilon_2 + R/u)} \tag{2.6}$$

where $2\gamma T = E_0(2m_0c^2 + E_0)(m_0c^2 + E_0)$, E_0 represents the kinetic energy of electrons entering the spectrometer, and $m_0c^2 = 511$ keV is the electron rest energy. If the spectrometer is to be reasonably compact, the value of R cannot exceed 10–20 cm and D is limited to a few micrometers per electron volt for $E_0 = 100$ keV.

2.2.1.1 The Effect of Fringing Fields

The SCOFF approximation is convenient for discussing the general properties of a magnetic prism and is useful in the initial stages of spectrometer design, but does not provide accurate predictions of the focusing. The effects of a spatially extended fringing field have been described by Enge (1964) as follows.

First of all, the exit beam is displaced in the radial direction compared to the SCOFF trajectory. This effect can be taken into account by shifting the magnet slightly in the $+x$-direction or by increasing the magnetic field by a small amount.

Second, the focusing power in the axial (y-) direction is decreased, whereas the radial (x-) focusing remains practically unaltered. As a result, either ε_1 or ε_2 must be increased (compared to the SCOFF prediction) in order to maintain double focusing. The net result is a slight increase in image distance; see Fig. 2.8.

A third effect of the extended fringing fields is to add a convex component of curvature to the entrance and exit edges of the magnet, the magnitude of this component varying inversely with the polepiece width w. Such curvature affects the spectrometer aberrations, as discussed in Section 2.2.2. Finally, the extended fringing field introduces a discrepancy between the "effective" edge of the magnet (which serves as a reference point for measuring object and image distances) and the actual "mechanical" edge, the former generally lying outside the latter.

To define the spatial extent of the fringing field, so that it can be properly taken into account in EFF calculations and is less affected by the surroundings of the spectrometer, plates made of a soft magnetic material ("mirror planes") are sometimes placed parallel to the entrance and exit edges, to "clamp" the field to a low value at the required distance from the edge. If the plate–polepiece separation is chosen as $g/2$, where g is the length of the polepiece gap in the y-direction, and the polepiece edges are beveled at 45° to a depth $g/2$ (see Fig. 2.9), the magnetic field decays almost linearly over a distance g along the optic axis. More importantly, the position, angle, and curvature of the magnetic field boundary more nearly coincide with those of the polepiece edge. However, the correspondence is not likely to be exact, partly because the fringing field penetrates to some extent into the holes

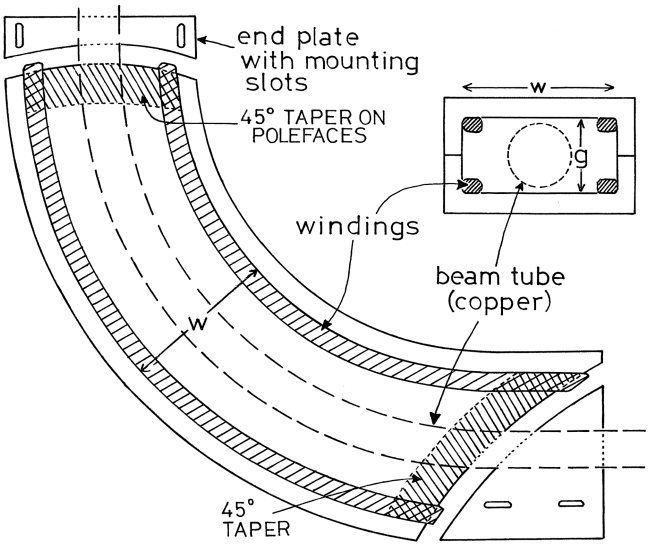

Fig. 2.9 Cross sections (in the x–z and x–y planes) through an aberration-corrected spectrometer with curved and tapered polepiece edges, soft-magnetic mirror plates, and window-frame excitation coils. The magnet operates in air, the vacuum being confined within a nonmagnetic "drift" tube. From Egerton (1980b), copyright Elsevier

that must be provided in the mirror plates to allow the electron beam to enter and leave the spectrometer (Fig. 2.9). The remaining discrepancy between the effective and mechanical edge depends on the polepiece gap, on the separation of the field clamps from the magnet, and on the radius of curvature of the edges (Heighway, 1975).

The effect of the fringing field on spectrometer focusing can be specified in terms of the gap length g and a shape parameter K_1 defined by

$$K_1 = \int_{-\infty}^{\infty} \frac{B_y(z')[B - B_y(z')]}{gB^2} dz' \tag{2.7}$$

where $b_y(z')$ is the y-component of induction at $y = 0$ and at a perpendicular distance z' from the polepiece edge; B is the induction between the polepieces within the interior of the spectrometer. The SCOFF approximation corresponds to $K_1 = 0$; the use of tapered polepiece edges and mirror plates, as specified above, gives $K_1 \approx 0.4$. If the fringing field is not clamped by mirror plates, the value of K_1 is higher: approximately 0.5 for a square-edged magnet and 0.7 for tapered polepiece edges (Brown et al., 1977). If the polepiece gap is large, a second coefficient K_2 may be necessary to properly describe the effect of the fringing field; however, its effect is small for $g/R < 0.3$ (Heighway, 1975).

2.2.1.2 Matrix Notation

Particularly when fringing fields are taken into account, the equations needed to describe the focusing properties of a magnetic prism become quite complicated. Their form can be simplified and the method of calculation made more systematic by using a matrix notation, as in the design of light-optical systems. The optical path between object and image is divided into sections and a *transfer matrix* written down for each section. The first stage of the electron trajectory corresponds to *drift* in a straight line through the field-free region between the object plane and the entrance edge of the magnet. The displacement coordinates (x, y, z) of an electron change, but not its angular coordinates $(x' = dx/dz,\ y' = dy/dz)$. Upon arrival at the entrance edge of the magnet, these four coordinates are therefore given by the following matrix equation:

$$\begin{pmatrix} x \\ x' \\ y \\ y' \end{pmatrix} = \begin{pmatrix} 1 & u & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & u \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ x'_0 \\ y_0 \\ y'_0 \end{pmatrix} \tag{2.8}$$

Here, x_0 and y_0 are the components of electron displacement at the object plane, x'_0 and y'_0 are the corresponding angular components, and the 4×4 square matrix is the transfer matrix for drift over a distance u (measured along the optic axis).

The electron then encounters the focusing action of the tilted edge of the magnet. In the SCOFF approximation, the focusing powers are $1/f_x = -(\tan \varepsilon)/R$ and

$1/f_y = (\tan\varepsilon)/R$. The focusing being of equal magnitude but opposite sign in the x- and y-directions, the magnet edge is equivalent to a quadrupole lens. Allowing for extended fringing fields, the corresponding transfer matrix is (Brown, 1967)

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ R^{-1}\tan\varepsilon_1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -R^{-1}\tan(\varepsilon_1 - \psi_1) & 1 \end{pmatrix} \tag{2.9}$$

where ψ_1 represents a correction for the extended fringing field, given by

$$\psi_1 \approx (g/R)K_1(1 + \sin^2\varepsilon_1)/\cos\varepsilon_1 \tag{2.10}$$

The third part of the trajectory involves bending of the beam within the interior of the magnet. As discussed in Section 2.1.1, the uniform magnetic field has a positive (convex) focusing action in the x-direction but no focusing action in the y-direction. The effect is equivalent to that of a dipole field, as produced by a *sector* magnet with $\varepsilon_1 = \varepsilon_2 = 0$. If ϕ is the bend angle, the corresponding transfer matrix can be written in the form (Penner, 1961)

$$\begin{pmatrix} \cos\phi & R\sin\phi & 0 & 0 \\ -R^{-1}\sin\phi & \cos\phi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.11}$$

Upon arrival at the exit edge of the prism, the electron again encounters an effective quadrupole, whose transfer matrix is specified by Eqs. (2.9) and (2.10) but with ε_2 substituted for ε_1. Finally, after leaving the prism, the electron drifts to the image plane, its transfer matrix being identical to that in Eq. (2.8) but with the object distance u replaced by the image distance v.

Following the rules of matrix manipulation, the five transfer matrices are multiplied together to yield a 4×4 transfer matrix that relates the electron coordinates and angles at the image plane (x_i, y_i, x'_i, and y'_i) to those at the object plane. However, the first-order properties of a magnetic prism can be specified more completely by introducing two additional parameters. One of these is the total distance or path length l traversed by an electron, which is of interest in connection with time-of-flight measurements but not relevant to dispersive operation of a spectrometer. The other additional parameter is the fractional momentum deviation δ of the electron, relative to that required for travel along the optic axis (corresponding to a kinetic energy E_0 and zero energy loss). This last parameter is related to the energy loss E by

$$\delta = -E/(2\gamma T) \tag{2.12}$$

where $2\gamma T = E_0(2m_0c^2 + E_0)/(m_0c^2 + E_0)$. The first-order properties of the prism are then represented by the equation

$$\begin{pmatrix} x_t \\ x'_{\mathrm{i}} \\ y_{\mathrm{i}} \\ y'_{\mathrm{i}} \\ l_{\mathrm{i}} \\ \delta_{\mathrm{i}} \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} & 0 & 0 & 0 & R_{16} \\ R_{21} & R_{22} & 0 & 0 & 0 & R_{26} \\ 0 & 0 & R_{33} & R_{34} & 0 & 0 \\ 0 & 0 & R_{44} & R_{44} & 0 & 0 \\ R_{51} & R_{52} & 0 & 0 & 1 & R_{56} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ x'_0 \\ y_0 \\ y'_0 \\ l_0 \\ \delta_0 \end{pmatrix} \tag{2.13}$$

Many of the elements in this 6×6 matrix are zero as a result of the mirror symmetry of the spectrometer about the x–z plane. Of the remaining coefficients, R_{11} and R_{33} describe the lateral image magnifications (M_x and M_y) in the x- and y-directions. In general, $R_{11} \neq R_{33}$, so the image produced by the prism suffers from rectangular distortion. For a real image, R_{11} and R_{33} are negative, denoting the fact that the image is inverted about the optic axis. R_{22} and R_{44} are the angular magnifications, approximately equal to the reciprocals of R_{11} and R_{33}, respectively.

Provided the spectrometer is double focusing and the value of the final drift length used in calculating the R-matrix corresponds to the image distance, R_{12} and R_{34} are both zero. If the spectrometer is not double focusing, $R_{12} = 0$ at the x-focus and the magnitude of R_{34} gives an indication of the length of the line focus in the y-direction. To obtain good energy resolution from the spectrometer, R_{12} should be zero at the energy-selection plane and R_{34} should preferably be small. The other matrix coefficient of interest in connection with energy-loss spectroscopy is $R_{16} = \partial x_i / \partial(\delta_0)$, which relates to the energy dispersion of the spectrometer. Using Eq. (2.12), the dispersive power $D = -\partial x_i / \partial E$ is given by

$$D = R_{16}/(2\gamma T) \tag{2.14}$$

The R-matrix of Eq. (2.13) can be evaluated by multiplication of the individual transfer matrices, provided the values of u, ε_1, ϕ, ε_2, K_1, g, and v are specified. Such tedious arithmetic is best done by computer, for example, by running the TRANSPORT program (Brown et al., 1977). This program[1] also computes second- and third-order focusing, the effects of other elements (e.g., quadrupole or sextupole lenses), of a magnetic field gradient or inhomogeneity, and of stray magnetic fields.

2.2.2 Higher Order Focusing

The matrix notation is well suited to the discussion and calculation of second-order properties of a magnetic prism. Using the same six coordinates (x, x', y, y', l, and δ), second derivatives in the form (for example) $\partial^2 x_{\mathrm{i}} / \partial x_0 \partial x'_0$ can be defined and arranged in the form of a $6 \times 6 \times 6$ T-matrix, analogous to the first-order R-matrix. Many of the $6^3 = 216$ second-order T-coefficients are zero or

[1] Available from http://www.slac.stanford.edu/pubs/slacreports/slac-r-530.html

are related to one another by midplane symmetry of the magnet. For energy-loss *spectroscopy*, where the beam diameter at the object plane (i.e., the source size) is small and where image distortions and off-axis astigmatism are of little significance, the most important second-order matrix elements are $T_{122} = \partial^2 x_i/\partial(x_0')^2$ and $T_{144} = \partial^2 x_i/\partial(y_0')^2$. These coefficients represent second-order *aperture* aberrations that increase the image width in the x-direction and therefore degrade the energy resolution, particularly in the case of a large spread of incident angles (x_0' and y_0').

Whereas the first-order focusing of a magnet boundary depends on its effective quadrupole strength (equal to $-\tan\ \varepsilon = -\partial z/\partial x$ in the SCOFF approximation), the second-order aperture aberration depends on the effective sextupole strength: $-(2\rho\cos^3\varepsilon)^{-1}$ in the SCOFF approximation (Tang, 1982a). The aberration coefficients can therefore be varied by adjusting the angle ε and curvature $\rho = \partial^2 z/\partial x^2$ of the boundary. Convex boundaries can only correct second-order aberration for electrons traveling in the x–y plane ($T_{122} = 0$), but if one boundary is made concave, the aberration for electrons traveling out of the radial plane can also be corrected ($T_{122} = T_{144} = 0$). Alternatively, the correction can be carried out by means of magnetic or electrostatic sextupole lenses placed before and after the spectrometer (Parker et al., 1978).

A second-order property that is of particular importance is the angle ψ between the dispersion plane (the plane of best chromatic focus for a point object) and the x-axis adjacent to the image; see Fig. 2.10. This tilt angle is related to the matrix element $T_{126} = \partial^2 x_i/\partial x_0'\partial(\delta)$ by[2]

$$\tan\psi = -T_{126}/(R_{22}R_{16}) \tag{2.15}$$

The condition $\psi = 0$ is desirable if lenses follow the spectrometer or if a parallel-recording detector is oriented perpendicular to the exit beam. More generally, adjustment of ψ allows control over the *chromatic* aberration of whole system, external lenses included. Another second-order coefficient of some relevance is T_{166}, which does not affect the energy resolution but specifies nonlinearity of the energy-loss axis.

The matrix method has been extended to third-order derivatives, including the effect of extended fringing fields (Matsuda and Wollnik, 1970; Matsuo and Matsuda, 1971; Tang, 1982a, b). Many of the 1296 third-order coefficients are zero as a result of the midplane symmetry, and only a limited number of the remaining ones are of interest for energy-loss spectroscopy. The coefficients $\partial^3 x_i/\partial(x_0')^3$ and $\partial^3 x_i/\partial x_0'\partial(y')^2$ represent aperture aberrations and (like T_{122} and T_{144}) may have the same or opposite signs (Scheinfein and Isaacson, 1984). Tang (1982a) pointed out that correction of second-order aberrations by curving the entrance and exit edges of the magnet can increase these third-order coefficients, so that the latter limit the energy resolution for entrance angles γ above 10 mrad. The chromatic

[2] If TRANSPORT is used to calculate the matrix elements, a multiplying factor of 1000 is required on the right-hand side of Eq. (2.15) as a result of the units (x in cm, x' in mrad, and δ in %) used in that program.

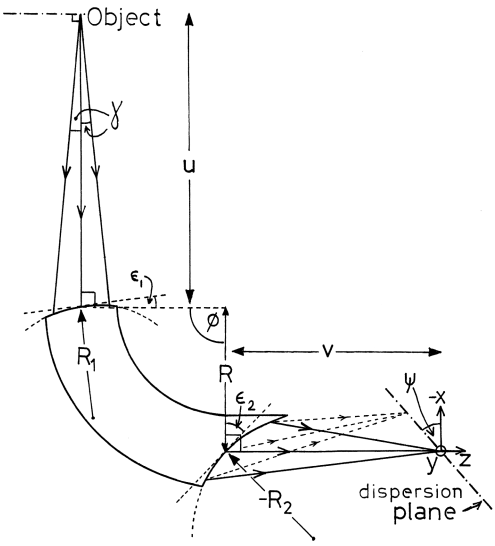

Fig. 2.10 Electron optics of a double-focusing spectrometer with curved polefaces. The *y*-axis and the applied magnetic field are perpendicular to the plane of the diagram. The polepiece-tilt angles (ε_1 and ε_2) refer to the central trajectory (the optic axis). Exit trajectories of energy-loss electrons are shown by *dashed lines*. From Egerton (1980b), copyright Elsevier

term $\partial^3 x_\mathrm{i}/\partial(\delta)^3$ causes additional nonlinearity of the energy-loss axis but is likely to be important only for energy losses of several kiloelectron volts. The coefficients $\partial^3 x_\mathrm{i}/\partial(x_0')^2\partial(\delta)$, $\partial^3 x_\mathrm{i}/\partial(y_0')^2\partial(\delta)$, and $\partial^3 x_\mathrm{i}/\partial(x_0')\partial(\delta)^2$ introduce tilt and curvature of the dispersion plane, which may degrade the energy resolution when a parallel-recording system is used.

If third-order aberrations are successfully corrected, for example, by the use of octupole lenses outside the spectrometer (Tang, 1982b; Krivanek et al., 2008), the energy resolution is limited by the fourth-order aberrations: $\partial^4 x_\mathrm{i}/\partial(x_0')^4$, $\partial^4 x_\mathrm{i}/\partial(y_0')^4$, and $\partial^4 x_\mathrm{i}/\partial(x_0')^2\partial(y_0')^2$. Fourth-order matrix theory has not been developed but ray-tracing programs can be used to predict the focusing of electrons. They work by evaluating the rate of change of momentum as $(-e)(\mathbf{v} \times \mathbf{B})$ and using this information to define an electron path, initially the optic axis. A trajectory originating from the center of the object plane but at a small angle relative to the optic axis is then evaluated, where this ray arrives back at the optic axis defines the Gaussian image plane. The positions on this plane of electrons with increasing angular deviation define an aberration figure (or spot diagram), from which aberration coefficients can be estimated. One program that implements ray tracing is SIMION (http://simion.com), which runs on Windows and Linux computers.

Spectrometer aberrations are particularly important in the case of core-loss spectroscopy involving higher energy ionization edges, where the angular range of the inelastic scattering can extend to tens of milliradians and where high collection efficiency is desirable to obtain adequate signal.

2.2.3 Spectrometer Designs

To illustrate the above concepts, we outline a procedure for designing a double-focusing spectrometer with aperture aberrations corrected to second order by

curving the entrance and exit edges. First of all, the prism angles are chosen so as to obtain suitable first-order focusing. As discussed below, the value of ε_1 should either be fairly large (close to 45°) or quite small (<10°, or even negative). Knowing the location of the spectrometer object point and the required bend radius R (which determines the energy dispersion D and the size and weight of the magnet), approximate values of ε_2 and v can be calculated using Eqs. (2.4) and (2.5). If either v or ε_2 turns out to be inconveniently large, a different value of ε_1 must be selected.

These first-order parameters are then refined to take account of extended fringing fields, requiring a knowledge of the integral K_1 (which depends on the shape of the polepiece corners and on whether magnetic field clamps are to be used) and the polepiece gap g (typically $0.1R$–$0.2R$). The TRANSPORT program uses a fitting procedure to find the exact image distance v corresponding to an x-focus ($R_{12} = 0$). The value of R_{34} will then be nonzero, indicating a line focus. If $|R_{34}|$ is excessive (>1 μm/mrad), either ε_1 or ε_2 is changed slightly to obtain a closer approach to double focusing. The dispersive power of the spectrometer is estimated from Eq. (2.6) or obtained more accurately using Eq. (2.14).

The next stage is to determine values of the edge curvatures R_1 and R_2 that make the second-order aperture aberrations zero. This is most easily done by recognizing that T_{122} and T_{144} both vary linearly with the edge curvatures. In other words,

$$-T_{122} = a_0 + a_1(R/R_1) + a_2(R/R_2) \tag{2.16}$$

$$-T_{144} = b_0 + b_1(R/R_1) + b_2(R/R_2) \tag{2.17}$$

where a_0, a_1, a_2, b_0, b_1, and b_2 are constants for a given first-order focusing. In general, a_0 is positive but a_1 and a_2 are negative; T_{122} can therefore be made zero with R_1 and R_2 both positive, implying convex entrance and exit edges. However, b_0, b_1, and b_2 are usually all positive so $T_{144} = 0$ requires that either R_1 or R_2 be negative, indicating a concave edge (Fig. 2.10). The required edge radii (R_1^* and R_2^*) are found empirically by using the matrix program to calculate T_{122} and T_{144} for three arbitrary pairs of R/R_1 and R/R_2, such as (0, 0), (0, 1), and (1, 0), generating six simultaneous equations that can be solved for a_0, a_1, a_2, b_0, b_1, and b_2. Then R_1^* and R_2^* are deduced by setting T_{122} and T_{144} to zero in Eqs. (2.16) and (2.17).

Not all spectrometer geometries yield reasonable values of R_1^* and R_2^*. For example, the completely symmetric case ($u = v = 2R$, $\varepsilon_1 = \varepsilon_2 = 26.6°$ for $\phi = 90°$ in the SCOFF approximation) gives $R_1^* = R_2^* = 0$, corresponding to infinite curvature. As $|\varepsilon_1 - \varepsilon_2|$ increases, the necessary edge curvatures must be kept reasonably low because the maximum effective width w^* of the polepieces at the entrance or exit edge is given by

$$w^* = 2\left|R^*\right|(1 - \sin|\varepsilon|) \tag{2.18}$$

for a concave edge and by $w^* = 2R^*(\cos\varepsilon)$ for a convex edge. In practice, the concave edge corresponds to the higher value of ε, so (for small R^*) Eq. (2.18) imposes an upper limit on the angular range (x') of electrons that can pass through

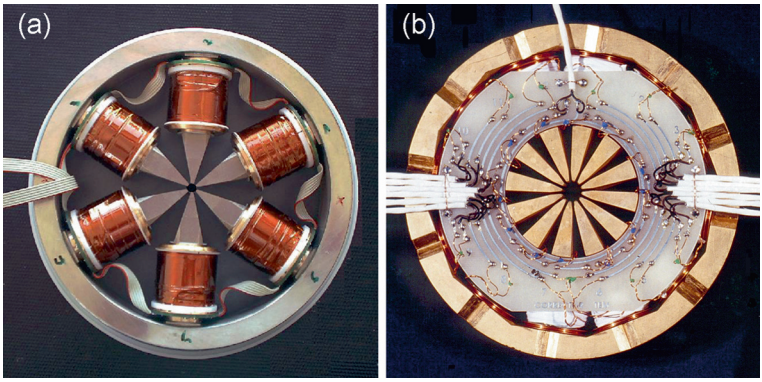

Fig. 2.11 (**a**) Magnetic sextupole and (**b**) multipole lens, used to correct the aberrations of a TEM lens or a spectrometer. Courtesy of Max Haider, CEOS GmbH

the magnet. This limitation is not present if aberrations are adjusted by means of external sextupoles. With the addition of external octupoles, third-order aberrations to be corrected (Krivanek et al., 2008). Such multipole devices are also used to correct the aberrations of axially symmetric imaging lenses; see Fig. 2.11.

In the above analysis, the object distance u and bend radius R were assumed to be fixed by the geometry of an electron microscope column and the space available for the spectrometer. If the ratio u/R can be varied, there is freedom to adjust a further second-order matrix element, such as T_{126}. Parker et al. (1978) showed that there can be two values of image (or object) distance for which $T_{126} = 0$, giving zero tilt of the dispersion plane.

Table 2.1 gives examples of aberration-corrected designs. The recent GIF Quantum spectrometer (Gubbens et al., 2010) uses a gradient field design to reduce

Table 2.1 Design parameters for aberration-corrected spectrometers

ϕ (deg)	ε_1 (deg)	ε_2 (deg)	u/R	v/R	R_1^*/R	R_2^*/R	g/R	Reference
60	14.64	18.07	∞	1.46	1.351	−2.671	0.07	Fields (1977)
90	0	45.0	1.45	2.16	0.807	−1.357	0.2	Egerton (1980b)
70	11.75	28.79	3.60	2.38	0.707	−0.603	0.137	Shuman (1980)
90	17.5	45.0	5.5	0.98	1.0	−0.496	0.125	Krivanek and Swann (1981)
66.6	−15	45.8	2.25	2.06	2.34	−1.30	0.18	Tang (1982a)
90	15.9	46.5	4.52	1.08	0.867	−0.500	0.19	Reichelt and Engel (1984)
80	14.6	35.1	3.5	1.82	0.728	−0.576	0.25	Scheinfein and Isaacson (1984)
90	16	47	6.2	0.9	1.0	−0.42	0.125	Krivanek et al. (1995)
90	0	0	10.9	0.7	∞	∞	0.227	Gubbens et al. (2010)

the poleface angles (ε_1 and ε_2). The pole faces are not curved; aberration correction (including partial fourth and fifth orders) is achieved by means of three external dodecupole (12-pole) lenses, one before the prism and two between the prism and the energy-selecting slit. The poles of these lenses can be individually excited to generate any combination of dipole, quadrupole, sextupole, and higher order elements, in order to control the prism focusing up to sixth order. Because these different elements share the same optic axis, alignment is easier than with separate elements. A further five dodecupoles (after the slit) project a spectrum or an energy-filtered image onto the CCD detector.

2.2.4 Practical Considerations

The main aim when designing an electron spectrometer is to achieve good energy resolution even in the presence of a large spread γ of entrance angles, enabling the spectrometer system to have a high collection efficiency (see Section 2.3). For $\gamma = 10$ mrad, correction of second-order aberrations allows a resolution $\approx$1 eV for energy losses up to 1 keV (Krivanek and Swann, 1981; Colliex, 1982; Scheinfein and Isaacson, 1984). The value of γ is limited by the internal diameter of the "drift" tube (Fig. 2.9), which is necessarily less than the magnet gap g, so the historical trend has been toward relatively large g/R ratios (see Table 2.1), even though this makes accurate calculation of the fringing-field properties more difficult (Heighway, 1975). The use of multipole elements, giving partial correction up to fifth order, can provide a resolution below 0.1 eV (Gubbens et al., 2010).

The energy range falling on the detector depends on the bend radius R and the dispersion D, which increases with decreasing beam energy E_0. In the standard (SR) version of the GIF Quantum spectrometer, the bend radius R has been reduced from 100 to 75 mm, allowing 2 keV range for 200 keV electrons or 682 eV at 60 keV. An even smaller value (50 mm) is scheduled for lower-voltage TEMs (15–60 keV) and a larger one (200 mm) for high-voltage operation (400–1250 keV).

Spectrometer designs such as those in Table 2.1 assume that the magnetic induction B within the magnet is uniform or (in the gradient field case) varies linearly with distance x from the optic axis. More generally, the induction might vary according to

$$B(x) = B(0)[1 - n(x/R) - m(x/R)^2 + \cdots] \tag{2.19}$$

where the coefficients $n, m, \ldots$ introduce multipole components in the focusing. In a gradient field spectrometer the value of n depends on the angle between the polefaces, which controls first-order focusing in the x- and y-directions, as an alternative to tilting the entrance and exit faces. Likewise, a sextupole component (dependent on m) could be deliberately added to control second-order aberrations (Crewe and Scaduto, 1982). However, the focusing properties are quite sensitive to the values of n and m; matrix calculations suggest that changing m by two parts in 10^{-6} will degrade the energy resolution by 1 eV, for $\gamma = 10$ mrad (Egerton, 1980b). Therefore

unintended variations must be avoided if a spectrometer is to behave as designed. A "*C*-core" magnet (where the magnetic field is generated by a coil connected by side arms to the polepieces) does not provide the required degree of uniformity, whereas a more symmetrical arrangement with *window frame* coils placed on either side of the gap (Fig. 2.9) can give a sufficiently uniform field, particularly if the separation between the planes of the two coils is carefully adjusted (Tang, 1982a).

Further requirements for field uniformity are that the magnetic material is sufficiently homogeneous and adequately thick. Homogeneity is achieved by annealing the magnet after machining and by choosing a material with high relative permeability μ and low coercivity at low field strength, such as mu-metal. Its minimum thickness t can be estimated by requiring the magnetic reluctance $(\propto w/t\mu)$ of each polepiece (in the x-direction) to be much less than the reluctance of the gap $(\propto g/w)$, giving

$$t >> w^2/(\mu g) \tag{2.20}$$

Equation (2.20) precludes the use of thin magnetic sheeting, which would otherwise be attractive in terms of reduced weight of the spectrometer.

It might appear that B-uniformity of 2×10^{-16} would require the polepiece gap to be uniform to within 2 μm over an x-displacement of 1 cm, for $g = 1$ cm. However, the allowable variation in field strength is that averaged over the whole electron trajectory, variations in the z-direction having less effect on the focusing. Also, provided they are small, the x- and x^2-terms in Eq. (2.19) can be corrected by external quadrupole and sextupole coils.

A substantial loss of energy resolution can occur if stray magnetic fields penetrate into the spectrometer. Field penetration can be reduced by enclosing the magnet and (more importantly) the entrance and exit drift spaces in a soft magnetic material such as mu-metal. Such screening is usually not completely effective but the influence of a remaining alternating field can be canceled by injecting a small alternating current into the spectrometer scan coils, as described in Section 2.2.5.

The magnetic induction within the spectrometer is quite low (<0.01 T for 100 keV operation) and can be provided by window frame coils of about 100 turns carrying a current of the order of 1 A. To prevent drift of the spectrum due to changes in temperature and resistance of the windings, the power supply must be *current* stabilized to within one part in 10^6 for 0.3-eV stability at 100-keV incident energy. Stability is improved if the power supply is left running continuously.

2.2.5 Spectrometer Alignment

Like all electron-optical elements, the magnetic prism performs to its design specifications only if it is correctly aligned relative to the incoming beam of electrons. Since the energy dispersion is small for high-energy electrons, this alignment is fairly critical if the optimum energy resolution is to be achieved.

2.2.5.1 Initial Alignment

When a spectrometer is installed for the first time, or if the alignment of the spectrometer or the microscope column has been disturbed, the electron beam may travel in a path that is far from the optic axis of the spectrometer (defined by the prism orientation and the value of the magnetic induction B). In this situation, a rough alignment of the system can be carried out in much the same way as alignment of an electron microscope column. Beam-limiting apertures, such as the spectrometer entrance aperture, are withdrawn and the entrance beam broadened, for example, by defocusing the illumination at the specimen plane. It may also be useful to sweep the magnetic field periodically by applying a fast ramp to the spectrometer scan coils, to deflect the exit beam over a range of several millimeters in the x-direction. Use of a two-dimensional detector, such as a phosphor screen and CCD camera, allows the exit beam to be located in both the x- and y-directions, especially if the energy-selecting slit is withdrawn. By alternately focusing and defocusing the pre-spectrometer lenses, it is possible to discover if the electron beam is passing through the center of the drift tube or is cut off asymmetrically by the tube walls or fixed apertures. To ensure that the beam travels close to the *mechanical* axis of the spectrometer, it may be desirable to shift or tilt the magnet so that the positions (on the phosphor screen) where the exit beam is cut off are symmetric with respect to the center of the detector.

2.2.5.2 Aberration Figure

For optimum performance from the spectrometer, the beam must travel close to the *magnetic* axis of the prism and the spectrometer focusing must be correct. The desirable conditions can be recognized from the *shape* of the beam at the detector plane. The first-order focusing is correct when the exit beam at the detector plane has minimum width in the direction of dispersion. This condition is normally adjusted by means of quadrupole elements placed before or after the spectrometer. The focusing can be set more accurately if the depth of focus is made small, by using a large spectrometer entrance aperture and adjusting the TEM lenses so that the circle of illumination is large enough at that plane.

If there were no aberrations and if the spectrometer were exactly double focusing, the exit beam would appear as a point or circle of very small diameter at the detector plane. Spectrometer aberrations spread the beam into an *aberration figure* that can be observed directly on a fluorescent screen if the entrance divergence and the aberration coefficients are large enough. Second-order aberrations produce a figure whose shape (Fig. 2.12) can be deduced from the equations

$$x_{\mathrm{i}} = T_{122}(x_0')^2 + T_{144}(y_0')^2 \tag{2.21}$$

$$y_{\mathrm{i}} = T_{324}x_0'y_0' \tag{2.22}$$

T_{122}, T_{144}, and T_{324} are matrix coefficients that represent second-order aperture aberrations; x_0' and y_0' represent the angular coordinates of an electron entering

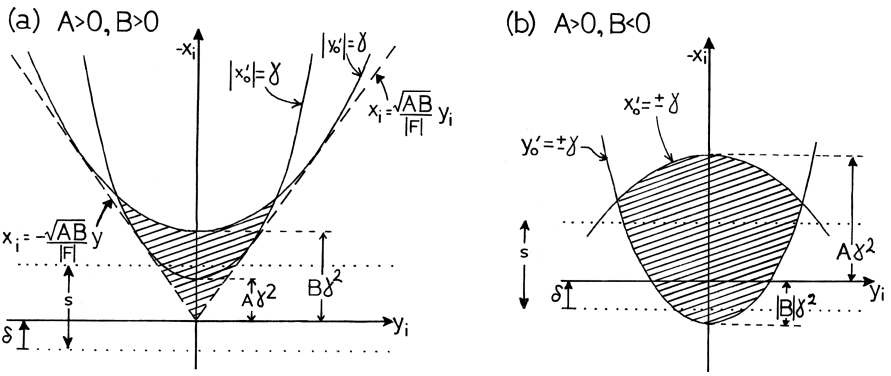

Fig. 2.12 Aberration figures of a properly aligned magnetic prism whose energy resolution is determined by second-order aberration coefficients $T_{122} = -A$ and $T_{144} = -B$ that are (**a**) of the same sign and (**b**) of opposite sign. The entrance angles $x_0{}'$ and $y_0{}'$ are assumed to be limited to values in the range $-\gamma$ to $+\gamma$ by a square entrance aperture; the result of a circular entrance aperture is similar except that the *top of figure* (**a**) is convex. *Dotted lines* indicate the position of the detector slit when recording an alignment figure. The matrix element T_{324} is denoted by F. From Egerton (1981b), copyright Elsevier

the spectrometer. For a fixed x_0' and a range of y_0' (or vice versa), the relationship between the image-plane coordinates x_i and y_i is a parabola. In practice, both x_0' and y_0' take a continuous range of values: $-\gamma$ to $+\gamma$, where γ is the maximum entrance angle (defined by a spectrometer entrance aperture, for example). The image-plane intensity is then represented by the shaded area in Fig. 2.12. When the magnet is correctly aligned, this figure is symmetric about the vertical (x-) axis; second-order aberrations are properly corrected when its width in the x-direction is a minimum. Because the aberration figure has very small dimensions, it is difficult to observe unless the spectrometer is followed by magnifying electron lenses.

2.2.5.3 Alignment (Nonisochromaticity) Figure

An alternative way of observing the aberration properties of a spectrometer is to place a narrow slit in its image plane and measure the electron flux through this slit by means of a single-channel detector (e.g., scintillator and photomultiplier), as in the case of serial recording of energy-loss spectra. Rather than scanning the exit beam across the slit, the entrance angle is varied by rocking the entrance beam about the spectrometer object point (Fig. 2.10). For a TEM fitted with a scanning attachment, the incident probe can be scanned over the specimen plane in the form of a two-dimensional raster; if the object plane of the spectrometer contains a diffraction pattern of the specimen (at the projector lens cross-over, for example), the beam entering the spectrometer is swept in angle in both the x- and y-directions. Applying voltages proportional to x_0' and y_0' to the horizontal and vertical channels of an oscilloscope and using the signal from the electron detector to modulate the

brightness of the oscilloscope beam (z-modulation), an *alignment* figure is obtained whose shape depends on the aberrations that directly affect the resolving power of the spectrometer.

An electron arriving at the image plane will pass through the detection slit provided

$$-\delta < x_0 < s - \delta \tag{2.23}$$

where s is the slit width in the x-direction and δ specifies the position of the aberration figure relative to the slit; see Fig. 2.12. For the case where second-order aberrations are dominant, the shape of the alignment figure is specified by Eqs. (2.21) and (2.23). If $\delta = 0$ and if T_{122} and T_{144} are both negative (as in the case of a typical straight-edged magnet), a solid ellipse is formed, whose dimensions depend on the values of T_{122}, T_{144}, and s (Fig. 2.13a). As the current in the spectrometer field coils is increased ($\delta > 0$), the pattern shrinks inward and eventually disappears; if the spectrometer excitation is decreased, the pattern expands in outline but develops a hollow center. If T_{122} and T_{144} were both positive, this same sequence would be observed as the spectrometer current were *decreased*.

When T_{122} and T_{144} are of opposite sign, the alignment figure consists of a pair of hyperbolas (Fig. 2.13b). If $T_{144} > 0$, the hyperbolas come together as the spectrometer current is increased and then separate in the y-direction, as in Fig. 2.14. The sequence would be reversed if T_{122} were the positive coefficient.

When third-order aberrations are dominant, the aberration figure has three lobes (Fig. 2.15) but retains its mirror plane symmetry about the x-axis, in accordance with the symmetry of the magnet about the x–z plane.

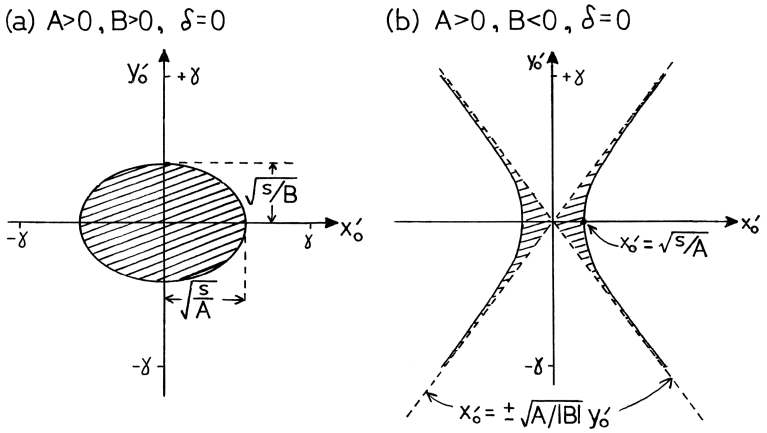

Fig. 2.13 Alignment figures of a magnetic prism whose energy resolution is limited by second-order aberration coefficients ($T_{122} = -A$ and $T_{144} = -B$) that are (**a**) of the same sign and (**b**) of opposite sign. In (**a**), the scan range 2γ is assumed to be larger than the major axis of the ellipse. From Egerton (1981b), copyright Elsevier

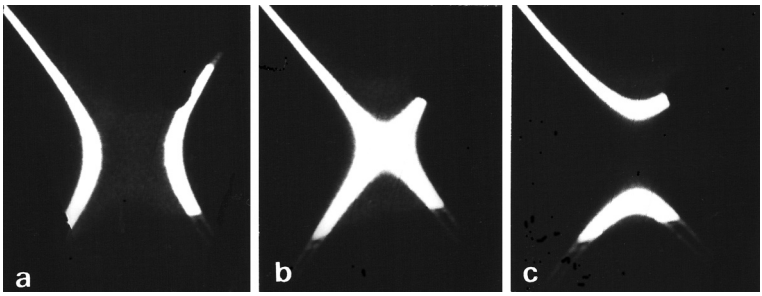

Fig. 2.14 Change in shape of the alignment figure as the spectrometer excitation is increased, for the case $T_{122} < 0$ and $T_{144} > 0$. In (**a**), $\delta < 0$; in (**b**), $\delta \approx 0$; in (**c**), $\delta > 0$. From Egerton (1981b), copyright Elsevier

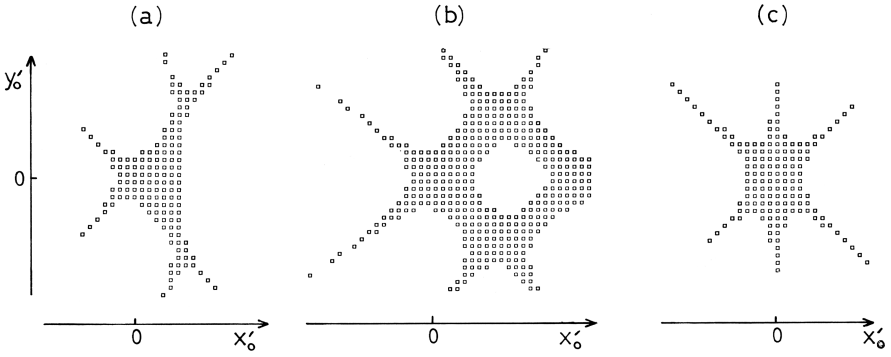

Fig. 2.15 Calculated alignment figures for the magnetic prism spectrometer having (**a**) third-order and residual second-order aberrations (with T_{122} and T_{144} of opposite sign), (**b**) fourth-order and residual third-order aberrations, and (**c**) pure third-order aberrations. From Scheinfein and Isaacson (1984), copyright SEM Inc., Illinois

Following from the above, some uses of the alignment figure are as follows.

(1) In order to optimize the energy resolution, the spectrometer should be mechanically or electrically aligned such that the figure is symmetrical about its x'_0 axis. The most sensitive alignment is the tilt of the magnet about the exit-beam direction but it is not easy to provide this rotation in the form of a single mechanical control.
(2) The alignment figure enables the currents in multipole coils to be adjusted to compensate residual aberrations. To maximize the collection efficiency of the spectrometer for a given energy resolution, the currents should be adjusted so that the pattern is as large in area and as near-circular as possible. During this adjustment, it may be necessary to change the prism current to prevent the display from disappearing or developing a hollow center.
(3) The symmetry of the alignment figure indicates the order of the uncorrected aberrations and the relative signs of the dominant aberration coefficients. The

absolute signs can be deduced from the change in the pattern as the spectrometer excitation is varied. The ratio of the coefficients can be estimated by measuring the aspect ratio (Fig. 2.13a) or angle between the asymptotes (Fig. 2.13b). Absolute magnitudes can be obtained if the display is calibrated in terms of entrance angle.

(4) If a spectrometer entrance aperture is inserted to limit the angular range of electrons entering the prism, its image should appear in outline on the display. To achieve the best combination of energy resolution and collection efficiency, the aperture is centered so that as little as possible of the alignment figure is cut off from the display.

(5) The influence of stray ac magnetic fields can be detected as a blurring or waviness of the edges of the alignment figure. Imperfections in the energy-selecting slit (due to mechanical irregularity or contamination) show up as a streaking of the pattern.

2.2.5.4 Stray-Field Compensation

Stray magnetic fields can easily affect the performance of an electron spectrometer. In the case of a magnetic prism attached to a conventional TEM, external fields can penetrate into the viewing chamber and deflect the electron beam before it enters the spectrometer. Slowly changes in field are minimized by installing a field compensation system and by ensuring that movable magnetic objects, such as steel chairs, are replaced by nonmagnetic ones. Some of the external interference comes from mains frequency fields and can be compensated by a simple circuit that applies mains frequency current of adjustable amplitude and phase to the spectrometer excitation coils (Egerton, 1978b). External fields are less likely to be troublesome if the TEM viewing chamber is made of a magnetically shielding material (such as soft iron) or if the viewing chamber is eliminated, as in some recent TEM designs.

2.3 The Use of Prespectrometer Lenses

The single-prism electron spectrometer fitted to a conventional (fixed-beam) TEM is located below the imaging lenses, so electrons emerging from the specimen pass through these lenses before reaching the spectrometer. Not surprisingly, the performance of the EELS system (energy resolution, collection efficiency, and spatial resolution of analysis) is affected by the properties of the TEM imaging lenses and the way in which they are operated.

The influence of TEM lenses on spectrometer performance was analyzed in a general way by Johnson (1980a, b), Egerton (1980a), and Krivanek et al. (1995). Good energy resolution requires that an electron-beam crossover of small diameter be placed at the spectrometer object plane. In practice, this crossover is either a low-magnification image of the specimen or a portion of its diffraction pattern (just the central beam, if a bright-field objective aperture is inserted). Since the spectrometer in turn images this crossover onto the EELS detector or the energy-selecting slit,

what is actually recorded represents a *convolution* of the energy-loss spectrum with the diffraction pattern or image of the specimen, sometimes called spectrum diffraction or spectrum image *mixing*. In order to prevent diffraction or image information from seriously contaminating or distorting the energy-loss spectrum, the dimensions of the image or diffraction pattern (at the spectrometer object plane) must be made small relative to the energy dispersion.

Some early spectrometer systems (Pearce-Percy, 1976; Joy and Maher, 1978; Egerton, 1978b) operated with the TEM projector lens turned off. Electrons were focused into a small crossover at the level of the TEM screen, which was also the spectrometer object plane. The region of specimen (diameter d) giving rise to the energy-loss spectrum was determined by the diameter of electron beam at the specimen or by inserting a selected area diffraction (SAD) aperture. The energy resolution available in this mode was analyzed by Johnson (1980a, b) and Egerton (1980a).

2.3.1 TEM Imaging and Diffraction Modes

Gatan spectrometers work with the projector lens on, as in normal TEM operation. The projector forms an optical crossover just below its lens bore, a distance h (typically 30–40 cm) above the TEM viewing screen, and this crossover acts as the object point O of the spectrometer; see Fig. 2.16. Because the final TEM lens is designed to produce a large diameter image or diffraction pattern, the solid angle

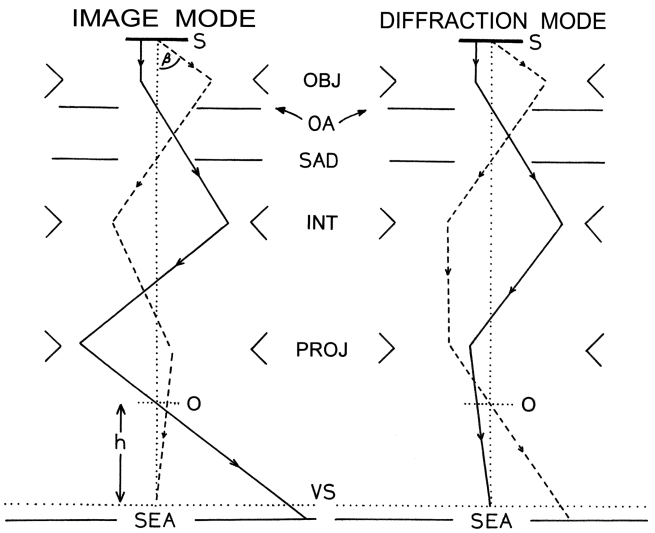

Fig. 2.16 Simplified optics for the image and diffraction modes of a conventional TEM. *S* represents the specimen; *OBJ*, *INT*, and *PROJ* represent the objective lens, intermediate-lens system, and final imaging (projector) lens. *O* and *VS* are the spectrometer object point and viewing screen; *OA*, *SAD*, and *SEA* are the objective, selected area diffraction and spectrometer entrance apertures

of divergence at O is large. Since electron optical brightness is conserved, the angle diameter product is constant; therefore, the crossover has a very small diameter.

If the TEM is operated in *image mode*, with an image of the specimen of magnification M on the viewing screen, the spectrometer is said to be *diffraction coupled* because the projector lens crossover then contains a small diffraction pattern of the specimen. The size of this diffraction pattern is represented by a camera length: $L_o = h/M$, and can be as small as 1 μm. The angular range of scattering allowed into the spectrometer (the collection semi-angle β) is controlled by varying the size of the objective lens aperture. The region of specimen giving rise to the energy-loss spectrum is determined by a spectrometer entrance aperture (SEA) and corresponds to a portion of the image close to the center of the TEM viewing screen (before the screen is lifted to allow electrons through to the spectrometer). More precisely, the diameter of analysis is $d = 2R/M'$, where R is the SEA radius and $M' = M(h'/h)$ is the image magnification at the SEA plane, h' being height of the projector lens crossover relative to the SEA.. Because of the large depth of field, an image that is in focus at the TEM screen is very nearly in focus at the SEA plane, so the SEA can act as an area-selecting aperture.

If the TEM is operated in *diffraction mode*, with a diffraction pattern of camera length L at the viewing screen, the spectrometer is *image coupled* because the projector crossover now contains an image of the illuminated area of the specimen. The image magnification at O is $M_o = h/L$ and is typically of the order of 1. Unless the objective aperture limits it to a smaller value, the collection semi-angle is $\beta = R/L'$, where $L' = L(h'/h)$ is the camera length at the SEA plane. To ensure that the SEA is centered on the optic axis, TEM diffraction shift controls have to be adjusted for maximum intensity of some sharp spectral feature. Alternatively, these controls can be used to select any desired region of the diffraction pattern for energy analysis. The area of specimen being analyzed is determined by the electron-beam diameter at the specimen or else by a selected area diffraction (SAD) aperture, if this aperture is inserted to define a smaller area.

The above considerations are based on first-order geometric optics. Although some objective lenses are corrected for spherical aberration (Hawkes, 2008), most TEM imaging lenses suffer from spherical and chromatic aberrations, whose practical consequences we now discuss.

2.3.2 Effect of Lens Aberrations on Spatial Resolution

Because of chromatic aberration, a TEM image cannot be in focus for all energy losses. Most of this aberration occurs at the *objective* lens, where the image-plane angular divergence is higher than in subsequent lenses (Reimer and Kohl, 2008). If the objective (chromatic aberration coefficient C_c, magnification M_o) is focused for zero-loss electrons, an electron with energy loss E and scattering angle θ arrives at the first image plane with a radial displacement $R = M_o\theta\Delta f$ relative to the optic axis, where $\Delta f = C_c(E/E_0)$ and E_0 is the incident energy. Because R is proportional to θ, the Lorentzian distribution of inelastic intensity $dJ/d\Omega$ per unit *solid* angle (Chapter 3) gives rise to a Lorentzian distribution of intensity dJ/dA per unit *area*

in the image plane. The equivalent intensity at the specimen plane, the chromatic *point-spread function* (PSF), is given by

$$\text{PSF} \propto \left(r^2 + r_E^2\right)^{-1} \tag{2.24}$$

Here r is a radial coordinate at the specimen and $r_E = \theta_E \Delta f$, where $\theta_E \approx E/(2E_0)$ is the characteristic angle of inelastic scattering.

In the case of inner-shell energy losses, the values of E and r_E can be large. However, a common procedure is to increase the TEM high voltage (by an amount E_1/e) so that, for some energy loss E_1 within the recorded range, chromatic aberration is zero (since these electrons have the same kinetic energy as the original zero-loss electrons). Chromatic broadening is minimized if E_1 corresponds to the *center* of the recorded range (width Δ), and for parallel-recording spectroscopy the *maximum* broadening (at either end of the range) corresponds to Eq. (2.24) with

$$r_E = \theta_E \Delta f \approx (E/2E_0)(\Delta/2)(C_\text{c}/E_0) \tag{2.25}$$

In the case of energy-filtered (EFTEM) imaging (or serial EELS) with an energy-selecting slit, the electron intensity is summed over the slit width Δ, which is small. Then Eq. (2.24) can be integrated over energies within the slit to give (Egerton and Crozier, 1997)

$$\text{PSF} \propto (1/r)\{\tan^{-1}[(r_\text{c}/r)E/(2\beta E_0)] - \tan^{-1}[E/(2\beta E_0)]\} \tag{2.26}$$

for $|r| < r_\text{c} = (\Delta/2)(\beta C_\text{c}/E_0)$ and zero otherwise. Here, β is the maximum scattering angle contributing to the data, determined by an objective lens aperture. This function (curves in Fig. 2.17) can be integrated over r and the radius r_{50} containing 50% of the electrons is found to be typically four to eight times smaller than the total radius r_c of the chromatic disk; see data points in Fig. 2.17. A computer program is available to evaluate Eq. (2.26) and remove the chromatic spreading by deconvolution (Lozano-Perez and Titchmarsh, 2007). Quantum mechanical imaging theory suggests (Schenner et al., 1995) that the above geometrical optics analysis underestimates the amount of blurring at low chromatic defocus because it does not include inelastic delocalization (Section 3.11).

A similar geometric optics treatment of the effect of objective lens spherical aberration (coefficient C_s) on a core-loss image gives

$$\text{PSF} \propto \left[r^2 + \theta_E^2\, C_\text{s}^{2/3} r^{4/3}\right]^{-1} \tag{2.27}$$

and r_{50} values are typically 2–10% of the total radius $r_\text{s} = C_\text{s}\beta^3$ (Egerton and Crozier, 1997). Spherical and chromatic aberration produce less spatial broadening when spectra or images are recorded from the valence-loss region, where E and θ_E are much smaller. For a TEM in which spherical or chromatic aberration of the imaging lenses are corrected, these sources of broadening would be absent.

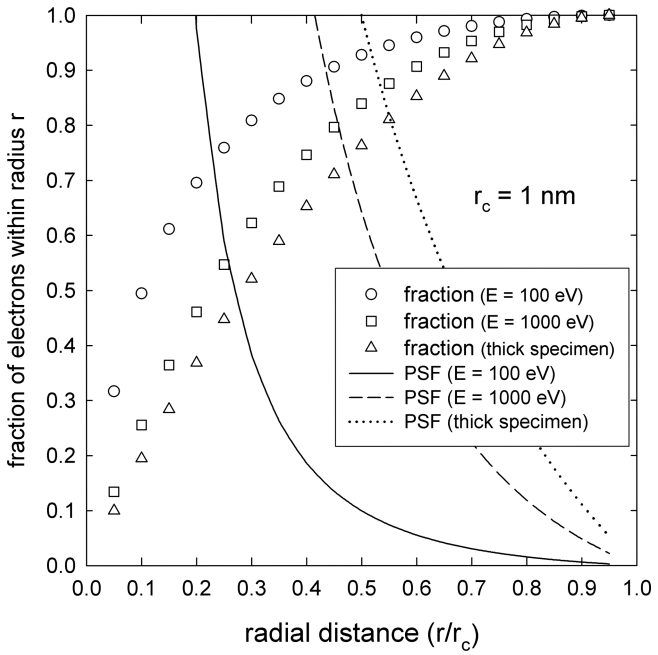

Fig. 2.17 Chromatic aberration point-spread function for a core-loss image of a TEM specimen, evaluated from Eq. (2.26) for $C_c = 2$ mm, $\beta = 10$ mrad, $\Delta = 20$, and $E_0 = 100$ keV. Open data points give the fraction of electron intensity contained within a radius r

Equations (2.26) and (2.27) determine the spatial resolution not only of core-loss EFTEM images but also of energy-loss spectra, if an area-selecting aperture (e.g., spectrometer entrance aperture, for TEM image mode) is used to define the region of analysis. However, this region can instead be defined by the electron beam, as with a finely focused probe. The probe diameter is determined by the electron-source size, diffraction at the aperture of the probe-forming lens, and spherical and chromatic aberration of that lens. The chromatic broadening is

$$\Delta r_c \approx C_c \, \alpha \, \Delta E_0 / E_0 \tag{2.28}$$

where ΔE_0 is the energy width of the illumination, often below 1 eV, and for a strong probe-forming lens ($C_c \approx 2$ mm) and $\alpha \approx 10$ mrad, $\Delta r_c \approx 0.2$ nm, so sub-nanometer probes are entirely practical. A similar argument applies to the resolution of a STEM image.

2.3.3 Effect of Lens Aberrations on Collection Efficiency

When a conventional TEM operates in image mode, lens aberrations produce a blurring of all image features, including the edge of the illumination disk. If the

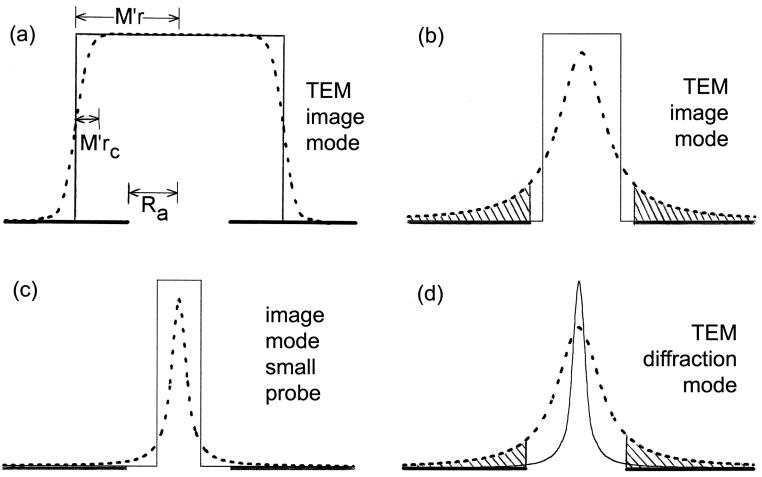

Fig. 2.18 Electron intensity at the plane of the spectrometer entrance aperture (SEA) for microscope image mode (**a**)–(**c**) and diffraction mode (**d**). Chromatic aberration of post-specimen lenses changes each *solid profile* into the *dashed one*. *Shaded areas* represent electrons that are rejected by the SEA as a result of this aberration

diameter of illumination on the TEM screen is much larger than the diameter of the spectrometer entrance aperture, this blurring occurs well outside the SEA perimeter (see Fig. 2.18a) and will not affect the inelastic signal collected by the aperture. Considering only the chromatic aberration (broadening r_c at the specimen plane, as discussed in the last section), this condition requires that

$$M'r > R_a + M'r_c \tag{2.29}$$

where M' is the final magnification at the SEA plane, r is the radius of illumination at the specimen, and R_a is the SEA radius. In other words, the magnified radius of illumination must exceed the SEA radius by an amount at least equal to the chromatic broadening in the image. Provided an objective aperture is used, r_c is normally below 1 μm and Eq. (2.29) can be satisfied by adjusting the condenser lenses so that the radius of illumination (at the TEM screen) is several times the SEA radius. Under these conditions, the loss of electrons (due to chromatic aberration) from points within the selected area is compensated by an equal gain from illuminated regions of specimen outside this area (Titchmarsh and Malis, 1989). However, this compensation is exact only if the current density is uniform within the disk of illumination and if the specimen is uniform in thickness and composition within this region.

If the illumination is focused so that its screen-level radius becomes comparable to that of the spectrometer entrance aperture, part of the aberration tails can be cut off by the aperture (Fig. 2.18b) and the collection efficiency of the spectrometer will be reduced. Since r_c is a function of energy loss, the decrease in collection efficiency

due to chromatic aberration is *energy dependent* and will change the elemental ratio deduced from measurements on two different ionization edges.

If r_c and r are small enough, however, the aberration tails may occur *within* the SEA (Fig. 2.18c) and the collection efficiency is unaffected. The necessary condition is

$$M'r + M'r_c < R_a \tag{2.30}$$

For $R_a = 3$ mm and $r = r_c = 100$ nm, Eq. (2.30) can be satisfied if $M' < 15,000$, assuming that the illumination can be accurately focused into the center of the spectrometer entrance aperture.

Equations (2.29) and (2.30) represent conditions for getting *no* loss of collection efficiency due to chromatic aberration. Because the radius r_{50} containing 50% of electrons is much less than r_c (see Fig. 2.17), the change in intensity and elemental ratio should be small until the aperture radius is reduced to $M'r_c/4$ in most cases.

If the TEM is operated in *diffraction mode*, chromatic aberration could change the distribution of inelastic intensity in the diffraction pattern and therefore the signal collected by the spectrometer entrance aperture. Here the major chromatic effect is likely to arise from microscope *intermediate* lenses but should be significant only for energy losses above 500 eV and analyzed areas (defined by the incident beam or SAD aperture) larger than 3 μm in diameter (Yang and Egerton, 1992). Errors in quantitative analysis should therefore be negligible in the case of sub-micrometer probes (Titchmarsh and Malis, 1989). This conclusion assumes that the imaging lenses are in good alignment, a condition that can be ensured by positioning the illumination (or SAD aperture) so that the voltage center of the *diffraction pattern* coincides with the center of the viewing screen.

If the energy-loss spectrum is acquired by *serial recording*, chromatic aberration effects are avoided by keeping the spectrometer at a fixed excitation and scanning through energy loss by applying a ramp signal to the microscope high-voltage generator. Likewise for *parallel recording*, the microscope voltage can be raised by an amount equal to the energy loss of interest, but chromatic aberration will reduce the collection efficiency for energy losses that differ from this value. In other words, the lens system acts as a bandpass filter and can introduce artifacts in the form of broad peaks in the spectrum (Kruit and Shuman, 1985b), particularly if the optic axis does not coincide with the SEA center (Yang and Egerton, 1992).

2.3.4 Effect of TEM Lenses on Energy Resolution

The resolution in an energy-loss spectrum depends on several factors: the energy spread ΔE_0 of the electrons before they reach the specimen (reflecting the energy width of the electron source and the Boersch effect), broadening ΔE_{so} due to the spectrometer, dependent on its electron optics, and the spatial resolution

s of the electron detector (or slit width for serial recording). Because these components are independent, they can be added in quadrature, giving the measured resolution ΔE as

$$(\Delta E)^2 \approx (\Delta E_0)^2 + (\Delta E_{so})^2 + (s/D)^2 \tag{2.31}$$

where D is the spectrometer dispersion. In general, ΔE_{so} varies with energy loss E, since both the spectrometer focusing and the angular width of inelastic scattering are E-dependent. As a result, the energy resolution at an ionization edge can be worse than that measured at the zero-loss peak.

The image produced by a double-focusing spectrometer is a *convolution* of the energy-loss spectrum with the image or diffraction intensity at the spectrometer object plane (the *mixing* effect referred to in Section 2.3). For an object of width d_o, an ideal spectrometer with magnification M_x (in the direction of dispersion) would produce an image of width $M_x d_o$. But if the spectrometer has aberrations of order n, the image is broadened by an amount $C_n \gamma^n$, where γ is the divergence semi-angle of the beam entering the spectrometer and the aberration coefficient C_n depends on the appropriate nth-order matrix elements (Section 2.2.2). The spectrometer resolution is then given by

$$(\Delta E_{so})^2 \approx (M_x d_o/D)^2 + (C_n \gamma^n/D)^2 \tag{2.32}$$

The values of d_o and γ depend on how the TEM lenses are operated and on the diameter of the spectrometer entrance aperture.

In the absence of an entrance aperture, the product $d_o\gamma$ would be constant, since electron-optical brightness is conserved (Reimer and Kohl, 2008). If the lens conditions are changed so as to reduce d_o and therefore decrease the source size contribution to ΔE_{so}, the value of γ and of the spectrometer aberration term must increase, and vice versa. As a result, there is a particular combination of d_o and γ that minimizes ΔE_{so}. This combination corresponds to optimum L_o or M_o at the spectrometer object plane and to optimum values of M or L at the TEM viewing screen; see Fig. 2.19.

The effect of a spectrometer entrance aperture (radius R_a, distance h' below O) is to limit γ to a value R_a/h', so that the second term in Eq. (2.32) cannot exceed a certain value. If the microscope is operated in image mode with the SEA defining the area of analysis ($M' d/2 > R_a$), $\Delta E_{so} \approx 1$ eV for a spectrometer with second-order aberrations corrected. But at very low screen magnifications or camera lengths, the energy resolution degrades, due to an increase in spectrometer object size; see Fig. 2.19. The only cure for this degradation is to reduce β (by using a small objective aperture) in image mode or to reduce the diameter d of the analyzed region in diffraction mode.

The above analysis neglects aberrations of the TEM lenses themselves, which could affect the energy resolution if a specimen image is present at the spectrometer object plane (Johnson, 1980a; Egerton, 1980a). In serial acquisition, TEM chromatic aberration is avoidable by scanning the high voltage; for parallel recording,

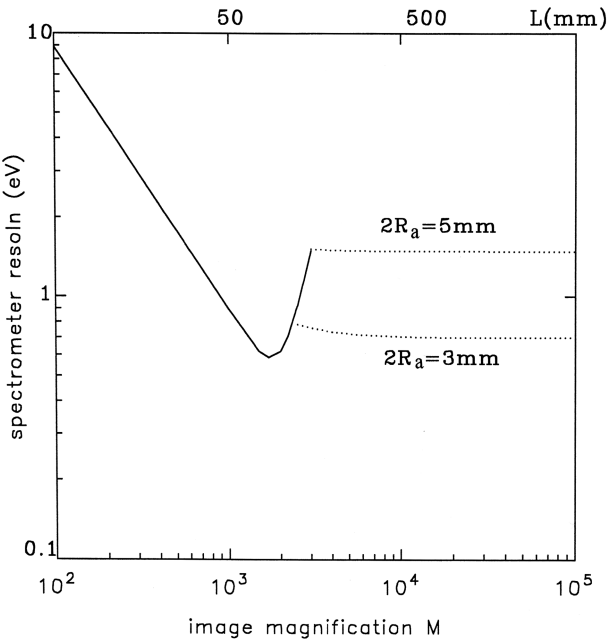

Fig. 2.19 Spectrometer resolution ΔE_{so} as a function of microscope magnification (for image mode) or camera length (for diffraction mode), calculated for an illumination diameter of $d = 1\mu m$ at the specimen, $\beta = 10$ mrad, and a spectrometer with $C_2 = 0$, $C_3 = 50$ m, and $D = 1.8\,\mu m/ev$ at $E_0 = 100$ keV. *Dotted lines* correspond to the situation in which the diameter of illumination (or the central diffraction disk) at the SEA plane exceeds the diameter ($2R_a$) of a 3 mm of a 5-mm spectrometer entrance aperture

it may be possible to bring each energy loss into focus by adjusting multipole lenses associated with the spectrometer (equivalent to adjusting the tilt ψ of the detector plane).

2.3.5 STEM Optics

If energy-loss spectroscopy is carried out in a dedicated scanning transmission electron microscope (STEM), there need to be no imaging lenses between the spectrometer and specimen. However, the specimen is immersed in the field of the probe-forming objective lens, whose post-field reduces the divergence of the electron beam entering the spectrometer by an angular compression factor $M = \beta/\gamma$; see Fig. 2.20a. This post-field creates a *virtual* image of the illuminated area of the specimen, which acts as the object point O for the spectrometer (Fig. 2.10). The spectrometer is therefore *image coupled* with an object-plane magnification equal to the angular compression factor.

Because there are no image-plane (area-selecting) apertures, the spatial resolution is defined by the incident probe diameter d. Although this diameter is affected

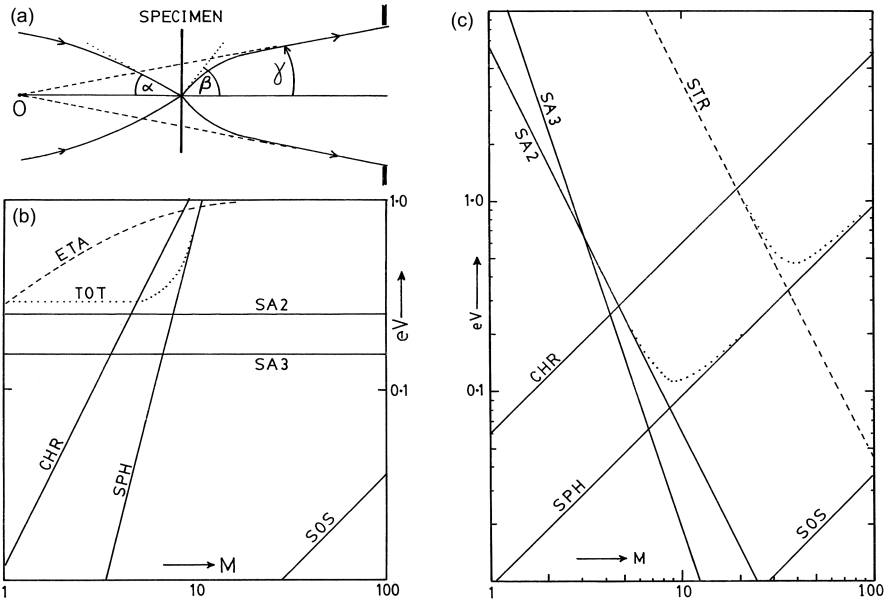

Fig. 2.20 (**a**) Geometry of a focused STEM probe, showing the angular compression ($M = \beta/\gamma$) introduced by the post-field. Point O acts as a virtual object for the spectrometer system. (**b**) Contributions to the energy resolution from spectrometer object size (SOS), from second- and third-order spectrometer aberrations (SA2 and SA3), and from spherical (SPH) and chromatic (CHR) aberrations, calculated as a function of M for $\gamma = 5$ mrad and $E = 400$ eV. Also shown is the signal collection efficiency for $E = 400$ eV, assuming chromatic aberration is corrected by prespectrometer optics. Instrumental parameters are for a VG-HB5 high-excitation polepiece ($C_s' = 1.65$ mm, $C_c' = 2$ mm, $d = 1$ nm, $E_0 = 100$ keV) and an aberration-corrected spectrometer with $C_2 = 2$ cm and $C_3 = 250$ cm, $M_x = 0.73$, $D = 2\,\mu$m/eV (Scheinfein and Isaacson, 1984). (**c**) Contributions to the energy resolution calculated for the case where M and γ are varied to maintain $\beta = 25$ mrad. The *dashed line* shows the resolution of an equivalent straight-edge magnet with uncorrected second-order aberrations ($C_2 = 140$ cm)

by spherical and chromatic aberration of the *pre*field, it is unaffected by aberration coefficients of the post-field; therefore, the *spatial* resolution is independent of energy loss. The focusing power of a magnetic lens is roughly proportional to the reciprocal of the electron kinetic energy (Reimer and Kohl, 2008). Therefore the angular compression of the post-field changes with energy loss, but only by about 1% per 1000 eV (for $E_0 = 100$ keV), so the corresponding variation in collection efficiency is unimportant.

Most STEM instruments now have post-specimen lenses in order to further compress the angular range and provide greater flexibility of operation. Chromatic aberration in these lenses can cause the collection efficiency to increase or decrease with energy loss, creating errors in quantitative analysis (Buggy and Craven, 1981; Craven et al., 1981).

The energy resolution of the spectrometer/post-field combination can be analyzed as for a TEM with image coupling. The contribution $MM_x d/D$, representing

geometric source size, is negligible if the incident beam is fully focused (a field-emission source allows values of d below 1 nm), but it becomes appreciable if the probe is defocused to several hundred nanometers or scanned over a similar distance with no descanning applied. Spherical and chromatic aberrations of the objective post-field (coefficients $C_s{}'$ and $C_c{}'$) broaden the spectrometer source size, resulting in contributions $MM_xC_s'\beta^3/D$ and $MM_xC_c'\beta(E/E_0)/D$, which become significant if M exceeds 10; see Fig. 2.20b. Finally, spectrometer aberration terms, of the form $(C_n/D)(\beta/M)^n$, can be made small by appropriate spectrometer design. Good resolution is possible with high collection efficiency ($\beta = 25$ mrad) if M is of the order of 10; see Fig. 2.20c. Correction of spherical aberration of the *imaging* lenses permits a collection angle of more than 100 mrad (Botton et al., 2010).

Correction of spherical aberration of the *probe-forming* lens has resulted in probe diameters below 0.1 nm but with significant increase in the incident beam convergence, which places extra demands on the performance of the spectrometer and post-specimen optics. Quadrupole and octupole elements have recently been used to correct the remaining aberrations of the Enfina spectrometer, besides allowing the camera length of recorded diffraction patterns to be adjusted (Krivanek et al., 2008).

2.4 Recording the Energy-Loss Spectrum

The energy-loss spectrum is recorded electronically as a sequence of *channels*, the electron intensity in each channel being represented by a number stored in computer memory. Historically, there have been two strategies for converting the intensity distribution into stored numbers.

In *parallel recording*, a position-sensitive electron detector simultaneously records all of the incident electrons, resulting in relatively short recording times and therefore drift and radiation damage to the specimen during spectrum acquisition. The original parallel-recording device was photographic film, whose optical-density distribution (after chemical development) could be digitized in a film scanner. Position-sensitive detectors (based on silicon diode arrays) now provide a more convenient option; the procedures involved are discussed in Section 2.5.

In *serial recording*, the spectrum at the image plane of the electron spectrometer is recorded by scanning it across a narrow slit placed in front of a single-channel electron detector. Because electrons intercepted by the slit are wasted, this method is inefficient, requiring longer recording times to avoid excessive statistical (shot) noise at high energy loss. But because the same detector is used to record all energy-loss channels, serial recording avoids certain artifacts (interchannel coupling and gain variations) that arise in parallel recording and is adequate or even preferable for recording low-loss spectra.

Regardless of the recording system employed, it is often necessary to scan or shift the spectrum to record different ranges of energy loss, and various ways of

doing this are discussed in Section 2.4.1. Electron scattering in front of the detector, leading to a "spectrometer background," is discussed in Section 2.4.2. The technique of coincidence counting, which can reduce the background to core-loss edges, is outlined in Section 2.4.3.

2.4.1 Spectrum Shift and Scanning

Several methods are available for scanning the energy-loss spectrum across an energy-selecting slit (as required for serial recording or EFTEM imaging) or for shifting it relative to the detector (often necessary in parallel recording).

(a) *Ramping the magnet*. The magnetic field in a single-prism spectrometer can be changed by varying the main excitation current or by applying a current ramp to a separate set of window-frame coils. Because the detector is stationary, the recorded electrons always have the same radius of curvature within the spectrometer. From Eq. (2.1), the magnetic induction required to record electrons of energy E_0 is

$$B = \gamma m_0 v/(eR) = (ecR)^{-1} E_0 (1 + 2m_0 c^2/E_0)^{1/2} \tag{2.33}$$

where $m_0 c^2 = 511$ keV is the electron rest energy. Even assuming ideal properties of the magnet ($B \propto$ ramp current), a linear ramp provides an energy axis that is slightly nonlinear due to the square-root term in Eq. (2.33); see Fig. 2.21. The nonlinearity is only about 0.5% over a 1000-eV scan (at $E_0 = 100$ keV) and can be avoided by using a digitally programmed power supply.

(b) *Pre- or post-spectrometer deflection*. Inductance of the magnet windings causes a lag between the change in B and the applied voltage, limiting the scan rate to typically 1 per second. Higher rates, which are convenient for adjusting the

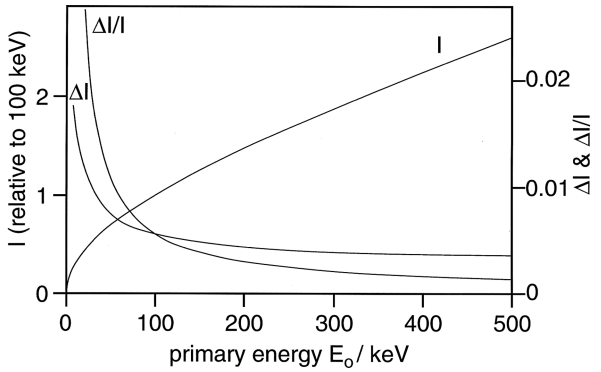

Fig. 2.21 Spectrometer current I (relative to its value for 100-keV electrons) for a fixed bend radius R, as a function of the primary energy E_0. Also shown is the change ΔI and fractional change $\Delta I/I$ in current required to compensate for a 1000-eV change in energy of the detected electrons, from which the incident energy E_0 can be measured. From Meyer et al. (1995), copyright Elsevier

position of the zero-loss peak and setting the width of the detection slit, are achieved by injecting a ramp signal into dipole coils located just before or after the spectrometer. Even faster deflection is possible by using electrostatic deflection plates (Fiori et al., 1980; Craven et al., 2002) and this technique has enabled near-simultaneous recording of low-loss and core-loss spectra on the same CCD camera (Scott et al., 2008; Gubbens et al., 2010).

(c) *Ramping the high voltage*. The spectrum can also be shifted or scanned by applying a signal to the feedback amplifier of the microscope's high-voltage generator, thereby changing the incident electron kinetic energy E_0. The spectrometer and detection slit then act as an energy filter that transmits electrons of a fixed kinetic energy, thereby minimizing the unwanted effects of chromatic aberration in the post-specimen lenses (Section 2.3). The difference in energy of the electrons passing through the *condenser* lenses results in a change in illumination focus, but this effect can be compensated by applying a suitable fraction of the scanning signal to the condenser lens power supply (Wittry et al., 1969; Krivanek et al., 1992).

(d) *Drift-tube scanning*. If the flight tube of a magnetic spectrometer is electrically isolated from ground, applying a voltage to it changes the kinetic energy of the electrons traveling through the magnet and shifts the energy-loss scale. The applied potential also produces a weak electrostatic lens at the entrance and exit of the drift tube, tending to defocus and possibly deflect the spectrum, but these effects appear to be negligible provided the internal diameter of the drift tube and its immediate surroundings are not too small (Batson et al., 1981). In the absence of an electrostatic lens effect, a given voltage applied to the drift tube will displace the energy-loss spectrum by the same number of electron volts, allowing the energy-loss axis to be accurately calibrated. Meyer et al. (1995) have shown that when a known voltage (e.g., 1000 eV) is applied to the drift tube and the zero-loss peak is returned to its original position by changing the magnet current I by an amount ΔI, a measurement of $\Delta I/I$ allows the accelerating voltage to be determined to an accuracy of about 50 V; see Fig. 2.21.

Particularly where long recording times are necessary, it is convenient to add several readouts in computer memory, a technique sometimes called multiscanning or signal averaging. The accumulated data can be regularly displayed, allowing broad features to be discerned after only a few scans, so that the acquisition can be aborted if necessary. Otherwise, the readouts are repeated until the signal/noise ratio (SNR) of the data becomes acceptable. For a given total time T of acquisition, the SNR is similar to that for a spectrum acquired in a single scan of duration T but the effect of instrumental instability is different. Any drift in high voltage or prism current in V_0 can be largely eliminated by shifting individual readouts so that a particular spectral feature (e.g., the zero-loss peak) always occurs at the same spectral channel (Batson et al., 1971; Egerton and Kenway, 1979; Kimoto and Matsui, 2002).

2.4.2 Spectrometer Background

The energy-loss spectrum of a thin specimen covers a large dynamic range, with the zero-loss peak having the highest intensity (e.g., Fig. 1.3). As a result, the zero-loss electrons can produce an observable effect when the zero-loss peak is at some distance from the detector. Stray electrons are generated by backscattering from whatever surface absorbs the zero-loss beam and some of these find their way (by multiple backscattering) to the electron detector and generate a *spectrometer* background that typically varies *slowly* with energy loss and is therefore more noticeable at higher energy loss. In the Gatan parallel-detection system, a beam trap is used to minimize this backscattering, resulting in a background that (in terms of intensity per eV) is over 10^6 times smaller than the integrated zero-loss intensity.

Because most of the stray electrons and x-rays are generated by the zero- and low-loss electrons, an instrumental background similar to that present in a real spectrum can be measured by carrying out serial acquisition with no specimen in the beam. The resulting *background spectrum* enables the spectrometer contribution to be assessed and if necessary subtracted from real spectral data (Craven and Buggy, 1984).

The magnitude of the spectrometer background can also be judged from the *jump ratio* of an ionization edge, defined as the ratio of maximum and minimum intensities just after and just before the edge. A very thin carbon foil (<10 nm for $E_0 = 100$ keV) provides a convenient test sample (Joy and Newbury, 1981); if the spectrum is recorded with a collection semi-angle β less than 10 mrad, a jump ratio of 15 or more at the *K*-edge indicates a low instrumental background (Egerton and Sevely, 1983).

A further test for spectrometer background is to record the *K*-ionization edge of a thin (<50 nm) aluminum or silicon specimen in the usual bright-field condition (collection aperture centered about the optic axis) and in dark field, where the collection aperture is shifted or the incident illumination tilted so that the central undiffracted beam is intercepted by the aperture. In the latter case, stray electrons and x-rays are generated mainly at the collection aperture and have little chance of reaching the detector, particularly if an objective-lens aperture acts as the collection aperture. As a result, the jump ratio of the edge may be higher in dark field, the amount of improvement reflecting the magnitude of the bright-field spectrometer background (Oikawa et al., 1984; Cheng and Egerton, 1985). If the spectrometer background is high, the measured jump ratio may actually increase with increasing specimen thickness (Hosoi et al., 1984; Cheng and Egerton, 1985), up to a thickness at which plural scattering in the specimen imposes an opposite trend (Section 3.5.4).

Although not usually a problem, electron scattering within the microscope column can also contribute to the instrumental background. For example, insertion of an area-selecting aperture has been observed to degrade the jump ratio of an edge, presumably because of scattering from the edge of the aperture (Joy and Maher, 1980a).

2.4.3 Coincidence Counting

If an energy-dispersive x-ray (EDX) detector is operated simultaneously with an energy-loss spectrometer, it is possible to improve the signal/background and (in principle) the signal/noise ratio of an ionization edge. By applying both detector outputs to a gating circuit that gives an output pulse only when an energy-loss electron and a characteristic x-ray photon of the same energy are received within a given time interval, the background to an ionization edge can be largely eliminated (Wittry, 1976). Some "false coincidences" occur, due to x-rays and energy-loss electrons generated in separate scattering events; their rate is proportional to the product of resolution time, x-ray signal, and energy-loss signal. To keep this contribution small, the incident beam current must be kept low, resulting in a low overall count rate. A small contribution from false coincidences can be recognized, since it has the same energy dependence as the energy-loss signal before coincidence gating, and subtracted (Kruit et al., 1984). A peak due to bremsstrahlung loss also appears in the coincidence energy-loss spectrum, at an energy just below the ionization-edge threshold (Fig. 2.22).

Measured coincidence rates have amounted to only a few counts per second (Kruit et al., 1984, Nicholls et al., 1984) but with recent improvements in the

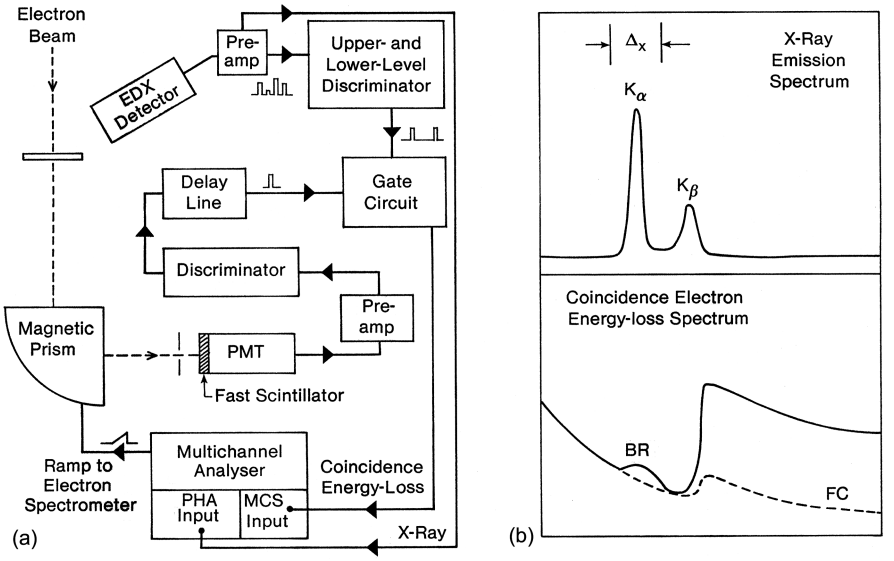

Fig. 2.22 (**a**) Scheme for simultaneous measurement of x-ray emission and coincident energy-loss electrons, here recorded serially. The energy window Δ_x for x-ray gating is selected by a dual-level discriminator. (**b**) The coincidence energy-loss spectrum contains a background FC due to false coincidences and a small peak BR arising from bremsstrahlung energy losses at the x-ray gating window, here chosen to match the lowest energy x-ray peak

collection efficiency of EDX detectors, coincidence counting could become useful and routine. Ideally, the energy-loss spectrum would be recorded with a parallel-detection system operating in an electron counting mode.

Particularly for light element detection, the x-ray detector could be replaced by an Auger electron detector as the source of the gating signal (Wittry, 1980). Alternatively, by using the energy-loss signal for gating, coincidence counting could be used to improve the energy resolution of the Auger spectrum (Cazaux, 1984; Haak et al., 1984). Coincidence between energy-loss events and secondary electron generation has yielded valuable information about the mechanism of secondary electron emission (Pijper and Kruit, 1991; Müllejans et al., 1993; Scheinfein et al., 1993).

2.4.4 Serial Recording of the Energy-Loss Spectrum

A serial recording system contains four components: (1) the detection slit, which selects electrons of a particular energy loss; (2) the electron detector; (3) a method of scanning the loss spectrum across the detection slit; and (4) a means of converting the output of the electron detector into binary numbers for electronic storage. These components are shown schematically in Fig. 2.23 and discussed below.

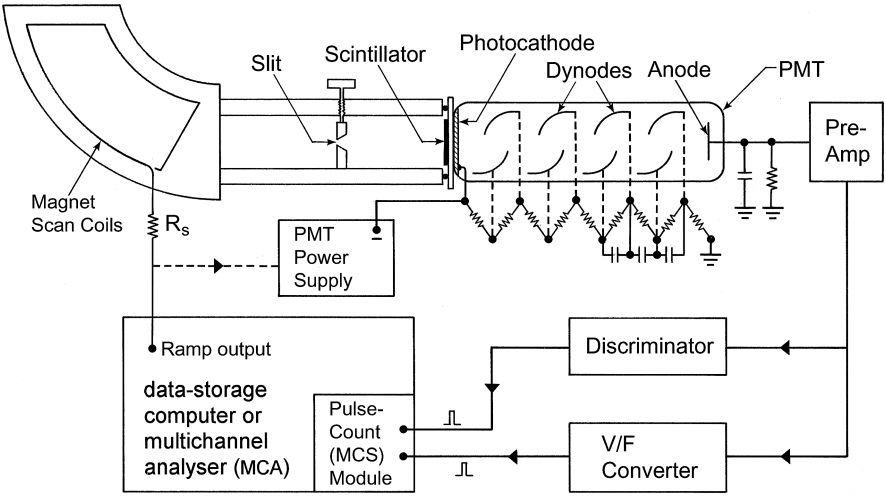

Fig. 2.23 A serial-acquisition system for energy-loss spectroscopy. An energy-selecting slit is located in the spectrometer image plane. Electrons that have passed through the slit cause luminescence in a scintillator; the light produced is turned into an electrical signal and amplified by a photomultiplier tube (PMT). The PMT output is fed into a multichannel scaling (MCS) circuit whose ramp output scans the spectrum across the slit, the value of resistor R_s determining the scan range and electron volt per channel

2.4.4.1 Design of a Detection Slit

Because the energy dispersion of an electron spectrometer is only a few micrometers per electron volt at 100 keV primary energy, the edges of any energy-selecting slit must be smooth (on a μm scale) over the horizontal width of the electron beam at the spectrometer image plane. For a double-focusing spectrometer, this width is in principle very small, but stray magnetic fields can result in appreciable y-deflection and cause a modulation of the detector output if the edges are not sufficiently straight and parallel (Section 2.2.5). The slit blades are usually constructed from materials such as brass or stainless steel, which are readily machinable, although gold-coated glass fiber has also been used (Metherell, 1971).

The slit width s (in the vertical direction) is normally adjustable to suit different circumstances. For examining fine structure present in the loss spectrum, a small value of s ensures good energy resolution of the recorded data. For the measurement of elemental concentrations from ionization edges, a larger value may be needed to obtain adequate signal ($\propto s$) and signal/noise ratio. For large s, the energy resolution becomes approximately s/D, where D is the dispersive power of the spectrometer.

The slit blades are designed so that electrons which strike them produce negligible response from the detector. It is relatively easy to prevent the incident electrons being transmitted through the slit blades, since the stopping distance (electron range) of 100-keV electrons is less than 100 μm for solids with atomic number greater than 14. On the other hand, x-rays generated when an electron is brought to rest are more penetrating; in iron or copper, the attenuation length of 60-keV photons being about 1 mm. Transmitted x-rays that reach the detector give a spurious signal (Kihn et al., 1980) that is independent of the slit opening s. In addition, half of the x-ray photons are generated in the backward direction and a small fraction of these are reflected so that they pass through the slit to the detector (Craven and Buggy, 1984). Even more important, an appreciable number of fast electrons that strike the slit blades are backscattered and after subsequent backreflection may pass through the slit and arrive at the detector. Coefficients of electron backscattering are typically in the range of 0.1–0.6, so the stray-electron signal can be appreciable.

The stray electrons and x-rays are observed as a *spectrometer* background to the energy-loss spectrum, resulting in reduced signal/background and signal/noise ratio at higher energy loss. Most of the energy-loss intensity occurs within the low-loss region, so most of the stray electrons and x-rays are generated close to the point where the zero-loss beam strikes the "lower" slit blade when recording higher energy losses.

Requirements for a low spectrometer background are therefore as follows. The slit material should be conducting (to avoid electrostatic charging in the beam) and thick enough to prevent x-ray penetration. For 100-keV operation, 5-mm thickness of brass or stainless steel appears to be sufficient. The angle of the slit edges should be close to 80° so that the zero-loss beam is absorbed by the full thickness of the slit material when recording energy losses above a few hundred electron volts. The defining edges should be in the same plane so that, when the slit is almost closed to the spectrometer exit beam, there is no oblique path available for scattered electrons

(traveling at some large angle to the optic axis) to reach the detector. The length of the slit in the horizontal (nondispersive) direction should be restricted, to reduce the probability of scattered electrons and x-rays reaching the detector. A length of a few hundred micrometers may be necessary to facilitate alignment and allow for deflection of the beam by stray dc and ac magnetic fields. Since the coefficient η of electron backscattering is a direct function of atomic number, the "lower" slit blade should be coated with a material such as carbon ($\eta \approx 0.05$). The easiest procedure is to "paint" the slit blades with an aqueous suspension of colloidal graphite, whose porous structure helps further in the absorbing scattered electrons. The lower slit blade should be flat within the region over which the zero-loss beam is scanned. Sharp steps or protuberances can give rise to sudden changes in the scattered electron background, which could be mistaken for real spectral features (Joy and Maher, 1980a). For a similar reason, the use of "spray" apertures in front of the energy-selecting slit should be avoided.

Moving the detector further away from the slits and minimizing its exposed area (just sufficient to accommodate the angular divergence of the spectrometer exit beam) decrease the fraction of stray electrons and x-rays that reach the detector (Kihn et al., 1980). Employing a scintillator whose thickness is just sufficient to stop fast electrons (but not hard x-rays) will further reduce the x-ray contribution. Finally, the use of an electron counting technique for the higher energy losses reduces the intensity of the instrumental background because many of the backscattered electrons and x-rays produce output pulses that fall below the threshold level of the discriminator circuit (Kihn et al., 1980).

2.4.4.2 Electron Detectors for Serial Recording

Serial recording has been carried out with solid-state (silicon diode) detectors but counting rates are limited (Kihn et al., 1980) and radiation damage is a potential problem at high doses (Joy, 1984a). Windowless electron multipliers have been used up to 100 Mcps but require excellent vacuum to prevent contamination (Joy, 1984a). Channeltrons give a relatively noisy output for electrons whose energy exceeds 10 keV (Herrmann, 1984). For higher energy electrons, the preferred detector for serial recording consists of a scintillator (emitting visible photons in response to each incident electron) and a photomultiplier tube (PMT), which converts some of these photons into electrical pulses or a continuous output current.

2.4.4.3 Scintillators

Properties of some useful scintillator materials are listed in Table 2.2. Polycrystalline scintillators are usually prepared by sedimentation of a phosphor powder from aqueous solution or an organic liquid such as chloroform, sometimes with an organic or silicate binder to improve the adhesive properties of the layer. Maximizing the signal/noise ratio requires an efficient phosphor of suitable thickness. If the thickness is less than the electron range, some kinetic energy of the incident electron is wasted; if the scintillator is too thick, light may be lost by

Table 2.2 Properties of several scintillator materials[a]

Material	Type	Peak wavelength (nm)	Principal decay constant (ns)	Energy conversion efficiency (%)	Dose for damage (mrad)
NE 102	Plastic	420	2.4, 7	3	1
NE 901	Li glass	395	75	1	10^3
ZnS(Ag)	Polycrystal	450	200	12	
P-47	Polycrystal	400	60	7	10^2
P-46	Polycrystal	550	70	3	$>10^4$
CaF_2(Eu)	Crystal	435	1000	2	10^4
YAG	Crystal	560	80		$>10^4$
YAP	Crystal	380	30	7	$>10^4$

[a]P-46 and P-47 are yttrium aluminum garnet (YAG) and yttrium silicate, respectively, each doped with about 1% of cerium. The data were taken from several sources, including Blasse and Bril (1967), Pawley (1974), Autrata et al. (1983), and Engel et al. (1981). The efficiencies and radiation resistance should be regarded as approximate

absorption or scattering within the scintillator, especially in the case of a transmission screen (Fig. 2.23). The optimum mass thickness for P-47 powder appears to be about 10 mg/cm^2 for 100-keV electrons (Baumann et al., 1981).

To increase the fraction of light entering the photomultiplier and prevent electrostatic charging, the entrance surface of the scintillator is given a reflective metallic coating. Aluminum can be evaporated directly onto the surface of glass, plastic, and single-crystal scintillators. In the case of a powder-layer phosphor, the metal would penetrate between the crystallites and reduce the light output, so an aluminum film is prepared separately and floated onto the phosphor or else the phosphor layer is coated with a smooth layer of plastic (e.g., collodion) before being aluminized.

The efficiency of many scintillators decreases with time as a result of irradiation damage (Table 2.2). This process is particularly rapid in plastics (Oldham et al., 1971), but since the electron penetrates only about 100 μm, the damaged layer can be removed by grinding and polishing. In the case of inorganic crystals, the loss of efficiency is due mainly to darkening of the material (creation of color centers), resulting in absorption of the emitted radiation, and can sometimes be reversed by annealing the crystal (Wiggins, 1978).

The decay time of a scintillator is of particular importance in electron counting. Plastics generally have time constants below 10 ns, allowing pulse counting up to at least 20 MHz. However, most scintillators have several time constants, extending sometimes up to several seconds. By shifting the effective baseline at the discriminator circuit, the longer time constants increase the "dead time" between counted pulses (Craven and Buggy, 1984).

P-46 (cerium-doped $Y_3Al_5O_{12}$) can be grown as a single crystal (Blasse and Bril, 1967; Autrata et al., 1983) and combines high quantum efficiency with excellent radiation resistance. The light output is in the yellow region of the spectrum, which is efficiently detected by most photomultiplier tubes. Electron counting up to a few megahertz is possible with this material (Craven and Buggy, 1984).

2.4.4.4 Photomultiplier Tubes

A photomultiplier tube contains a photocathode (which emits electrons in response to incident photons), several "dynode" electrodes (that accelerate the photoelectrons and increase their number by a process of secondary emission), and an anode that collects the amplified electron current so that it can be fed into a preamplifier; see Fig. 2.23. To produce photoelectrons from *visible* photons, the photocathode must have a low work function and cesium antimonide is a popular choice, although single-crystal semiconductors such as gallium arsenide have also been used.

The spectral response of a PMT depends on the material of the photocathode, its treatment during manufacture, and on the type of glass used in constructing the tube. Both sensitivity and spectral response can change with time as gas is liberated from internal surfaces and becomes adsorbed on the cathode. Photocathodes whose spectral response extends to longer wavelengths tend to have more "dark emission," leading to a higher dark current at the anode and increased output noise. The dark current decreases by typically a factor of 10 when the PMT is cooled from room temperature to –30°C, but is increased if the cathode is exposed to room light (even with no voltages applied to the dynodes) or to strong light from a scintillator and can take several hours to return to its original value.

The dynodes consist of a staggered sequence of electrodes with a secondary electron yield of about 4, giving an overall gain of 10^6 or more if there are 10 electrodes. Gallium phosphide has been used for the first dynode, giving the higher secondary electron yield, improved signal/noise ratio, and easier discrimination against noise pulses in the electron counting mode (Engel et al., 1981).

The PMT anode is usually operated at ground potential, the photocathode being at a negative voltage (typically −700 to −1500 V) and the dynode potentials supplied by a chain of low-noise resistors (Fig. 2.23). For *analog operation*, where the anode signal is treated as a continuous current, the PMT acts as an almost ideal current generator, the negative voltage developed at the anode being proportional to the load resistor and (within the linear region of operation) to the light input. Linearity is said to be within 3% provided the anode current does not exceed one-tenth of that flowing through the dynode resistance chain (Land, 1971). The electron gain can be controlled over a wide range by varying the voltage applied to the tube. Since the gain depends sensitively on this potential, the voltage stability of the power supply needs to be an order of magnitude better than the required stability of the output current.

An electron whose energy is 10 keV or more produces some hundreds of photons within a typical scintillator. Even allowing for light loss before reaching the photocathode, the resulting negative pulse at the anode is well above the PMT noise level, so energy-loss *electrons* can be individually *counted*. The maximum counting frequency is determined by the decay time of the scintillator, the characteristics of the PMT, and the output circuitry. To ensure that the dynode potentials (and secondary electron gain) remain constant during the pulse interval, capacitors are placed across the final dynode resistors (Fig. 2.23). To maximize the pulse amplitude and avoid overlap of output pulses, the anode time constant R_1C_1 must be less than the average

time between output pulses. The capacitance to ground C_l is kept low by locating the preamplifier close to the PMT (Craven and Buggy, 1984).

2.4.5 DQE of a Single-Channel System

In addition to the energy-loss signal (S), the output of an electron detector contains noise (N). The quality of the signal can be expressed in terms of a signal-to-noise ratio: SNR = S/N. However, some of this noise is already present within the electron beam in the form of *shot noise*; if the mean number of electrons recorded in a given time is n, the actual number recorded under the same conditions follows a Poisson distribution whose variance is $m = n$ and whose standard deviation is $\sqrt{m}$, giving an inherent signal/noise ratio: $(\text{SNR})_\text{i} = n/\sqrt{m} = \sqrt{n}$. The noise performance of a *detector* is represented by its detective quantum efficiency (DQE):

$$\text{DQE} \equiv [\text{SNR}/(\text{SNR})_\text{i}]^2 = (\text{SNR})^2/n \tag{2.34}$$

For an "ideal" detector that adds no noise to the signal: $\text{SNR} = (\text{SNR})_\text{i}$, giving DQE = 1. In general, DQE is not constant for a particular detector but depends on the incident electron intensity (Herrmann, 1984).

The measured DQE of a scintillator/PMT detector is typically in the range of 0.5–0.9 for incident electron energies between 20 and 100 keV (Pawley, 1974; Baumann et al., 1981; Comins and Thirlwall, 1981). One reason for DQE < 1 concerns the statistics of photon production within the scintillator and photoelectron generation at the PMT photocathode. Assuming the Poisson statistics, it can be shown that DQE is limited to a value given by

$$\text{DQE} \leq p/(1+p) \tag{2.35}$$

where p is the average number of photoelectrons produced for each incident fast electron (Browne and Ward, 1982). For optimum noise performance, p is kept high by using an efficient scintillator, metallizing its front surface to reduce light losses and providing an efficient light path to the PMT. However, Eq. (2.35) shows that the DQE is only seriously degraded if p falls below about 10.

DQE is also reduced as a result of the statistics of electron multiplication within the PMT and dark emission from the photocathode. These effects are minimized by using a material with a high secondary electron yield (e.g., GaP) for the first dynode and by using pulse counting of the output signal to discriminate against the dark current. In practice, the pulse–height distributions of the noise and signal pulses overlap (Engel et al., 1981), so that even at its correct setting a discriminator rejects a fraction f of the signal pulses, reducing the DQE by the factor $(1 - f)$. The overlap occurs as a result of a high-energy tail in the noise distribution and because some signal pulses (e.g., due to electrons that are backscattered within the scintillator) are weaker than the others. In electron counting mode, the discriminator setting

therefore represents a compromise between loss of signal and increase in detector noise, both of which reduce the DQE.

If the PMT is used in analog mode together with a V/F converter (see Section 2.4.6), DQE will be slightly lower than for pulse counting with the discriminator operating *at its optimum setting*. In the low-loss region of the spectrum, lower DQE is unimportant since the signal/noise ratio is adequate. If an A/D converter used in conjunction with a filter circuit whose time constant is comparable to the dwell time per channel, there is an additional noise component. Besides variation in the number of fast electrons that arrive within a given dwell period, the contribution of a given electron to the sampled signal depends on its time of arrival (Tull, 1968) and the DQE is halved compared to the value obtained using a V/F converter, which integrates the charge pulses without the need of an input filter.

The preceding discussion relates to the DQE of the electron detector alone. When this detector is part of a serial-acquisition system, one can define a detective quantum efficiency $(\mathrm{DQE})_{\mathrm{syst}}$ for the recording system as a whole, taking n in Eq. (2.34) to be the number of electrons analyzed by the spectrometer during the acquisition time, rather than the number that passes through the detection slit. At any instant, the detector samples only those electrons that pass through the slit (width s) and so, evaluated over the entire acquisition, the fraction of analyzed electrons that reach the detector is $s/\Delta x$, where Δx is the image-plane distance over which the spectrometer exit beam is scanned. The overall DQE in serial mode can therefore be written as

$$(\mathrm{DQE})_{\mathrm{syst}} = (s/\Delta x)(\mathrm{DQE})_{\mathrm{detector}} \tag{2.36}$$

The energy resolution ΔE in the recorded data cannot be better than s/D, so Eq. (2.35) can be rewritten in the form

$$(\mathrm{DQE})_{\mathrm{syst}} \leq (\Delta E/E_{\mathrm{scan}})(\mathrm{DQE})_{\mathrm{detector}} \tag{2.37}$$

where E_{scan} is the energy width of the recorded data. Typically, ΔE is in the range 0.2–2 eV while E_{scan} may be in the range 100–5000 eV, so the overall DQE is usually below 1%. In a serial detection system, $(\mathrm{DQE})_{\mathrm{syst}}$ can always be improved by widening the detection slit, but at the expense of degraded energy resolution.

2.4.6 Serial-Mode Signal Processing

We now discuss methods for converting the output of a serial-mode detector into numbers stored in computer memory. The detector is assumed to consist of a scintillator and PMT, although similar principles apply to solid-state detectors.

2.4.6.1 Electron Counting

Photomultiplier tubes have low noise and high sensitivity; some can even count *photons*. Within a suitable scintillator, a high-energy electron produces a rapid burst

of light containing *many* photons, so it is relatively easy to detect and count individual electrons, resulting in a one-to-one relationship between the stored counts per channel and the number of energy-loss electrons reaching the detector. The intensity scale is then linear down to low count rates and the sensitivity of the detector is unaffected by changes in PMT gain arising from tube aging or power-supply fluctuations and should be independent of the incident electron energy.

In order to extend electron counting down to low arrival rates, a *lower level* discriminator is used to eliminate anode signals generated by stray light, low-energy x-rays, or noise sources within the PMT (mainly dark emission from the photocathode). If the PMT has single-photon sensitivity, dark emission produces discrete pulses at the anode, each containing G electrons, where G is the electron gain of the dynode chain. To accurately set the discriminator threshold, it is useful to measure the distribution of pulse amplitudes at the output of the PMT preamplifier, using an instrument with pulse–height analysis (PHA) facilities or a fast oscilloscope. The pulse–height distribution should contain a maximum (at zero or low pulse amplitude) arising from noise and a second maximum that represents signal pulses; the discriminator threshold is placed between the peaks (Engel et al., 1981).

A major limitation of pulse counting is that (owing to the distribution of decay times of the scintillator) the maximum count rate is only a few megahertz for P-46 (Ce-YAG) and of the order of 20 MHz for a plastic scintillator (less if the scintillator has suffered radiation damage), rates that correspond to an electron current below 4 pA. Since the incident beam current is typically in the range of 1 nA to 1 μA, alternative arrangements are usually necessary for recording the low-loss region of the spectrum.

2.4.6.2 Analog/Digital Conversion

At high incident rates, the charge pulses produced at the anode of a PMT merge and the preamplifier output becomes a continuous current or voltage, whose level is related to the electron flux falling on the scintillator. There are two ways of converting this voltage into binary form for digital storage. One is to feed the preamplifier output into a voltage-to-frequency (V/F) converter (Maher et al., 1978; Zubin and Wiggins, 1980). This is essentially a voltage-controlled oscillator; its output consists of a continuous train of pulses that can be counted using the same scaling circuitry as employed for electron counting. The output frequency is proportional to the input voltage between (typically) 10 μV and 10 V, providing excellent linearity over a large dynamic range. V/F converters are available with output rates as high as 20 MHz. Unfortunately the output frequency is slightly temperature dependent, but this drift can be accommodated by providing a "zero-level" frequency-offset control that is adjusted from time to time to keep the output rate down to a few counts per channel with the electron beam turned off. The minimum output rate should not fall to zero, since this condition could change to a lack of response at low electron intensity, resulting in recorded spectra whose channel contents vanish at some value of the energy loss (Joy and Maher, 1980a). Any remaining background within each

spectrum (measured, for example, to the left of the zero-loss peak) is subtracted digitally in computer memory.

An alternative method of digitizing the analog output of a PMT is via an analog-to-digital (A/D) converter. The main disadvantage is limited dynamic range and the fact that, whereas the V/F converter effectively integrates the detector output over the dwell time per data channel, an A/D converter may sample the voltage level only once per channel. To eliminate contributions from high-frequency noise, the PMT output must therefore be smoothed with a time constant approximately equal to the dwell period per channel (Egerton and Kenway, 1979; Zubin and Wiggins, 1980), which requires resetting the filter circuit each time the dwell period is changed. Even if this is done, the smoothing introduces some smearing of the data between adjacent channels. The situation is improved by sampling the data many times per channel and taking an average.

2.5 Parallel Recording of Energy-Loss Data

A parallel-recording system utilizes a *position-sensitive* detector that is exposed to a broad range of energy loss. Because there is no energy-selecting slit, the detective quantum efficiency (DQE) of the recording *system* is the same as that of the *detector*, rather than being limited by Eq. (2.36). As a result, spectra can be recorded in shorter times and with less radiation dose than with serial acquisition, for the same noise content. These advantages are of particular importance for the spectroscopy of ionization edges at high energy loss, where the electron intensity is low.

Photographic film was the earliest parallel-recording medium. With suitable emulsion thickness, the DQE exceeds 0.5 over a limited exposure range (Herrmann, 1984; Zeitler, 1992). Its disadvantages are a limited dynamic range, the need for chemical processing, and the fact that densitometry is required to obtain quantitative data.

Modern systems utilize a silicon diode array, such as a photodiode array (PDA) or charge-coupled diode (CCD) array. These two types differ in their internal mode of operation, but both provide a pulse-train output that can be fed to an electronic data storage system, just as in serial acquisition. A one-dimensional (linear) PDA was used in the original Gatan PEELS spectrometer but subsequent models use two-dimensional (area) CCD arrays. Appropriate readout circuitry allows the same array to be used for recording spectra, TEM images, and diffraction patterns.

2.5.1 Types of Self-Scanning Diode Array

The first parallel-recording detectors to be used with energy-loss spectrometers were linear photodiode arrays, containing typically 1024 silicon diodes. They were later replaced by two-dimensional charge-coupled diode (CCD) arrays, which are also used in astronomy and other optics applications. In the CCD, each diode is initially charged to the same potential and this charge is depleted by electrons and holes

created by the incident photons, in proportion to the local photon intensity. To read out the remaining charge on each diode, charge packets are moved into a transfer buffer and then to an output electrode (Zaluzec and Strauss, 1989; Berger and McMullan, 1989), during which time the electron beam is usually blanked.

The slow-scan arrays used for EELS or TEM recording differ somewhat from those found in TV-rate video cameras. Their pixels are larger, allowing more charge to be stored per pixel and giving a larger dynamic range. The frame-transfer buffer can be eliminated, allowing almost the entire areas of the chip to be used for image recording and therefore a larger number of pixels (at least 1k × 1k is common). Finally, they are designed to operate below room temperature (e.g., –20°C) so that dark current and readout noise are reduced, which also improves the sensitivity and dynamic range.

The Gatan Enfina spectrometer, frequently used with STEM systems, employs a rectangular CCD array (1340 × 100 pixels). The number of pixels in the nondispersive direction, which are summed (binned) during readout, can be chosen to suit the required detector sensitivity, readout time, and dynamic range. Summing all 100 pixels provides the highest sensitivity. The Gatan imaging filter (Section 2.6.1) contains a square CCD array fiber optically coupled to a thin scintillator. This equipment can be used to record either energy-loss spectra or energy-filtered images or diffraction patterns, depending on the settings of the preceding quadrupole lenses.

2.5.2 Indirect Exposure Systems

Diode arrays are designed as light-optical sensors and are used as such, together with a conversion screen (imaging scintillator), in an *indirect exposure* system. Figure 2.24 shows the general design of a system that employs a thin scintillator,

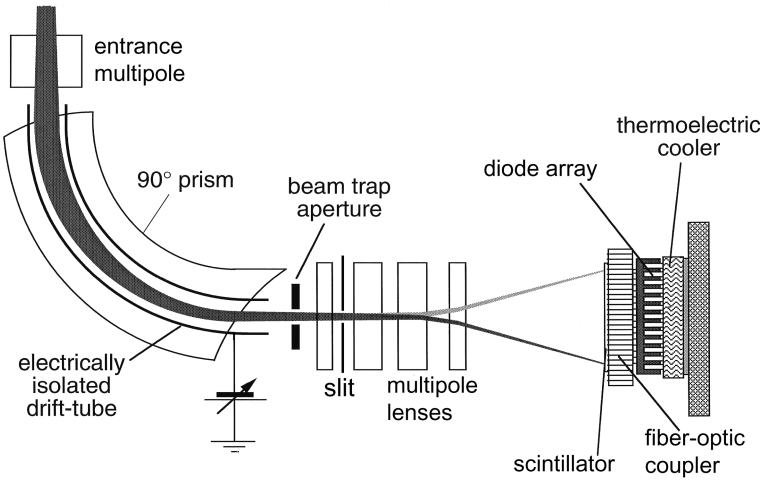

Fig. 2.24 Schematic diagram of a parallel-recording energy-loss spectrometer, courtesy of O. Krivanek

coupled by fiber-optic plate to a thermoelectrically cooled diode array. To provide sufficient energy resolution to examine fine structure in a loss spectrum, the spectrometer dispersion is increased by multipole lenses. The main components of an indirect exposure system will now be discussed in sequence.

2.5.2.1 Magnifying the Dispersion

The spatial resolution (interdiode spacing) of a typical CCD is 14 μm, while the energy dispersion of a compact magnetic spectrometer is only a few micrometers per electron volt. To achieve a resolution of 1 eV or better, it is therefore necessary to magnify the spectrum onto the detector plane. A round lens can be used for this purpose (Johnson et al., 1981b) but it introduces a magnification-dependent rotation of the spectrum unless compensated by a second lens (Shuman and Kruit, 1985).

A magnetic quadrupole lens provides efficient and rotation-free focusing in the vertical (dispersion) plane but does not focus in the horizontal direction, giving a line spectrum. In fact, a line spectrum is preferable because it involves lower current density and less risk of radiation damage to the scintillator. The simplest system consists of a single quadrupole (Egerton and Crozier, 1987) but using several allows the horizontal width to be controlled. Other quadrupole designs (Scott and Craven, 1989; Stobbs and Boothroyd, 1991; McMullan et al., 1992) allow the spectrometer to form crossover at which an energy-selecting slit can be introduced in order to perform energy-filtered imaging in STEM or fixed-beam mode.

Gatan spectrometers use several multipoles, allowing the final dispersion to be varied and (in the GIF system) an energy-filtered image or diffraction pattern to be projected onto the CCD array if an energy-selecting slit is inserted. The image is formed in CTEM mode, without raster scanning of the specimen, using the two-dimensional imaging properties of a magnetic prism (see Section 2.1.2). However, this operation requires corrections for image distortion and aberrations, hence the need for a complicated system of multipole lenses (Fig. 2.25), made possible by computer control of the currents in the individual multipoles. As seen in Fig. 2.26,

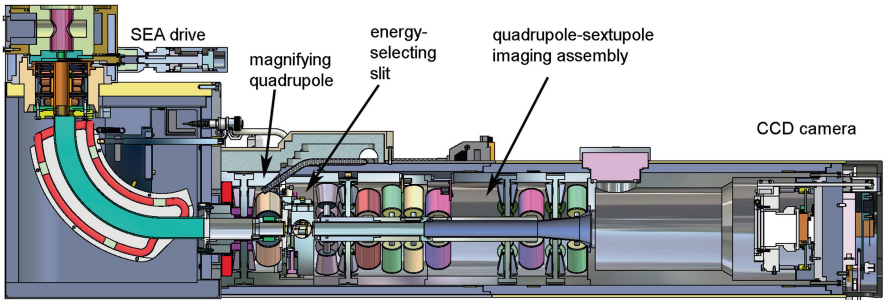

Fig. 2.25 Cross section of a Gatan GIF Tridiem energy-filtering spectrometer (model 863), showing the variable entrance aperture, magnetic prism with window-frame coils, multipole lens system, and CCD camera. Courtesy of Gatan

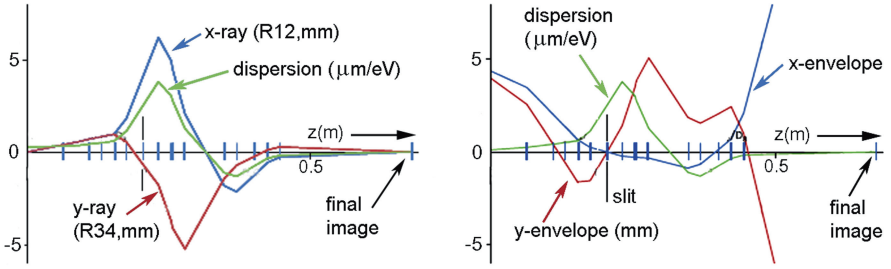

Fig. 2.26 GIF Quantum electron optics, from specimen image inside the magnetic prism ($z = 0$) to final image on the CCD camera ($z = 0.68$ m). *Vertical bars* indicate individual multipole elements and position of the energy-selecting slit. From Gubbens et al. (2010), copyright Elsevier

there is a crossover in both x- and y-directions at the energy-selecting slit (double-focusing condition) and the dispersion is zero at the final image plane (achromatic image of the specimen) or its diffraction pattern.

2.5.2.2 Conversion Screen

The fluorescent screen used in a parallel-recording system performs the same function as the scintillator in a serial-recording system, with similar requirements in terms of sensitivity and radiation resistance, but since it is an *imaging* component, spatial resolution and uniformity are also important. Good resolution is achieved by making the scintillator thin, which reduces lateral spreading of the incident electron beam. Resolution is specified in terms of a point-spread function (PSF), this being the response of the detector to an electron beam of small diameter (less than the interdiode spacing). The modulation transfer function (MTF) is the Fourier transform of the PSF and represents the response of the scintillator to sinusoidally varying illumination of different spatial frequencies.

Uniformity is most easily achieved by use of an amorphous material, such as NE 102 plastic or a single crystal such as CaF_2, NaI, or Ce-doped yttrium aluminum garnet (YAG). Since organic materials and halides suffer radiation damage under a focused electron beam, YAG has been a common conversion screen material in parallel-recording spectrometers (Krivanek et al., 1987; Strauss et al., 1987; Batson, 1988; Yoshida et al., 1991; McMullan et al., 1992) and CCD camera electron-imaging systems (Daberkow et al., 1991; Ishizuka, 1993; Krivanek and Mooney, 1993).

Single-crystal YAG is uniform in its light-emitting properties, emits yellow light to which silicon diode arrays are highly sensitive, and is relatively resistant to radiation damage (see Table 2.2). It can be thinned to below 50 μm and polished by standard petrographic techniques. The YAG can be bonded directly to a fiber-optic plate, using a material of high refractive index to ensure good transmission of light in the forward direction. Even so, some light is multiply reflected between the two surfaces of the YAG and may travel some distance before entering the array, giving rise to extended tails on the point-spread function; see Fig. 2.27a. These tails are

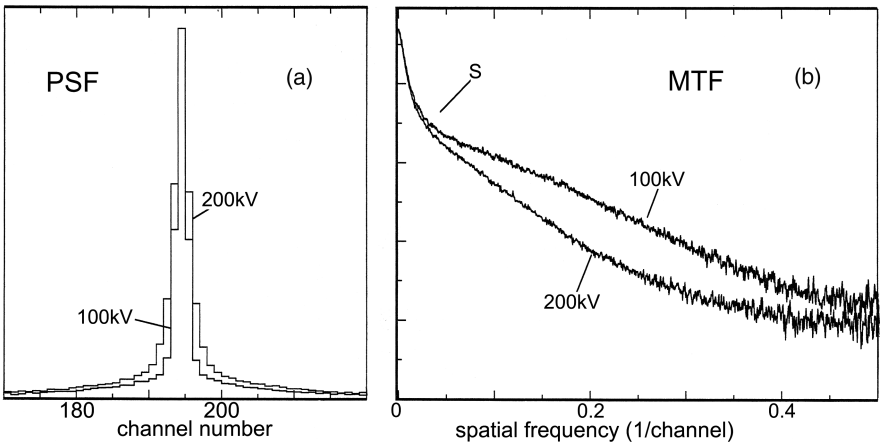

Fig. 2.27 (**a**) Point-spread function for a photodiode array, showing the narrow central peak and extended tails. (**b**) Modulation transfer function, evaluated as the square root of the PSF power spectrum; the rapid fall to the shoulder *S* arises from the PSF tails. From Egerton et al. (1993), copyright Elsevier

reduced by incorporating an antireflection coating between the front face of the YAG and its conducting coating. Long-range tails on the PSF are indicated by the low-frequency behavior of the MTF; see Fig. 2.27b. Many measurements of MTF and DQE have been made on CCD-imaging systems (Kujiwa and Krahl, 1992; Krivanek and Mooney, 1993; Zuo, 2000; Meyer and Kirkland, 2000; Thust, 2009; McMullan et al., 2009a; Riegler and Kothleitner, 2010). The noise properties and DQE of a parallel-recording system are discussed further in Section 2.5.4.

Powder phosphors can be more efficient than YAG, and light scattering at grain boundaries reduces the multiple internal reflection that gives rise to the tails on the point-spread function. Variations in light output between individual grains add to the fixed-pattern noise of the detection system (Daberkow et al., 1991), which is usually removed by computer processing.

Backscattering of electrons from the conversion screen may reduce the DQE of the system and subsequent scattering of these electrons back to the scintillator adds to the *spectrometer* background (Section 2.4.2).

2.5.2.3 Light Coupling to the Array

A convenient means of transferring the conversion-screen image to the diode array is by imaging (coherent) fiberoptics. The resulting optical system requires no focusing, has a good efficiency of light transmission, has no field aberrations (e.g., distortion), and is compact and rigid, minimizing the sensitivity to mechanical vibration. The fiber-optic plate can be bonded with transparent adhesive to the scintillator and with silicone oil to the diode array; minimizing the differences in refractive index reduces light loss by internal reflection at each interface.

Fiber-optic coupling is less satisfactory for electrons of higher energy (>200 keV), some of which penetrate the scintillator and cause radiation damage (darkening) of the fibers or generate x-rays that could damage a nearby diode array. Some electrons are backscattered from the fiber plate, causing light emission into adjacent diodes and thereby augmenting the tails on the response function (Gubbens et al., 1991). These problems are avoidable by using a self-supporting scintillator and glass lenses to transfer the image from the scintillator to the array. Lens optics allows the sensitivity of the detector to be varied (by means of an aperture stop) and makes it easier to introduce magnification or demagnification, so that the resolution of the conversion screen and the detector can be matched in order to optimize the energy resolution and DQE (Batson, 1988). However, the light coupling is less efficient, resulting in decreased noise performance of the system.

2.5.3 Direct Exposure Systems

Although diode arrays are designed to detect visible photons, they also respond to charged particles such as electrons. A single 100-keV electron generates about 27,000 electron–hole pairs in silicon, well above CCD readout noise, allowing a directly exposed array to achieve high DQE at low electron intensities. At *very* low intensity (less than one electron/diode within the integration period) there is the possibility of operation in an electron counting mode.

This high sensitivity can be a disadvantage, since the saturation charge of even a large-aperture photodiode array is equivalent to only a few hundred directly incident electrons, giving a dynamic range of $\approx 10^2$ for a single readout. However, the sensitivity can be reduced by shortening the integration time and accumulating a large number of readouts, thereby increasing the dynamic range (Egerton, 1984). But to record the *entire* spectrum with a reasonable incident beam current (>1 pA), some form of dual system is needed, either using serial recording to record the low-loss region (Bourdillon and Stobbs, 1986) or using fast beam switching and a dual integration time on a CCD array (Gubbens et al., 2010).

Direct exposure involves some risk of radiation damage to the diode array. To prevent rapid damage to field-effect transistors located along the edge of a photodiode array, Jones et al. (1982) masked this area from the beam. Even then, radiation damage can cause a gradual increase in dark current, resulting in increased diode shot noise and reduced dynamic range (Shuman, 1981). The damage mechanism is believed to involve creation of electron–hole pairs within the SiO_2 passivating layer covering the diodes (Snow et al., 1967) and has been reported to be higher at 20-keV incident energy compared to 100 keV (Roberts et al., 1982). When bias voltages are removed, the device may recover, especially if the electron beam is left on (Egerton and Cheng, 1982).

Since the dark current diminishes with decreasing temperature, cooling the array reduces the symptoms of electron-beam damage. Measurements on a photodiode array cooled to $-30°C$ suggested an operating lifetime of at least 1000 h (Egerton and Cheng, 1982). Jones et al. (1982) reported no observable degradation for an

array kept at −40°C, provided the current density was below 10 μA/m^2. Diode arrays can be operated at temperatures as low as −150°C for low-level photon detection (Vogt et al., 1978).

Irradiation of a chemically thinned device from its back surface was proposed long ago as a way of avoiding oxide-charge buildup (Imura et al., 1971; Hier et al., 1979). More recently, electron counting and event-processed modes are being explored in connection with direct recording of low-intensity electron images on thinned CMOS devices (Vos et al., 2009; McMullan et al., 2009b) so there appears to be renewed hope for the direct-recording concept.

2.5.4 DQE of a Parallel-Recording System

Detective quantum efficiency (DQE) is a measure of the quality of a recording system: how little noise it adds to the electron image. In the case of a parallel-recording system, DQE is usually taken to be a function of spatial frequency in the recorded image. Here we present a simplified analysis in which DQE is represented as a single number, together with an interchannel mixing parameter that accounts for the width of the point-spread function (PSF).

Consider a one- or two-dimensional array that is uniformly irradiated by fast electrons, N being the average number recorded by each element during the integration period. Random fluctuations (electron-beam shot noise) contribute a root-mean-square (rms) variation between channels of magnitude $N^{1/2}$, according to the Poisson statistics. The PSF of the detector may be wider than the interdiode spacing, so electrons arriving at a given location are spread over several diode channels, reducing the recorded electron-beam shot noise to

$$N_\mathrm{b} = N^{1/2}/s \tag{2.38}$$

where s is a smoothing (or mixing) factor that can be determined experimentally (Yoshida et al., 1991; Ishizuka, 1993; Egerton et al., 1993). It represents degradation of the spatial or energy resolution resulting from light spreading in the scintillator and any interchannel coupling in the array.

In the case of *indirect* recording, N fast electrons generate (on average) N_p photons in the scintillator. If all electrons followed the same path, the statistical variation in the number of photons produced would be $(N_\mathrm{p})^{1/2}$, but in practice, each electron behaves differently. For example, some penetrate only a short distance before being backscattered and produce significantly fewer photons. Allowing for channel mixing and dividing by p so that the photon noise component N_p is expressed in units of fast electrons, we must therefore write

$$N_\mathrm{p} = s^{-1}N^{1/2}(\sigma_p/p) \tag{2.39}$$

where σ_p is the actual root-mean-square (rms) variation in light output. Monte Carlo simulations and measurements of the height distribution of photon pulses have given $\sigma_\mathrm{p}/p \approx 0.31$ for a YAG scintillator that is thick enough to absorb the incident

electrons, and $\sigma_p/p \approx 0.59$ for a 50-μm YAG scintillator exposed to 200-keV electrons (Daberkow et al., 1991).

Each photon produces an electron–hole pair in the diode array, but random variation (shot noise) in the diode leakage current and electronic noise (whose components may include switching noise, noise on supply and ground lines, video-amplifier noise, and digitization error of an A/D converter) add a total readout noise N_r, expressed here in terms of fast electrons. It is also possible to include a fixed-pattern term N_v, representing the rms fractional variation v in gain between individual diode channels, which arises from differences in sensitivity between individual diodes, nonuniformities of the optical coupling, and variations in sensitivity of the scintillator. Adding all noise components in quadrature, the total noise N_t is given by

$$N_t^2 = N_b^2 + N_p^2 + N_r^2 + N_v^2 \tag{2.40}$$

From Eqs. (2.38), (2.39), and (2.40), the signal/noise ratio (SNR) of the diode array output is

$$\text{SNR} = N/N_t = N(N/s^2 + N\sigma_p^2 p^{-2} s^{-2} + N_r^2 + v^2 N^2)^{1/2} \tag{2.41}$$

SNR increases with signal level N, tending asymptotically to $1/v$. As an illustration, measurements on a PEELS detector based on a Hamamatsu S2304 photodiode array gave $N_r = 60$ and a limiting SNR of 440, implying $v = 0.23\%$ (Egerton et al., 1993). The shape of the SNR versus N_t curve could be fitted with $s = 5$, as shown by the solid curve in Fig. 2.28a.

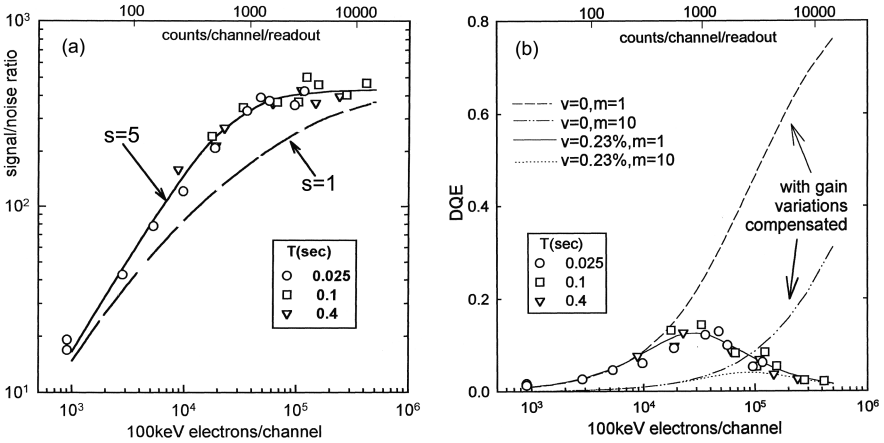

Fig. 2.28 (**a**) SNR and (**b**) DQE for a PEELS detector, as a function of signal level (up to saturation). The experimental data are for three different integration times and are matched by Eqs. (2.41) and (2.44) with $s = 5$. The parameter m represents the number of readouts. From Egerton et al. (1993), copyright Elsevier

As in the case of a single-channel system, the quality of the detector is represented by its DQE, which can be defined as

$$\mathrm{DQE} = [(\mathrm{SNR})/(\mathrm{SNR})_\mathrm{i}]^2 \tag{2.42}$$

According to *this* definition, $(\mathrm{SNR})_\mathrm{i}$ is the signal/noise ratio of an ideal detector that has the *same energy resolution* (similar PSF) but with $N_\mathrm{p} = N_\mathrm{r} = N_\mathrm{v} = 0$, so that

$$(\mathrm{SNR})_\mathrm{i} = N/N_\mathrm{b} = sN^{1/2} \tag{2.43}$$

Making use of Eqs. (2.41), (2.42), and (2.43), the detective quantum efficiency is

$$\mathrm{DQE} = s^{-2}(SNR)^2/N = (1 + \sigma_\mathrm{p}^2/p^2 + s^2N_\mathrm{r}^2/N + v^2s^2N)^{-1} \tag{2.44}$$

Electron-beam shot noise is represented by the first term in parentheses in Eq. (2.44); the other terms cause DQE to be less than 1. For low N, the third term becomes large and DQE is reduced by readout noise. At large N, the fourth term predominates and DQE is reduced as a result of gain variations. Between these extremes, DQE reaches a maximum, as illustrated in Fig. 2.28b. Note that if the smoothing effect of the PSF is ignored, the *apparent* DQE, if defined as $(\mathrm{SNR})^2/N$, can exceed unity.

Provided the gain variations are reproducible, they constitute a fixed pattern that can be removed by signal processing (Section 2.5.5), making the last term in Eq. (2.44) zero. The detective quantum efficiency then increases to a limiting value at large N, equal to $(1 + \sigma_\mathrm{p}^2/p^2)^{-1} \approx 0.9$ for a 50-μm YAG scintillator (thick enough to absorb 100-keV electrons), and this high DQE allows the detection of small fractional changes in electron intensity, corresponding to elemental concentrations below 0.1 at.% (Leapman and Newbury, 1993).

2.5.4.1 Multiple Readouts

If a given recording time T is split into m periods, by accumulating m readouts in computer memory, the electron-beam and diode array shot noise components are in principle unaltered, since they depend only on the total integration time. The noise due to gain variations, which depends on the total recorded signal, should also be the same. But if readout noise (exclusive of diode shot noise) adds independently at each readout, its total is augmented by a factor of $m^{1/2}$, increasing the $N_\mathrm{r}{}^2$ term in Eq. (2.44) by a factor of m. As a result, the DQE will be lower, as shown for a photodiode array in Fig. 2.28b.

Modern CCD arrays have a low readout noise: 50 diode electrons or less for the array used in a GIF Quantum spectrometer (Gubbens et al., 2010). However, an array cannot be read out instantaneously; a readout time of 115 ms has been quoted for an Enfina spectrometer (Bosman and Keast, 2008). During this time, the beam in a Gatan system is usually blanked but the specimen is still irradiated, possibly undergoing radiation damage. So for a given specimen irradiation time,

increasing number of readouts reduces the spectrum-recording time, lowering the SNR and DQE. If beam blanking is not applied, gain variations between different spectral channels are increased; however, the situation can be improved by binned gain averaging (Bosman and Keast, 2008); see Section 2.5.5.

One advantage of multiple readouts is increased dynamic range. The *minimum* signal that can be detected is $S_{\min} = FN_t$, where N_t is the total noise and F is the minimum acceptable signal/noise ratio, often taken as 5 (Rose, 1970). Since some of the noise components are not increased by multiple readouts, the total noise increases by a factor *less* than $m^{1/2}$. The *maximum* signal that can be recorded in m readouts is

$$S_{\max} = m(Q_{sat} - I_d T/m) = mQ_{\text{sat}} - I_d T \tag{2.45}$$

where Q_{sat} is the diode-saturation charge and I_d is the diode thermal-leakage current. $S_{\max}$ is increased by more than a factor of m compared to the largest signal ($Q_{\text{sat}} - I_d T$) that can be recorded in the same time with a single readout. Therefore the dynamic range $S_{\max}/S_{\min}$ of the detector is increased by *more than* a factor of $m^{1/2}$ by use of multiple readouts. A further advantage of multiple readouts is that spectrum drift can be compensated inside the data-recording system.

The choice of m is therefore a compromise. To obtain adequate dynamic range with the least penalty in terms of readout noise, the integration time per readout should be adjusted so that the array output almost saturates at each readout. This procedure minimizes the number of readouts in a given specimen irradiation time, optimizing the signal/noise ratio and minimizing any radiation damage to the specimen.

2.5.5 Dealing with Diode Array Artifacts

The extended tails of the detector point-spread function distort all spectral features recorded with a diode array. Since the tails contain only low spatial frequencies, they can be largely removed by Fourier ratio deconvolution, taking the zero-loss peak as representing the detector PSF (Mooney et al., 1993). To avoid noise amplification, the central (Gaussian) portion of the zero-loss peak can be used as a reconvolution function; see Section 4.1.2.

Some kinds of diode array suffer from incomplete discharge: each readout contains a partial memory of previous ones. Under these conditions, it is advisable to discard several readouts if acquisition conditions, such as the beam current or the region of specimen analyzed, are suddenly changed. Longer term memory effects arise from electrostatic charging within the scintillator, resulting in a local increase in sensitivity in regions of high electron intensity. This condition is alleviated by prolonged exposure to a broad, undispersed electron beam.

The thermal leakage current varies slightly between individual diodes, so the dark current background is not quite constant across a spectrum. This effect is usually removed by recording the array output with the electron beam blanked off. Even so,

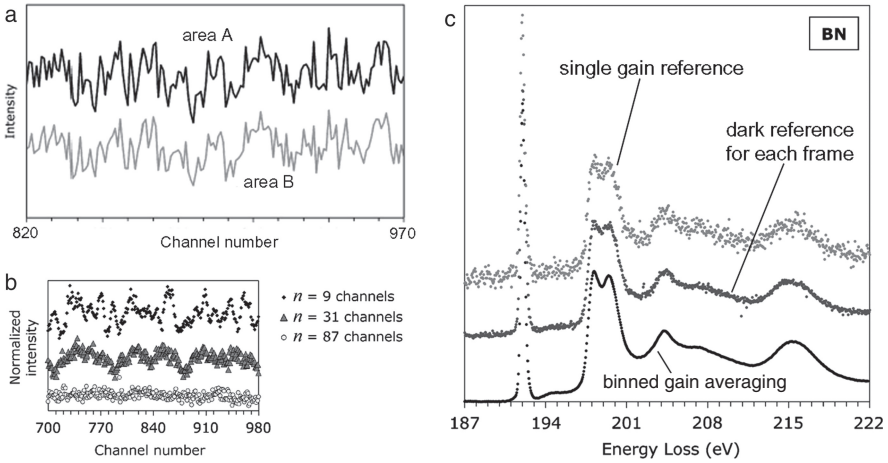

Fig. 2.29 (**a**) Summations of 170 spectra from two areas of the field of view (A and B) with no specimen present, showing the high degree of correlation of the noise. (**b**) Result of gain averaging 2025 spectra with different amounts *n* of spectral shift. (**c**) Boron nitride *K*-edge spectra acquired with no CCD binning and a single gain reference or a separate gain reference for each readout, compared with the result of binned gain averaging (*lower curve*). From Bosman and Keast (2008), copyright Elsevier

artifacts remain because the sensitivity of each diode varies slightly and (for indirect recording) the phosphor screen may be nonuniform in its response. The result is a correlated or *fixed-pattern* noise, the same in each readout, which may considerably exceed the electron-beam shot noise; see Fig. 2.29a.

2.5.5.1 Gain Normalization

Variations in sensitivity (gain) across the array are most conveniently dealt with by dividing each spectral readout by the response of the array when illuminated by a broad undispersed beam. This *gain normalization* or *flat-fielding* procedure can achieve a precision of better than 1% for two-dimensional (CCD) detectors (Krivanek et al., 1995) and it therefore works well enough in many cases. However, other methods, originally developed to correct for gain variations in linear photodiode arrays, may be necessary when weak fine-structural features must be extracted from EELS data.

2.5.5.2 Gain Averaging

First applied to low-energy HREELS (Hicks et al., 1980), this method of eliminating the effect of sensitivity variations involves scanning a spectrum along the array in one-channel increments and performing a readout at each scan voltage. The resulting spectra are electronically shifted into register before being added together in data memory. Within the range of energy loss sampled by *all* diodes of the array, the

effect of gain variations should exactly cancel. If the array contains N elements and gain-corrected data are required in M channels, $N+M$ spectra must be accumulated. Because M may be greater than N, the recorded spectral range can be greater than that covered by the array. However, some electrons fall outside the array and are not recorded, so the system DQE is reduced by a factor of $N/(N+M)$ compared to parallel recording of $N+M$ spectra which are stationary on the array. Batson (1988) implemented this method for a 512-channel photodiode array with M up to 300, substantial computing being required to correct for the fact that the scan step was not equal to the interdiode spacing. The same procedure has been used with a smaller number of readouts ($M < N$), in which case the gain variations are reduced but not eliminated.

2.5.5.3 Binned Gain Averaging

Bosman and Keast (2008) reported a variant of the above procedure, in which spectra are recorded from a two-dimensional CCD detector while a linear ramp voltage is applied to the spectrometer drift tube. For fast readout, the spectra are binned, i.e., all CCD elements corresponding to a given spectral channel are combined in an on-chip register during readout, with no dark current or gain corrections applied. Using the same readout settings, another series of N spectra is acquired without illuminating the detector, all of which are summed and divided by N to give an average spectrum that is subtracted from each individual spectrum of the first data set. This procedure ensures a dark reference with high SNR (including any difference in efficiency between the quadrants, in a four-quadrant detector). Finally, the dark-corrected spectra from the first data set are aligned (by Gaussian fitting to a prominent spectral feature or by using the drift-tube voltage as the required shift) and summed, or used for EELS mapping. The dynamic range of each readout is limited by the capacity of the register pixels but increases when the readouts are combined. The result can be a dramatic reduction in the noise content of a spectrum; see Fig. 2.29c.

2.5.5.4 Iterative Gain Averaging

Another possibility is to record M spectra, each of the form $J_m(E)G(E)$, with successive energy shifts Δ between each spectrum, then electronically shift them back into register and add them to give a single spectrum:

$$\bar{S}_1(E) = \frac{1}{M}\sum_{m=0}^{m=M} J_m(E)G(E - m\Delta) \tag{2.46}$$

where the gain variations are spread out and reduced in amplitude by a factor $\approx M^{1/2}$. Each original spectrum is then divided by $\bar{S}_1(E)$ and the result averaged over all M spectra

$$G_1(E) = \frac{1}{M} \sum_m \frac{S_m(E)}{\bar{S}_1(E)} = \sum_m \frac{J_m(E)\, G(E)}{\bar{S}_1(E)} \tag{2.47}$$

to give a first estimate $G_1(E)$ of the gain profile of the detector so that each original spectrum can be corrected for gain variations:

$$S_m^1(E) = \frac{S_m(E)}{G_1(E)} = \frac{J_m(E)\, G(E)}{G_1(E)} \tag{2.48}$$

The process is then repeated, with the M spectra $S_m(E)$ replaced by $S_m^1(E)$, to obtain revised data $\bar{S}_2(E)$, $G_2(E)$, and $S_m^2(E)$, and this procedure repeated until the effect of gain variations becomes negligible. Schattschneider and Jonas (1993) analyzed this method in detail, showing that the variance due to gain fluctuations is inversely proportional to the square root of the number of iterations.

2.5.5.5 Energy-Difference Spectra

By recording two or three spectra, $J_1(E - \varepsilon)$, $J_2(E)$, and $J_3(E + \varepsilon)$, displaced in energy by applying a small voltage ε to the spectrometer drift tube, first-difference FD(E) or second-difference SD(E) spectra can be computed as

$$\mathrm{FD}(E) = J_1(E - \varepsilon) - J_2(E) \tag{2.49}$$

$$\mathrm{SD}(E) = J_1(E - \varepsilon) - 2J_2(E) + J_3(E + \varepsilon) \tag{2.50}$$

Writing the original spectrum as $J(E) = G(E)[A + BE + C(E)]$, where $G(E)$ represents gain modulation by the detector, gives (for small ε)

$$\begin{aligned} \mathrm{FD}(E) &= G(E)[-B\varepsilon + C(E - \varepsilon) - C(E)] \\ &\approx G(E)[-B\varepsilon + \varepsilon^{-1}(dC/dE)] \end{aligned} \tag{2.51}$$

$$\begin{aligned} \mathrm{SD}(E) &= G(E)[C(E - \varepsilon) - 2C(E) + C(E + \varepsilon)] \\ &\approx G(E)[\varepsilon^{-2}(d^2C/dE^2)] \end{aligned} \tag{2.52}$$

Because component A is absent from FD(E), gain modulation of any constant background is removed when forming a first-difference spectrum (Shuman and Kruit, 1985). Likewise, gain modulation of any linearly varying "background" component is removed when forming SD(E). Consequently, the signal/background ratio of genuine spectral features is enhanced. The resulting spectra, which resemble first and second derivatives of the original data, are highly sensitive to spectral fine structure, making quantitative treatment of the data more difficult. FD(E) can be integrated digitally, the integration constant A being estimated by matching to the original spectra. The *statistical* noise content (excluding gain variations) in the jth channel

of the integrated spectrum $J(j)$ is increased relative to a directly acquired spectrum by a factor of $[2(j-1)]^{1/2}$ for $\varepsilon = 1$ channel.

2.5.5.6 Dynamic Calibration Method

Shuman and Kruit (1985) proposed a gain-correction procedure in which the gain $G(j)$ of the jth spectral channel is calculated from two difference spectra, $J_1(i)$ and $J_2(i)$, shifted by one channel:

$$\frac{G(j)}{G(1)} = \prod_{i=2}^{i=j} J_1(i) \Bigg/ \prod_{i=1}^{i=j-1} J_2(i) \tag{2.53}$$

where Π represents a product of the contents $J(i)$ of all channels between the stated limits of i. By multiplying one of the original spectra by $G(j)$, the effect of gain variations is removed, although the random-noise content of $J(j)$ is increased by a factor of $[2(j-1)]^{1/2}$.

2.6 Energy-Selected Imaging (ESI)

As discussed in Section 2.1, the information carried by inelastic scattering can be acquired and displayed in several ways, one of these being the energy-loss spectrum. In a transmission electron microscope, the volume of material giving rise to the spectrum can be made very small by concentrating the incident electrons into a small-diameter probe. EELS can be used to quantitatively analyze this small volume, the TEM image or diffraction pattern being used to define it relative to its surroundings (Fig. 2.30a). This process can then be repeated for other regions of the specimen, either manually or by raster scanning a small probe across the specimen and collecting a spectrum from every pixel, as in STEM spectrum imaging; see Fig. 2.30c. However, it is sometimes useful to display some spectral feature, such as the ionization edge representing a single chemical element, simultaneously, using the imaging or diffraction capabilities of a conventional TEM. An image-forming spectrometer is used as a filter that accepts energy losses within a specified range, giving an energy-selected image or diffraction pattern. We now discuss several instrumental arrangements that achieve this capability.

2.6.1 Post-column Energy Filter

As discussed in Section 2.1.1, the magnetic prism behaves somewhat like an electron lens, producing a chromatically dispersed image of the spectrometer object O at the plane of an energy-selecting slit or diode array detector. A conventional TEM not only provides a suitable small-diameter object at its projector lens crossover but also produces a magnified image or diffraction pattern of the specimen at the level of

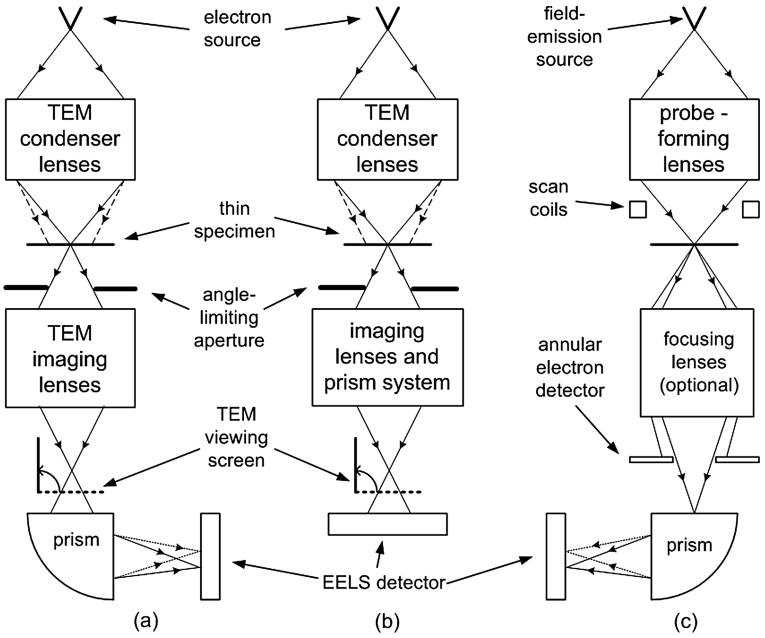

Fig. 2.30 Three common types of instrument used for EELS and energy-filtered imaging: (**a**) imaging filter below the TEM column, (**b**) in-column filter, and (**c**) STEM system. In an instrument dedicated to STEM, the arrangement is often inverted, with the electron source at the bottom

the viewing screen, closer to the spectrometer entrance. The spectrometer therefore forms a second image, further from its exit, which corresponds to a magnified view of the specimen or its diffraction pattern. This second image will be *energy filtered* if an energy-selecting slit is inserted at the plane of the first image (the energy-loss spectrum). Slit design is discussed in Section 2.4.4.1.

In general, the energy-filtered image suffers from several defects. Its magnification is different in the *x*- and *y*-directions, although such rectangular distortion can easily be corrected electronically if the image is recorded into a computer. It exhibits axial astigmatism (the spectrometer is double focusing only at the spectrum plane), which is correctable by using the dipole-coil stigmators (Shuman and Somlyo, 1982). More problematic, it may be a *chromatic* image, blurred in the *x*-direction by an amount dependent on the width of the energy-selecting slit. Even so, by using a 20-μm slit (giving an energy resolution of 5 eV) and a single-prism spectrometer of conventional design, Shuman and Somlyo (1982) obtained an image with a spatial resolution of 1.5 nm over a 2-μm field of view at the specimen plane.

The Gatan imaging filter (GIF) uses post-spectrometer multipoles to correct for these image defects and form a good-quality image or diffraction pattern of the specimen, which is recorded by a CCD camera. Alternatively, the energy-selecting slit can be withdrawn and the GIF lens system used to project an energy-loss spectrum, allowing the same camera to be used for spectroscopy, as in Fig. 2.30a. The electron

optics is discussed briefly in Section 2.5.2. Since a modern TEM usually has STEM capabilities, the GIF can also be used to acquire spectrum images.

2.6.2 In-Column Filters

The prism–mirror and omega filters (Section 2.1.2) were developed specifically for producing EFTEM images and diffraction patterns. Located in the middle of a TEM column, they are designed as part of a complete system rather than as add-on attachments to a microscope. Following the construction of such systems in several laboratories (Castaing et al., 1967; Henkelman and Ottensmeyer, 1974a,b; Egerton et al., 1974; Zanchi et al., 1977a; Krahl et al., 1978), energy-selecting microscopes are now produced commercially (Egle et al., 1984; Bihr et al., 1991; Tsuno et al., 1997; Koch et al., 2006). Figure 2.31 shows the electron optics of a Zeiss omega-filter system; the operating modes of this kind of instrument have been described by Reimer (1991).

Because of the symmetry of these multiple-deflection systems, the image or diffraction pattern produced is achromatic and has no distortion or axial astigmatism. Midplane symmetry also precludes second-order aperture aberrations and image-plane distortion (Rose, 1989). Reduction of axial aberration in the *energy-selecting plane* and of field astigmatism and tilt of the *final image* requires sextupoles or curved pole faces (Rose and Pejas, 1979; Jiang and Ottensmeyer, 1993), or optimization of the optical parameters (Lanio, 1986; Krahl et al., 1990). Without such measures, the selected energy loss changes over the field of view, so that with a narrow energy-selecting slit and no specimen in the microscope, the zero-loss electrons would occupy a limited area on the final screen, equivalent to the alignment pattern of a single magnetic prism (Section 2.2.5). Or with an aluminum specimen, a low-magnification energy-selected image exhibits concentric rings corresponding to multiples of the plasmon energy (Zanchi et al., 1977a).

The Zeiss SESAM instrument represents a recent version of this concept (Koch et al., 2006). It uses a Schottky source, an *electrostatic* Omega-type filter as a monochromator and a magnetic MANDOLINE filter as the spectrometer and energy filter; see Fig. 2.32. To minimize vibrations, the TEM column is suspended at the level of the objective lens. A comparable project, the JEOL MIRAI instrument, uses a dual Wien filter as the monochromator and has achieved an energy resolution of 0.14 eV at 200 kV accelerating voltage (Mukai et al., 2004).

2.6.3 Energy Filtering in STEM Mode

Energy filtering is relatively straightforward in the case of a scanning transmission electron microscope, where the incident electrons are focused into a very small probe that is scanned over the specimen in a raster pattern. A filtered image can be viewed simply by directing the output of a single-channel electron detector (as

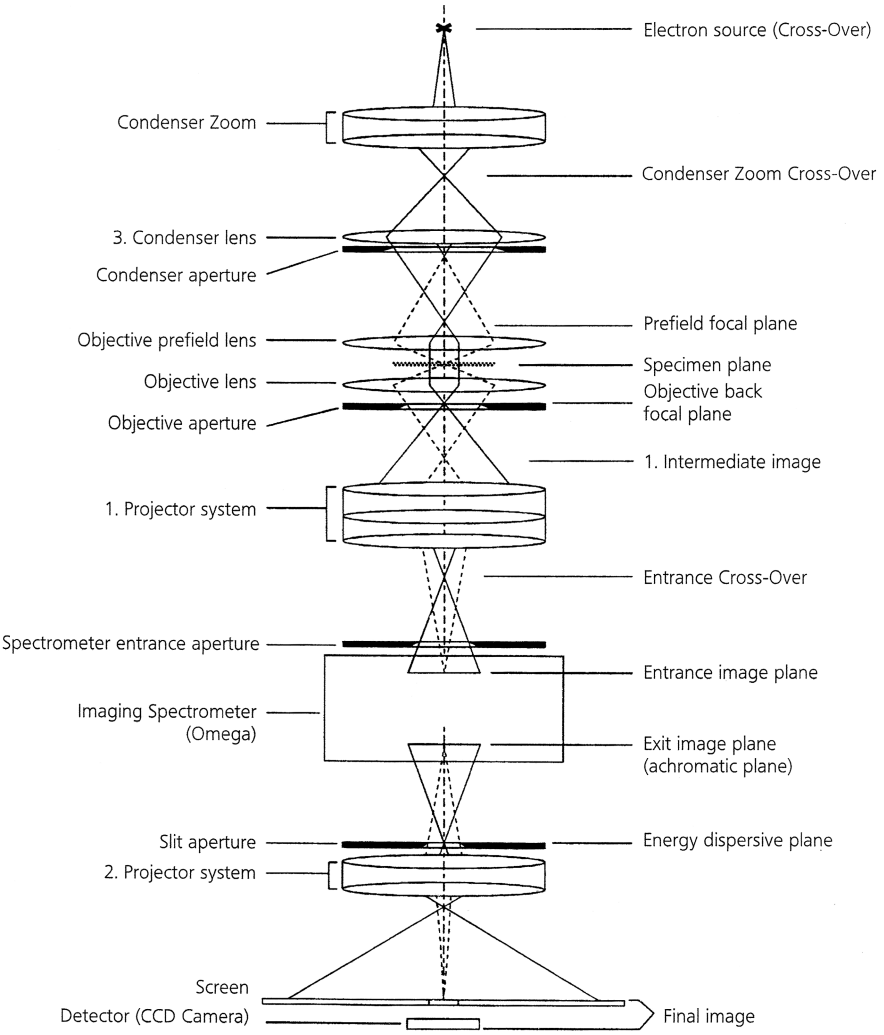

Fig. 2.31 Zeiss EM912 energy-filtering microscope: *solid lines* show field-defining rays, *dashed lines* represent image-defining trajectories. The Omega filter produces a unit-magnification achromatic image of the specimen (or its diffraction pattern) created by the first group of post-specimen lenses and generates (at the plane of the energy-selecting slit) a unit-magnification energy-dispersed image of its entrance crossover. From Carl Zeiss Topics, Issue 4, p. 4

used in serial EELS) to an image display monitor. It is preferable to *digitally* scan the probe and to digitize the detector signal so that it can be stored in computer memory, allowing subsequent image processing.

In STEM mode, the *spatial* resolution of a filtered image depends on the incident probe diameter (spot size) and beam broadening in the specimen, but

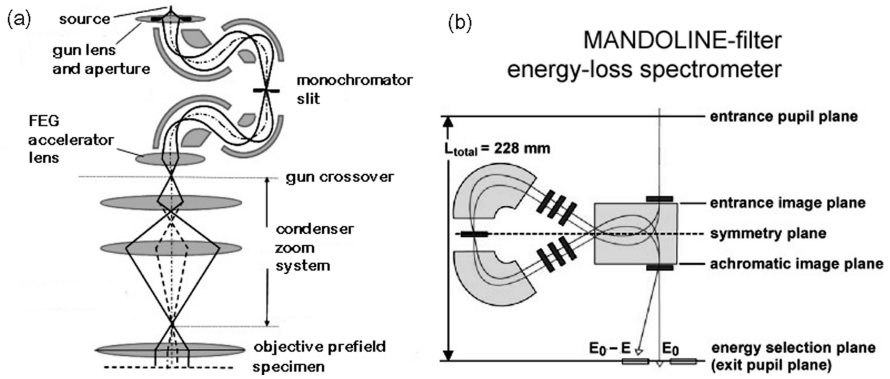

Fig. 2.32 Components of the Zeiss SESAM instrument: (**a**) electrostatic omega filter (dispersion $\approx$ 12 μm/eV at the midplane slit) and (**b**) MANDOLINE filter, whose dispersion at the energy-selecting plane exceeds 6 μm/eV

is independent of the spectrometer. For an instrument fitted with a Schottky or field-emission source, the probe diameter can be below 1 nm or less than 0.1 nm for an aberration-corrected probe-forming lens (Krivanek et al., 2008).

If electron lenses are present between the specimen and spectrometer, the mode of operation of these lenses does affect the *energy* resolution (Section 2.3.4). Also, the scanning action of the probe can result in a corresponding motion of the spectrum, due to movement of the spectrometer object point (image coupling) or through spectrometer aberrations (for diffraction coupling). Different regions of the image then correspond to different energy loss, limiting the field of view for a given energy window. The field of view d can be increased by widening the energy-selecting slit, but this limits the energy resolution to $MM_x d/D$ (see Section 2.3.5). A preferable solution is to *descan* the electron beam by applying the raster signal to dipole scan coils located after the specimen (Fig. 2.33). Ideally, the field of view is then unlimited and the energy resolution is the same as for a stationary probe. As there is no equivalent of descanning in a fixed-beam TEM, the use of STEM mode makes it easier to obtain good energy resolution in lower magnification energy-filtered images.

The STEM spectrometer can also be used to obtain an energy-selected *diffraction pattern,* by using x- and y-deflection coils to scan the latter across the spectrometer entrance aperture. However, this method is inefficient, since a large proportion of the electrons are rejected by the angle-defining aperture. If the required energy resolution is $\Delta\beta$ and the scan range is $\pm\beta$, the collection efficiency (and system DQE) is $(\Delta\beta/\beta)^2/4$. Shorter recording times and lower radiation dose are possible by using an imaging filter to process the whole diffraction pattern simultaneously (Krivanek et al., 1994), making the recording of core-loss diffraction patterns more feasible (Botton, 2005).

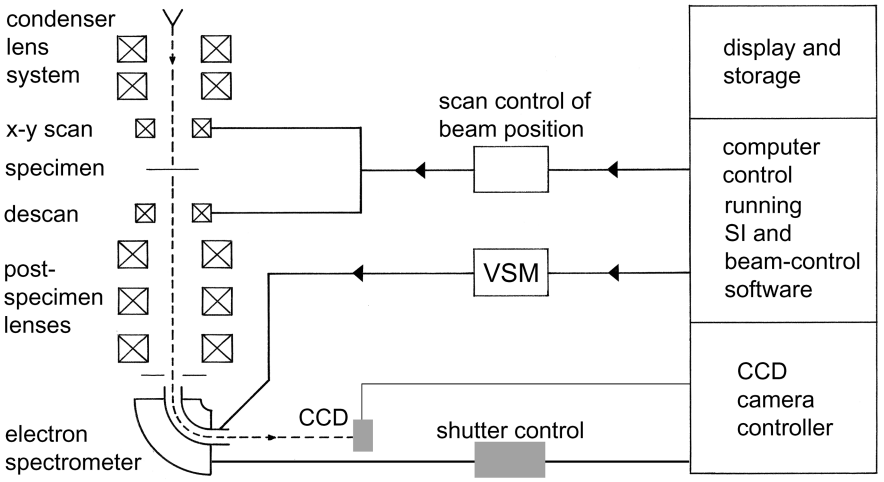

Fig. 2.33 Scheme for STEM spectrum imaging (Gubbens et al., 2010). A focused probe is scanned across the specimen and (for lower magnification work) descanned using post-specimen coils. At each pixel, an extended range of the energy-loss spectrum, chosen by the voltage voltage-scan module (VSM), is acquired by a parallel-recording spectrometer. An electrostatic shutter within the spectrometer allows fast switching between the low-loss and core-loss regions of the spectrum

2.6.4 Spectrum Imaging

With a parallel-recording system attached to a scanning-transmission electron microscope (STEM), the energy-loss spectrum can be read out at each picture point, creating a four-dimensional data array that corresponds to electron intensity within a three-dimensional (x, y, E) data cube (Fig. 2.1). The number of intensity values involved is large: 500 M for a 512 × 512 image, in the case of a 2048-channel spectrum. If each spectral intensity is recorded to a depth of 16 bits (64K gray levels), the total information content is then 0.5 Gbyte if the data are stored as integers (without data compression). However, advances in electronics have made the acquisition and storage of such data routine (Gubbens et al., 2010). To be efficient, the process requires good synchronization between the control computer, the CCD camera, and the probe-scanning and voltage-offset modules; see Fig. 2.33.

An equivalent data set can be obtained from an imaging filter, by reading out a series of energy-selected images at different energy loss, sometimes called an image spectrum (Lavergne et al., 1992). Since electrons intercepted by the energy-selecting slit do not contribute, this process is less efficient than spectrum imaging and involves a higher electron dose to the specimen, for the same information content. However, it may involve a shorter recording time, if the incident beam current divided by the number of energy-selected images exceeds the probe current used in the spectrum imaging; see Section 2.6.5. Common problems are undersampling (dependent on the energy shift between readouts) and a loss of energy resolution

(due to the width of the energy-selecting slit) but these can be addressed by FFT interpolation and deconvolution methods (Lo et al., 2001).

The main attraction of the spectrum image concept is that more spectroscopic data are recorded and can be subsequently processed to extract information that might otherwise have been lost. Examples of such processing include the calculation of local thickness, pre-edge background subtraction, deconvolution, multivariate analysis, and Kramers–Kronig analysis. The resulting information can be displayed as line scans (Tencé et al., 1995) or two-dimensional images of specimen thickness, elemental concentration, complex permittivity, and bonding information (Hunt and Williams, 1991; Botton and L'Esperance, 1994; Arenal et al., 2008). In addition, instrumental artifacts such as gain nonuniformities and drift of the microscope high voltage or beam current can be corrected by post-acquisition processing.

The acquisition time of a spectrum image is often quite long. In the past, the minimum pixel time has been limited by the array readout time, but that has recently been reduced from 25 to 1 ms (Gubbens et al., 2010). A *line spectrum*, achieved by scanning an electron probe in a line and recording a spectrum from each pixel, can be acquired more rapidly and is often sufficient for determining elemental profiles. Similar data can be obtained in fixed-beam TEM mode, with broad beam illuminating a slit introduced at the entrance of a double-focusing spectrometer. The long direction (y) of the slit corresponds to the nondispersive direction in the spectrometer image plane, allowing the energy-loss intensity $J(y,E)$ to be recorded by a CCD camera. One advantage here is that all spectra are acquired simultaneously, so specimen drift does not distort the information obtained, although it may result in loss of spatial resolution.

A common form of energy-filtered image involves selecting a range of energy loss (typically 10 eV or more in width) corresponding to an inner-shell ionization edge. Since each edge is characteristic of a particular element, the core-loss image contains information about the spatial distribution of elements present in the specimen. However, each ionization edge is superimposed on a spectral background arising from other energy-loss processes. To obtain an image that represents the characteristic loss intensity alone, the background contribution I_b within the core-loss region of the spectrum must be subtracted, as in the case of spectroscopy (Section 4.4). The background intensity often decreases smoothly with energy loss E, approximating to a power law form $J(E) = AE^{-r}$, where A and r are parameters that can be determined by examining $J(E)$ at energy losses just below the ionization threshold (Section 4.4.2). Unfortunately, both A and r can vary across the specimen, as a result of changes in thickness and composition (Leapman and Swyt, 1983; Leapman et al., 1984c), in which case a separate estimation of I_b is required at each picture element (pixel).

In the case of STEM imaging, where each pixel is measured sequentially, local values of A and r can be obtained through a least-squares or two-area fitting to the pre-edge intensity recorded over several channels preceding the edge. With electrostatic deflection of the spectrometer exit beam and fast electronics, the necessary data processing may be done within each pixel dwell period ("on the fly") and the system can provide a live display of the appropriate part of the spectrum (Gorlen

et al., 1984), values of A and r being obtained by measuring two or more energy-loss channels preceding the edge (Jeanguillaume et al., 1978). By recording a complete line of the picture before changing the spectrometer excitation, the need for fast beam deflection is avoided, background subtraction and fitting being done off-line after storing the images (Tencé et al., 1984).

In the case of energy filtering in a fixed-beam TEM, the simplest method of background subtraction is to record one image at an energy loss just below the ionization edge of interest and subtract some constant fraction of its intensity from a second image recorded just above the ionization threshold. First done by photographic recording (Ottensmeyer and Andrew, 1980; Bazett-Jones and Ottensmeyer, 1981), the subtraction process is more accurate and convenient with CCD images. Nevertheless, this simple procedure assumes that the exponent r that describes the *energy dependence* of the background is constant across the image or that the background is a constant fraction of the core-loss intensity. In practice, r varies as the local composition or thickness of the specimen changes (Fig. 3.35), making this two-window method of background subtraction potentially inaccurate (Leapman et al., 1984c). Variation of r is taken into account in the *three-window method* by electronically recording *two* background-loss images at slightly different energy loss and determining A and r at each pixel. However, the reduction in systematic error of background fitting comes at the expense of an increased statistical error (Section 4.4.3), so a longer recording time is needed for an acceptable signal/noise ratio (Leapman and Swyt, 1983; Pun and Ellis, 1983).

2.6.4.1 Influence of Diffraction Contrast

Even if the background intensity is correctly subtracted, a core-loss image may be modulated by diffraction (aperture) contrast, arising from variations in the amount of *elastic* scattering intercepted by the angle-limiting aperture. A simple test is to examine an *unfiltered* bright-field image, recorded at Gaussian focus using the same collection aperture; the intensity modulation in this image is a measure of the amount of diffraction contrast. Several methods have been proposed for removing this aperture contrast, in order to obtain a true elemental map:

(a) Dividing the core-loss intensity I_k by a pre-edge background level, to form a jump-ratio image (Johnson, 1979).
(b) Dividing by the intensity of a low-loss (e.g., first-plasmon) peak, which is also modulated by diffraction contrast.
(c) Dividing I_k by the intensity I_l, measured over an equal energy window in the low-loss region of the spectrum (Egerton, 1978a; Butler et al., 1982). According to Eq. (4.65), the ratio I_k/I_l is proportional to areal density (number of atoms of an analyzed element per unit area of the specimen).
(d) Taking a ratio of the core-loss intensities of two elements, giving an image that represents their elemental ratio; see Eq. (4.66).
(e) Using conical rocking-beam illumination, which varies the angle of incidence over a wide range (Hofer and Warbichler, 1996).

(f) Recording the filtered images with a large collection semi-angle, so that almost all the inelastic and elastic scattering enters the spectrometer (Egerton, 1981d; Muller et al., 2008).

Methods (a), (b), and (d) have the advantage that they correct also for variations in specimen thickness (if the thickness is not too large), giving an image intensity that reflects the *concentration* of the analyzed element (atoms/volume) rather than its areal density.

2.6.5 Comparison of Energy-Filtered TEM and STEM

To compare the advantages of the fixed-beam TEM and STEM procedures of energy filtering, we will assume equal collection efficiency (same β^*; see Section 4.5.3), similar spectrometer performance in terms of energy resolution over the field of view, and electron detectors with similar noise properties. We will take the spatial resolution to be the same in both methods, a resolution below 1 nm being achievable (for some specimens) using either procedure.

Consider first *elemental mapping* with a TEM imaging filter, in comparison with a STEM system operating with an energy-selecting slit (as in *serial* recording). At any instant, a single energy loss is recorded, energy-selecting slits being present in both cases; see Fig. 2.34. For two-window background fitting, three complete images are acquired for each element in either mode, as discussed in the previous section. For the same amount of information, the *electron dose* is therefore the *same* in both methods. The only difference is that the electron dose is delivered continuously in the fixed-beam TEM and for a small fraction of the frame time in STEM mode.

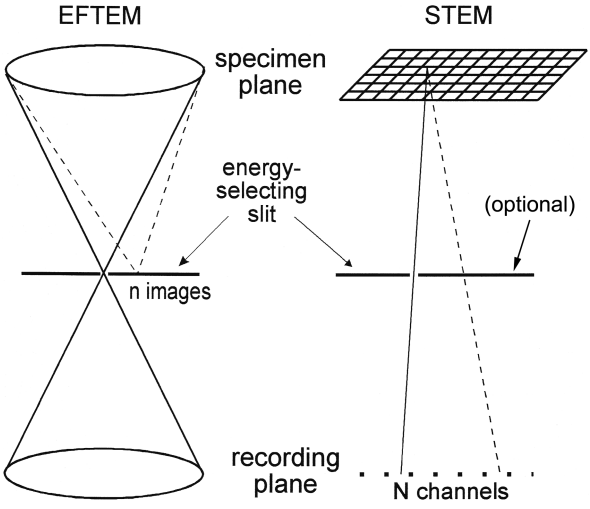

Fig. 2.34 Comparison of TEM and STEM modes of energy-selected imaging. The STEM system can exist *with* an energy-selecting slit and single-channel detector (serial recording spectrometer) or *without* the slit, allowing the use of parallel recording (spectrum image mode)

In the STEM case, the dose *rate* (current density in the probe) can be considerably higher than for broad-beam TEM illumination. Whether an equal dose implies the same amount of radiation damage to the specimen then depends on whether the radiation damage is dose rate dependent. For polymer specimens, temperature rise in the beam can increase the amount of structural damage for a given dose (Payne and Beamson, 1993). In some inorganic oxides, mass loss (hole drilling) occurs only at the high current densities such as are possible in a field-emission STEM (see Section 5.7.5). For such specimens, the STEM procedure would be more damaging, for the same recorded information.

The *recording time* for a single-element map is generally longer in STEM, since the probe current (even with a field-emission source) is typically below 1 nA, whereas the beam current in a conventional TEM can be as high as 1 μA. STEM recording of a 512 × 512-pixel image may take as much as 1 h to achieve adequate statistics, which is inconvenient and requires specimen drift correction. The usual solution is to reduce the amount of information recorded by decreasing the number of pixels. In a conventional TEM, the CCD camera offers typically 2k × 2k pixels, so EFTEM can provide a higher *information rate*.

When a *parallel-recording* spectrometer is used in the STEM, the three images required for each element are recorded simultaneously, reducing both the time and the dose by a factor of 3. This factor becomes 3*n* in a case where *n* ionization edges are recorded simultaneously. STEM with a parallel-recording spectrometer should therefore produce less radiation damage than an energy-filtering TEM, unless dose rate effects occur that outweigh this advantage.

The advantage of STEM is increased further when an extended range of energy loss is recorded by the diode array, as in spectrum imaging. If use is made of the information recorded by all *N* detector channels, a given electron dose to the specimen yields *N* times as much information as in EFTEM, where *N* energy-selected images would need to be acquired sequentially to form an image spectrum of equal information content. STEM spectrum imaging also makes it easier to perform on- or off-line correction for specimen drift, so that long recording times (although inconvenient) do not necessarily compromise the spatial resolution of analysis. The acquisition time in STEM could actually be shorter if the product NI_p, where I_p is the probe current, exceeds the beam current used for EFTEM imaging.

Both scanning and fixed-beam modes of operation are possible in a conventional TEM equipped with probe scanning and (preferably) a field-emission source and parallel-recording spectrometer. With such an instrument, single-element imaging is likely to be faster in EFTEM mode, but in the case of multielement imaging or the need for data covering a large range of energy loss, the STEM mode is more efficient in terms of specimen dose and (possibly) acquisition time.

The above arguments assume that the whole of the imaged area is uniformly irradiated in STEM mode, just as it is in EFTEM. However, if the STEM probe size were kept constant and the magnification reduced, only a fraction of the pixel size would be irradiated by the probe; radiation damage would be higher than necessary, with regions between scan lines or probe positions (for a digital scan) left unirradiated. In other words, the specimen would be *undersampled* by the probe. One

solution is to increase the probe diameter as the magnification is lowered, although this may still leave some of the specimen unirradiated because the probe is circular and the pixels are square. An equally good solution, in the case of a digitally generated scan, is to use sub-pixel scanning: for an image having $N \times N$ pixels, waveforms whose frequency is N times larger and amplitude N times smaller are added to the line and frame scans. The very small probe is then scanned in a small square raster, covering each pixel area, so that severe undersampling of the specimen is avoided.

2.6.6 Z-Contrast and Z-Ratio Imaging

The scanning-transmission electron microscope has the advantage of being able to efficiently collect electrons that are scattered through large angles, by use of a high-angle annular dark-field (HAADF) detector. This high-angle elastic scattering approximates to Rutherford elastic scattering (Section 3.1.2), which ideally has a Z^2 dependence on atomic number Z, although slightly lower exponents are common (Section 3.1.6). The STEM therefore provides *Z-contrast* images, useful for imaging clusters of heavy atoms on a lower-Z substrate (e.g., catalyst studies) and atomic columns in crystalline specimen (Pennycook and Jesson, 1991; Krivanek et al., 2010). Channeling of the incident electrons limits the lateral spreading of the beam in a crystalline specimen, allowing core-loss spectroscopy to identify the atomic number of each column (Browning et al., 1993b, 1999; Pennycook et al., 1995a; Varela et al., 2004). This spectroscopy can be precisely simultaneous with the structural imaging because it makes use of electrons that pass through the central hole of the HAADF detector, as in Fig. 2.35.

Z-ratio imaging refers to a technique first used by Crewe et al. (1975) to display images of single atoms of Hg and U on a thin-carbon (<10 nm) support film. The ADF signal I_d was divided by the total-inelastic signal I_i recorded

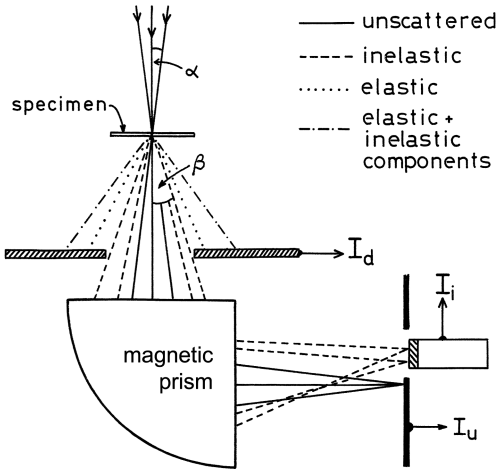

Fig. 2.35 *Z*-contrast imaging in STEM mode. The inner radius of the annular detector subtends a semi-angle β at the specimen, the outer radius being considerably larger. An electron spectrometer separates the inelastic and unscattered electrons

through a spectrometer, using a wide energy-selecting slit that simply eliminated the unscattered component I_u; see Fig. 2.35. A single-channel detector, such as a scintillator/PMT combination, is sufficient to measure I_i. Because variations in mass thickness of the support contribute almost equally to I_d and I_i, their unwanted effect largely disappears in the image formed from the ratio $R = I_d/I_i$. Taking a ratio also cancels any fluctuations in probe current but does involve some increase in noise, particularly if the specimen is very thin.

Because of its relatively broad angular distribution (Section 3.2.1), elastic scattering provides the main contribution to I_d, whereas the main contribution to I_i comes from electrons that suffer only inelastic scattering. If the specimen is amorphous and very thin, the elastic and inelastic scattering are both proportional to specimen thickness and to the appropriate atomic cross sections σ_e and σ_i. Assuming that *all* elastic scattering is recorded by the dark-field detector and *all* inelastic scattering is captured by the spectrometer, the ratio signal is

$$R = I_d/I_i = \sigma_e/\sigma_i = \lambda_i/\lambda_e \tag{2.54}$$

where λ_e and λ_i are mean free paths for elastic and inelastic scattering. The intensity in the ratio image is then proportional to the local value of the elastic/inelastic scattering ratio, which in theory and in practice (Section 3.2.1) is proportional to atomic number Z.

In the case of specimens that are thicker than about $\lambda_i/2$, the dark-field and inelastic signals are given more accurately (Lamvik and Langmore, 1977; Egerton, 1982c) by

$$I_d = I[1 - \exp(-t/\lambda_e)] \tag{2.55}$$

$$I_i = I\exp(-t/\lambda_e)[1 - \exp(-t/\lambda_i)] \tag{2.56}$$

where I is the incident beam current and t is the local thickness of the specimen. The exponential functions in Eqs. (2.55) and (2.56) occur because of plural scattering, as a result of which neither I_i nor I_d is proportional to specimen thickness. For $t/\lambda_i < 1$, this nonlinearity can be removed by digital processing (Jeanguillaume et al., 1992). For $t/\lambda_i > 1$, I_i reaches a maximum and then decreases with increasing thickness, as many of the inelastically scattered electrons also undergo elastic scattering and therefore contribute to I_d rather than to I_i.

Z-ratio imaging has been used to increase the contrast of thin (30-nm) sections of stained and unstained tissues (Carlemalm and Kellenberger, 1982; Reichelt et al., 1984). If the sample is crystalline and the ADF detector accepts *medium-angle* scattering, both the elastic and inelastic images are strongly influenced by diffraction contrast, which may increase rather than cancel when the ratio is taken (Donald and Craven, 1979).

http://www.springer.com/978-1-4419-9582-7

Electron Energy-Loss Spectroscopy in the Electron Microscope

Egerton, R.F.

2011, XII, 491 p., Hardcover
ISBN: 978-1-4419-9582-7