ELECTRON MICROSCOPES

In 1873, it was proven by Ernst Abbe that the resolving power of a light microscope will always be limited by the wavelength of the light, which is of the order of 1 μm, so that there could be no hope to visualize much smaller objects such as atomic scale structures. (In the 1980s, near-field optical scanning techniques were developed that can bring the resolution down by two orders of magnitude.) Fifty years later, a new impulse was given to the problem by the hypothesis of Louis De Broglie about the wave nature of particles so that other

particles could also serve as "light." In 1931 Ernst Ruska developed the first transmission microscope TEM that uses electrons instead of photons. In 1986 Ernst Ruska was awarded the Nobel Prize for his pioneering work. Electrons are the best candidates since they can easily be generated by a heated filament or extracted from a point by an electric field and they are easily deflected by electric and magnetic fields. When accelerated to, say, 100 keV, their wavelength is much smaller (3 pm = 3×10^{-12} m) than that of visible light. They can also be detected on a photoplate, a fluorescent screen, or an electronic camera. On the other hand, they can only propagate in vacuum and they can only penetrate through very thin objects ($<10^3$ nm), so that vacuum and specimen preparation techniques are crucial.

In past decades, electron microscopy has matured to an indispensable tool for materials and biomedical research. Numerous are the Nobel Prizes for research work in which electron microscopy revealed crucial evidence. Despite the very small wavelength of the electrons, it has only recently been possible to visualize individual atoms. The reason for this is that magnetic lenses inevitably suffer from aberrations that, contrary to glass lenses in a light microscope, cannot easily be corrected. Recently the problem of compensating the spherical aberration has become solved on the research level.

When individual atoms, the building blocks of nature, can be revealed, electron microscopy enters into a new era. Indeed, then it becomes possible to determine the atomic structure of matter quantitatively and accurately, even for objects for which only limited prior knowledge is available. Although the goal is not yet achieved, the progress is very encouraging. In the future, as materials science evolves into materials design, and mesostructures into nanostructures, the role of electron microscopy will become even more important.

In recent years we have also seen the development of related techniques such as scanning transmission electron microscopy (STEM), electron holography, ptychography, and add-on techniques that use complementary information such as X-ray analysis (EDX), and energy filtering. There is now a growing tendency to incorporate all these techniques into one versatile instrument under full computer control, where they are all considered as different ways to obtain complementary information, mainly of chemical nature.

The development of scanning electron microscopy (SEM) has taken place parallel to that of TEM. The first SEM was constructed by Manfred von Ardenne in the 1930s. The SEM is conceptually simpler than the TEM. With the aid of magnetic lenses, the electron beam is focused into a small point that is scanned over the surface of the object. The backscattered electrons, or other secondary signals, can then be detected and synchronously displayed on a cathode-ray-tube (CRT) screen. The SEM is relatively easy to use and does not require extensive specimen preparation and is also less expensive than the TEM. On the other hand, it can only be used to study the surface area and, because of the beam spread in the object, its resolution is rather limited and quantitative interpretation of the images is more difficult.

TRANSMISSION ELECTRON MICROSCOPY

The Instrument

A transmission electron microscope is very similar to a light microscope (Fig. 1) in which the light bulb is replaced by an

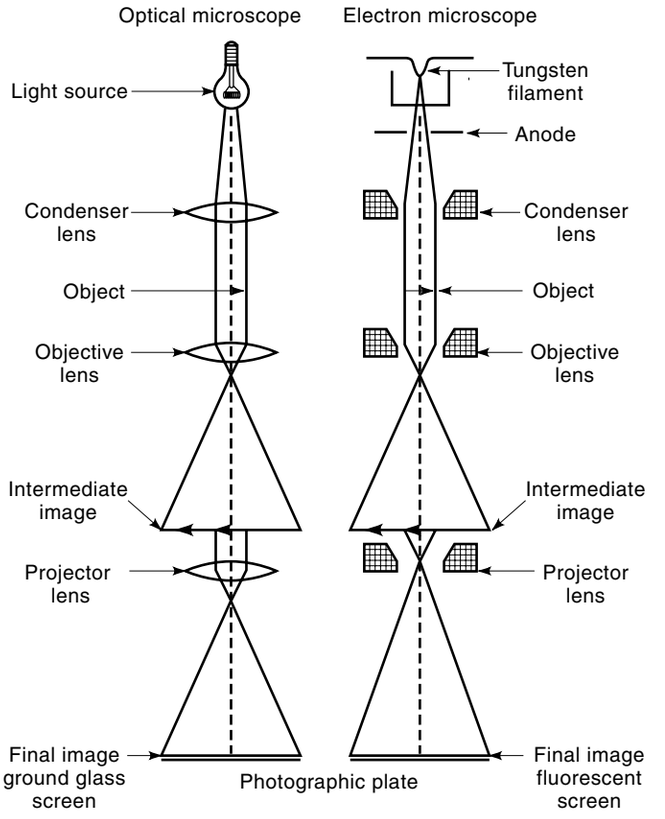


Figure 1. Comparison between an electron microscope (right) and a light microscope (left). All the components of the light microscope have their analog in the electron microscope. The acceleration electrons have a wave character and act as light.

electron source and the glass lenses by magnetic (or electric) lenses. The focusing of the lenses can be varied by changing the lens currents so that the setting of the microscope (focusing, magnification, etc.) can easily be altered. The condenser lens shapes the electron flow into a parallel beam. The specimen is a very thin object mounted in a holder that, through an airlock system, can be brought into the electron beam. Holders exist with a variety of degrees of freedom (translation, rotation, tilt, temperature, stress). The objective lens produces an enlarged and rotated image, which in its turn, through a system of intermediate and projector lenses, is further magnified and projected onto a fluorescent screen, a photoplate, or an electronic camera.

Another mode of operation is not to image the object but to image the focal plane of the objective lens. In this way one can observe the electron-diffraction pattern directly onto the fluorescent screen.

The resolving power of the electron microscope is mainly determined by the quality of the objective lens. With modern instruments, a resolution of about 0.15 nm can be reached.

In modern electron microscopes the microscope is operated under computer control and the alignment can be done automatically. The visual electron source, which is a thermal filament, is increasingly replaced by a field-emission source, from which the electrons are extracted by an electric field and which yields a much higher brightness.

Image Formation

Intuitive Description. The imaging process in the electron microscope can be sketched in a very simple way as follows (Fig. 2). The object is illuminated by a plane wave. Due to the interaction of the electrons with the atoms of the object, the electron wave is modulated by the structure of the object. By propagating this object exit wave through the electron microscope to the image plane this electron wave is blurred, that is, each point of the wave is spread into the so-called point-spread function or impulse response function. The wave as well as the point-spread function is a complex valued function with an amplitude and a phase component. Finally, however, in the image plane only the intensity is recorded so that the phase information is lost. Full recording of this phase information can be done using holographic techniques (discussed later).

Wave Optical Description. As in the light microscope, the formation of the image must be described by wave optics (Fig. 3). The electron wave is diffracted by the object, and each diffracted beam is focused by the objective lens into one point in the focal plane. The electron wave in the focal plane can be considered as the Fourier transform of the exit wave of the object. The high spatial frequencies (small details) correspond to beams far from the optical axis and the small spatial frequencies (large details) are located close to the optical axis. In case the object is periodical, such as is the case of a perfect crystal, the Fourier transform is discrete, that is, the diffraction pattern in the back focal plane consists of discrete spots. These are the different diffracted beams. In case the object is aperiodic, the image in the focal plane shows diffuse intensity. In the second step of the imaging process, the points in the focal plane act again as sources of waves, which interfere in the image plane. The image wave found in the image plane is the reverse Fourier transform of the wave found in the focal plane; therefore, under ideal imaging conditions with no imaging aberrations or apertures, the image wave should be identical to the object exit wave. In short, the whole imaging process is wave optically considered as a sequence of two Fourier transforms, the first one leading from the object plane to the back focal plane of the objective lens, and the second one from the focal plane to the image plane. In practice, however, the objective lens has intrinsic aberrations. The most important ones are the spherical aberration and the defocus. As a consequence, the Fourier components of the electron wave in Fourier space suffer from hampering phase shifts arising from these aberrations that are more important for the high spatial frequencies and that blur the image.

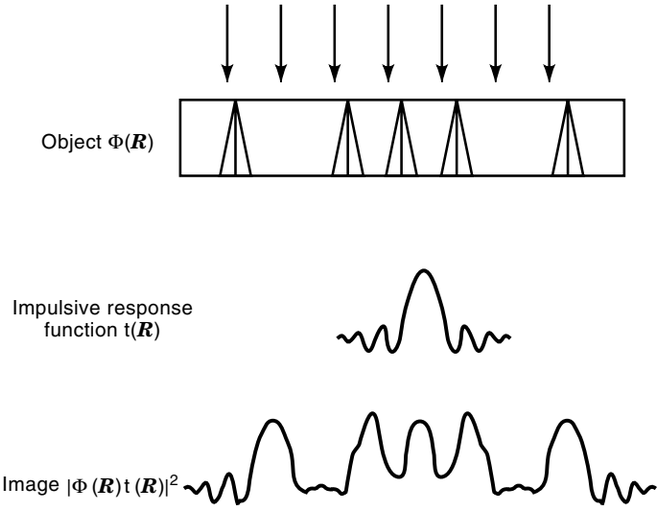


Figure 2. Schematical representation of the imaging process. The electron wave is modulated by the object and then blurred by the electron microscope.

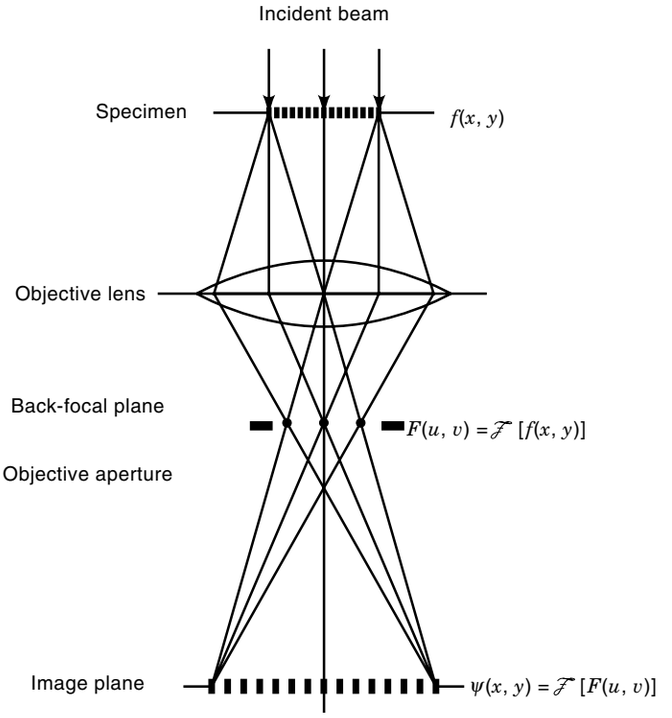


Figure 3. Wave optical representation of the image formation by the objective lens in a transmission electron microscope. The corresponding mathematical operations are indicated [see text with $\boldsymbol{R} = (x, y)$ and $\boldsymbol{G} = (u, v)$].

Usually the Fourier components, falsified by more than a tolerable threshold, are masked out to minimize the blurring effect. On the other hand, one has to find a compromise because, for high-resolution imaging, as many Fourier components as possible should be admitted to the image. In a sense, one can then consider the image wave as a truncated Fourier reconstruction of the object exit wave with only a limited number of Fourier components, which may be falsified by the aberrative phase shift. The image formed by interference of the residual waves cannot be interpreted easily.

The resolution, in a sense the smallest detail present in the image, is then determined by the highest spatial frequency that is allowed to contribute to the image formation. One distinguishes between the point resolution, which is the highest spatial frequency that contributes without extra phase shift and which gives rise to directly interpretable details, and the information limit, which is the highest spatial frequency in the image formation but with a phase shift that complicates a direct interpretation of the details.

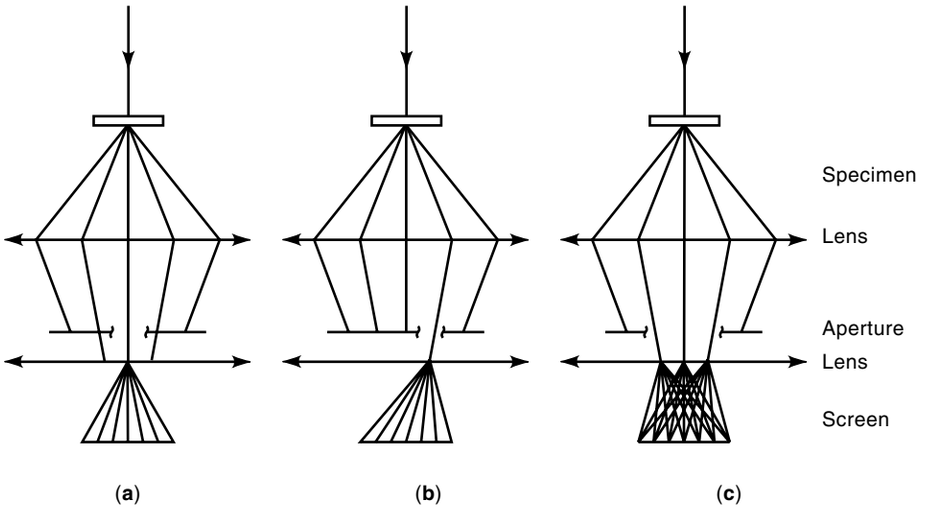


Figure 4. Different imaging modes. (a) Bright-field imaging: only the transmitted beam contributes to the image. (b) Dark field: only one diffracted beam contributes to the image. (c) High-resolution imaging: the image is formed by interference of many diffracted beams.

For special investigations of crystalline objects, it is useful to include only one diffracted beam for imaging (Fig. 4). If the central beam is used, one obtains a so-called bright field image (BF); if only a diffracted beam is included, a so-called dark field image results. In these cases, since there is no interference between diffracted beams, the image contrast can more easily be interpreted in terms of the number of electrons scattered in the corresponding particular direction. However, high resolution cannot be obtained with these techniques.

In any case, for the interpretation of the images, it is equally important to understand the interaction between the electrons and the object.

Mathematical Formulation

Transfer Function. By interaction with the object, the object exit wave $\psi(\boldsymbol{R})$ is modulated both in amplitude and phase. Hence $\psi(\boldsymbol{R})$ is complex. $\boldsymbol{R}$ is the position vector in the object exit plane. According to Fraunhofer's diffraction theory the diffracted wave in the direction given by the reciprocal vector $\boldsymbol{g}$ (or spatial frequency) is given by the Fourier transform of the object function, that is,

$$\phi(\boldsymbol{g}) = F_g \psi(\boldsymbol{R}) \tag{1}$$

The wave in the back focal plane of the objective lens is then the Fourier transform of the object wave. If one directly images the focal plane, one can see the diffraction pattern, given by $|\phi(\boldsymbol{g})|^2$. If the object is periodic, such as a crystal, the diffraction pattern will consist of sharp spots. A continuous object will give rise to a continuous diffraction pattern. The second stage of the imaging process is described by an inverse Fourier transform that reconstructs the object function $\psi(\boldsymbol{R})$ (usually enlarged) in the image plane (Fig. 3). The intensity in the image plane is then given by $|\psi(\boldsymbol{R})|^2$.

In practice, by inserting an aperture in the focal plane of the objective lens, it is possible to obtain an image in which only selected beams contribute. On passing through the objective lens, each electron beam $\boldsymbol{g}$ undergoes a phase shift and an amplitude reduction (damping). Hence the transfer function takes the form

$$T(\boldsymbol{g}) = A(\boldsymbol{g}) \exp[-i\chi(\boldsymbol{g})] D(\boldsymbol{g}) \tag{2}$$

$A(\boldsymbol{g})$ describes the effect of the beam-selecting aperture and the damping caused by incoherent effects (such as fluctuations in voltage and lens currents). $\chi(\boldsymbol{g})$ is the phase shift, given by

$$\chi(\boldsymbol{g}) = \frac{\pi}{2}(C_s \lambda^3 g^4 + 2\epsilon\lambda g^2) \tag{3}$$

with C_s the spherical aberration constant of the objective lens, ϵ the focus distance, and λ the electron wavelength. The wave function at the image plane is now given by the inverse Fourier transform

$$\psi_{\text{im}}(\boldsymbol{R}) = F_{\boldsymbol{R}}^{-1} T(\boldsymbol{g}) \phi(\boldsymbol{g}) \tag{4}$$

and the image intensity by

$$I_{\text{im}}(\boldsymbol{R}) = \|\psi_{\text{im}}(\boldsymbol{R})\|^2 \tag{5}$$

Equation (4) is called the coherent approximation; it is valid for thin objects. In general, the expressions are more complicated. One then has to consider the image as an incoherent superposition of images that are slightly different, due to fluctuations in the conditions of the electron microscope. For a more general description we refer the reader to Ref. 1.

The total image formation process is illustrated in Fig. 2.

Impulse Response Function. As intuitively described earlier the image transfer can also be described in real space as a blurring effect, and as also follows from Eq. (4) using the convolution theorem,

$$\Phi(\boldsymbol{R}) = \psi(\boldsymbol{R}) * t(\boldsymbol{R}) \tag{6}$$

where $\psi(\boldsymbol{R})$ is the object wave in real space and $t(\boldsymbol{R})$ is the Fourier transform of the transfer function. For an hypothetical ideal pointlike object, the object wave $\psi_{\text{im}}(\boldsymbol{R})$ would be a δ function or "impulse" $[\psi(\boldsymbol{R}) = \delta(\boldsymbol{R})]$ so that $\Phi_{\text{im}}(\boldsymbol{R}) = t(\boldsymbol{R})$, that is, the microscope would reveal $t(\boldsymbol{R})$, which therefore is called impulse response function. If the transfer function would be constant (i.e., perfectly flat) in the whole spatial frequency range, the impulse response function would be a δ function so that $\psi_{\text{im}}(\boldsymbol{R}) = \psi(\boldsymbol{R})$, that is, the wave function in the image plane represents exactly the wave function of the object. In a

sense the microscope is perfect. However, in practice the transfer function cannot be made constant as is shown in Fig. 5. The impulse response function is still peaked as shown in Fig. 6.

Hence, as follows from Eq. (6), the object wave $\psi(\boldsymbol{R})$ is smeared out (blurred) over the width of the peak. This width can then be considered as a measure for the resolution in the sense as originally defined by Rayleigh. The width of this peak is the inverse of the width of the constant plateau of the transfer function in Fig. 5. From another point of view one can argue that if all the spatial frequencies have the same phase shift, the information is transferred forward and keeps a point to point relation to the object.

However, the information beyond this plateau is still contributing to the image but with a wrong phase. It is scattered outside the peak of the impulse response function and it is thus redistributed over a larger area in the image plane.

Phase-Contrast Microscopy. In an ideal microscope, the image wave function would exactly represent the object wave function and the image of a pure phase object function would show no contrast. This can be compared with imaging a glass plate with variable thickness in an ideal light microscope.

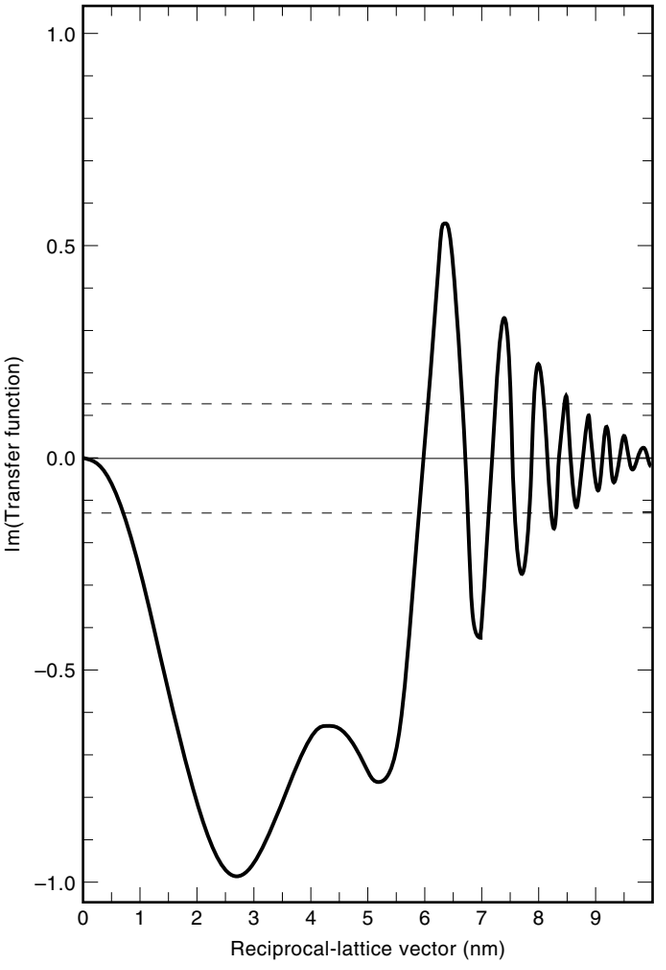


Figure 5. Transfer function (imaginary part) of a 300 keV electron microscope at optimum defocus.

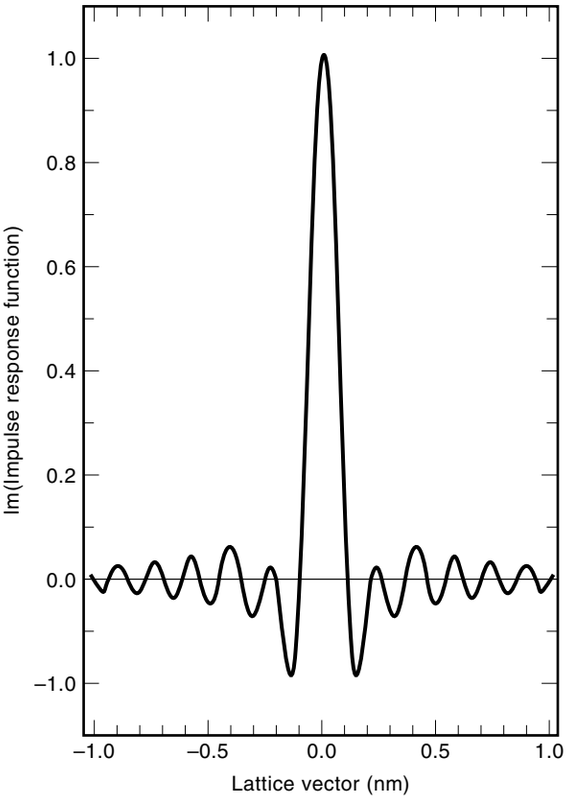


Figure 6. Impulse response function (imaginary part) corresponding to the transfer function of Fig. 5.

If the phases of the diffracted beam (Fourier components) are shifted over $\pi/2$ with respect to the central beam the image contrast would directly reveal the phase of the object. In light microscopy, phase contrast can be obtained by inserting an annular quarter-wavelength plate in the focal plane of the objective lens.

In electron microscopy, phase-contrast imaging can be achieved by making the transfer function as constant as possible. From Eq. (3) it is clear that phase shifts occur due to spherical aberration and defocus. However, the effect of spherical aberration, which, in a sense, makes the objective lens too strong for the most inclined beams, can be compensated somewhat by slightly underfocussing the lens.

The focus setting is called the optimum focus or Scherzer focus and its value can be calculated as

$$\epsilon = -1.2 C_s^{1/2} \lambda^{1/2} \tag{7}$$

The transfer function for this situation is depicted in Fig. 5. The phase shift $\chi(\boldsymbol{g})$ is nearly equal to $-\pi/2$ with respect to the central beam for a large range of spatial coordinates, up to the value

$$|\boldsymbol{g}| = 1.5 C_s^{-1/4} \lambda^{-3/4} \tag{8}$$

The image then reveals directly the phase of the object. Now a thin material object acts as a phase object in which the phase is proportional to the electrostatic potential of the atoms projected along the viewing direction. Hence, if the ob-

ject would be very thin, optimum focus imaging would directly reveal atom columns as dark dots and empty spaces as light areas. However, this argument only holds for spatial frequencies that are within the range given by Eq. (8). Furthermore the thickness up to which an object can be considered as a weak phase object is very small (e.g., 1 nm) and is rarely met in practice.

Resolution. One has to distinguish between point resolution (or structural resolution) as the finest detail that can be interpreted in terms of the structure and the information limit, which is the finest detail that can be resolved by the instrument, irrespective of a possible interpretation.

The point resolution can be obtained from the inverse of the maximal spatial frequency equation (8) as

$$\rho_s = 0.65 C_s^{1/4} \lambda^{3/4} \tag{9}$$

The point resolution is also equal to the "width" of the impulse response function (Fig. 6). The information beyond ρ_s is transferred with a nonconstant phase and, as a consequence, is redistributed over a larger image area.

The information limit ρ_I can be defined as the finest detail that can be resolved by the instrument. This limit is mainly determined by incoherent effects (fluctuations on the energy of the electrons and on the lens currents, divergences of the illuminating beams). The information limit is usually smaller than the point resolution. Typical values are λ = 2 pm (300 keV), C_s = 1 mm, ρ_s = 0.2 nm, ρ_I = 0.13 nm. The point resolution can be improved by reducing the incoherent effects (e.g., energy and current fluctuations), for instance, by using a field emission source and a stable voltage supply.

The Specimen

Specimens for high-resolution electron microscopy are prepared using the same techniques as for usual transmission electron microscopy, that is, ion beam milling, chemical and electrolytical thinning, cleavage, crushing, and so on. The only requirements are that the specimen should be sufficiently thin, that is, less than about 10 nm. Furthermore, the specimen should be sufficiently clean and free of contamination. Depending on the type of specimen one can use different thinning methods such as ion beam milling, electrochemical etching, crushing, and dimpling.

For details of specimen preparation we refer the reader to Ref. 2.

Crystalline specimens are oriented with a zone axis parallel to the incident beam so that all the diffracted beams of that zone are excited simultaneously and maximal information is present in the image. This is shown in Fig. 7. In this situation, the electrons propagate parallel to a zone axis, that is, parallel to the atom rows, which makes a possible interpretation of the images in terms of the projected structure meaningful. As an example, Fig. 8 shows a drawing of a model of a dislocation viewed along a zone axis.

It is also possible, using an aperture placed in the focal plane of the objective lens, to select a particular set of beams so that the images contain specific information. If the central beam is not included, these images are called dark-field images.

After finding a suitably thin part with the proper orientation, one has to adjust the focus. When the specimen is very thin, the zero focus corresponds to minimal contrast.

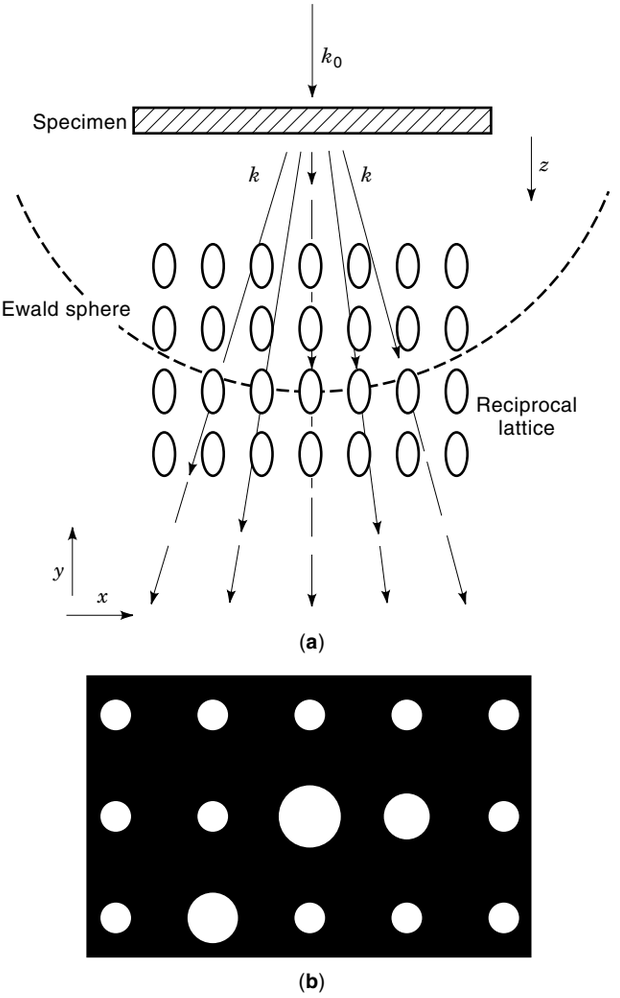


Figure 7. Formation of the diffraction pattern. The simultaneously excited electron beams can be used for image formation.

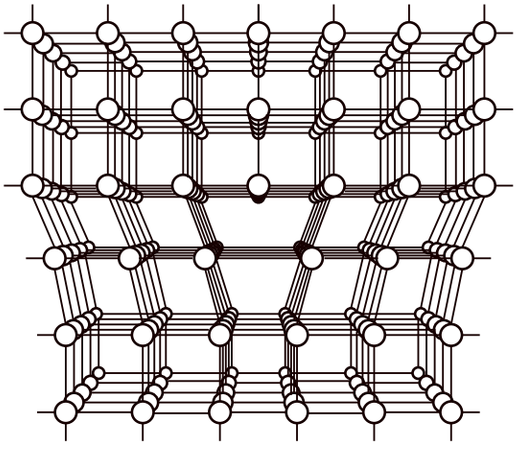


Figure 8. Structure model of a crystal containing a dislocation viewed along the atomic columns.

In practice, one takes a series of images at gradually different focus settings, recorded approximately around the optimum defocus. This is called a through-focus series. When dealing with a specimen that is unstable in the electron beam, the specimen can be completely destroyed within a period of seconds. Here a minimal exposure technique has to be used.

Electron Diffraction

Diffraction Mode. When the focal length of the intermediate lens is increased, by weakening its excitation, so as to make the back focal plane of the objective lens coincide with the object plane of the projector lens, a magnified image of the diffraction pattern is projected on the screen. Neither specimen orientation nor selected area are hereby changed; the diffraction pattern is thus representative of the selected area. It should be noted that the area selected in the diffraction mode is much larger than the field of view for imaging.

Geometry of the Diffraction Pattern. The diffraction conditions can be formulated in two different ways: either emphasizing direct space or reciprocal space. In direct space the diffraction conditions are known as *Bragg's law* (3). Attention is focused on sets of parallel lattice planes with interplanar spacing d_H, where $\boldsymbol{H}$ represents the three Miller indices of the considered family of lattice planes. "Reflection" will occur for certain "glancing" angles θ_H satisfying Bragg's law; $2d_H \sin \theta_H = n\lambda$, where λ is the wavelength of the radiation used and n is an integer. This relation expresses the condition that the difference in path length between waves scattered by two successive lattice planes is an integer number of wavelengths (Fig. 9). For 100 kV electrons $\lambda \simeq 0.004$ nm and hence the Bragg angles are very small, of the order of a few degrees at most.

In terms of reciprocal space the *Ewald diffraction condition* (4) can be formulated as $\boldsymbol{k}_H = \boldsymbol{k}_0 + \boldsymbol{B}_H$, where $\boldsymbol{k}_0$ is the wave vector of the incident wave and $\boldsymbol{k}_H$ that of the scattered wave and $\boldsymbol{B}_H$ is a reciprocal-lattice vector $\boldsymbol{B}_H = h_1\boldsymbol{b}_1 + h_2\boldsymbol{b}_2 + h_3\boldsymbol{b}_3$. The $\boldsymbol{b}_j$ are the base vectors of the reciprocal lattice, defined in terms of the base vectors $\boldsymbol{a}_i$ of the direct lattice by the relations $\boldsymbol{a}_i \cdot \boldsymbol{b}_j = \delta_{ij}$ (Kronecker δ) $(i, j = 1, 2, 3)$. The Ewald condition expresses the conservation of energy and linear momentum on elastic scattering; its physical content is the same as that of Bragg's law. Geometrically it expresses the condition that a sphere with radius $|\boldsymbol{k}| = 1/\lambda$ and center C in $-\boldsymbol{k}_0$ (Ewald's sphere) must pass through a point $\boldsymbol{H}$ of the reciprocal lattice for diffraction to occur. The scattered wave direction is then obtained by joining the center of Ewald's sphere with the excited reciprocal-lattice point $\boldsymbol{H}$ (Figs. 7 and 10).

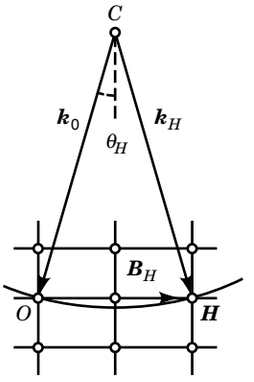


Figure 10. Ewald construction of diffracted wave. A diffracted wave is formed if Ewald's sphere ($\approx$ plane) intersects a reciprocal-lattice node $\boldsymbol{H}$. $\boldsymbol{k}_0$ is the wave vector of the incident wave, $\boldsymbol{k}_H$ is the wave vector of diffracted wave, and $\boldsymbol{B}_H$ is the reciprocal-lattice vector belonging to node $\boldsymbol{H}$.

In electron microscopy the specimen has to be a thin foil since electrons are strongly "absorbed" in solids. This causes an anisotropic relaxation of the Bragg condition along the foil normal. Diffraction now occurs also for conditions deviating somewhat from the exact Bragg angles. This is represented in reciprocal space as a change of the sharp spots into infinitely thin rods, called "relrods" perpendicular to the foil surfaces. These relrods have an intensity profile represented in Fig. 11 according to the kinematical theory of electron diffraction (see the Appendix).

Since the radius of Ewald's sphere $(1/\lambda)$ is large as compared to the mesh size of the reciprocal lattice $(1/d_H)$, the sphere can be approximated by a plane. The diffraction pattern can thus be considered as the projection of an undeformed planar section of the reciprocal lattice.

The distance s_H by which Ewald's sphere misses the reciprocal-lattice node $\boldsymbol{H}$ is called the *excitation error*. It is a vector parallel to the foil normal that connects the reciprocal-lattice

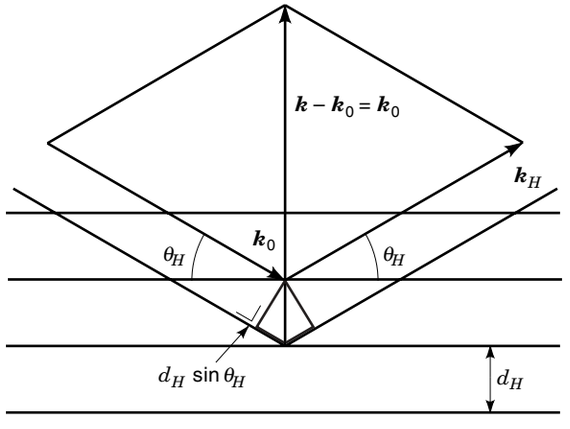


Figure 9. "Reflection" of electron waves by successive lattice planes leading to the Bragg condition.

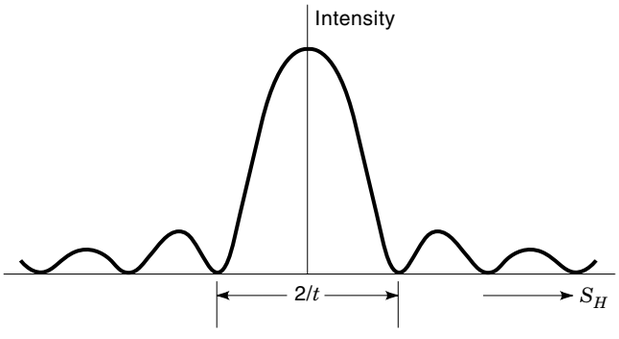


Figure 11. Intensity profile along a "relrod" for a foil with thickness t.

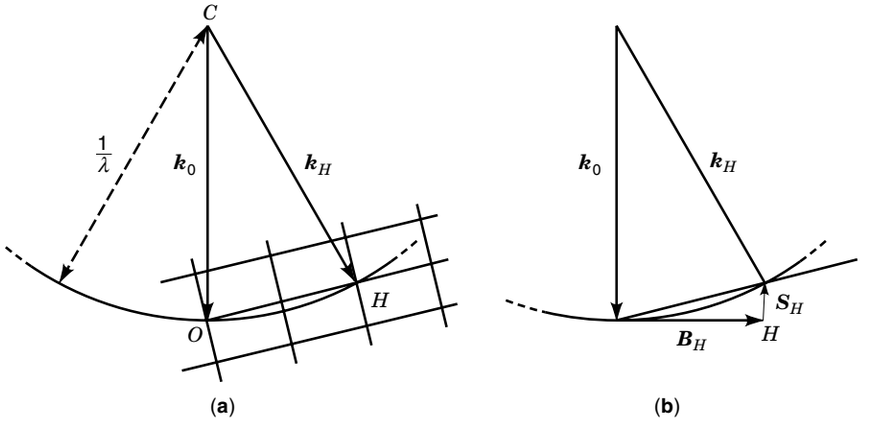


Figure 12. Illustration of the diffraction geometry in reciprocal space, defining the excitation error s_H. (a) $s_H = 0$. (b) $s_H < 0$.

node and the intersection point with Ewald's sphere. It is by convention called positive when pointing in the sense of the propagating electrons, that is, when the node is inside Ewald's sphere. In Fig. 12(b), s_H is negative.

Convergent Beam Electron Diffraction. By converging the incident beam into a cone, the diffraction spots become disks with a fine structure of intensity variations. This way of diffraction is highly sensitive to small changes in the crystal structure. It can be used very efficiently to determine the space group symmetry of a crystal, to measure very accurately the unit cell dimensions, thermal expansion coefficients, local strain, and crystal potentials. For more detail, we refer to Ref. 5.

Imaging

Diffraction-Contrast Imaging

Principles. Diffraction-contrast images are preferably obtained under two-beam conditions since this allows an easier interpretation. The specimen is oriented such that apart from the direct beam only one scattered beam is strong. The selector aperture allows us to select either the direct beam or the intense scattered beam. The selected diffraction spot is then highly magnified by the lenses. The image obtained in this way is a highly magnified map of the intensity distribution in the corresponding diffraction spot. If the direct beam is magnified this is called a *bright-field image* [Fig. 13(a)]; if the diffracted beam is selected, the image is a *dark-field image* [Fig. 13(b)]. Since the electrons propagate within a narrow column by diffracting back and forth, such columns can be considered as picture elements forming the image (Fig. 14). This is the so-called column approximation (6,7).

The amplitude of either the scattered or the transmitted beam emerging in a point on the exit surface of the foil is the sum of the contributions of all volume elements along a column parallel to the incident beam. This amplitude depends on the variation of the excitation error s_H along the column, which itself depends on the local orientation of the diffracting planes. In a perfect foil s is constant and it is the same along all the columns; the image is an area of uniform intensity. If

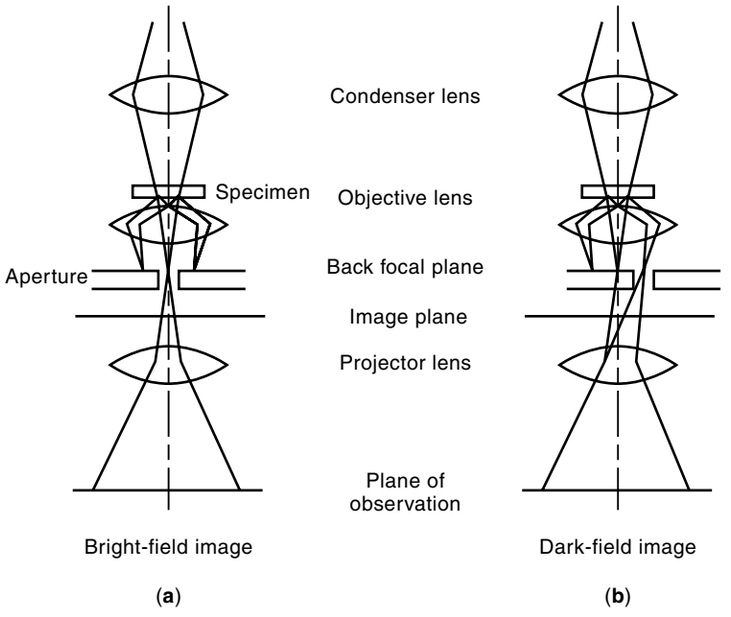


Figure 13. Ray paths for the (a) bright-field mode and (b) dark-field mode.

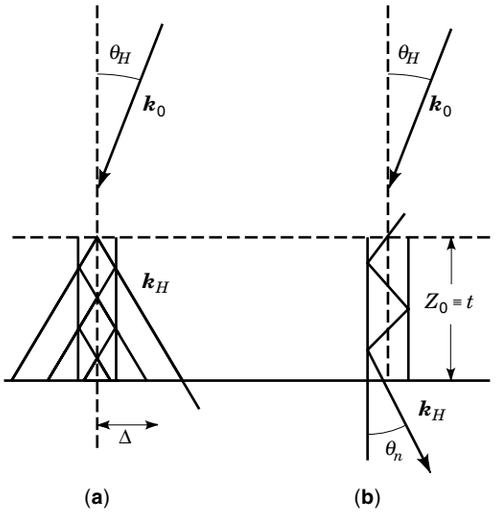


Figure 14. Repeated reflection of electron waves in a column of crystal. According to the (a) kinematical theory and (b) dynamical theory.

a strain pattern is present in the foil s will vary differently along different columns and as a result an image is formed. Diffraction-contrast images are thus very sensitive to strain patterns but they do not reveal the crystal structure.

Examples

Interface Contrast. Interfaces situated in planes inclined with respect to the foil surfaces divide the foil into two wedge-shaped parts. These two parts are related by a translation if the interface is a stacking fault or an antiphase boundary with $\boldsymbol{R} = \boldsymbol{R}_0$; they differ in orientation, that is, in s when the interface is a domain boundary (Fig. 15).

In both cases sets of parallel fringes are formed, which are parallel to the closest intersection line of the interface and the surface. Their intensity profiles are different, however. Planar interfaces parallel to the foil surfaces only cause a brightness difference in the faulted area.

In Fig. 16 the two types of fringes are compared. The most important feature is the nature, bright (B) or dark (D), of the edge fringes (i.e., the first and the last fringe). Bright-field fringe patterns produced at stacking faults have edge fringes of the same nature, whereas in the dark -field fringe pattern the edge fringes have opposite nature. The reverse is true for domain boundary fringes; the edge fringes are of opposite nature in the bright-field image, whereas in the dark-field image the nature of the two edge fringes is the same.

From the nature of the first and last fringes one can conclude, for instance, whether a stacking fault in a face-centered cubic crystal is either *intrinsic* (i.e., of the type . . . *abcabab*c. . .) (8) or *extrinsic* (i.e., of the type . . . *abcabacabc*. . .). Figure 16, right refers to a domain boundary whereas Fig. 19, left is due to a stacking fault.

Dislocation Contrast (9,10). The contrast produced at dislocation lines can be understood by noting that the reflecting lattice planes in the regions on two opposite sides of the dislocation line are tilted in the opposite sense. Hence the Bragg condition for reflection is differently affected on the two sides of the line. On one side, the diffracted intensity may be enhanced because the Bragg condition is better satisfied (s is smaller), whereas it is decreased on the other side because s is larger, leading to a black–white line contrast shown schematically in Fig. 17 for the case of an edge dislocation. In this schematic representation the line thickness is proportional to the local beam intensity. In bright-field images dislocation lines are thus imaged as dark lines, slightly displaced from the actual position of the dislocation line towards the "image side."

This model implies that imaging in reflections associated with families of lattice planes that are not deformed by the presence of the dislocation will not produce a visible line image; the image is then said to be *extinct* (Fig. 18). The extinction condition can to a good approximation be formulated as $\boldsymbol{H} \cdot \boldsymbol{b} = 0$, where $\mathbf{b}$ is the Burgers vector of the dislocation. If extinction occurs for two different sets of lattice planes with normals $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$, the direction of the Burgers vector $\boldsymbol{b}$ is parallel to $\boldsymbol{H}_1 \times \boldsymbol{H}_2$.

Images of dislocations can be simulated accurately by numerically solving the equations that describe the dynamical scattering of the electrons in the object (see the Appendix).

Fast computer programs (10) have been developed to calculate such images for various strain fields and various diffraction conditions. An example of the agreement between the observed and the computed image that can be achieved is shown in Fig. 19 after Ref. 10.

Weak-Beam Images (11). The width of the bright peak that images a dislocation in the dark-field imaging mode decreases with increasing s. This effect is exploited systematically in the weak-beam method, which allows one to image the dislocations as very fine bright lines on a dark background, using a reflection that is only weakly excited, that is, for which s is

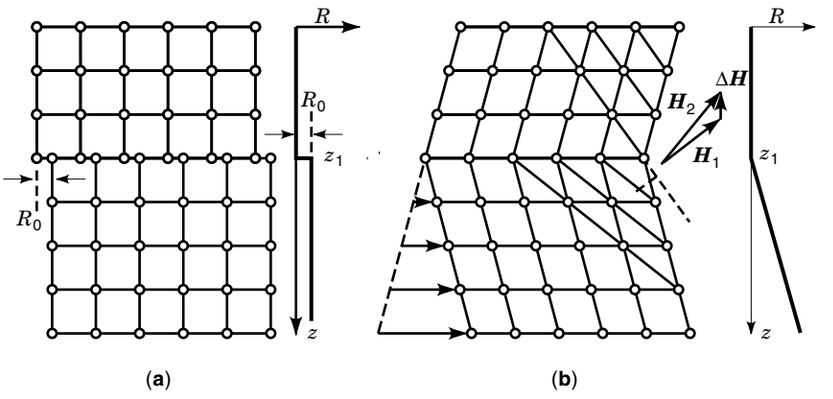


Figure 15. Displacement fields of planar interfaces: (a) stacking fault and (b) domain boundary.

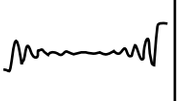
	$\alpha = 2\pi \boldsymbol{H} \bullet \boldsymbol{R}_0$					$\delta = S_1 t_{\boldsymbol{H}_1} - S_2 t_{\boldsymbol{H}_2}$			
	BF		DF			BF		DF	
	F	L	F	L		F	L	F	L
$\sin \alpha > 0$	B	B	B	D	$\delta > 0$	B	D	B	B
$\alpha \neq \pi$	–	–	–	–	–	–	–	–	–
$\sin \alpha > 0$	D	D	D	B	$\delta < 0$	D	B	D	D
									

Figure 16. Comparison of fringe pattern characteristics due to a stacking fault (left) and to a domain boundary (right). The abbreviations F, L, B, and D denote first fringe, last fringe, bright, and dark, respectively.

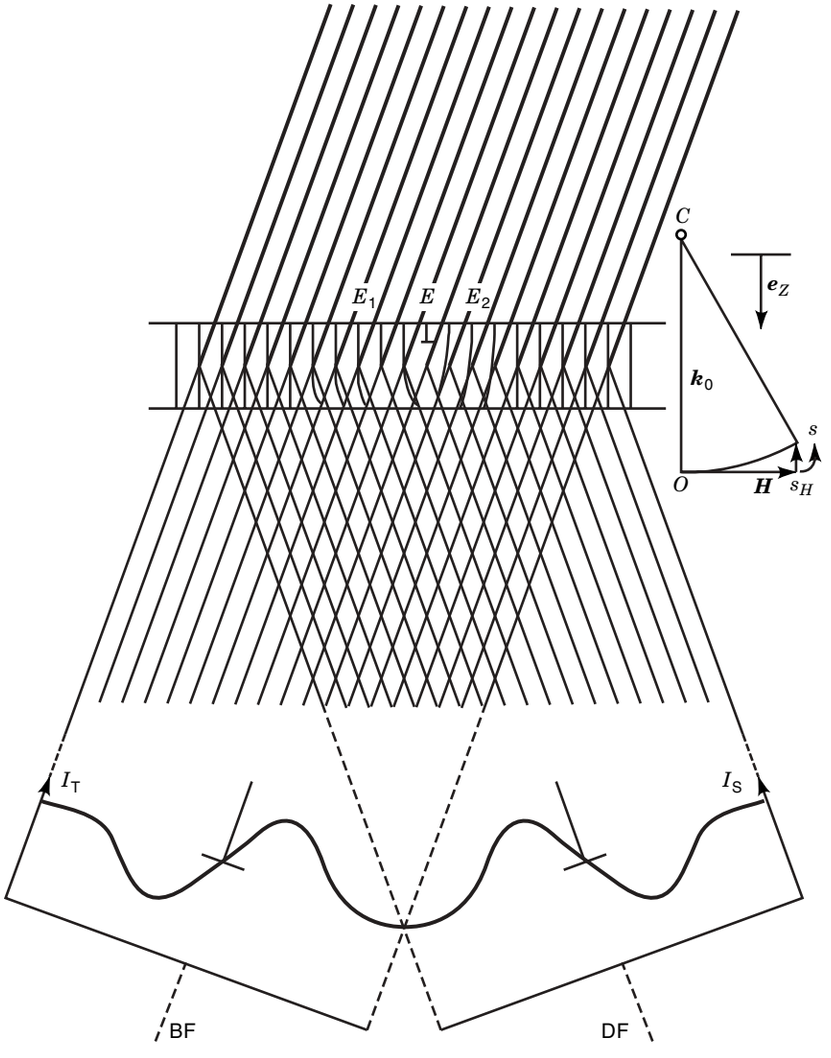


Figure 17. Image formation at an edge dislocation according to the kinematical approximation. The thickness of the lines is a measure for the intensity of the beams. I_T is the intensity of the transmitted beam and I_S of the scattered beam.

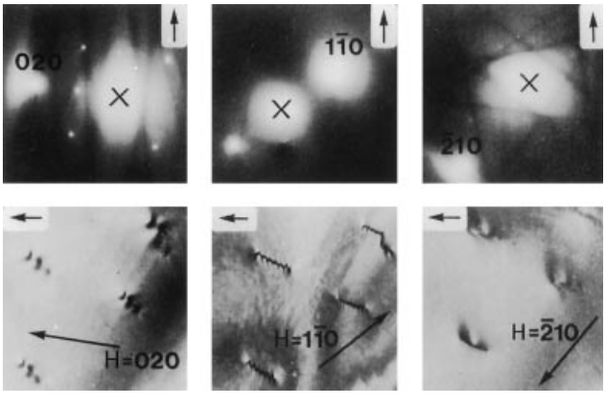


Figure 18. Three images of the same area made under two-beam conditions using three different diffraction vectors $\boldsymbol{H}_1 = 020$, $\boldsymbol{H}_2 = 1\bar{1}0$, $\boldsymbol{H}_3 = \bar{2}10$. Note the extinctions.

large. Unfortunately the contrast is weak and long exposure times are required (Fig. 20).

High-Resolution Imaging

Images of Thin Objects. A very thin object acts as a phase object, in which the phase is proportional to the projected potential along the electron path. The reason is that, in an electrostatic potential, the electron changes speed, which results in a phase shift. Then the exit wave of the object can be written as

$$\psi(\boldsymbol{R}) \approx 1 + i\sigma V_p(\boldsymbol{R}) \tag{10}$$

$V_p(\boldsymbol{R})$ is the projected potential of the object.

In the phase-contrast mode [Appendix A1, Eq. (18)], the phase shift of $\pi/2$ changes i into -1 so that the image intensity is

$$I(\boldsymbol{R}) \approx 1 - 2\sigma V_p(\boldsymbol{R}) \tag{11}$$

The image contrast of a thin object is proportional to its electrostatic potential $V_p(\boldsymbol{R})$ projected along the direction of incidence.

Building-Block Structures. It often happens that a family of crystal structures exists, of which all members consist of a stacking of the simple building blocks but with a different stacking sequence. This is, for instance, the case in mixed-layer compounds, polytypes, and periodical twins, but also periodic interfaces such as antiphase boundaries and crystallographic shear planes can be considered as mixed-layer systems. If the blocks are larger than the resolution of the microscope, each block will show its characteristic contrast. In this way, stacking of the blocks can be directly "read" from the image. The relation between image and structure is called the image code.

An example is shown in Fig. 21 for the case of the binary alloy Au_4Mn in which Au or Mn atoms are located on two sublattices of a single fcc lattice. In this image the Mn atoms are visualized as bright dots. The Au atoms are not visible. This kind of image can be interpreted unambiguously.

Interpretation using Image Simulation. In most cases, however, the image cannot easily be decoded in terms of the object

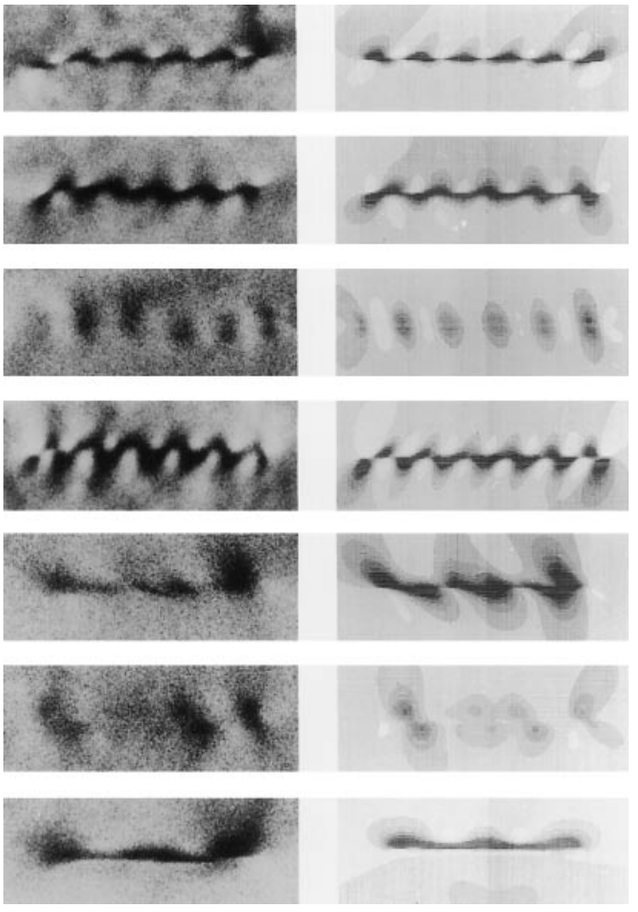


Figure 19. Comparison of observed (left) and computer simulated (right) images of dislocations. Note the excellent correspondence (after Ref. 9).

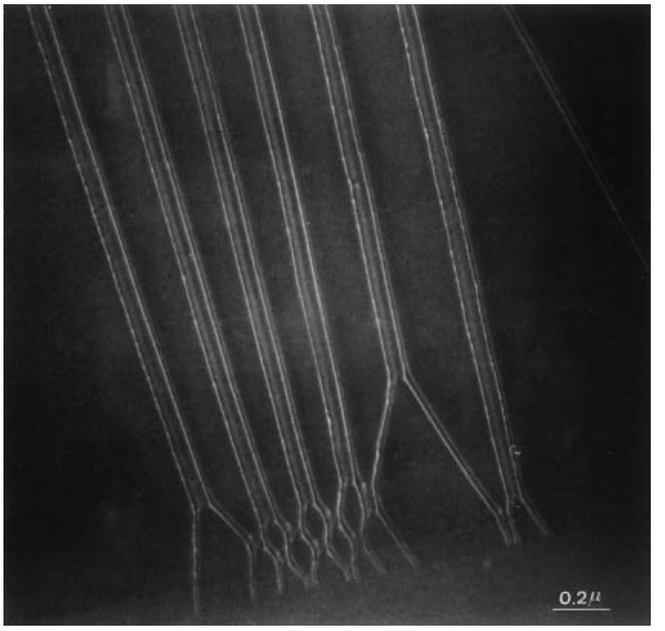


Figure 20. Weak beam image of dislocations in $RuSe_2$.

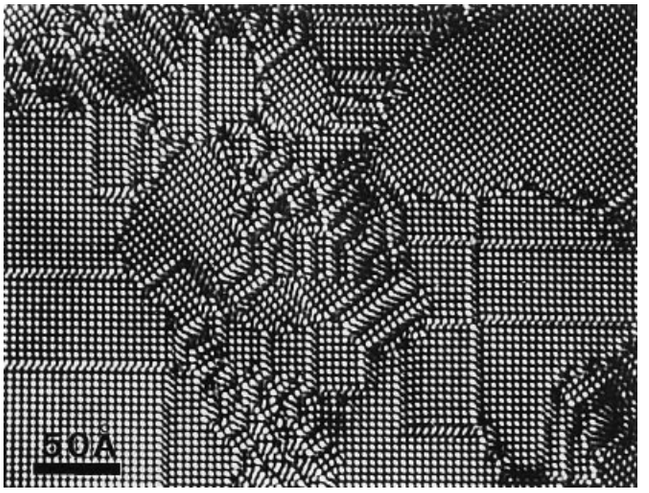


Figure 21. Dark-field superlattice image of Au_4Mn. Orientation and translation variants are revealed. (Courtesy of G. Van Tendeloo.)

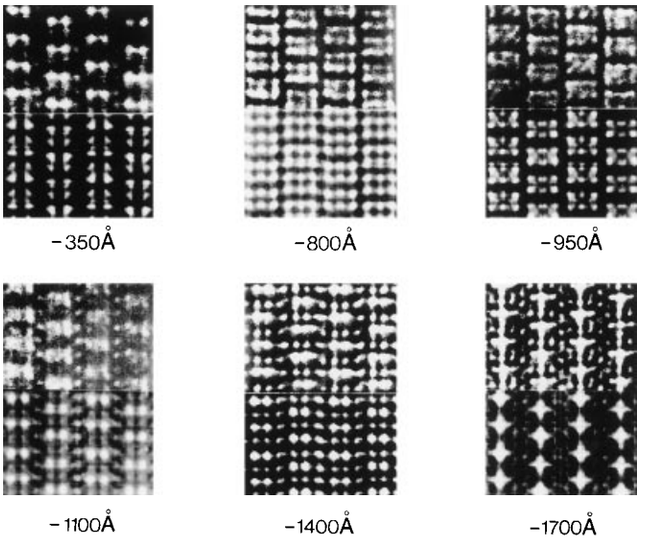


Figure 23. Comparison of experimental images (upper parts) and computer simulated images (lower parts) for $Ti_2Nb_{10}O_{29}$ as a function of defocus. (Courtesy S. Iijima.)

structure, making interpretation difficult, especially at very high resolution, where the image contrast can vary drastically with the focus distance. As a typical and historical example, structure images obtained by Iijima for the complex oxide $Ti_2Nb_{10}O_{25}$ with a point resolution of approximately 0.35 nm are shown in Fig. 22 (upper parts). The structure as reproduced schematically in Fig. 23 consists of a stacking of corner- or face-sharing NbO_6 octahedra with the titanium atoms in tetrahedral positions. High-resolution images are taken at different focus values, causing the contrast to change drastically. The best resemblance with the X-ray structure can be obtained near the optimum Scherzer defocus, which is -90 nm in this particular case. However, the interpretation of such high-resolution images never appears to be trivial. The only solution remains in the comparison of the experimental images with those calculated for various trial structures. The results of such a calculation using the model of Fig. 23 are also shown in Fig. 22 (lower parts) and show a close resemblance with the experimental images. Image simulation, however, is a very tedious technique that has to deal with a number of unknown parameters (specimen thickness, exact focus, beam convergence, etc.). Furthermore, the comparison is often done visually. As a consequence, the technique can only be used if the number of plausible models is very limited.

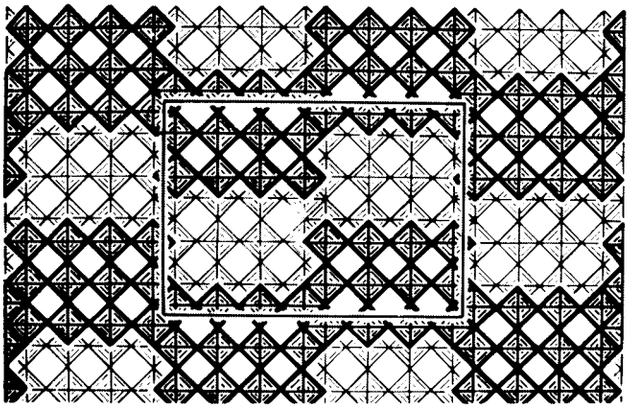


Figure 22. Schematic representation of the unit cell of $Ti_2Nb_{10}O_{29}$ consisting of corner-sharing NbO_6 octahedra with Ti atoms in tetrahedral sites.

Holographic Reconstruction Methods. Due to the complexity of the imaging process, the information about the structure of the object is scrambled in the image. The structural information can be extracted directly from the images using holographic reconstruction methods. These methods aim at undoing the image process, that is, going back from image to object. Such a procedure consists of three steps. First one has to reconstruct the electron wave in the image plane. Then one has to reconstruct the exit wave at the object and from this one has to deduce the projected structure of the object. In a recorded image, which shows only intensities, the phase information is lost. Hence, the reconstruction of the whole image wave is a typical phase problem that can only be solved using holographic methods. Basically two methods are workable. One method is *off-axis holography* (12). Here the electron beam is split in two waves by means of a biprism, which essentially is an electrostatically charged wire. One wave crosses the object so as to produce an enlarged image. The other wave (reference wave) passes by the object through the vacuum and interferes with the image wave in the image plane. In this way the high-resolution image is modulated by the interference fringes. From the position of the fringes one can then determine the phase of the electron wave. The other method is the *focus variation* method (1). The image wave is calculated by computer processing a series of images taken at different focus settings. Figure 24 shows an experimentally reconstructed exit wave for $YBa_2Cu_4O_8$. From this the structure of the object can be deduced.

Quantitative Structure Determination. Ideally quantitative extraction of information should be done as follows. One has a model for the object, for the electron object interaction, for the microscope transfer and for the detection, that is, all the ingredients needed to perform a computer simulation of the

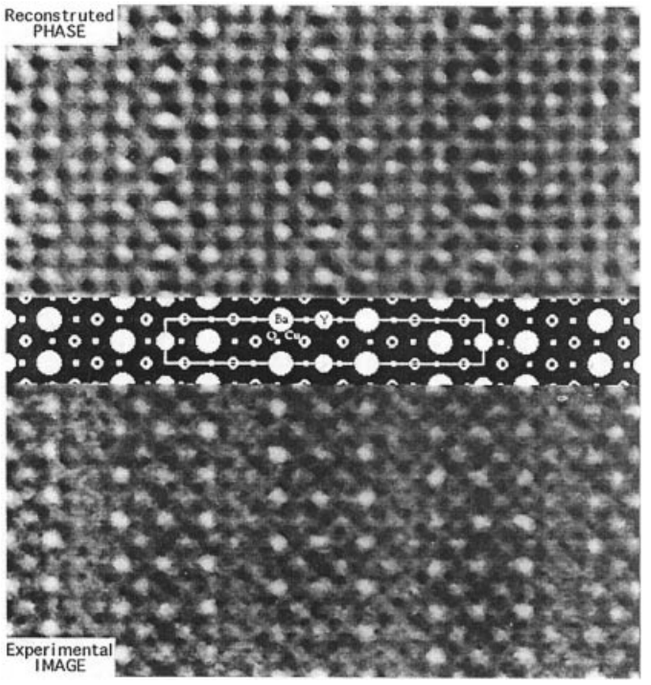


Figure 24. Experimentally reconstructed exit wave for YBa_2Cu_{48}. Top, reconstructed phase; center: structure model; bottom: experimental image.

experiment. The object model that describes the interaction with electrons consists of the assembly of the electrostatic potentials of the constituting atoms. Also the imaging process is characterized by a number of parameters such as defocus, spherical aberration, and voltage. These parameters can either be known a priori with sufficient accuracy or not, in which case they have to be determined from the experiment. The model parameters can be estimated from the fit between the theoretical images and the experimental images. What one really wants is not only the best estimate for the model parameters but also their standard deviation (error bars), a criterion for the goodness of fit, and a suggestion for the best experimental setting. This requires a correct statistical analysis of the experimental data. The goodness of the fit between model and experiment has to be evaluated using a criterion such as likelihood, mean square difference, or R factor (as with X-ray crystallography). For each set of parameters of the model, one can calculate this goodness of fit, so as to yield a fitness function in parameter space.

In principle, the search for the best parameter set is then reduced to the search for optimal fitness in parameter space. This search can only be done in an iterative way as given in the schematic in Fig. 25. First one has a starting model, that is, a starting value for the object and imaging parameters $\{a_n\}$. From these one can calculate the experimental images. This is a classical image simulation. (Note that the experi-

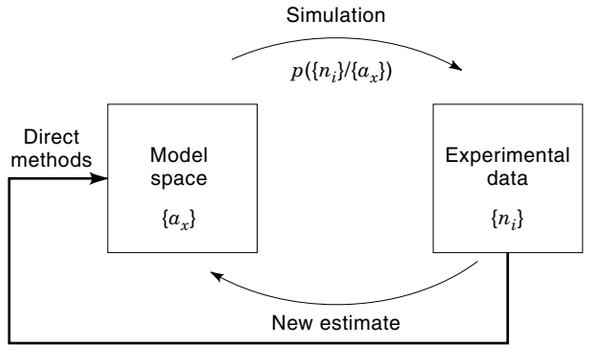


Figure 25. Scheme for the refinement procedure.

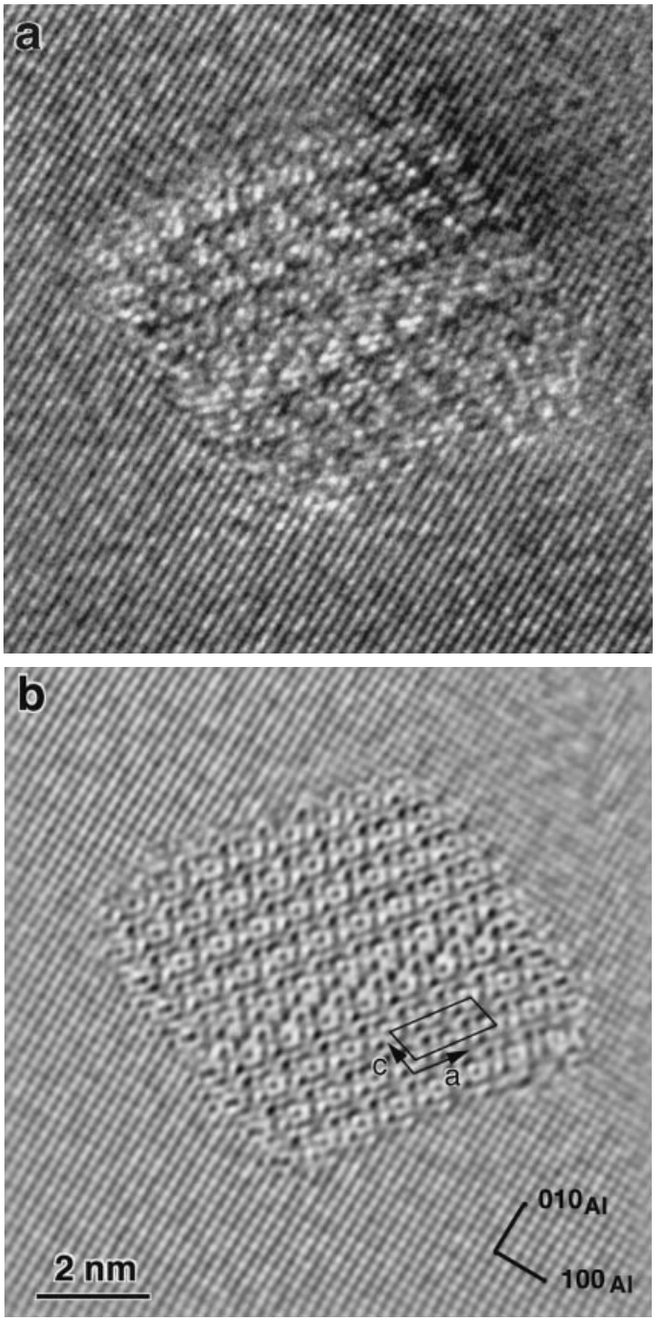


Figure 26. HREM image (a) and phase of the experimentally reconstructed exit wave (b) of an Mg/Si precipitate in an Al matrix. (Courtesy H. Zandbergen (14).

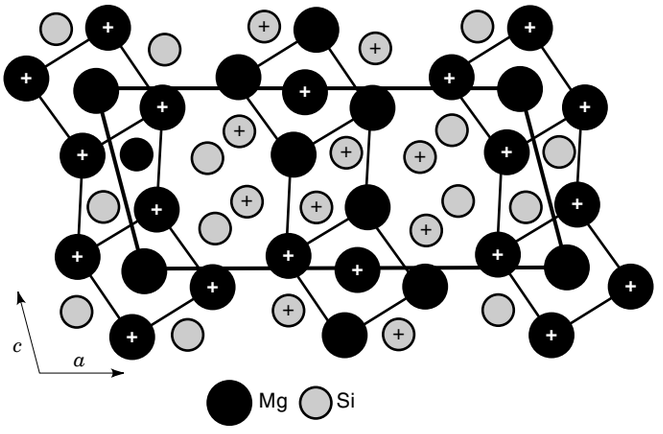


Figure 27. Structure model with MSLS from the fitting procedure described in the text. [Courtesy H. Zandbergen (14).]

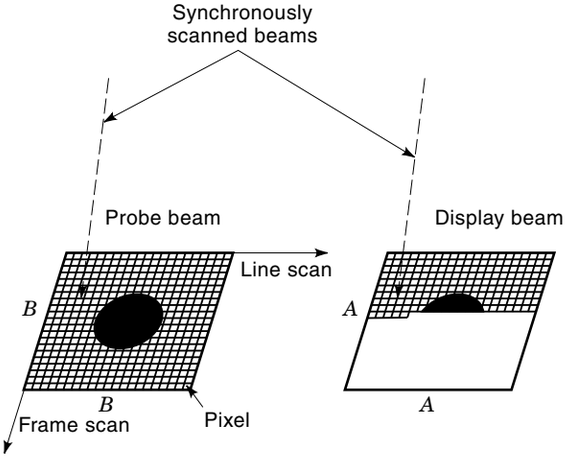


Figure 28. Schematic illustration of the basic mapping principle of the scanning electron microscope. [Courtesy D. Joy (19).]

mental data can also be a series of images and/or diffraction patterns.) From the mismatch between experimental and simulated images one can obtain a new estimate for the model parameters (for instance, using a gradient method), which can then be used for the next iteration. This procedure is repeated until the optimal fitness (i.e., optimal match) is reached.

The refinement procedure needs a good starting model to guarantee convergence. Such a model can be derived from holographic reconstruction. The refinement can also be done using experimental electron diffraction patterns.

An application of such refinement is shown in Figs. 26 and 27. Figure 26 (left) shows an HREM image of a Mg/Si precipitate in an Al matrix (13). Figure 26 (right) shows the phase of the exit wave, which is reconstructed experimentally using the focus variation method. From this an approximate structure model can be deduced. From different precipitates and different zones, electron diffraction patterns could be obtained, which were used simultaneously for a final fitting.

For each diffraction pattern the crystal thickness as well as the local orientation was also treated as a fittable parameter. An overview of the results is shown in Table 1. The obtained R factors are of the order of 5%, which is well below the R factors using kinematical refinement that do not account for the dynamical electron scattering. Figure 27 shows the structure obtained after refinement. Details of this study have been published by Zandbergen et al. (14).

SCANNING ELECTRON MICROSCOPY

The SEM is a mapping, rather than an imaging device (Fig. 28) and so is a member of the same class of instruments as the facsimile machine, the scanning probe microscope, and the confocal optical microscope (19). The sample is probed by a beam of electrons scanned across the surface. Radiations from the specimen stimulated by the incident beam are detected, amplified, and used to modulate the brightness of a second beam of electrons scanned, synchronously with the first beam, across a cathode-ray-tube display. If the area scanned on the display tube is $A \times A$ and the corresponding area scanned on the sample is $B \times B$, then the linear magnification $M = A/B$. The magnification is therefore geometric in origin and may be changed by varying the area scanned on the sample. The arrangement makes it possible for a wide range of magnifications to be obtained and allows rapid changes of magnification since no alterations to the electron-optical system are required. There is no rotation between object and image planes, and once the instrument has been focused on a given area the focus need not be changed when the magnification is varied. To a first approximation the size of the finest detail visible in the image will be set by the size of the probe scanning the specimen. Multiple detectors can be used to collect several signals simultaneously that can then be displayed individually or combined in perfect register with each other. It is this probability in particular that makes the

Table 1. Results of Structure Refinement Using Electron Diffraction Data

Zone	Number of Observed Reflections	Thickness (nm)	Crystal Misorientation			R Value (%)	
			h	k	l	MSLS	Kinematic
[010]	50	6.7(5)	8.3	0	−2.3	3.0	3.7
[010]	56	15.9(6)	2.6	0	−1.8	4.1	8.3
[010]	43	16.1(8)	−1.7	0	0.3	0.7	12.4
[010]	50	17.2(6)	−5.0	0	−1.0	1.4	21.6
[010]	54	22.2(7)	−5.9	0	2.5	5.3	37.3
[001]	72	3.7(3)	−3.9	4.5	0	4.1	4.5
[001]	52	4.9(6)	3.6	−1.9	0	6.8	9.3

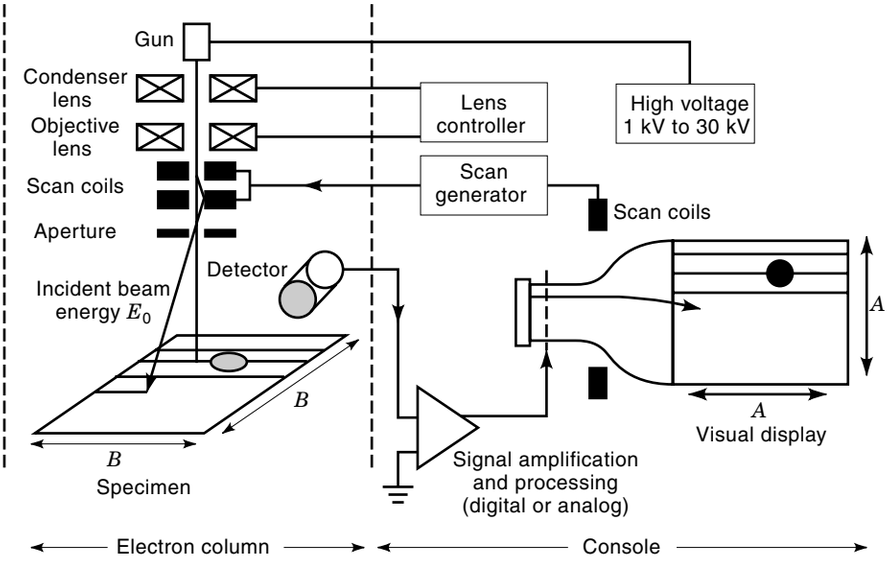


Figure 29. Basic components of the scanning electron microscope. [Courtesy D. Joy (19).]

SEM so useful a tool since multiple views of a sample, in different imaging modes, can be collected and compared in a single pass of the beam.

Figure 29 shows the basic components of a SEM. These can be divided into two main categories: the electron-optical and detector systems and the scanning processing and display systems. The electron-optical components are often described as being the "column" of the instrument while the other items are the "console" of the machine. The source of electrons is the gun, which produces electrons either by thermal emission, from tungsten or lanthanum hexaboride cathodes, or from a field-emission source. These electrons are then accelerated to an energy in the range from 500 eV to 30 keV. The beam of electrons leaving the gun is then focused onto the specimen by one or more condenser lenses. Although either electrostatic or electromagnetic lenses could be employed all modern SEMs use electromagnetic lenses. Typically, the final objective lens has been of the pinhole design with the sample sitting outside the field of the lens since this arrangement gives good physical access to the specimen. However, in this arrangement the specimen is 10 mm to 20 mm away from the lens, which must therefore be of long focal length and correspondingly high aberration coefficients. In modern, high-performance instruments it is now common to use an immersion lens (15), in which the sample sits inside the lens at the center of the lens field, or a "snorkel" lens (16) in which the magnetic field extends outside of the lens to envelop the sample. Although the immersion lens gives very good performance and, by making the sample part of the lens structure, ensures mechanical stability, the amount of access to the specimen is limited. The snorkel lens, on the other hand, combines both good electron-optical characteristics with excellent access for detectors and stage mechanisms.

The coils that scan the beam are usually incorporated within the objective lens. A double-scan arrangement is often employed in which one sets of coils scans the beam through some angle θ from the axis of the microscope while a second set scans the beam through an angle 2θ in the opposite direction. In this way all scanned beams pass through a single point on the optic axis allowing for the placement of a defining aperture without any constriction of the scanned area. The scan pattern, produced on the specimen, is usually square in shape and is made up of 1000 horizontal lines, each containing 1000 individual scanned points or *pixels*. The final image frame thus contains 10^6 pixels, although for special activities such as focusing or alignment frames containing only 256 $\times$ 256 pixels may be used.

Increasingly the detector output is passed through an analog-to-digital convertor (ADC) and then handled digitally rather than as an analog video signal. This permits images to be stored, enhanced, combined, and analyzed using either an internal or an external computer. While the majority of the images are still recorded onto photographic film, digital images can be stored directly to magnetic or magneto-optic disks, and hard-copy output of the images can then be obtained using laser or dye sublimation printers. Typically scan repetition rates ranging from 15 or 20 frames/s ("TV rate") to one frame in 30 s to 60 s ("photographic rate") are provided. In addition individual pixels or arrays of pixels within an image field may be accessed if required.

In the case of the SEM the attainable resolution is determined by a number of factors, including the diameter of the electron-beam probe that can be generated, the current I_b contained in that probe, the magnification of the image, and the type of imaging mode that is being used. Over most of the operating energy range (5 keV to 30 keV) of the SEM, the probe size and beam current are related by an expression of the form (17)

$$d = C_s^{1/4}\lambda^{3/4}\left(1 + \frac{I_b}{\beta\lambda^2}\right)^{3/8} \tag{12}$$

where λ is the wavelength of the electrons ($\lambda \approx 1.226E_0^{-1/2}$ nm, where E_0 is the incident electron energy in eV), β is the brightness of the electron gun in $A \cdot cm^{-2} \cdot sr^{-1}$, and C_s is the spherical aberration coefficient of the objective lens.

Finally, if the gun brightness is further increased to 10^8 $A \cdot cm^{-2} \cdot sr^{-1}$ by using a field-emission source (18), then the factor is close to unity for both modes of operation considered.

For a modern SEM C_s is typically a few millimeters; thus minimum probe sizes of 1 nm or 2 nm are available. At low beam energies (below 5 keV) additional effects including the energy spread of electrons in the beam must also be considered, but the general conclusions discussed previously remain correct.

Modes of Operation

Secondary-Electron Imaging. Secondary electrons (SE) are those electrons emitted by the specimen, under irradiation by the beam, which have energies between 0 eV and 50 eV. Because of their low energy the SE only travel relatively short distances in the specimen (3 mm to 10 mm) and thus they emerge from a shallow "escape" region beneath the surface. There are two cases in which an SE can be generated and subsequently escape from the specimen: first, when an incident electron passes downward through the escape depth, and second as a backscattered electron leaves the specimen and again passes through the escape region. Secondary electrons produced in the first type of event are designated SE1, and because they are generated at the point where the incident beam enters the specimen, it is these that carry high-resolution information. The other secondary electrons are called SE2, and these come from a region the size of which is of the order of the incident beam range in the sample.

Secondary-electron imaging is the most common mode of operation of the SEM. The reason for this is that secondary electrons are easy to collect and they carry information about the surface topography of the specimen. Information about surface chemistry and magnetic and electric fields may also be obtainable on suitable specimens. SE images can usually be interpreted readily without specialized knowledge and they yield a spatial resolution of 1 nm or better. Examples of typical SE images are shown in Figs. 30 and 31. The light and shadow effects together with the very large depth of focus enhance the 3D aspects of the surface structure.

Another imaging mode is voltage constrast, which is illustrated in Fig. 32. Here large regions of uniform bright and dark contrast correspond to regions that have a negative and positive voltage with respect to ground.

Backscattered Electrons. Backscattered electrons (BSE) are defined as being those electrons emitted from the specimen with energies between 50 eV and the incident beam energy E_0. Because the yield of BSE varies with the atomic number of the specimen the contrast of the images is related to the atomic number of the object.

Figure 31. High-resolution image of magnetic disk media surface recorded at 30 keV in a JEOL JSM 890 field-emission SEM. [Courtesy D. Joy (19).]

Other Imaging Modes. With a SEM it is possible to measure the current through the object as induced by the imaging electron beam [electron-beam-induced current (EBIC)]. This signal gives information about the electron-hole pair carriers in a semiconductor such as those at p–u junctions. In cathode luminescence, one detects the fluorescence radiation that is due to irradiation by the incident beam. This is a very sensitive technique that gives information about the impurities in semiconductors.

For more information we refer the reader to Ref. 19.

SCANNING TRANSMISSION ELECTRON MICROSCOPY

In principle a STEM can be considered as a SEM in which the object is transparent for the high energy electrons and in which the detector is placed behind the object. As in a STEM,

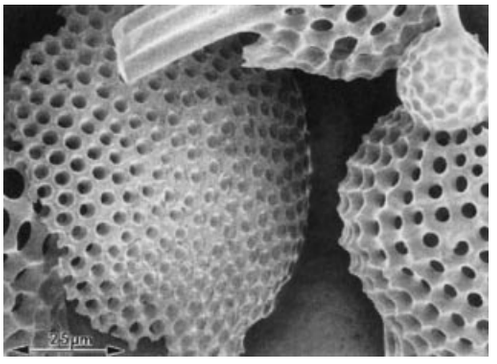


Figure 30. Secondary-electron images of radiolarium. Recorded in Hitachi S-4500 field emission SEM at 5 keV beam energy. Magnification: 800×. [Courtesy D. Joy (19).]

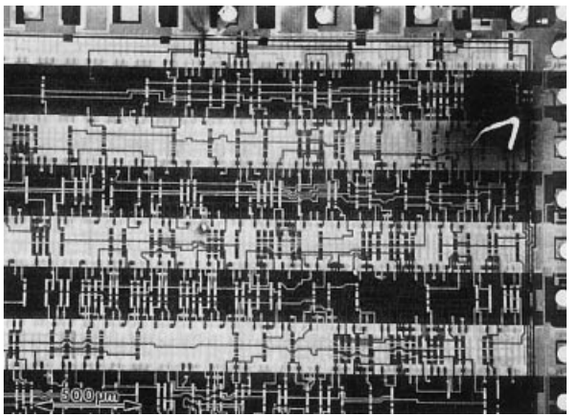


Figure 32. Voltage contrast from an integrated circuit. Recorded at 5 keV in a Hitachi S-800 field-emission SEM. [Courtesy D. Joy (19).]

a fine electron probe, formed by using a strong objective electron lens to demagnify a small source, is scanned over the specimen in a two-dimensional raster [Fig. 33(a)]. The electron probe is necessarily convergent: the convergence angle is, ideally, inversely proportional to the minimum probe size that determines the microscope resolution. On any plane after the specimen, a convergent beam electron-diffraction pattern is formed. Some part of this diffraction pattern is collected in a detector, creating a signal, which is displayed on a cathode-ray-tube screen to form the image using a raster scan matched to that which deflects the incident electron beam (20,21).

Dark-field images, obtained with an annular detector in a STEM instrument, showed the first clear electron microscopy images of individual heavy atoms (22). From that time, STEM has developed as an important alternative to conventional, fixed-beam transmission electron microscopy (CTEM), with special advantages for many purposes.

The use of a field emission gun (FEG) for high-resolution STEM is necessary to provide sufficient signal strength for viewing or recording images in a convenient time period. Because the FEG source has a brightness that is a factor of 10^4 or 10^3 greater than that of a W hairpin filament, the total current in the electron beam is greater when beam diameters of less than about 10 nm are produced. The current in a beam of 1 nm diameter is typically about 1 nA. As suggested by Fig. 33(b), the essential components of a STEM imaging system are the same as those for a CTEM instrument, with the electrons traveling in the opposite direction. In this diagram condensor and projector lenses have been omitted, and only the essential objective lens, which determines the imaging characteristics, is included. The STEM detector replaces the CTEM electron source. The STEM gun is placed in the detector plane of the CTEM, and the scanning system effectively translates the STEM source to cover the CTEM recording plate. When one uses a detector with a hole to eliminate the unscattered electron beam, the imaging is effectively incoherent so that the image contrast can be interpreted directly in terms of the atomic number of the constituting atoms. This imaging mode is therefore called Z contrast imaging (20). Figure 34 shows a STEM image of a tilt boundary in silicon in which the local atomic configuration can be seen directly in the images.

The strength of STEM as compared to TEM is that a variety of signals may be obtained in addition to the bright-field or dark-field signals derived from the elastic scattering of electrons in the specimen. STEM instruments are visually fitted with an energy-loss spectrometer. Energy filtered images reveal compositional information. For more information we refer to Refs. 20 and 24.

APPENDIX A. ELECTRON-DIFFRACTION THEORIES

Phase Object Approximation

We will now follow a classical approach. The nonrelativistic expression for the wavelength of an electron accelerated by an electrostatic potential E is given by

$$\lambda = \frac{h}{\sqrt{2meE}} \tag{A.1}$$

with h the Planck constant, m the electron mass and e the electron charge.

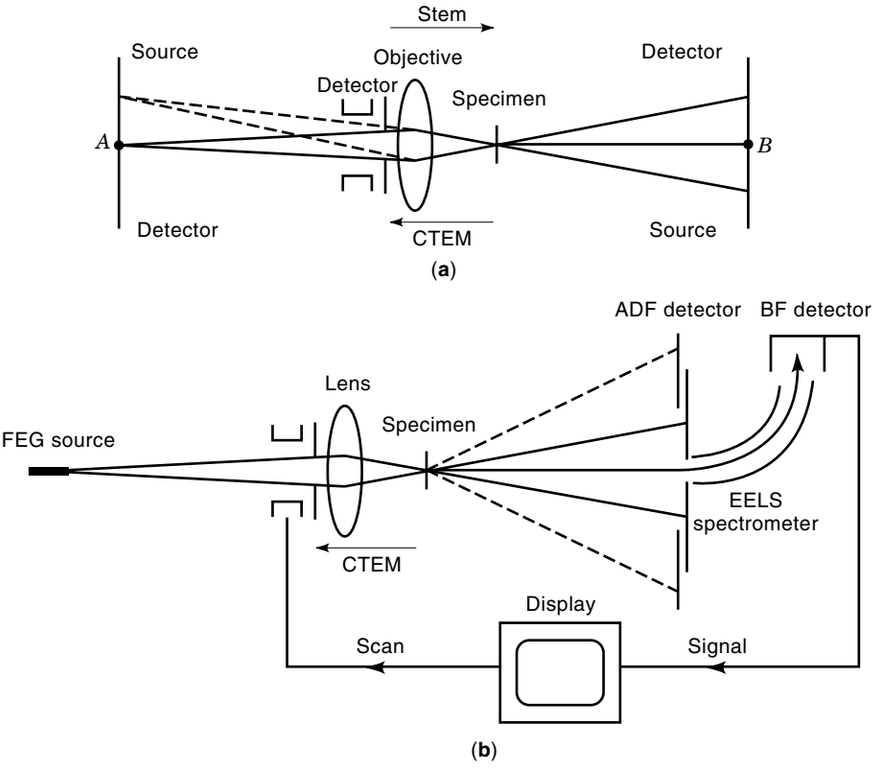


Figure 33. (a) Diagram of the essential components of a STEM instrument. (b) Diagram suggesting the reciprocity relationship between STEM (electrons going from left to right) and CTEM (electrons going from right to left). [Courtesy J. Cowley (20).]

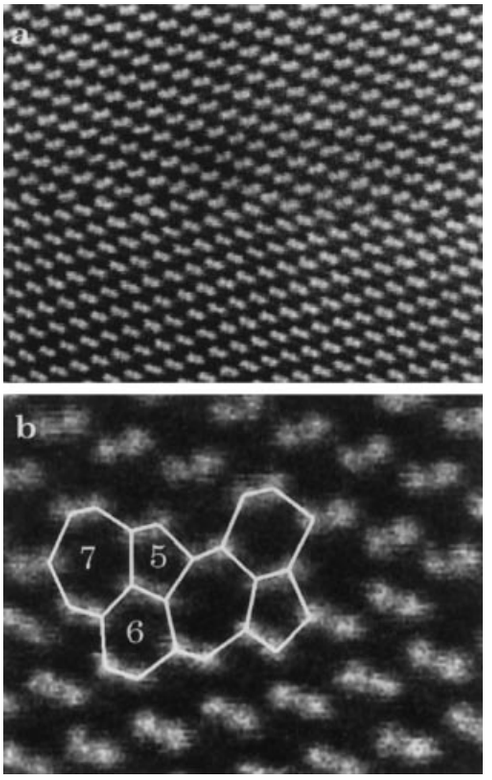


Figure 34. Σ = 9,{221}(100) symmetric tilt boundary in silicon viewed along the [110] direction showing its five- and seven-membered ring structure. ADF = annular dark field; EELS = electron energy loss spectroscopy. [Courtesy S. Pennycook (21).]

During the motion through an object with local potential $V(x, y, z)$ the wavelength will vary with the position of the electron as

$$\lambda^1(x, y, z) = \frac{h}{\sqrt{2me[E + V(x, y, z)]}} \tag{A.2}$$

For thin phase objects and large accelerating potentials the assumption can be made that the electron keeps traveling along the z direction so that by propagation through a slice dz the electron suffers a phase shift.

$$\begin{aligned} d_\phi(x, y, z) &= 2\pi \frac{dz}{\lambda^1} - 2\pi \frac{dz}{\lambda} \\ &= 2\pi \frac{dz}{\lambda} \left(\frac{\sqrt{E + V(x, y, z)}}{\sqrt{E}} - 1 \right) \\ &\simeq \sigma V(x, y, z)\, dz \end{aligned}$$

with

$$\sigma = \pi / \lambda E \tag{A.3}$$

Therefore the total phase shift is given by

$$\phi(x, y) = \sigma \int V(x, y, z)\, dz = \sigma V_\mathrm{p}(x, y) \tag{A.4}$$

where $V_\mathrm{p}(x, y)$ represents the potential of the specimen projected along the z direction.

Under this assumption the specimen acts as a pure phase object with transmission function

$$\psi(x, y) = \exp[i\sigma V_\mathrm{p}(x, y)] \tag{A.5}$$

In case the object is very thin, one has

$$\psi(x, y) \approx 1 + i\sigma V_\mathrm{p}(x, y) \tag{A.6}$$

This is the weak-phase approximation.

The effect of all processes prohibiting the electrons from contributing to the image contrast, including the use of a finite aperture, can in a first approximation be represented by a projected absorption function in the exponent of Eq. (A.5) so that

$$\psi(x, y) = \exp[i\sigma V_\mathrm{p}(x, y) - \mu(x, y)] \tag{A.7}$$

or

$$\psi(\boldsymbol{R}) = \exp[i\sigma V_\mathrm{p}(\boldsymbol{R}) - mu(\boldsymbol{R})] \tag{A.8}$$

with $\boldsymbol{R} = (x, y)$ the vector in the plane perpendicular to z.

Kinematical Theory

As follows from Eq. (1) a diffraction pattern can be calculated from the Fourier transforms of the exit wave $\psi(\boldsymbol{R})$. However, even for a simple approximation such as Eq. (A.8) the Fourier transform is not expressed in a simple analytical form. In order to derive a simpler, albeit approximated, expression for the diffraction pattern it is more convenient to describe the diffraction process directly in Fourier space.

According to the kinematical diffraction theory electrons are scattered in the specimen only and moreover the incident beam is not depleted by scattering. Each atom (scattering center) thus sees the same incident beam amplitude. This approximation is excellent in neutron diffraction, justified in X-ray diffraction, but it is poor in electron diffraction because the atomic scattering cross sections for electrons are relatively much larger than those for the other forms of radiation. The kinematical approximation is therefore only applicable to very thin crystals (a few nanometers for most materials) or for very large deviations from the exact Bragg condition (large s). It allows one to compute the amplitude of the diffracted beam only since the incident beam remains undepleted. Qualitative conclusions from the kinematical theory are nevertheless usually in agreement with the observations.

A crystal is made up of identical unit cells, regularly arranged at the basic lattice nodepoints given by

$$\boldsymbol{A_L} = l_1\boldsymbol{a}_1 + l_2\boldsymbol{a}_2 + l_3\boldsymbol{a}_3 \tag{A.9}$$

(where l_j is an integer). In each unit cell, a number N of atoms is found at the relative positions $\boldsymbol{\rho}_k$ (k = 1, . . . , N). Mathematically speaking, the whole crystal is made up by convolution of one unit cell with the basic crystal lattice. Atom positions are thus $\boldsymbol{r}_j = \boldsymbol{A_L} + \boldsymbol{\rho}_k$ and they depend on four indices l_1, l_2, l_3, and k.

Let $\boldsymbol{k}_0$ represent the wave vector of the incident wave and $\boldsymbol{k}$ that of the diffracted wave; then at large distance, i.e., in the Fraunhofer approximation, the phase difference between a wave diffracted by an atom at the origin and an atom at $\boldsymbol{r}_j$

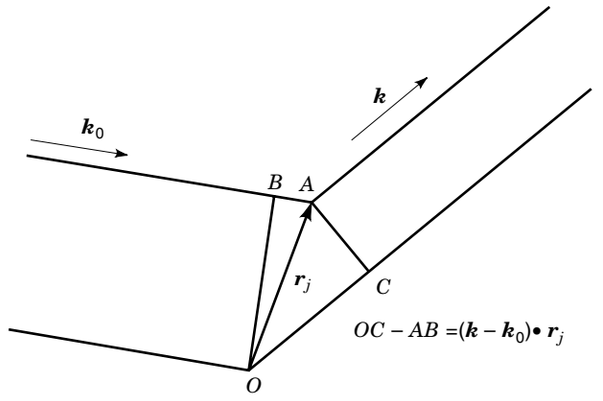


Figure 35. Illustrating the path difference $OC - AB$ between waves diffracted by an atom at the origin and an atom at $\boldsymbol{r}_j$.

is given by $2\pi(\boldsymbol{k} - \boldsymbol{k}_0)\cdot\boldsymbol{r}_j$ and the scattered amplitude $A(\boldsymbol{k})$ along the direction of $\boldsymbol{k}$ (Fig. 35)

$$A(\boldsymbol{k}) = \sum_j f_j \exp[2\pi i(\boldsymbol{k} - \boldsymbol{k}_0)\cdot\boldsymbol{r}_j] \tag{A.10}$$

This amplitude will exhibit maxima if all exponents are integer multiples of $2\pi i$; maxima will thus occur if $(\boldsymbol{k} - \boldsymbol{k}_0)\cdot\boldsymbol{r}_L =$ integer, which implies that $\boldsymbol{k} - \boldsymbol{k}_0$ must be a reciprocal-lattice vector

$$\boldsymbol{k} - \boldsymbol{k}_0 = \boldsymbol{B}_H \equiv h_1\boldsymbol{b}_1 + h_2\boldsymbol{b}_2 + h_3\boldsymbol{b}_3 \tag{A.11}$$

(where h_j are integers and $\boldsymbol{b}_i$ are base vectors). This is Ewald's condition as discussed in the section on electron diffraction.

However, $A(\boldsymbol{k})$ will also be different from zero if the diffraction condition is not exactly satisfied, that is, if Ewald's sphere misses closely a reciprocal-lattice node by a vector $\boldsymbol{s}$, called the excitation error. This vector is parallel to the foil normal and connects the reciprocal-lattice node with the intersection point with Ewald's sphere; by convention s is positive when the reciprocal-lattice node is inside Ewald's sphere and negative when outside (Fig. 7). One can now set

$$\boldsymbol{k} - \boldsymbol{k}_0 = \boldsymbol{B}_\mathrm{H} + \boldsymbol{s} \tag{A.12}$$

and

$$A_H = \sum_j f_j \exp[2\pi i(\boldsymbol{B}_H + \boldsymbol{s})\cdot\boldsymbol{r}_j] \tag{A.13}$$

and with $\boldsymbol{r}_j = \boldsymbol{A}_L + \boldsymbol{\rho}_k$.

$$A_H = \sum_L \sum_k f_k \exp[2\pi i(\boldsymbol{B}_H + \boldsymbol{s})\cdot(\boldsymbol{A}_L + \boldsymbol{\rho}_k)] \tag{A.14}$$

Neglecting $\boldsymbol{s}\cdot\boldsymbol{\rho}_k$ as compared to the other terms and noting that $\boldsymbol{B}_H\cdot\boldsymbol{A}_L$ is always an integer, this can be written as

$$A_H = F_H \sum_L \exp(2\pi i\mathbf{s}\cdot\mathbf{A}_L) \tag{A.15}$$

where the structure factor F_H is defined as

$$F_H = \sum_k f_k \exp[2\pi i(\boldsymbol{B}_H\cdot\boldsymbol{\rho}_k)] \tag{A.16}$$

Equation (A.15) is in fact a triple sum over the indices $\boldsymbol{L}(l_1, l_2, l_3)$. If $\boldsymbol{s}$ is written as a vector in terms of the reciprocal-space base vectors b_j:

$$\boldsymbol{s} = s_1\boldsymbol{b}_1 + s_2\boldsymbol{b}_2 + s_3\boldsymbol{b}_3 \tag{A.17}$$

one has

$$\boldsymbol{s}\cdot\boldsymbol{A}_L = l_1s_1 + l_2s_2 + l_3s_3 \tag{A.18}$$

The triple sum can be expressed as the product of three single sums of geometrical progressions. Calling N_1, N_2, and N_3 the numbers of unit cells along the three lattice directions $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$, one obtains, neglecting an irrelevant phase factor

$$\begin{aligned} A_H &= F_H \sum_{l_1=0}^{N_1-1}\sum_{l_2=0}^{N_2-1}\sum_{l_3=0}^{N_3-1} \exp[2\pi i(l_1s_1 + l_2s_2 + l_3s_3)] \\ &= F_H \frac{\sin\pi s_1N_1}{\sin\pi s_1}\frac{\sin\pi s_2N_2}{\sin\pi s_2}\frac{\sin\pi s_3N_3}{\sin\pi s_3} \end{aligned} \tag{A.19}$$

This is the well-known *von Laue interference function* (10) (Fig. 36), which describes the dependence of the scattered amplitude on the deviation parameter s. The sine functions in the denominators can be approximated by their arguments, since these are always small. We further note that for large N one has $\sin(\pi N_s)/N\pi_s \sim \delta(s)$ with $\delta(s) = 0$ for $s \neq 0$ and $\delta(s) = 1$ for $s = 0$. We can then write, neglecting irrelevant phase factors,

$$A_H = F_H\delta(s_1)\delta(s_2)\delta(s_3)(\Omega/V_a) \tag{A.20}$$

where Ω is the volume of the crystal and V_a the volume of the unit cell ($\Omega = N_1N_2N_3a_1a_2a_3$; $V_a = a_1a_2a_3$). For a parallelopiped-shaped crystal block one often introduces the components of $\boldsymbol{s}$ along the three mutually perpendicular edges of the block with unit vectors $\boldsymbol{e}_x$, $\boldsymbol{e}_y$, and $\boldsymbol{e}_z$:

$$\boldsymbol{s} = s_x\boldsymbol{e}_x + s_y\boldsymbol{e}_y + s_z\boldsymbol{e}_z \tag{A.21}$$

One can then rewrite Eq. (A.19) in terms of s_x, s_y, and s_z as

$$A_H = F_H \frac{\sin\pi s_xN_1a_1}{\sin\pi s_xa_1}\frac{\sin\pi s_yN_2a_2}{\sin\pi s_ya_2}\frac{\sin\pi s_zN_3a_3}{\sin\pi s_za_3} \tag{A.22}$$

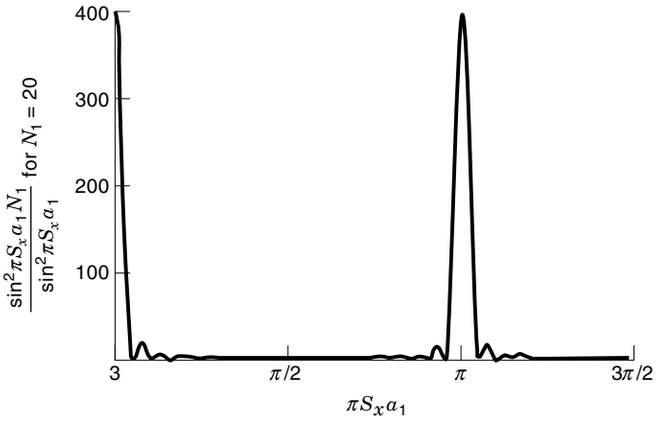


Figure 36. von Laue interference function describing the dependence of the scattered intensity on the excitation error s_x.

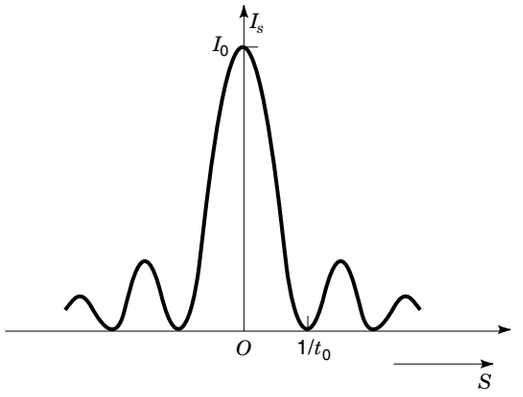


Figure 37. Rocking curve for a foil with thickness t_0 according to the kinematical theory.

Hereby use was made of the relations valid for a parallelopiped

$$s_1 = s_x a_1, \qquad s_2 = s_y a_2, \qquad s_z = s_z a_3 \tag{A.23}$$

Equation (A.22) is only true if N_1, N_2, and N_3 are sufficiently large. However, the foils used in transmission electron microscopy are only large in two dimensions, that is, along x and y, the foil being only a small number N_3 of unit cells thick along z. In such a foil one thus obtains

$$A_{\boldsymbol{H}} = F_{\boldsymbol{H}} \frac{\Omega}{V_a} \delta(s_x)\delta(s_y) \frac{\sin \pi s_z N_3 a_3}{\pi s_z a_3} \tag{A.24}$$

Introducing the specimen thickness $N_3 a_3 = t$ and assuming s_x, $s_y = 0$ and calling $s_z = s$ one finds

$$A_H = \frac{\sin \pi s t}{s t_H} \tag{A.25}$$

per surface area where $t_H = \pi/F_H$; t_H is called the *extinction distance*. This result is interpreted as meaning that the sharp reciprocal-lattice nodes, characteristic of a large crystal, become rods in the case of a thin plate, as already mentioned before. These rods are perpendicular to the foil plane and have a weight profile given by $\sin(\pi s t)/s t_{\boldsymbol{H}}$. The corresponding intensity is given by (Fig. 37)

$$I_H = \frac{\sin^2 \pi s t}{(s t_{\boldsymbol{H}})^2} \tag{A.26}$$

it is called the *rocking curve*. An intensity can be associated with each intersection point of the Ewald sphere with this rod (called relrod), that is, with each s value, the intensity being given by the value of the function at the intersection point.

Another way to interpret these results is the following: In the Fraunhofer approximation, the diffracted wave is expressed by the Fourier transform of the electrostatic potential of the crystal.

A crystal can be considered as a convolution product of two factors: The convolution theorem then states that the diffracted wave is given by the product of the respective Fourier transforms of these two factors [e.g., Eq. (A.15)]. The Fourier transform of the lattice function yields delta functions at the reciprocal nodepoints, which describe the directions of the diffracted beams. The amplitudes of these beams are then given by the Fourier transforms of the potential of one unit cell, i.e., the structure factors [e.g., Eq. (A.20)]. The conservation of energy also requires that the wavevectors of the diffracted beams should all have constant length, or, that the reciprocal nodes should lie on a sphere, the Ewald sphere. In case the object is a thin crystal slab, it can be described as the product of an infinite crystal with a slab function that is equal to 1 inside the slab and 0 elsewhere. In that case, the diffraction pattern is given by the convolution product of the diffraction pattern of the infinite crystal with the Fourier transform of the slab function. Then each reciprocal node is smeared along a line perpendicular to the slab, with a factor given by a sinc function of the form in Eq. (A.26).

Note that if the Ewald sphere would be flat, the diffracted wave can be derived from the Fourier transform of Eq. (A.8) provided the phase is weak. This means that, apart from the Ewald sphere, the weak phase object theory and the kinematical theory are equivalent.

Equation (A.15) can also be understood intuitively in terms of the column approximation along the z direction. The amplitude of the wave diffracted by the volume element Δz_n at level z_n in the column (measured from the entrance face) is given by $\Delta A_H = F_H \Delta z_n \exp(2\pi i s z_n)$ or in differential form

$$dA_H = F_H \exp(2\pi i s z)\, dz \tag{A.27}$$

The amplitude at the exit face of the column is then given by the sum

$$A_{\boldsymbol{H}} = F_{\boldsymbol{H}} \sum_n \exp(2\pi i s z_n) \Delta z_n \tag{A.28}$$

which, if s = const, can be approximated by the integral

$$A_{\boldsymbol{H}} = F_{\boldsymbol{H}} \int_0^t \exp(2\pi i s z)\, dz \tag{A.29}$$

or

$$A_{\boldsymbol{H}} = (F_{\boldsymbol{H}} \sin 2\pi s t)/\pi s \tag{A.30}$$

which is consistent with Eq. (A.26), though not identical.

In the complex plane the sum equation (A.28) can be represented by an amplitude-phase diagram (7) (Fig. 38); it consists of the vector sum of elementary vectors, all of the same length, each representing the amplitude diffracted by a unit cell. Successive unit cells along a column in a perfect crystal diffract with constant phase differences,that is, the corresponding vectors enclose constant angles. The diagram is a regular polygon that can be approximated by a circle with radius $F_{\boldsymbol{H}}/2\pi s$. The length of the arc of the circle is equal to the column length. The amplitude diffracted by the column is given by the length of the segment connecting the endpoints P' and P of the arc. It is clear that the amplitude will be zero if the column length (i.e., the foil thickness t) is an integer number of complete circles long. The maximum amplitude is equal to the diameter of the circle, that is, to $F_{\boldsymbol{H}}/\pi s$. Along deformed columns the amplitude-phase diagrams become curved (spiral shaped) since the angle between successive segments is no longer constant.

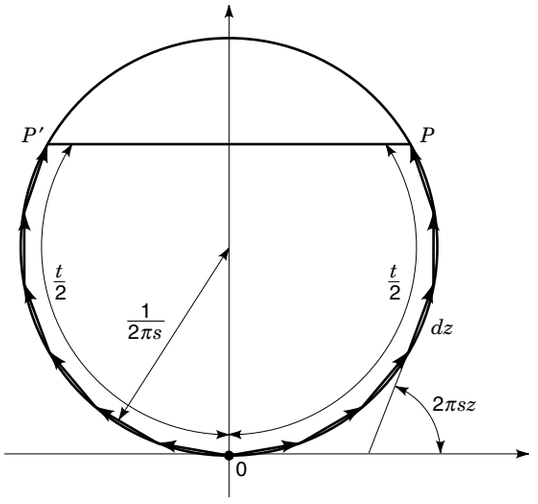


Figure 38. Complex plane construction of the amplitude-phase diagram for a perfect foil.

Two-Beam Dynamical Theory for Perfect Crystals

The dynamical theory takes into account that a scattered beam can act in turn as an incident beam and be scattered again in the interior of the crystal. The simplest case that can analytically be discussed and which moreover is relevant for image formation in the diffraction contrast mode is the two-beam case. Next to the incident beam only one beam is strongly excited (has small s). This scattered beam is then again an incident under the Bragg angle for the same set of lattice planes and can thus be scattered again. This interplay between incident and scattered beam tends to obliterate the destruction between incident and scattered beam in the interior of the crystal; it limits strongly the lateral spread of an incident Bragg diffracted electron beam and justifies the column approximation used in image calculations of defects according to the diffraction contrast mode. The dynamical theory is applicable to "thick" crystals provided also absorption is taken into account. It allows to compute the amplitudes of the transmitted beam as well as of the diffracted beam for a single Bragg reflection.

We will include all usual approximations in the model already from the onset. These approximations are as follows: (i) ideal two-beam situation, (ii) no absorption, and (iii) column approximation. Within a column along z, perpendicular to the foil surface, we describe the interplay between the transmitted beam represented by the plane wave $\phi_0(z)\ \exp(2\pi i \boldsymbol{k}_0 \cdot \boldsymbol{r})$ and the scattered beam represented by $\phi_{\boldsymbol{H}}(z)\ \exp(2\pi i \boldsymbol{k} \cdot r)$ (two-beam approximation). The complex amplitudes ϕ_0 and $\phi_{\boldsymbol{H}}$ depend on z only (column approximation). Within the slice dz at the level z behind the interface we express that the transmitted beam amplitude results from the interference between the twice transmitted beam with amplitude $\phi_0(z)\phi_0(dz)$ and the doubly scattered beam of which the amplitude is $\phi_{\boldsymbol{H}}(z)\phi_{-\boldsymbol{H}}(dz)$ [Fig. 39(a)]. The minus sign in $-\boldsymbol{H}$ means that reflection takes place from the $-\boldsymbol{H}$ side of the set of lattice planes. We thus obtain

$$\phi_0(z+dz) = \phi_0(z)\phi_0(dz) + \phi_{\boldsymbol{H}}(z)\phi_{-\boldsymbol{H}}(dz) \tag{A.31}$$

The slice dz being arbitrarily thin the kinematical approximation [e.g., Eq. (A.27)] can be applied rigorously to $\phi_0(dz)$ ($\equiv 1$, no beam depletion) and $\phi_{-\boldsymbol{H}}(dz) = \pi i/t_{\boldsymbol{H}}\ \exp(-2\pi i s z)dz$, where the factor i results from the phase change on scattering and where the structure amplitude $F_{\boldsymbol{H}}$ has been expressed in terms of the extinction distance $t_{\boldsymbol{H}}$. Note that changing $\boldsymbol{H} \rightarrow -\boldsymbol{H}$ also changes the sign of $\boldsymbol{s}$ [Fig. 39(b)].

Similarly the scattered beam amplitude results from the interference between (1) the transmitted beam, which is subsequently scattered in dz, and (2) the scattered beam, which is subsequently transmitted through dz [Fig. 39(a)]. This leads to the relation

$$\phi_{\boldsymbol{H}}(z+dz) = \phi_0(z)\phi_{\boldsymbol{H}}(dz) + \phi_{\boldsymbol{H}}(z)\phi_0(dz) \tag{A.32}$$

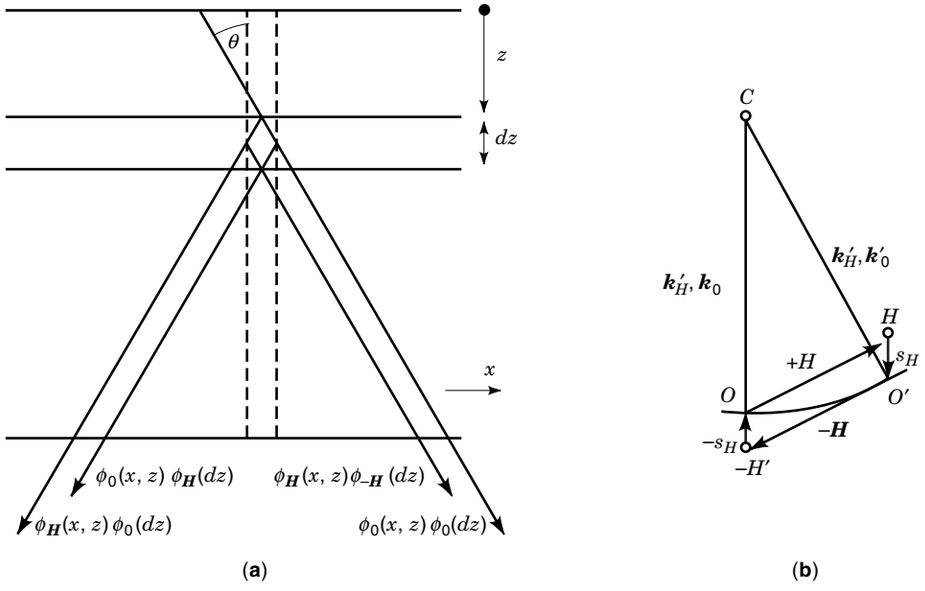


Figure 39. Schematic representation of the interfering waves during dynamical diffraction in the two-beam case. (a) Right: transmitted beam, left: scattered beam. (b) Changing $H \rightarrow -H$ changes also $s \rightarrow -s$.

where again $\phi_0(dz) = 1$ and $\phi_{\boldsymbol{H}}(dz) = (\pi i/t_H)[(\exp(2\pi isz)]\ dz$. The two equations (A.31) and (A.32) can be transformed into differential equations by noting that quite generally

$$\phi(z + dz) - \phi(z) = \frac{d\phi}{dz}\, dz \tag{A.33}$$

One thus obtains the following set of coupled differential equations:

$$\begin{aligned} \frac{d\phi_0}{dz} &= [(\pi i/t_{-\boldsymbol{H}}) \exp(2\pi isz)]\phi_H(z) \\ \frac{d\phi_H}{dz} &= [(\pi i/t_{\boldsymbol{H}}) \exp(-2\pi isz)]\phi_0(z) \end{aligned} \tag{A.34}$$

in centro-symmetrical crystals $t_{-\boldsymbol{H}} = t_{\boldsymbol{H}}$. An alternative system is obtained by the substitution

$$\phi_0 = T, \qquad \phi_{\boldsymbol{H}} = S \exp(-2\pi isz) \tag{A.35}$$

which only changes the phase of the amplitudes but not the resulting intensities. One obtains

$$\begin{aligned} \frac{dT}{dz} &= (\pi i/t_{-\boldsymbol{H}})S \\ \frac{dS}{dz} &= 2\pi isS + (\pi i/t_{\boldsymbol{H}})\, T \end{aligned} \tag{A.36}$$

These are the *Darwin-Howie-Whelan equations* (10,11) of the two-beam dynamical diffraction theory.

The solution for a perfect crystal (i.e., s is constant) is easily obtained by the standard procedure used to solve systems of coupled first order differential equations; one finds

$$\begin{aligned} T &= [\cos(\pi\sigma_{\boldsymbol{H}}z) - i(s_{\boldsymbol{H}}/\sigma_{\boldsymbol{H}}) \sin(\pi\sigma_{\boldsymbol{H}}z)] \exp(\pi is_{\boldsymbol{H}}z) \\ S &= [(i/\sigma_{\boldsymbol{H}}t_{\boldsymbol{H}}) \sin \pi\sigma_{\boldsymbol{H}}z] \exp(\pi is_{\boldsymbol{H}}z) \end{aligned} \tag{A.37}$$

where

$$\sigma_{\boldsymbol{H}}^2 = (1 + s_{\boldsymbol{H}}^2 t_{\boldsymbol{H}}^2)/t_{\boldsymbol{H}}^2 \tag{A.38}$$

The scattered intensity is thus given by the square modulus of S

$$I_S = SS^* = [\sin^2(\pi\sigma_{\boldsymbol{H}}z)]/(\sigma_{\boldsymbol{H}}t_{\boldsymbol{H}})^2 \tag{A.39}$$

where S^* denotes the complex conjugate of S and $I_T = 1 - I_S$ since absorption is neglected. Formula (A.39) is the homolog of formula (A.26), found in the kinematical approximation. Note that the depth period, which is $1/s_{\boldsymbol{H}}$ in the kinematical case, now becomes $1/\sigma_{\boldsymbol{H}}$. There is no longer a divergence for $s_{\boldsymbol{H}} \to 0$.

Equations (A.34) describe the periodic transfer of electrons from the transmitted beam into the scattered beam and vice versa; this effect is called the *Pendellösung effect* because of its similarity with the behavior of two coupled pendulums or two coupled oscillating circuits.

Equations (A.37) describes the periodic depth variations of the diffracted and transmitted intensity, as well as the variation as a function of the excitation error s. Equation (A.39) is called the *rocking curve.* In an undeformed wedge-shaped specimen the depth variation gives rise to *thickness extinction contours,* which are parallel to the cutting edge of the wedge (Fig. 40). In a bent plane-parallel specimen the lines of constant s give rise to equi-inclination or *bent contours* (Fig. 41).

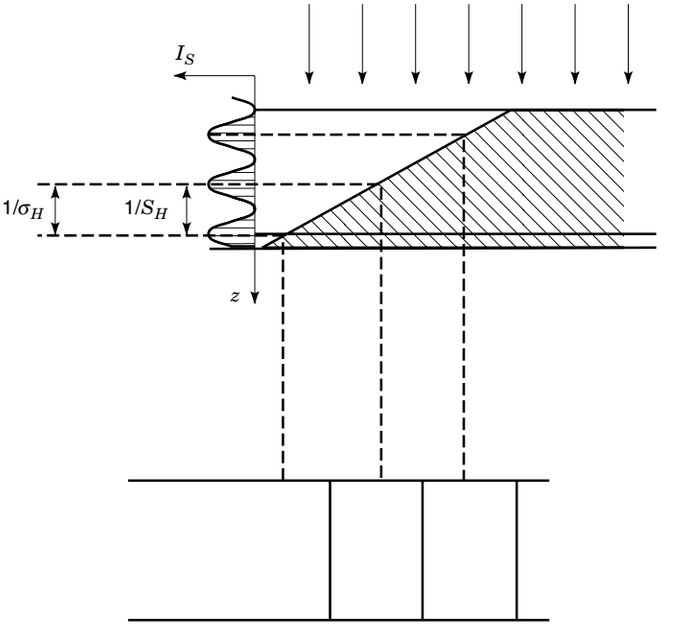


Figure 40. Illustration of the formation of thickness extinction contours.

It can be shown that taking absorption into account the shape of the rocking curve becomes asymmetric in s for the

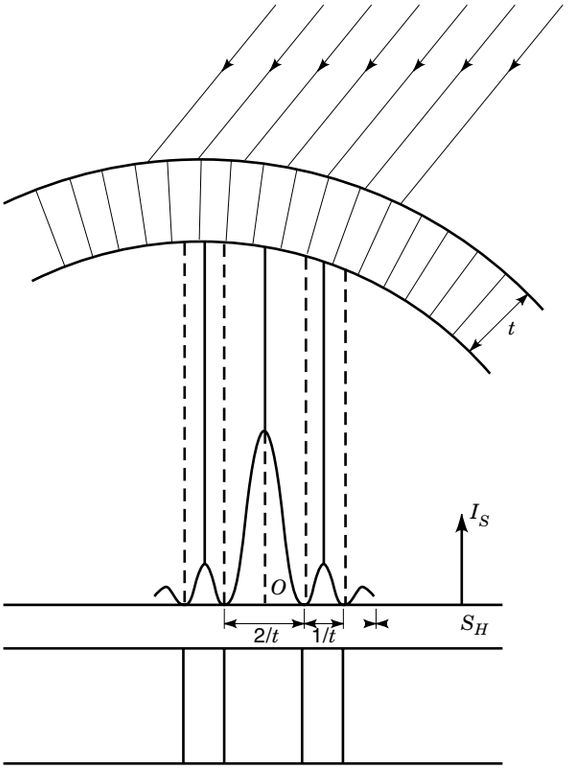


Figure 41. Illustration of the formation of equi-inclination (bent) contours.

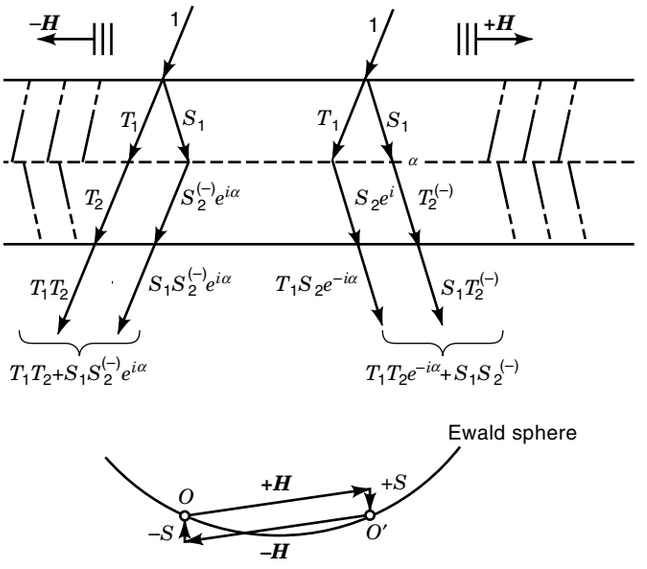


Figure 42. Schematic representation of the interfering waves in the case of a foil containing a planar interface. Left: transmitted amplitude; right: scattered amplitude.

transmitted beam (Fig. 42) whereas it remains symmetric for the scattered beam. [A similar effect occurs in x-ray diffraction: the *Bormann effect* (22).] The steep slope of the transmitted intensity in the vicinity of $s = 0$ is exploited in imaging strain fields due to dislocations and other defects.

Two-Beam Dynamical Theory for Faulted Crystals

Displacement Fields of Defects. In transmission electron microscopy defects are characterized by their displacement fields $R(\boldsymbol{r})$. The simplest example is the stacking fault, for which $R(\boldsymbol{r})$ is a step function, $R = 0$ for $z < z_1$, and $R = R_0$ for $z_1 < z < z_0$, z_1 being the level of the stacking fault plane in the foil and z_0 being the foil thickness. The exit part of the foil is displaced over a vector $\boldsymbol{R}_0$ with respect to the entrance part (Fig. 15). At the level of the interface the diffracted beam undergoes a relative phase shift given by

$$\alpha = 2\pi \boldsymbol{H} \cdot \boldsymbol{R}_0 \tag{A.40}$$

whereas the transmitted beam is unaffected. The amplitude T_S of the transmitted beam for the foil containing a stacking fault parallel to the foil plane can thus be formulated as (Fig. 42, left)

$$T_S = T_1 T_2 + S_1 S_2^- e^{i\alpha} \tag{A.41}$$

The expressions T_1, T_2, S_1, S_2 refer to the amplitudes for perfect foils. The indices 1 and 2 refer to the entrance part and the exit part, respectively; the minus sign indicates that the excitation error is $-\boldsymbol{s}$ in the corresponding expression because the diffraction vector is $-\boldsymbol{H}$.

Similarly the diffracted beam amplitude can be expressed as (Fig. 42, right)

$$S_S = T_1 S_2 e^{-i\alpha} + S_1 T_2^- \tag{A.42}$$

The meaning of Eq. (A.41) is obvious; it expresses that the transmitted beam results from the interference between the doubly transmitted beam and the doubly scattered beam. The factor $\exp(i\alpha)$ takes into account the phase shift over α of the beam scattered by the exit part. Equation (A.42) has a similar meaning. In Eq. (A.42) the phase shift is $-\alpha$ because the phase shifts are opposite in sign for S_2 and $S_2^{(-)}$. Detailed expressions can be obtained by replacing T_1, T_2, S_1, S_2 by their explicit expressions in Eq. (A.37).

If the fault plane is inclined with respect to the foil plane, the phase change α takes place at a level z_1, which now depends on position x along the foil. For instance, in Fig. 43, z_1 becomes a linear function of x. As a result T_S and S_S become quasiperiodic functions not only of z_1, but also of x.

For $s = 0$ the depth period is equal to t_H; for $s \neq 0$ it becomes $1/\sigma_H$, where σ_H is given by Eq. (A.38).

Strained Crystals. Strain fields and lattice defects are characterized by their displacement fields $\boldsymbol{R}(\boldsymbol{r})$: the atom that was at $\boldsymbol{r}$ before deformation will be found at $\boldsymbol{r} + \boldsymbol{R}(\boldsymbol{r})$ after deformation. A twin boundary with a small twinning vector (domain boundary) parallel to the foil plane at the level z_1 (Fig. 15) can, for instance, be represented by the displacement field $R = 0$ for $z < z_1$ and $R = kz$ for $z > z_1$.

A pure screw dislocation can be described by the function $\boldsymbol{R} = \boldsymbol{b}(\theta/2\pi)$, where θ is the azimuth angle measured in the plane perpendicular to b; all displacements are parallel to $\boldsymbol{b}$.

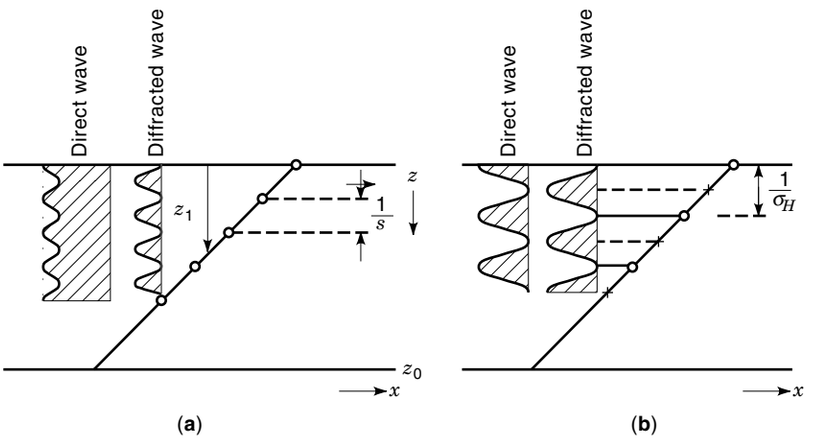


Figure 43. Cross section of foil containing a stacking fault in an inclined plane: illustrating the formation of stacking-fault fringes. (a) According to the kinematical theory $s \neq 0$; (b) according to the dynamical theory $s = 0$.

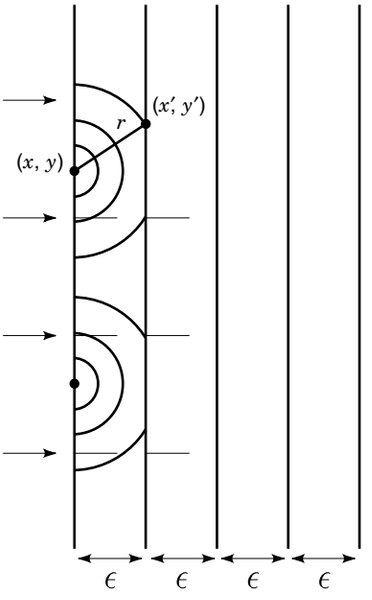


Figure 44. Schematic representation of the propagation effect of electrons between successive slices of thickness ϵ.

The Darwin-Howie-Whelan equations (A.36) can be adapted to the case of a deformed crystal by the substitution

$$s \Rightarrow s_{\text{eff}} = s + H\frac{dR}{dz} \tag{A.43}$$

The Multislice Method

The two-beam dynamical treatment is insufficient for the general situation in which high resolution images are taken with the incident beam along a zone axis where many diffracted beams are involved. Therefore the multislice method was developed as a numerical method to compute the exit wave of an object.

Although the multislice formula can be derived from quantum-mechanical principles, we follow a simplified version (23) of the more intuitive original optical approach (24). A more rigorous treatment is given in the next section.

Consider a plane wave, incident on a thin specimen foil and nearly perpendicular to the incident beam direction z. If the specimen is sufficiently thin, we can assume the electron to move approximately parallel to z so that the specimen acts a pure phase object with transmission function equation (A.5)

$$\psi(x, y) = \exp[i\sigma V_{\text{p}}(x, y)] \tag{A.44}$$

A thick specimen can now be subdivided into thin slices, perpendicular to the incident beam direction. The potential of each slice is projected into a plane that acts as a two-dimensional phase object. Each point (x, y) of the exit plane of the first slice can be considered as a Huyghens source for a secondary spherical wave with amplitude $\psi(x, y)$ (Fig. 44).

Now the amplitude $\psi(x', y')$ at the point (x', y') of the next slice can be found by the superposition of all spherical waves of the first slice, that is, by integration over x and y, yielding

$$\psi(x', y') = \int \exp[i\sigma V_{\text{p}}(x, y)] \times \frac{\exp(2\pi ikr)}{r}\,dx\,dy \tag{A.45}$$

When $|x - x'| \ll \epsilon |y - y'| \ll \epsilon$, with ϵ the slice thickness, the Fresnel approximation can be used, that is,

$$\begin{aligned} r &= \sqrt{(x-x')^2 + (y-y')^2 + \epsilon^2} \\ &\approx \epsilon\left(1 + \frac{(x-x')^2}{2\epsilon^2} + \frac{(y-y')^2}{2\epsilon^2}\right) \end{aligned} \tag{A.46}$$

so that

$$\begin{aligned} \psi(x', y') \approx &\frac{\exp(2\pi ik\epsilon)}{\epsilon} \\ &\times \int \exp[i\sigma V_{\text{p}}(x, y)] \\ &\times \exp\left(i\frac{\pi k}{\epsilon}[(x-x')^2 + (y-y')^2]\right) dx\,dy \end{aligned} \tag{A.47}$$

which, apart from constant factors, can be written as a convolution product:

$$\begin{aligned} \psi(x, y) = &\epsilon[i\sigma V_{\text{p}}(x, y)] \\ &\times \exp\left(\frac{i\pi k(x^2 + y^2)}{\epsilon}\right) \end{aligned} \tag{A.48}$$

where the convolution product of two functions is defined as (in one dimension)

$$f(x) * g(x) = \int f(x')g(x - x')\,dx' \tag{A.49}$$

If the wave function at the entrance face is $\psi(x, y, 0)$, instead of a plane wave one has for the wave function at the exit face

$$\begin{aligned} \psi(x, y, \epsilon) = &\{\psi(x, y, 0)\exp[i\sigma V_{\text{p}}(x, y)]\} \\ &\times \exp\left(\frac{i\pi k(x^2 + y^2)}{\epsilon}\right) \end{aligned} \tag{A.50}$$

This is the Fresnel approximation in which the emerging spherical wavefront is approximated by a paraboloidal wavefront.

The propagation through the vacuum gap from one slice to the next is thus described by a convolution product in which each point source of the previous slice contributes to the wave function in each point of the next slice. The motion of an electron through the whole specimen can now be described by an alternation of phase object transmissions (multiplications) and vacuum propagations (convolutions). In the limit of the slice thickness ϵ tending to zero, this multislice expression converges to the exact solution of the nonrelativistic Schrödinger equation in the forward-scattering approximation.

In the original multislice method one used the Fourier transform of Eq. (A.50) where the real space points (x, y) are transformed into diffracted beams $\boldsymbol{g}$ and where convolution and normal products are interchanged, that is,

$$\begin{aligned} \psi(\boldsymbol{g}, \epsilon) = &[\psi(\boldsymbol{g}, 0)\exp(i\sigma V_{\boldsymbol{g}})] \\ &\times \exp\left(\frac{i\pi g^2\epsilon}{k}\right) \end{aligned} \tag{A.51}$$

where V_g are the structure factors (Fourier transforms of the unit-cell potential).

The wave function at the exit face of the crystal can now be obtained by successive application of Eq. (A.50) or (A.51). This can either be done in real space [Eq. (A.50)] or in recipro-

cal space [Eq. (A.51)]. The major part of the computing time is required for the calculation of the convolution product, which is proportional to N^2 [N is the number of sampling points (real space) or beams (reciprocal space)].

Since the Fourier transform of a convolution product yields a normal product (with calculation time proportional to N) a large gain in speed can be obtained by alternatingly performing the propagation in reciprocal space and the phase object transmission in real space (23). In this way the computing time is devoted to the Fourier transforms and is proportional to $N \log_2 N$.

Another way of increasing the speed is in the so-called real-space method (24). Here the whole calculation is done in real space using Eq. (A.50) but the forward scattering of the electrons is exploited so as to calculate the convolution effect of the propagation only in a limited number of adjacent sampling points. In this way, the calculation time is proportional to N. This method does not require a periodic crystal and is thus suitable for calculation of crystal defects.

Electron Channeling

The multislice method is an efficient method to compute numerically the exit wave of an object. However it observes interesting physical aspects of dynamical electron scattering. The channelling theory is more approximate (1,25) (although improvements are currently being made) but it is simple and it gives much physical insight.

Electron Wave. Consider an isolated column of atoms, parallel to the electron beam. If we assume that the fast electron in the direction of propagation (z axis) behaves as a classical particle with velocity $v = hk/m$ we can consider the z axis as a time axis with

$$t = \frac{mz}{kh} \tag{A.52}$$

Hence we can start from the time-dependent Schrödinger equation

$$-\frac{\hbar}{i}\frac{\partial \psi}{\partial t}(\boldsymbol{R}, t) = H\psi(\boldsymbol{R}, t) \tag{A.53}$$

with

$$H = -\frac{\hbar^2}{2m}\Delta_{\boldsymbol{R}} - eU(\boldsymbol{R}, t) \tag{A.54}$$

with $U(\boldsymbol{R}, t)$ the electrostatic crystal potential, m and k the relativistic electron mass and wavelength, and $\Delta_{\boldsymbol{R}}$ the Laplacian operator acting in the plane ($\boldsymbol{R}$) perpendicular to z.

Using Eq. (A.52) we then have

$$\frac{\partial \psi(\boldsymbol{R}, z)}{\partial z} = \frac{i}{4\pi k}[\Delta_{\boldsymbol{R}} + V(\boldsymbol{R}, z)]\psi(\boldsymbol{R}, z) \tag{A.55}$$

with

$$V(\boldsymbol{R}, z) = \frac{2me}{h^2}U(\boldsymbol{R}, z) \tag{A.56}$$

This is the well-known high-energy equation in real space, which can also be derived from the stationary Schrödinger equation in the forward-scattering approximation (22).

If we now consider the depth proportional to the time, the dynamical equation (A.55) represents the walk of an electron in the two-dimensional projected potential of the columns. The solution can be expanded in eigenfunctions (eigenstates) of the Hamiltonian

$$\psi(\boldsymbol{R}, z) = \sum_n C_n \phi_n(\boldsymbol{R}) \exp\left(-i\pi \frac{E_n}{E}\frac{z}{\lambda}\right) \tag{A.57}$$

where

$$H\phi_n v(\boldsymbol{R}) = E_n \phi_n(\boldsymbol{R}) \tag{A.58}$$

with the Hamiltonian

$$H = -\frac{\hbar}{2m}\Delta_{\boldsymbol{R}} - eU(\boldsymbol{R}) \tag{A.59}$$

$U(\boldsymbol{R})$ is the projected potential of the column

$$E = \frac{h^2k^2}{2m} \tag{A.60}$$

E is the incident electron energy, and λ is the electron wavelength. For $E_n < 0$ the eigenstates are bound to the column. We now rewrite Eq. (A.57) as

$$\begin{aligned}\psi(\boldsymbol{R}, z) = &\sum_n C_n \phi_n(\boldsymbol{R}) + \sum_n C_n \phi_n(\boldsymbol{R}) \\ &\times \left[\exp\left(-i\pi \frac{E_n}{E}\frac{z}{\lambda}\right) - 1\right]\end{aligned} \tag{A.61}$$

The coefficients C_n are determined from the boundary condition

$$\sum_n C_n \phi_n(\boldsymbol{R}) = \psi(\boldsymbol{R}, 0) \tag{A.62}$$

In case of plane-wave incidence one thus has

$$\sum_n C_n \phi_n(\boldsymbol{R}) = 1 \tag{A.63}$$

so that

$$\psi(\boldsymbol{R}, z) = 1 + \sum_n C_n \phi_n(\boldsymbol{R}) \left[\exp\left(-i\pi \frac{E_n}{E}\frac{z}{\lambda}\right) - 1\right]$$

Only states will appear in the summation for which

$$|E_n| \geq \frac{E\lambda}{z} \tag{A.64}$$

These are bound states with deep energy levels that are localized near the column cores. In practice if the column does not consist of heavy atoms and the distance between columns is not too close (e.g., larger than 0.1 nm) only one eigenstate will appear, which can be compared to the 1s state of an atom.

We then have

$$\begin{aligned}\psi(\boldsymbol{R}, z) = &1 + C\phi(\boldsymbol{R}) \\ &\times \left[\exp\left(-i\pi \frac{E}{E_0}\frac{z}{\lambda}\right) - 1\right]\end{aligned} \tag{A.65}$$

A very interesting consequence of this description is that, since the state ϕ is very localized at the atom core, the wave function for the total object can be expressed as a superposition of the individual column functions ϕ_i so that Eq. (A.65) in that case becomes

$$\psi(\boldsymbol{R}, z) = 1 + \sum_i C_i \phi_i(\boldsymbol{R} - \boldsymbol{R}_i) \times \left[\exp\left(-i\pi \frac{E}{E_0} \frac{z}{\lambda}\right) - 1\right] \tag{A.66}$$

where the summation runs over all the atomic columns of the object parallel to the electron beam.

The interpretation of Eq. (A.66) is simple. Each column i acts as a channel in which the wave function oscillates periodically with depth. The periodicity is related to the "weight" of the column, that is, proportional to the atomic number of the atoms in the column and inversely proportional to their distance along the column. The importance of these results lies in the fact that they describe the dynamical diffraction for larger thicknesses than the usual phase-grating approximation and that they require only the knowledge of one function ϕ_i per column (which can be set in tabular form similar to atom scattering factors or potentials). Furthermore, even in the presence of dynamical scattering, the wave function at the exit face still retains a one-to-one relation with the configuration of columns for perfect crystals as well as for defective crystals provided they consist of columns parallel to the electron beam. Hence this description is very useful for interpreting high-resolution images.

Diffraction Pattern. Fourier transforming the wave function equation (A.66) at the exit face of the object yields the wave function in the diffraction plane, which can be written as

$$\psi(\boldsymbol{g}, z) = \sum_i \sum_i \exp(-2\pi i \boldsymbol{g} \cdot \boldsymbol{R}_i) F_i(\boldsymbol{g}, t) \tag{A.67}$$

In a sense the simple kinematical expression for the diffraction amplitude holds, provided the scattering factor for the atoms is replaced by a dynamical scattering factor for the columns, which is defined by

$$F_i(\boldsymbol{g}, z) = \left[\exp\left(\frac{-i\pi E_i}{E} \frac{z}{\lambda}\right) - 1\right] C_i f_i(\boldsymbol{g}) \tag{A.68}$$

with $f_i(\boldsymbol{g})$ the Fourier transform of $\phi_i(\boldsymbol{R})$. It is clear that the dynamical scattering factor varies periodically with depth. This periodicity may be different for different columns.

In case of a monatomic crystal, all F_i are identical. Hence $\psi(\boldsymbol{g}, z)$ varies perfectly periodically with depth. In a sense the electrons are periodically transferred from the central beam to the diffracted beams and back. The periodicity of this dynamical oscillation (which can be compared with the Pendelösung effect) is called the dynamical extinction distance. It has, for instance, been observed in Si(111). An important consequence of Eq. (A.67) is the fact that the diffraction pattern can still be described by a kinematical type of expression so that existing results and techniques (e.g., extinction rules) that have been based on the kinematical theory remain valid to some extent for thicker crystals in zone orientation.

BIBLIOGRAPHY

1. S. Amelinckx et al. (eds.), Electron microscopy, in *Handbook of Microscopy,* Weinheim: VCH, 1997, vol. 1, chap. IV.
2. D. W. Robards and A. J. Wilson, *Procedures in Electron Microscopy,* Chichester: Wiley, 1993.
3. W. L. Bragg, *Nature,* **124**: 125, 1929.
4. P. P. Ewald, *Ann. Phys.,* **54**: 519, 1917.
5. J. Steeds, Convergent beam electron diffraction, in S. Amelinckx et al. (eds.), *Handbook of Microscopy, Methods I,* Weinheim: VCH, 1997, vol. 1, chap. IV.1.5.
6. S. Takagi, *Acta Crystallogr.,* **15**: 1311, 1962.
7. P. B. Hirsch, A. Howie, and M. J. Whelan, *Philos. Trans. R. Soc.,* **A 252**: 499, 1960.
8. S. Amelinckx and J. Van Landuyt, in S. Amelinckx, R. Gevers, and J. Van Landuyt (eds.), *Diffraction and Imaging Techniques in Material Science,* Amsterdam: North Holland, 1978, p. 107.
9. P. B. Hirsch et al., *Electron Microscopy of Thin Crystals,* London: Butterworths, 1965.
10. P. Humble, in S. Amelinckx, R. Gevers, and J. Van Landuyt (eds.), *Diffraction and Imaging Techniques in Material Science,* Amsterdam: North Holland, 1978, p. 315.
11. D. J. H. Cockayne, in S. Amelinckx, R. Gevers, and J. Van Landuyt (eds.), *Diffraction and Imaging Techniques in Material Science,* Amsterdam: North Holland, 1978.
12. H. Lichte, Electron holography methods, in S. Amelinckx et al. (eds.), *Handbook of Microscopy, Methods I,* Weinheim: VCH, 1997, chap. IV.1.8.
13. J. Jansen et al., *Acta Crystallogr.,* **A 54**: 91, 1998.
14. H. W. Zandbergen, S. Anderson, and J. Jansen, *Science,* **12**: 1221, 1997.
15. T. Nagatani et al., *Scan. Microsc.,* **1**: 901, 1987.
16. T. E. Mulvey and C. D. Newman, *Inst. Phys. Conf. Ser.,* **18**: 16, 1973.
17. K. C. A. Smith, in O. Johari (ed.), *Proc. 5th Annu. SEM Symp.,* Chicago: IITRI, 1972, p. 1.
18. J. I. Goldstein et al., *Scanning Electron Microscopy and X-ray Microanalysis,* New York: Plenum, 1992, Chap. 2.
19. S. Amelinckx et al. (eds.), *Scanning Electron Microscopy,* D. C. Joy, Scanning reflection electron microscopy, in *Handbook of Microscopy, Methods II,* Weinheim: VCH, 1997, chapter IV.2.1.
20. J. Cowley, Scanning transmission electron microscopy, in S. Amelinckx et al. (eds.), *Handbook of Microscopy, Methods II,* Weinheim: VCH, 1997, vol. 1, chap. IV.2.2.
21. S. Pennycook, Scanning transmission electron microscopy, *Z* contrast, in S. Amelinckx et al. (eds.), *Handbook of Microscopy, Methods II,* Weinheim: VCH, 1997, Vol. 1, Chap. IV.2.3.
22. A. V. Crewe, J. Wall, and J. Langmore, *Science,* **168**: 1333, 1970.
23. D. Van Dyck, in P. Hawkes (ed.), *Advances in Electronics and Electron Physics,* New York: Academic, 1985.
24. J. M. Cowley and A. F. Moodie, *Acta Crystallogr.,* **10**: 609, 1957.
25. K. Ishizuka and N. Uyeda, *Acta Crystallogr. A* 33: 740, 1977.
26. K. Kambe, G. Lempfuhl, and F. Fujimoto, *Z. Naturforsch.* **29a**: 1034, 1974.

D. VAN DYCK
S. AMELINCKX
University of Antwerp

ELECTRON MICROSCOPY. See MICROSCOPE IMAGE PROCESSING AND ANALYSIS.

ELECTRON MULTIPLIERS. See PHOTOMULTIPLIERS.