International Tables for Crystallography (2006). Vol. C, Section 4.3.8, pp. 419–429.

Large gaps at the dispersion surface are associated with strong inner reflections – and a strong dynamical effect of two-beam-like character. The absolute magnitude of the gap – or its inverse, the extinction distance – can be obtained in different ways. Early measurements were based on the split of diffraction spots from a wedge, see Lehmpfuhl (1974), or the corresponding fringe periods measured in bright- and dark-field micrographs (Ando, Ichimiya & Uyeda, 1974). The most precise and applicable large-gap methods are based on the refinement of the fringe pattern in CBED discs from strong reflections, as developed by Goodman & Lehmpfuhl (1967) and Voss, Lehmpfuhl & Smith (1980). In recent years, this technique has been developed to high perfection by means of filtered CBED patterns, see Spence & Zuo (1992) and papers referred to therein. See also Chapter 8.8.

The gap at the dispersion surface can also be obtained directly from the split observed at the crossing of a weak Kikuchi line with a strong band. Gjønnes & Høier (1971) showed how this can be used to determine strong low-order reflections. High voltage may improve the accuracy (Terasaki, Watanabe & Gjønnes, 1979). The sensitivity of the intersecting Kikuchi-line (IKL) method was further increased by the use of CBED instead of Kikuchi patterns (Matsuhata, Tomokiyo, Watanabe & Eguchi, 1984; Taftø & Gjønnes, 1985). In a recent development, Høier, Bakken, Marthinsen & Holmestad (1993) have measured the intensity distribution in the CBED discs around such intersections and have refined the main structure factors involved.

Two-dimensional rocking curves collected by CBED patterns around the axis of a dense zone are complicated by extensive many-beam dynamical interactions. The Bristol–Bath group (Saunders, Bird, Midgley & Vincent, 1994) claim that the strong dynamic effects can be exploited to yield high sensitivity in refinement of low-order structure factors. They have also developed procedures for *ab initio* structure determination based on zone-axis patterns (Bird & Saunders, 1992), see Chapter 8.8.

Determination of phase invariants. It has been known for some time (*e.g* Kambe, 1957) that the dynamical three-beam case contains information about phase. As in the X-ray case, measurement of dynamical effects can be used to determine the value of triplets (Zuo, Høier & Spence, 1989) and to determine phase angles to better than one tenth of a degree (Zuo, Spence, Downs & Mayer, 1993) which is far better than any X-ray method. Bird (1990) has pointed out that the phase of the absorption potential may differ from the phase of the real potential.

Thickness is an important parameter in electron-diffraction experiments. In structure-factor determination based on CBED patterns, thickness is often included in the refinement. Thickness can also be determined directly from profiles connected with large gaps at the dispersion surface (Goodman & Lehmpfuhl, 1967; Blake, Jostsons, Kelly & Napier, 1978; Glazer, Ramesh, Hilton & Sarikaya, 1985). The method is based on the outer part of the fringe profile, which is not so sensitive to the structure factor. The intensity minimum of the ith fringe in the diffracted disc occurs at a position corresponding to the excitation error s_i and expressed as

$$(s_i^2 + 1/\varepsilon_g^2)t^2 = n_i^2, \qquad (4.3.7.9)$$

where n_i is a small integer describing the order of the minimum. This equation can be arranged in two ways for graphic determination of thickness. The commonest method appears to be to plot $(s_i/n_i)^2$ against $1/n_i{}^2$ and then determine the thickness from the intersection with the ordinate axis (Kelly, Jostsons, Blake & Napier, 1975). Glazer *et al.* (1985) claim that the method originally proposed by Ackermann (1948), where s_i^2 is plotted against n_i and the thickness is taken from the slope, is more accurate. In both cases, the outer part of the rocking curve is emphasized; exact knowledge of the gap is not necessary for a good determination of thickness, provided the assumption of a two-beam-like rocking curve is valid.

4.3.8. Crystal structure determination by high-resolution electron microscopy

(By J. C. H. Spence and J. M. Cowley)

4.3.8.1. *Introduction*

For the crystallographic study of real materials, high-resolution electron microscopy (HREM) can provide a great deal of information that is complementary to that obtainable by X-ray and neutron diffraction methods. In contrast to the statistically averaged information that these other methods provide, the great power of HREM lies in its ability to elucidate the detailed atomic arrangements of individual defects and the microcrystalline structure in real crystals. The defects and inhomogeneities of real crystals frequently exert a controlling influence on phase-transition mechanisms and more generally on all the electrical, mechanical, and thermal properties of solids. The real-space images that HREM provides (such as that shown in Fig. 4.3.8.1) can give an immediate and dramatic impression of chemical crystallography processes, unobtainable by other methods. Their atomic structure is of the utmost importance for

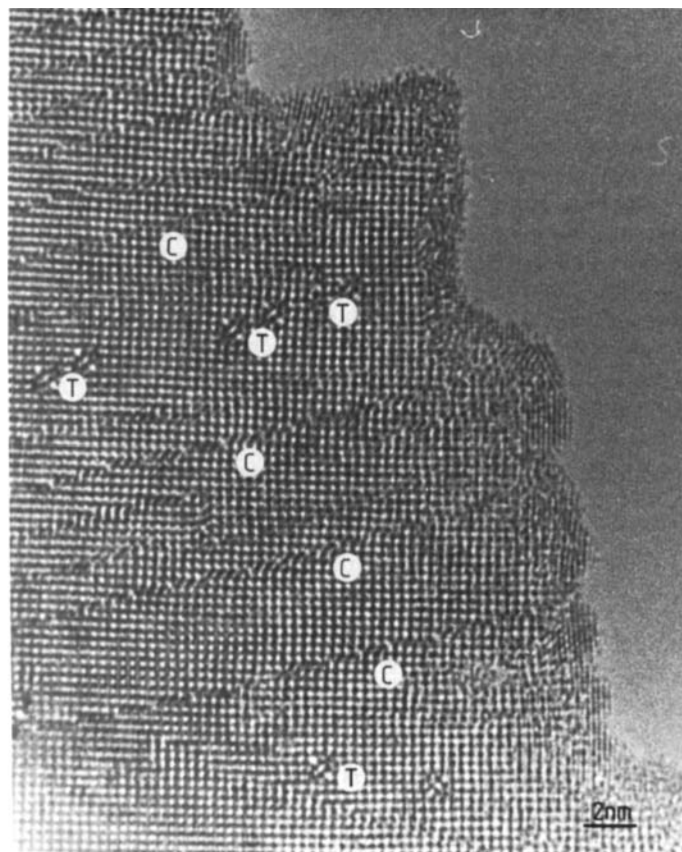

Fig. 4.3.8.1. Atomic resolution image of a tantalum-doped tungsten trioxide crystal (pseudo-cubic structure) showing extended crystallographic shear-plane defects (C), pentagonal-column hexagonal-tunnel (PCHT) defects (T), and metallization of the surface due to oxygen desorption (JEOL 4000EX, crystal thickness less than 200 Å, 400 kV, $C_s = 1$ mm). Atomic columns are black. [Smith, Bursill & Wood (1985).]

Copyright © 2006 International Union of Crystallography

an understanding of the properties of real materials. The HREM method has proven powerful for the determination of the structure of such defects and of the submicrometre-sized microcrystals that constitute many polyphase materials.

In summary, HREM should be considered the technique of choice where a knowledge of microcrystal size, shape or morphology is required. In addition, it can be used to reveal the presence of line and planar defects, inclusions, grain boundaries and phase boundaries, and, in favourable cases, to determine atomic structure. Surface atomic structure and reconstruction have also been studied by HREM. However, meaningful results in this field require accurately controlled ultra-high-vacuum conditions. The determination of the atomic structure of point defects by HREM so far has proven extremely difficult, but this situation is likely to change in the near future.

The following sections are not intended to review the applications of HREM, but rather to provide a summary of the main theoretical results of proven usefulness in the field, a selected bibliography, and recommendations for good experimental practice. At the time of writing (1997), the point resolution of HREM machines lies between 1 and 2 Å.

The function of the objective lens in an electron microscope is to perform a Fourier synthesis of the Bragg-diffracted electron beams scattered (in transmission) by a thin crystal, in order to produce a real-space electron image in the plane r. This electron image intensity can be written

$$|\psi(\mathbf{r})|^2 = \left| \int \Psi(\mathbf{u}) \exp\{2\pi i \mathbf{u} \cdot \mathbf{r}\} P(\mathbf{u}) \exp\{i\chi(\mathbf{u})\}\, \mathrm{d}\mathbf{u} \right|^2, \quad (4.3.8.1)$$

where $\Psi(\mathbf{u})$ represents the complex amplitude of the diffracted wave after diffraction in the crystal as a function of the reciprocal-lattice vector $\mathbf{u}$ [magnitude $(2\sin\theta)/\lambda$] in the plane perpendicular to the beam, so that the wavevector of an incident plane wave is written $\mathbf{K}_0 = \mathbf{k}_z + 2\pi\mathbf{u}$. Following the convention of Section 2.5.1 in *IT* B (1992), we write $|\mathbf{K}_0| = 2\pi\lambda^{-1}$. The function $\chi(\mathbf{u})$ is the phase factor for the objective-lens transfer function and $P(\mathbf{u})$ describes the effect of the objective aperture:

$$P(\mathbf{u}) = \begin{cases} 1 & \text{for } |\mathbf{u}| < u_0 \\ 0 & \text{for } |\mathbf{u}| \geq u_0. \end{cases}$$

For a periodic object, the image wavefunction is given by summing the contributions from the set of reciprocal-lattice points, $\mathbf{g}$, so that

$$|\psi(\mathbf{r})|^2 = \left| \sum_{\mathbf{g}} \Psi_{\mathbf{g}} \exp\{2\pi i \mathbf{g} \cdot \mathbf{r}\} P(\mathbf{g}) \exp\{i\chi(\mathbf{g})\} \right|^2. \quad (4.3.8.2)$$

For atomic resolution, with $u_0 \approx 1\text{Å}^{-1}$, it is apparent that, for all but the simplest structures and smallest unit cells, this synthesis will involve many hundreds of Bragg beams. A scattering calculation must involve an even larger number of beams than those that contribute resolvable detail to the image, since, as described in Section 2.5.1 in *IT* B (1992), all beams interact strongly through multiple coherent scattering. The theoretical basis for HREM image interpretation is therefore the dynamical theory of electron diffraction in the transmission (or Laue) geometry [see Chapter 5.2 in *IT* B (1992)]. The resolution of HREM images is limited by the aberrations of the objective electron lens (notably spherical aberration) and by electronic instabilities. An intuitive understanding of the complicated effect of these factors on image formation from multiply scattered Bragg beams is generally not possible. To provide a basis for understanding, therefore, the following section treats the simplified case of few-beam 'lattice-fringe' images, in order to expose the relationship between the crystal potential, its structure factors, electron-lens aberrations, and the electron image.

Image formation in the transmission electron microscope is conventionally treated by analogy with the Abbe theory of coherent optical imaging. The overall process is subdivided as follows. (*a*) The problem of beam–specimen interaction for a collimated kilovolt electron beam traversing a thin parallel-sided slab of crystal in a given orientation. The solution to this problem gives the elastically scattered dynamical electron wavefunction $\psi(\mathbf{r})$, where $\mathbf{r}$ is a two-dimensional vector lying in the downstream surface of the slab. Computer algorithms for dynamical scattering are described in Section 4.3.6. (*b*) The effects of the objective lens are incorporated by multiplying the Fourier transform of $\psi(\mathbf{r})$ by a function $T(\mathbf{u})$, which describes both the wavefront aberration of the lens and the diffraction-limiting effects of any apertures. The dominant aberrations are spherical aberration, astigmatism, and defect of focus. The image intensity is then formed from the modulus squared of the Fourier transform of this product. (*c*) All partial coherence effects may be incorporated by repeating this procedure for each of the component energies and directions that make up the illumination from an extended electron source, and summing the resulting intensities. Because this procedure requires a separate dynamical calculation for each component direction of the incident beam, a number of useful approximations of restricted validity have been developed; these are described in Subsection 4.3.8.4. This treatment of partial coherence assumes that a perfectly incoherent effective source can be identified. For field-emission HREM instruments, a coherent sum (over directions) of complex image wavefunctions may be required.

General treatments of the subject of HREM can be found in the texts by Cowley (1981) and Spence (1988). The sign

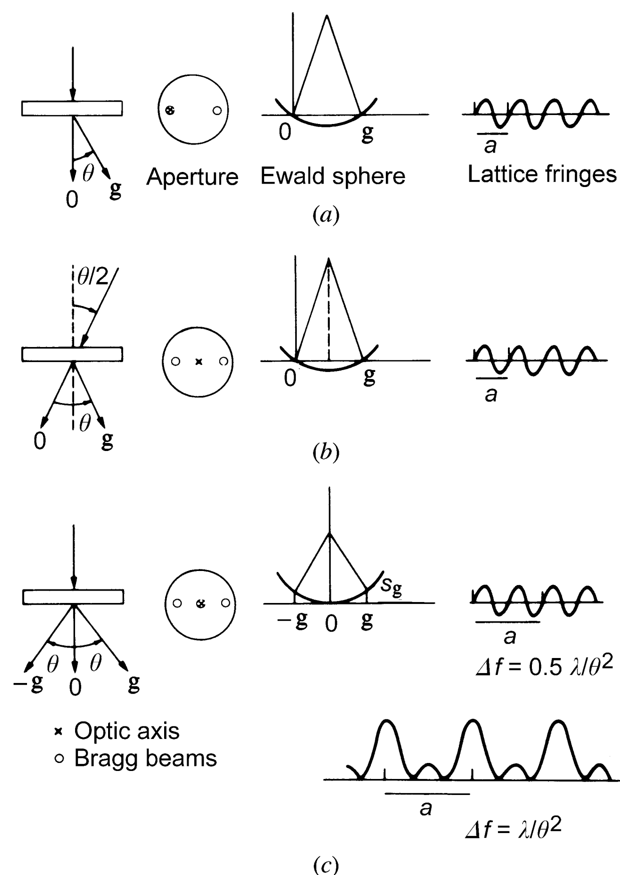

Fig. 4.3.8.2. Imaging conditions for few-beam lattice images. For three-beam axial imaging shown in (*c*), the formation of half-period fringes is also shown.

conventions used throughout the following are consistent with the standard crystallographic convention of Section 2.5.1 of *IT* B (1992), which assumes a plane wave of form $\exp\{-i(\mathbf{k}\cdot\mathbf{r}-\omega t)\}$ and so is consistent with X-ray usage.

4.3.8.2. *Lattice-fringe images*

We consider few-beam lattice images, in order to understand the effects of instrumental factors on electron images, and to expose the conditions under which they faithfully represent the scattering object. The case of two-beam lattice images is instructive and contains, in simplified form, most of the features seen in more complicated many-beam images. These fringes were first observed by Menter (1956) and further studied in the pioneering work of Komoda (1964) and others [see Spence (1988) for references to early work]. The electron-microscope optic-axis orientation, the electron beam, and the crystal setting are indicated in Fig. 4.3.8.2. If an objective aperture is used that excludes all but the two beams shown from contributing to the image, equation (4.3.8.2) gives the image intensity along direction $\mathbf{g}$ for a centrosymmetric crystal of thickness t as

$$I(x,t)=|\Psi_0(t)|^2+|\Psi_{\rm g}(t)|^2 + 2|\Psi_0||\Psi_{\rm g}|\cos\{2\pi x/d_{\rm g}+\chi(u_{\rm g})\eta_{\rm g}(t)-\eta_0(t)\}. \quad (4.3.8.3)$$

The Bragg-diffracted beams have complex amplitudes $\Psi_{\rm g}(t)=|\Psi_{\rm g}(t)|\exp\{i\eta_{\rm g}(t)\}$. The lattice-plane period is $d_{\rm g}$ in direction $\mathbf{g}$ [Miller indices (hkl)]. The lens-aberration phase function, including only the effects of defocus Δf and spherical aberration (coefficient C_s), is given by

$$\chi(u_{\rm g})=(2\pi/\lambda)\{(\Delta f\lambda^2u_{\rm g}^2/2)+C_s\lambda^4u_{\rm g}^4/4\}. \quad (4.3.8.4)$$

The effects of astigmatism and higher-order aberrations have been ignored. The defocus, Δf, is negative for the objective lens weakened (*i.e.* the focal length increased, giving a bright first Fresnel-edge fringe). The magnitude of the reciprocal-lattice vector $u_g=d_g^{-1}=(2\sin\theta_B)/\lambda$, where θ_B is the Bragg angle. If these two Bragg beams were the only beams excited in the crystal (a poor approximation for quantitative work), their amplitudes would be given by the 'two-beam' dynamical theory of electron diffraction as

$$\begin{aligned}\Psi_0(t)&=\{\cos[\pi t(1+w^2)^{1/2}/\xi_{\rm g}]+iw(1+w^2)^{-1/2}\\&\quad\times\sin[\pi t(1+w^2)^{1/2}/\xi_{\rm g}]\}\exp(-i\pi s_{\rm g}t)\\ \Psi_{\rm g}(t)&=i(1+w^2)^{-1/2}\sin[\pi t(1+w^2)^{1/2}/\xi_{\rm g}]\\&\quad\times\exp(-\pi i s_{\rm g}t),\end{aligned} \quad (4.3.8.5)$$

where $\xi_{\rm g}$ is the two-beam extinction distance, $V_{\rm g}=\pi/(\sigma\xi_{\rm g})$ is a Fourier coefficient of crystal potential, $s_{\rm g}$ is the excitation error (see Fig. 4.3.8.2), $w=s_{\rm g}\xi_{\rm g}$, and the interaction parameter σ is defined in Section 2.5.1 of *IT* B (1992).

The two-beam image intensity given by equation (4.3.8.3) therefore depends on the parameters of crystal thickness (t), orientation ($s_{\rm g}$), structure factor ($V_{\rm g}$), objective-lens defocus Δf, and spherical-aberration constant C_s. We consider first the variation of lattice fringes with crystal thickness in the two-beam approximation (Cowley, 1959; Hashimoto, Mannami & Naiki, 1961). At the exact Bragg condition ($s_{\rm g}=0$), equations (4.3.8.5) and (4.3.8.3) give

$$I(x,t)=1-\sin(2\pi t/\xi_{\rm g})\sin[2\pi x/d+\chi(u_{\rm g})]. \quad (4.3.8.6)$$

If we consider a wedge-shaped crystal with the electron beam approximately normal to the wedge surface and edge, and take x and $\mathbf{g}$ parallel to the edge, this equation shows that sinusoidal lattice fringes are expected whose contrast falls to zero (and reverses sign) at thicknesses of $t_n=n\xi_{\rm g}/2$. This apparent abrupt translation of fringes (by $d/2$ in the direction x) at particular thicknesses is also seen in some experimental many-beam images. The effect of changes in focus (due perhaps to variations in lens current) is seen to result in a translation of the fringes (in direction x), while time-dependent variations in the accelerating voltage have a similar effect. Hence, time-dependent variations of the lens focal length or the accelerating voltage result in reduced image contrast (see below). If the illumination makes a small angle $\alpha=\lambda u'$ with the optic axis, the intensity becomes

$$\begin{aligned}I(x,\alpha)&=|\Psi_0|^2+|\Psi_{\rm g}|^2+2|\Psi_{\rm g}||\Psi_0|\\&\quad\times\cos[\chi(-u_{\rm g}-u')-\chi(u')+2\pi x/d+\eta_{\rm g}(t)-\eta_0(t)].\end{aligned}$$

For a uniformly intense line source subtending a semiangle θ_c, the total lattice-fringe intensity is

$$I(x)=(1/\theta_c)\textstyle\int I(x,\alpha)\,{\rm d}\alpha.$$

The resulting fringe visibility $C=(I_{\max}-I_{\min})/(I_{\max}+I_{\min})$ is proportional to $C=(\sin\beta)/\beta$, where $\beta=2\pi\Delta f\theta_c/d$. The contrast falls to zero for $\beta=\pi$, so that the range of focus over which fringes are expected is $\Delta z=d/\theta_c$. This is the approximate depth of field for lattice images due to the effects of the finite source size alone.

The case of three-beam fringes in the axial orientation is of more practical importance [see Fig. 4.3.8.2(*b*)]. The image intensity for $\Psi_{\rm g}=\Psi_{-\rm g}$ and $s_{\rm g}=s_{-\rm g}$ is

$$\begin{aligned}I(x,t)&=|\Psi_0|^2+2|\Psi_{\rm g}|^2+2|\Psi_{\rm g}|^2\cos(4\pi x/d)\\&\quad+4|\Psi_0||\Psi_{\rm g}|\cos(2\pi x/d)\\&\quad\times\cos[\chi(u_{\rm g})+\eta_{\rm g}(t)-\eta_0(t)].\end{aligned} \quad (4.3.8.7)$$

The lattice image is seen to consist of a constant background plus cosine fringes with the lattice spacing, together with cosine fringes of half this spacing. The contribution of the half-spacing fringes is independent of instrumental parameters (and therefore of electronic instabilities if $\theta_c=0$). These fringes constitute an important HREM image artifact. For kinematic scattering, $\eta_{\rm g}(t)-\eta_0(t)=-\pi/2$ and only the half-period fringes will then be seen if $\chi(u_{\rm g})=n\pi$, or for focus settings

$$\Delta f=n\lambda^{-1}u_{\rm g}^{-2}-C_s\lambda^2u_{\rm g}^2/2. \quad (4.3.8.8)$$

Fig. 4.3.8.2(*c*) indicates the form of the fringes expected for two focus settings with differing half-period contributions. As in the case of two-beam fringes, dynamical scattering may cause Ψ_0 to be severely attenuated at certain thicknesses, resulting also in a strong half-period contribution to the image.

Changes of 2π in $\chi(u_{\rm g})$ in equation (4.3.8.7) leave $I(x,t)$ unchanged. Thus, changes of defocus by amounts

$$\Delta f_f=2n/(\lambda u_{\rm g}^2) \quad (4.3.8.9)$$

or changes in C_s by

$$\Delta C_s=4n/(\lambda^3u_{\rm g}^4) \quad (4.3.8.10)$$

yield identical images. The images are thus periodic in both Δf and C_s. This is a restricted example of the more general phenomenon of n-beam Fourier imaging discussed in Subsection 4.3.8.3.

We note that only a single Fourier period will be seen if Δf_f is less than the depth of field Δz. This leads to the approximate condition $\Theta_c>\lambda/d$, which, when combined with the Bragg law, indicates that a single period only of images will be seen when adjacent diffraction discs just overlap.

The axial three-beam fringes will coincide with the lattice planes, and show atom positions as dark if $\chi(u_{\mathbf{g}}) = (2n - 1/2)\pi$ and $\eta_0(t) - \eta_{\mathbf{g}}(t) = -\pi/2$. This total phase shift of $-\pi$ between Ψ_0 and the scattered beams is the desirable imaging condition for phase contrast, giving rise to dark atom positions on a bright background. This requires

$$C_s = (4n - 1)/(\lambda^3 u_{\mathbf{g}}^4) - 2\Delta f/(\lambda^2 u_{\mathbf{g}}^2)$$

as a condition for identical axial three-beam lattice images for $n = 0, 1, 2, \ldots$. This family of lines has been plotted in Fig. 4.3.8.3 for the (111) planes of silicon. Dashed lines denote the locus of 'white-atom' images (reversed contrast fringes), while the dotted lines indicate half-period images. In practice, the depth of field is limited by the finite illumination aperture θ_c, and few-beam lattice-image contrast will be a maximum at the stationary-phase focus setting, given by

$$\Delta f_0 = -C_s \lambda^2 u_{\mathbf{g}}^2. \qquad (4.3.8.11)$$

This choice of focus ensures $\nabla\chi(u) = 0$ for $u = u_{\mathbf{g}}$, and thus ensures the most favourable trade-off between increasing θ_c and loss of fringe contrast for lattice planes **g**. Note that Δf_0 is not equal to the Scherzer focus Δf_s (see below). This focus setting is

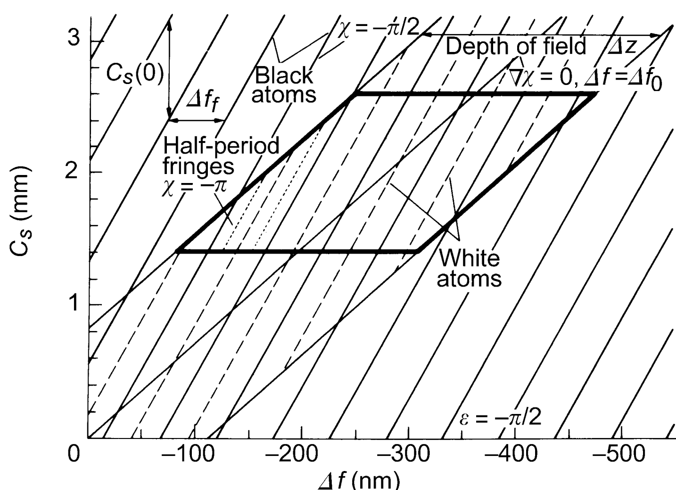

Fig. 4.3.8.3. A summary of three- (or five-) beam axial imaging conditions. Here, Δf_f is the Fourier image period, Δf_0 the stationary-phase focus, $C_s(0)$ the image period in C_s, and a scattering phase of $-\pi/2$ is assumed. The lines are drawn for the (111) planes of silicon at 100 kV with $\theta_c = 1.4$ mrad.

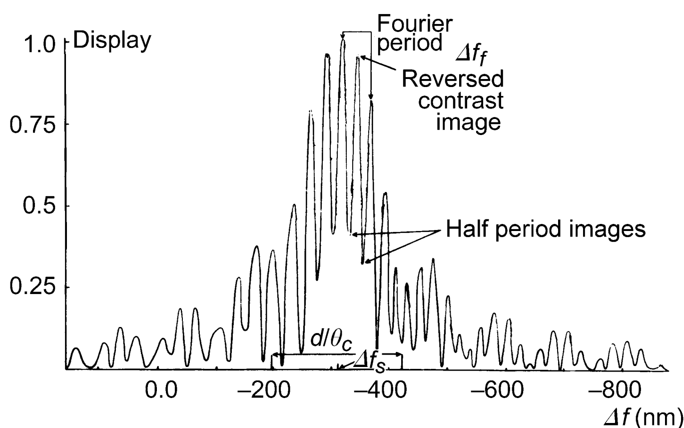

Fig. 4.3.8.4. The contrast of few-beam lattice images as a function of focus in the neighbourhood of the stationary-phase focus [see Olsen & Spence (1981)].

also indicated on Fig. 4.3.8.3, and indicates the instrumental conditions which produce the most intense (111) three- (or five-) beam axial fringes in silicon. For three-beam axial fringes of spacing d, it can be shown that the depth of field Δz is approximately

$$\Delta z = (\ln 2)^{1/2} d/\theta_c \pi. \qquad (4.3.8.12)$$

This depth of field, within which strong fringes will be seen, is indicated as a boundary on Fig. 4.3.8.3. Thus, the finer the image detail, the smaller is the focal range over which it may be observed, for a given illumination aperture θ_c.

Fig. 4.3.8.4 shows an exact dynamical calculation for the contrast of three-beam axial fringes as a function of Δf in the neighbourhood of Δf_0. Both reversed contrast and half-period fringes are noted. The effects of electronic instabilities on lattice images are discussed in Subsection 4.3.8.3. It is assumed above that θ_c is sufficiently small to allow the neglect of any changes in diffraction conditions (Ewald-sphere orientation) within Θ_c. Under a similar approximation but without the approximations of transfer theory, Desseaux, Renault & Bourret (1977) have analysed the effect of beam divergence on two-dimensional five-beam axial lattice fringes.

When two-dimensional patterns of fringes are considered, the Fourier imaging conditions become more complex (see Subsection 4.3.8.3), but half-period fringe systems and reversed-contrast images are still seen. For example, in a cubic projection, a focus change of $\Delta f_f/2$ results in an image shifted by half a unit cell along the cell diagonal. It is readily shown that

$$\exp[i\chi(\Delta f)] = \exp[i\chi(\Delta f + \Delta f_f)]$$

if $\Delta f_f = 2na^2/\lambda + 2mb^2/\lambda$ when n, m are integers and a and b are the two dimensions of any orthogonal unit cell that can be chosen for $\Psi_p(x, y)$. Thus, changes in focus by $\Delta f_f(n, m)$ produce identical images in crystals for which such a cell can be chosen, regardless of the number of beams contributing (Cowley & Moodie, 1960).

For closed-form expressions for the few-beam (up to 10 beams) two-dimensional dynamical Bragg-beam amplitudes $\Psi_{\mathbf{g}}$ in orientations of high symmetry, the reader is referred to the work of Fukuhara (1966).

4.3.8.3. *Crystal structure images*

We define a crystal structure image as a high-resolution electron micrograph that faithfully represents a projection of a crystal structure to some limited resolution, and which was obtained using instrumental conditions that are independent of the structure, and so require no *a priori* knowledge of the structure. The resolution of these images is discussed in Subsection 4.3.8.6, and their variation with instrumental parameters in Subsection 4.3.8.4.

Equation (4.3.8.2) must now be modified to take account of the finite electron source size used and of the effects of the range of energies present in the electron beam. For a perfect crystal we may write, as in equation (2.5.1.36) in *IT* B (1992),

$$I_T(\mathbf{r}) = \iint |\psi(\mathbf{u}', \Delta f, \mathbf{r})|^2 \, G(\mathbf{u}') B(\Delta f, \mathbf{u}') \, \mathrm{d}\mathbf{u}' \, \mathrm{d}\Delta f \qquad (4.3.8.13a)$$

for the total image intensity due to an electron source whose normalized distribution of wavevectors is $G(\mathbf{u}')$, where $\mathbf{u}'$ has components u_1, v_1, and which extends over a range of energies corresponding to the distribution of focus $B(\Delta f, \mathbf{u})$. If χ is also assumed to vary linearly across θ_c and changes in the diffraction conditions over this range are assumed to make only negligible changes in the diffracted-beam amplitude $\Psi_{\mathbf{g}}$, the expression for a Fourier coefficient of the total image intensity $I_T(\mathbf{r})$ becomes

$$I_{\mathbf{g}} = \sum_{\mathbf{h}} \Psi_{\mathbf{h}} \exp\{-i\chi(\mathbf{h})\}\gamma\{\nabla\chi(\mathbf{h}) - \nabla\chi(\mathbf{h}-\mathbf{g})\} \times \psi^*_{\mathbf{h}-\mathbf{g}} \exp\{i\chi(\mathbf{h}-\mathbf{g})\}\beta\{\tfrac{1}{2}(\mathbf{h}^2 - |\mathbf{h}-\mathbf{g}|^2)^2\}, \quad (4.3.8.13b)$$

where $\gamma(\mathbf{h})$ and $\beta(\mathbf{g})$ are the Fourier transforms of $G(\mathbf{u}')$ and $B(\Delta f, \mathbf{u}')$, respectively.

For the imaging of very thin crystals, and particularly for the case of defects in crystals, which are frequently the objects of particular interest, we give here some useful approximations for HREM structure images in terms of the continuous projected crystal potential

$$\varphi(x, y) = (1/t)\int_0^t \varphi(x, y, z)\,\mathrm{d}z,$$

where the projection is taken in the electron-beam direction. A brief summary of the use of these approximations is included in Section 2.5.1 of *IT* B (1992) and computing methods are discussed in Subsection 4.3.8.5 and Section 4.3.4.

The projected-charge-density (PCD) approximation (Cowley & Moodie, 1960) gives the HREM image intensity (for the simplified case where $C_s = 0$) as

$$I(x, y) = 1 + (\Delta f \lambda\sigma/2\pi\varepsilon_0\varepsilon)\rho_p(x, y), \quad (4.3.8.13c)$$

where $\rho_p(x, y)$ is the projected charge density for the specimen (including the nuclear contribution) and is related to $\varphi_p(x, y)$ through Poisson's equation. Here, $\varepsilon_0\varepsilon$ is the specimen dielectric constant. This approximation, unlike the weak-phase-object approximation (WPO), includes multiple scattering to all orders of the Born series, within the approximation that the component of the scattering vector is zero in the beam direction (a 'flat' Ewald sphere). Contrast is found to be proportional to defocus and to $\rho_p(x, y)$. The failure conditions of this approximation are discussed by Lynch, Moodie & O'Keefe (1975); briefly, it fails for $\chi(u_0) > \pi/2$ (and hence if C_s, Δf or u_0 becomes large) or for large thicknesses t ($t < 7$ nm is suggested for specimens of medium atomic weight and $\lambda = 0.037$ Å). The PCD result becomes increasingly accurate with increasing accelerating voltage for small C_s.

The WPO approximation has been used extensively in combination with the Scherzer-focus condition (Scherzer, 1949) for the interpretation of structure images (Cowley & Iijima, 1972). This approximation neglects multiple scattering of the beam electron and thereby allows the application of the methods of linear transfer theory from optics. The image intensity is then given, for plane-wave illumination, by

$$\begin{aligned} I(x, y) &= 1 + 2\sigma\varphi_p(x, y) * \mathcal{F}\{\sin\chi(u, v)P(u, v)\} \\ &= 1 + 2\sigma\varphi_p(x, y) * S(x, y), \end{aligned} \quad (4.3.8.14)$$

where $\mathcal{F}$ denotes Fourier transform, $*$ denotes convolution, and u and v are orthogonal components of the two-dimensional scattering vector $\mathbf{u}$. The function $S(x, y)$ is sharply peaked and negative at the 'Scherzer focus'

$$\Delta f = \Delta f_s = 1.2(C_s\lambda)^{1/2} \quad (4.3.8.15a)$$

and the optimum objective aperture size

$$\theta_0 = 1.5(\lambda/C_s)^{1/4}. \quad (4.3.8.15b)$$

It forms the impulse response of an electron microscope for phase contrast. Contrast is found to be proportional to φ_p and to the interaction parameter σ, which increases very slowly with accelerating voltage above about 500 keV. The point resolution [see Subsection 2.5.1.9 of *IT* B (1992) and Subsection 4.3.8.6] is conventionally defined from equation (4.3.8.15*b*) as λ/θ_0, or

$$d_p = 0.66\,C_s^{1/4}\lambda^{3/4}. \quad (4.3.8.16)$$

The occurrence of appreciable multiple scattering, and therefore of the failure of the WPO approximation, depends on specimen thickness, orientation, and accelerating voltage. Detailed comparisons between accurate multiple-scattering calculations, the PCD approximation, and the WPO approximation can be found in Lynch, Moodie & O'Keefe (1975) and Jap & Glaeser (1978). As a very rough guide, equation (4.3.8.14) can be expected to fail for light elements at 100 keV and thicknesses greater than about 5.0 nm. Multiple-scattering effects have been predicted within single atoms of gold at 100 keV.

The WPO approximation may be extended to include the effects of an extended source (partial spatial coherence) and a range of incident electron-beam energies (temporal coherence). General methods for incorporating these effects in the presence of multiple scattering are described in Subsection 4.3.8.5. Under the approximations of linear imaging outlined below, it can be shown (Wade & Frank, 1977; Fejes, 1977) that $\sin\chi(u, v)P(u, v)$ in equation (4.3.8.14) may be replaced by

$$\begin{aligned} A'(\mathbf{u}) &= P(\mathbf{u})\exp[i\chi(\mathbf{u})]\exp(-\pi^2\Delta^2\lambda^2\mathbf{u}^4/2)\gamma(\nabla\chi/2\pi) \\ &= P(\mathbf{u})\exp[i\chi(\mathbf{u})]\exp(i\pi^2\Delta^2\lambda^2\mathbf{u}^4/2)\exp(-\pi^2u_0^2q) \end{aligned} \quad (4.3.8.17)$$

if astigmatism is absent. Here, $\mathbf{u} = u\mathbf{i} + v\mathbf{j}$ and $|\mathbf{u}| = 2\theta/\lambda = (u^2 + v^2)^{1/2}$. In addition, $\gamma(\mathbf{u}')$ is the Fourier transform of the source intensity distribution (assumed Gaussian), so that $\gamma(\nabla\chi/2\pi)$ is small in regions where the slope of $\chi(\mathbf{u}')$ is large, resulting in severe attenuation of these spatial frequencies. If the illuminating beam divergence Θ_c is chosen as the angular half width for which the distribution of source intensity falls to half its maximum value, then

$$\theta_c = \lambda u_0(\ln 2)^{1/2}. \quad (4.3.8.18)$$

The quantity q is defined by

$$q = (C_s\lambda^3\mathbf{u}^3 + \Delta f\lambda\mathbf{u})^2 + T2,$$

where $T2$ expresses a coupling between the effects of partial spatial coherence and temporal coherence. This term can frequently be neglected under HREM conditions [see Wade & Frank (1977) for details]. The damping envelope due to chromatic effects is described by the parameter

$$\Delta = C_cQ = C_c\{[\sigma^2(V_0)]/V_0^2 + [4\sigma^2(I_0)]/I_0^2 + [\sigma^2(E_0)]/E_0^2\}^{1/2}, \quad (4.3.8.19)$$

where $\sigma^2(V_0)$ and $\sigma^2(I_0)$ are the variances in the statistically independent fluctuations of accelerating voltage V_0 and objective-lens current I_0. The r.m.s. value of the high voltage fluctuation is equal to the standard deviation $\sigma(V_0) = [\sigma^2(V_0)]^{1/2}$. The full width at half-maximum height of the energy distribution of electrons leaving the filament is

$$\Delta E = 2(2\ln 2)^{1/2}\sigma(E_0) = 2.355[\sigma^2(E_0)]^{1/2}. \quad (4.3.8.20)$$

Here, C_c is the chromatic aberration constant of the objective lens.

Equations (4.3.8.14) and (4.3.8.17) indicate that under linear imaging conditions the transfer function for HREM contains a chromatic damping envelope more severely attenuating than a Gaussian of width

$$U_0(\Delta) = [2/\pi\lambda\Delta]^{1/2},$$

which is present in the absence of any objective aperture $P(\mathbf{u})$. The resulting resolution limit

$$d_i = [\pi\lambda\Delta/2]^{1/2} \qquad (4.3.8.21)$$

is known as the information resolution limit (see Subsection 4.3.8.6) and depends on electronic instabilities and the thermal-energy spread of electrons leaving the filament. The reduction in the contribution of particular diffracted beams to the image due to limited spatial coherence is minimized over those extended regions for which $\nabla\chi(\mathbf{u})$ is small, called passbands, which occur when

$$\Delta f_n = [C_s\lambda(8n+3)/2]^{1/2}. \qquad (4.3.8.22)$$

The Scherzer focus Δf_s corresponds to $n = 0$. These passbands become narrower and move to higher $\mathbf{u}$ values with increasing n, but are subject also to chromatic damping effects. The passbands occur between spatial frequencies U_1 and U_2, where

$$U_{1,2} = C_s^{-1/4}\lambda^{-3/4}\{[(8n+2)/2]^{1/2} \pm 1\}^{1/2}. \qquad (4.3.8.23)$$

Their use for extracting information beyond the point resolution of an electron microscope is further discussed in Subsection 4.3.8.6.

Fig. 4.3.8.5 shows transfer functions for a modern instrument for $n = 0$ and 1. Equations (4.3.8.14) and (4.3.8.17) provide a simple, useful, and popular approach to the interpretation of HREM images and valuable insights into resolution-limiting factors. However, it must be emphasized that these results apply only (amongst other conditions) for $\Phi_0 \gg \Phi_{\mathbf{g}}$ (in crystals) and therefore do not apply to the usual case of strong multiple electron scattering. Equation (4.3.8.13*b*) does not make this approximation. In real space, for crystals, the alignment of columns of atoms in the beam direction rapidly leads to phase changes in the electron wavefunction that exceed $\pi/2$, leading to the failure of equation (4.3.8.14). Accurate quantitative comparisons of experimental and simulated HREM images must be based on equation (4.3.8.13*a*), or possibly (4.3.8.13*b*), with $\psi(\mathbf{u}', \Delta f, \mathbf{r})$ obtained from many-beam dynamical calculations of the type described in Subsection 4.3.8.5.

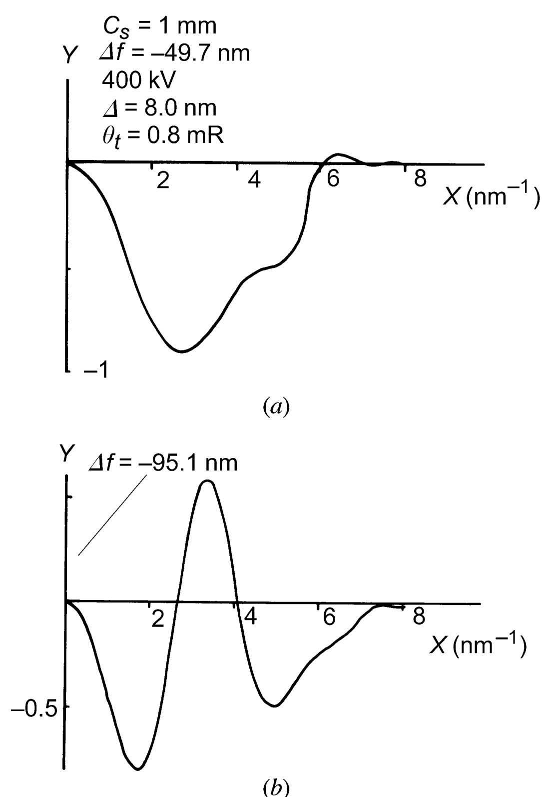

Fig. 4.3.8.5. (*a*) The transfer function for a 400 kV electron microscope with a point resolution of 1.7 Å at the Scherzer focus; the curve is based on equation (4.3.8.17). In (*b*) is shown a transfer function for similar conditions at the first 'passband' focus [$n = 1$ in equation (4.3.8.22)].

For the structure imaging of specific types of defects and materials, the following references are relevant. (i) For line defects viewed parallel to the line, d'Anterroches & Bourret (1984); viewed normal to the line, Alexander, Spence, Shindo, Gottschalk & Long (1986). (ii) For problems of variable lattice spacing (*e.g.* spinodal decomposition), Cockayne & Gronsky (1981). (iii) For point defects and their ordering, in tunnel structures, Yagi & Cowley (1978); in semiconductors, Zakharov, Pasemann & Rozhanski (1982); in metals, Fields & Cowley (1978). (iv) For interfaces, see the proceedings reported in *Ultramicroscopy* (1992), Vol. 40, No. 3. (v) For metals, Lovey, Coene, Van Dyck, Van Tendeloo, Van Landuyt & Amelinckx (1984). (vi) For organic crystals, Kobayashi, Fujiyoshi & Uyeda (1982). (vii) For a general review of applications in solid-state chemistry, see the collection of papers reported in *Ultramicroscopy* (1985), Vol. 18, Nos. 1–4. (viii) Radiation-damage effects are observed at atomic resolution by Horiuchi (1982).

4.3.8.4. *Parameters affecting HREM images*

The *instrumental parameters* that affect HREM images include accelerating voltage, astigmatism, optic-axis alignment, focus setting Δf, spherical-aberration constant C_s, beam divergence θ_c, and chromatic aberration constant C_c. Crystal parameters influencing HREM images include thickness, absorption, ionicity, and the alignment of the crystal zone axis with the beam, in addition to the structure factors and atom positions of the sample. The accurate measurement of electron wavelength or accelerating voltage has been discussed by many workers, including Uyeda, Hoier and others [see Fitzgerald & Johnson (1984) for references]. The measurement of Kikuchi-line spacings from crystals of known structure appears to be the most accurate and convenient method for HREM work, and allows an overall accuracy of better than 0.2% in accelerating voltage. Fluctuations in accelerating voltage contribute to the chromatic damping term Δ in equation (4.3.8.19) through the variance $\sigma^2(V_0)$. With the trend toward the use of higher accelerating voltages for HREM work, this term has become especially significant for the consideration of the information resolution limit [equation (4.3.8.21)].

Techniques for the accurate measurement of astigmatism and chromatic aberration are described by Spence (1988). The displacement of images of small crystals with beam tilt may be used to measure C_s; alternatively, the curvature of higher-order Laue-zone lines in CBED patterns has been used. The method of Budinger & Glaeser (1976) uses a similar dark-field image-displacement method to provide values for both Δf and C_s, and appears to be the most convenient and accurate for HREM work. The analysis of optical diffractograms initiated by Thon and co-workers from HREM images of thin amorphous films provides an invaluable diagnostic aid for HREM work; however, the determination of C_s by this method is prone to large errors, especially at small defocus. Diffractograms provide a rapid method for the determination of focus setting (see Krivanek, 1976) and in addition provide a sensitive indicator of specimen movement, astigmatism, and the damping-envelope constants Δ and Θ_c.

Misalignment of the electron beam, optic axis, and crystal axis in bright-field HREM work becomes increasingly important with increasing resolution and specimen thickness. The first-order effects of optical misalignment are an artifactual translation of spatial frequencies in the direction of misalignment by an amount proportional to the misalignment and to the square of spatial frequency. The corresponding phase shift is not observable in diffractograms. The effects of astigmatism on transfer functions for inclined illumination are discussed in Saxton (1978).

The effects of misalignment of the beam with respect to the optic axis are discussed in detail by Smith, Saxton, O'Keefe, Wood & Stobbs (1983), where it is found that all symmetry elements (except a mirror plane along the tilt direction) may be destroyed by misalignment. The maximum allowable misalignment for a given resolution δ in a specimen of thickness t is proportional to

$$\alpha = \delta/8t. \qquad (4.3.8.24)$$

Misalignment of a crystalline specimen with respect to the beam may be distinguished from misalignment of the optic axis with respect to the beam by the fact that, in very thin crystals, the former does not destroy centres of symmetry in the image.

The use of known defect point-group symmetry (for example in stacking faults) to identify a point in a HREM image with a point in the structure and so to resolve the black or white atomic contrast ambiguity has been described (Olsen & Spence, 1981). Structures containing screw or glide elements normal to the beam are particularly sensitive to misalignment, and errors as small as 0.2 mrad may substantially alter the image appearance.

A rapid comparison of images of amorphous material with the beam electronically tilted into several directions appears to be the best current method of aligning the beam with the optic axis, while switching to convergent-beam mode appears to be the most effective method of aligning the beam with the crystal axis. However, there is evidence that the angle of incidence of the incident beam is altered by this switching procedure.

The effects of misalignment and choice of beam divergence Θ_c on HREM images of crystals containing dynamically forbidden reflections are reviewed by Nagakura, Nakamura & Suzuki (1982) and Smith, Bursill & Wood (1985). Here the dramatic example of rutile in the [001] orientation is used to demonstrate how a misalignment of less than 0.2 mrad of the electron beam with respect to the crystal axis can bring up a coarse set of fringes (4.6 Å), which produce an image of incorrect symmetry, since these correspond to structure factors that are forbidden both dynamically and kinematically.

Crystal thickness is most accurately determined from images of planar faults in known orientations, or from crystal morphology for small particles. It must otherwise be treated as a refinement parameter. Since small crystals (such as MgO smoke particles, which form as perfect cubes) provide such an independent method of thickness determination, they provide the most convincing test of dynamical imaging theory. *The ability to match the contrast reversals and other detailed changes in HREM images as a function of either thickness or focus (or both) where these parameters have been measured by an independent method gives the greatest confidence in image interpretation.* This approach, which has been applied in rather few cases [see, for example, O'Keefe, Spence, Hutchinson & Waddington (1985)] is strongly recommended. The tendency for n-beam dynamical HREM images to repeat with increasing thickness in cases where the wavefunction is dominated by just two Bloch waves has been analysed by several workers (Kambe, 1982).

Since electron scattering factors are proportional to the difference between atomic number and X-ray scattering factors, and inversely proportional to the square of the scattering angle (see Section 4.3.1), it has been known for many years that the low-order reflections that contribute to HREM images are extremely sensitive to the distribution of bonding electrons and so to the degree of ionicity of the species imaged. This observation has formed the basis of several charge-density-map determinations by convergent-beam electron diffraction [see, for example, Zuo, Spence & O'Keefe (1988)]. Studies of ionicity effects on HREM imaging can be found in Anstis, Lynch, Moodie & O'Keefe (1973) and Fujiyoshi, Ishizuka, Tsuji, Kobayashi & Uyeda (1983).

The depletion of the elastic portion of the dynamical electron wavefunction by inelastic crystal excitations (chiefly phonons, single-electron excitations, and plasmons) may have dramatic effects on the HREM images of thicker crystals (Pirouz, 1974). For image formation by the elastic component, these effects may be described through the use of a complex 'optical' potential and the appropriate Debye–Waller factor (see Section 2.5.1). However, existing calculations for the absorption coefficients derived from the imaginary part of this potential are frequently not applicable to lattice images because of the large objective apertures used in HREM work. It has been suggested that HREM images formed from electrons that suffer small energy losses (and so remain 'in focus') but large-angle scattering events (within the objective aperture) due to phonon excitation may contribute high-resolution detail to images (Cowley, 1988). For measurements of the imaginary part of the optical potential by electron diffraction, the reader is referred to the work of Voss, Lehmpfuhl & Smith (1980), and references therein. All evidence suggests, however, that for the crystal thicknesses generally used for HREM work (t < 200Å) the effects of 'absorption' are small.

In summary, the general approach to the matching of computed and experimental HREM images proceeds as follows (Wilson, Spargo & Smith, 1982). (i) Values of Δ, Θ_c, and C_s are determined by careful measurements under well defined conditions (electron-gun bias setting, illumination aperture size, specimen height as measured by focusing-lens currents, electron-source size, *etc*). These parameters are then taken as constants for all subsequent work under these instrumental conditions (assuming also continuous monitoring of electronic instabilities). (ii) For a particular structure refinement, the parameters of thickness and focus are then varied, together with the choice of atomic model, in dynamical computer simulations until agreement is obtained. Every effort should be made to match images as a function of thickness and focus. (iii) If agreement cannot be obtained, the effects of small misalignments must be investigated (Smith *et al.*, 1985). Crystals most sensitive to these include those containing reflections that are absent due to the presence of screw or glide elements normal to the beam.

4.3.8.5. *Computing methods*

The general formulations for the dynamical theory of electron diffraction in crystals have been described in Section 5.2 of *IT* B (1992). In Section 4.3.6, the computing methods used for calculating diffraction-beam amplitudes have been outlined.

Given the diffracted-beam amplitudes, Ψ_g, the image is calculated by use of equations (4.3.8.2), including, when appropriate, the modifications of (4.3.8.13*b*).

The numerical methods that can be employed in relation to crystal-structure imaging make use of algorithms based on (i) matrix diagonalization, (ii) fast Fourier transforms, (iii) real-space convolution (Van Dyck, 1980), (iv) Runge-Kutta (or similar) methods, or (v) power-series evaluation. Two other solutions, the Cowley–Moodie polynomial solution and the

Feynman path-integral solution, have not been used extensively for numerical work. Methods (i) and (ii) have proven the most popular, with (ii) (the multislice method) being used most extensively for HREM image simulations. The availability of inexpensive array processors has made this technique highly efficient. A comparison of these two N-beam methods is given by Self, O'Keefe, Buseck & Spargo (1983), who find the multislice method to be faster (time proportional to $N \log_2 N$) than the diagonalization method (time proportional to N^2) for $N > 16$. Computing space increases roughly as N^2 for the diagonalization method, and as N for the multislice. The problem of steeply inclined boundary conditions for multislice computations has been discussed by Ishizuka (1982).

In the Bloch-wave formulation, the lattice image is given by

$$I(\mathbf{r}) = \sum_{i,j} \sum_{\mathbf{h},\mathbf{g}} C_0^{(i)} C_0^{(j)} C_{\mathbf{g}}^{(i)} C_{\mathbf{h}}^{(j)} \exp\{i[2\pi(\gamma^{(i)} - \gamma^{(j)})t + 2\pi(\mathbf{g} - \mathbf{h}) \cdot \mathbf{r} - \chi(\Delta f, C_s, \mathbf{g}) + \chi(\Delta f, C_s, \mathbf{h})]\}, \quad (4.3.8.25)$$

where $C_{\mathbf{g}}^{(i)}$ and $\gamma^{(i)}$ are the eigenvector elements and eigenvalues of the structure matrix [see Hirsch, Howie, Nicholson, Pashley & Whelan (1977) and Section 4.3.4].

Using modern personal computers or workstations, it is now possible to build efficient single-user systems that allow interactive dynamical structure-image calculations. Either an image intensifier or a cooled scientific grade charge-coupled device and single-crystal scintillator screen may be used to record the images, which are then transferred into a computer (Daberkow, Herrman, Liu & Rau, 1991). This then allows for the possibility of automated alignment, stigmation and focusing to the level of accuracy needed at 0.1 nm point resolution (Krivanek & Mooney, 1993). An image-matching search through trial structures, thickness and focus parameters can then be completed rapidly. Where large numbers of pixels, large dynamic range and high sensitivity are required, the Image Plate has definite advantages and so should find application in electron holography and biology (Shindo, Hiraga, Oikawa & Mori, 1990).

For the calculation of images of defects, the method of periodic continuation has been used extensively (Grinton & Cowley, 1971). Since, for kilovolt electrons traversing thin crystals, the transverse spreading of the dynamical wavefunction is limited (Cowley, 1981), the complex image amplitude at a particular point on the specimen exit face depends only on the crystal potential within a cylinder a few ångströms in diameter, erected about that point (Spence, O'Keefe & Iijima, 1978). The width of this cylinder depends on accelerating voltage, specimen thickness, and focus setting (see above references). Thus, small overlapping 'patches' of exit-face wavefunction may be calculated in successive computations, and the results combined to form a larger area of image. The size of the 'artificial superlattice' used should be increased until no change is found in the wavefunction over the central region of interest. For most defects, the positions of only a few atoms are important and, since the electron wavefunction is locally determined (for thin specimens at Scherzer focus), it appears that very large calculations are rarely needed for HREM work. The simulation of profile images of crystal surfaces at large defocus settings will, however, frequently be found to require large amounts of storage.

A new program should be tested to ensure that (*a*) under approximate two-beam conditions the calculated extinction distances for small-unit-cell crystals agree roughly with tabulated values (Hirsch *et al.*, 1977), (*b*) the simulated dynamical images have the correct symmetry, (*c*) for small thickness, the Scherzer-focus images agree with the projected potential, and (*d*) images and beam intensities agree with those of a program known to be correct. The damping envelope (product representation) [equation (4.3.8.17)] should only be used in a thin crystal with $\Phi_0 > \Phi_{\mathbf{g}}$; in general, the effects of partial spatial and temporal coherence must be incorporated using equation (4.3.8.13*a*) or (4.3.8.13*b*), depending on whether variations in diffraction conditions over θ_c are important. Thus, a separate multislice dynamical-image calculation for each component plane wave in the incident cone of illumination may be required, followed by an incoherent sum of all resulting images.

The outlook for obtaining higher resolution at the time of writing (1997) is broadly as follows. (1) The highest point resolution currently obtainable is close to 0.1 nm, and this has been obtained by taking advantage of the reduction in electron wavelength that occurs at high voltage [equation (4.3.8.16)]. A summary of results from these machines can be found in *Ultramicroscopy* (1994), Vol. 56, Nos. 1–3, where applications to fullerenes, glasses, quasicrystals, interfaces, ceramics, semiconductors, metals and oxides and other systems may be found. Fig. 4.2.8.6 shows a typical result. High cost, and the effects of radiation damage (particularly at larger thickness where defects with higher free energies are likely to be found), may limit these machines to a few specialized laboratories in the future. The attainment of higher resolution through this approach depends on advances in high-voltage engineering. (2) Aberration coefficients may be reduced if higher magnetic fields can be produced in the pole piece, beyond the saturation flux of the specialized iron alloys currently used. Research into superconducting lenses has therefore continued for many years in a few laboratories. Fluctuations in lens current are also eliminated by this method. (3) Electron holography was originally developed for the purpose of improving electron-microscope resolution, and this approach is reviewed in the following section. (4) Electron–optical correction of aberrations has been under study for many years in work by Scherzer, Crewe, Beck, Krivanek, Lanio, Rose and others – results of recent experimental tests are described in Haider & Zach (1995) and Krivanek, Dellby, Spence, Camps & Brown (1997). The attainment of 0.1 nm point resolution is considered feasible. Aberration correctors will also provide benefits other than increased resolution, including greater space in the pole piece for increased sample tilt and access to X-ray detectors, *etc*.

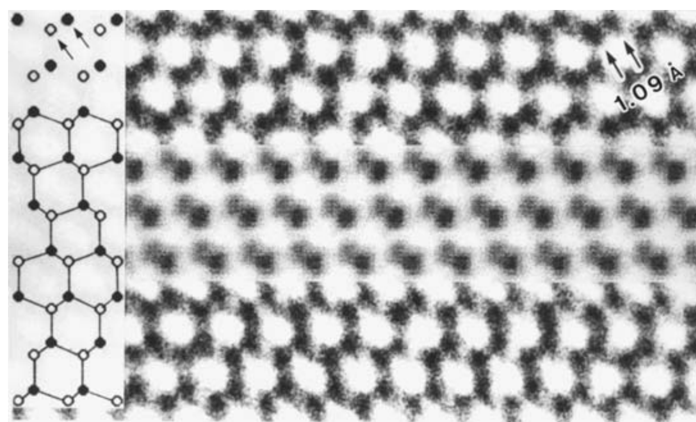

Fig. 4.3.8.6. Structure image of a thin lamella of the 6*H* polytype of SiC projected along [110] and recorded at 1.2 MeV. Every atomic column (darker dots) is separately resolved at 0.109 nm spacing. The central horizontal strip contains a computer-simulated image; the structure is sketched at the left. [Courtesy of H. Ichinose (1994).]

The need for resolution improvement beyond 0.1 nm has been questioned – the structural information retrievable by a single HREM image is always limited by the fact that a projection is obtained. (This problem is particularly acute for glasses.) Methods for combining different projected images (particularly of defects) from the same region (Downing, Meisheng, Wenk & O'Keefe, 1990) may now be as important as the search for higher resolution.

4.3.8.6. *Resolution and hyper-resolution*

Since the resolution of an instrument is a property of the instrument alone, whereas the ability to distinguish HREM image features due to adjacent atoms depends on the scattering properties of the atoms, the resolution of an electron microscope cannot easily be defined [see Subsection 2.5.1.9 in *IT* B (1992)]. The Rayleigh criterion was developed for the incoherent imaging of point sources and cannot be applied to coherent phase contrast. Only for very thin specimens of light elements for which it can be assumed that the scattering phase is $-\pi/2$ can the straightforward definition of point resolution d_p [equation (4.3.8.16)] be applied. In general, the dynamical wavefunction across the exit face of a crystalline sample bears no simple relationship to the crystal structure, other than to preserve its symmetry and to be determined by the 'local' crystal potential. The use of a dynamical '*R* factor' between computed and experimental images of a known structure has been suggested by several workers as the basis for a more general resolution definition.

For weakly scattering specimens, the most satisfactory method of measuring either the point resolution d_p or the information limit d_i [see equation (4.3.8.21)] appears to be that of Frank (1975). Here two successive micrographs of a thin amorphous film are recorded (under identical conditions) and the superimposed pair used to obtain a coherent optical diffractogram crossed by fringes. The fringes, which result from small displacements of the micrographs, extend only to the band limit d_i^{-1} of information common to both micrographs, and cannot be extended by photographic processing, noise, or increased exposure. By plotting this band limit against defocus, it is possible to determine both Δ and θ_c. As an alternative, for thin crystalline samples of large-unit-cell materials, the parameters Δ, θ_c, and C_s can be determined by matching computed and experimental images of crystals of known structure. It is the specification of these parameters (for a given electron intensity and wavelength) that is important in describing the performance of high-resolution electron microscopes. We note that certain conditions of focus or thickness may give a spurious impression of ultra-high resolution [see equations (4.3.8.7) and (4.3.8.8)].

Within the domain of linear imaging, implying, for the most part, the validity of the WPO approximation, many forms of image processing have been employed. These have been of particular importance for crystalline and non-crystalline biological materials and include image reconstruction [see Section 2.5.4 in *IT* B (1992)] and the derivation of three-dimensional structures from two-dimensional projections [see Section 2.5.5 in *IT* B (1992)]. For reviews, see also Saxton (1980*a*), Frank (1980), and Schiske (1975). Several software packages now exist that are designed for image manipulation, Fourier analysis, and cross correlation; for details of these, see Saxton (1980*a*) and Frank (1980). The theoretical basis for the WPO approximation closely parallels that of axial holography in coherent optics, thus much of that literature can be applied to HREM image processing. Gabor's original proposal for holography was intended for electron microscopy [see Cowley (1981) for a review].

The aim of image-processing schemes is the restoration of the exit-face wavefunction, given in equation (4.3.8.13*a*). The reconstruction of the crystal potential $\varphi_p(\mathbf{r})$ from this is a separate problem, since these are only simply related under the approximation of Subsection 4.3.8.3. For a non-linear method that allows the reconstruction of the dynamical image wavefunction, based on equation (4.3.8.13*b*), which thus includes the effects of multiple scattering, see Saxton (1980*b*).

The concept of holographic reconstruction was introduced by Gabor (1948, 1949) as a means of enhancing the resolution of electron microscopes. Gabor proposed that, if the information on relative phases of the image wave could be recorded by observing interference with a known reference wave, the phase modification due to the objective-lens aberrations could be removed. Of the many possible forms of electron holography (Cowley, 1994), two show particular promise of useful improvements of resolution. In what may be called in-line TEM holography, a through-focus series of bright-field images is obtained with near-coherent illumination. With reference to the relatively strong transmitted beam, the relative phase and amplitude changes due to the specimen are derived from the variations of image intensity (see Van Dyck, Op de Beeck & Coene, 1994). The tilt-series reconstruction method also shows considerable promise (Kirkland, Saxton, Chau, Tsuno & Kawasaki, 1995).

In the alternative off-axis approach, the reference wave is that which passes by the specimen area in vacuum, and which is made to interfere with the wave transmitted through the specimen by use of an electrostatic biprism (Möllenstedt & Düker, 1956). The hologram consists of a modulated pattern of interference fringes. The image wavefunction amplitude and phase are deduced from the contrast and lateral displacements of the fringes (Lichte, 1991; Tonomura, 1992). The process of reconstruction from the hologram to give the image wavefunction may be performed by optical-analogue or digital methods and can include the correction of the phase function to remove the effects of lens aberrations and the attendant limitation of resolution. The point resolution of electron microscopes has recently been exceeded by this method (Orchowski, Rau & Lichte, 1995).

The aim of the holographic reconstructions is the restoration of the wavefunction at the exit face of the specimen as given by equation (4.3.8.13*a*). The reconstruction of the crystal potential $\varphi(\mathbf{r})$ from this is a separate problem, since the exit-face wavefunction and $\varphi(\mathbf{r})$ are simply related only under the WPO approximations of Subsection 4.3.8.3. The possibility of deriving reconstructions from wavefunctions strongly affected by dynamical diffraction has been considered by a number of authors (for example, Van Dyck *et al.*, 1994). The problem does not appear to be solvable in general, but for special cases, such as perfect thin single crystals in exact axial orientations, considerable progress may be possible.

Since a single atom, or a column of atoms, acts as a lens with negative spherical aberration, methods for obtaining super-resolution using atoms as lenses have recently been proposed (Cowley, Spence & Smirnov, 1997).

4.3.8.7. *Alternative methods*

A number of non-conventional imaging modes have been found useful in electron microscopy for particular applications. In scanning transmission electron microscopy (STEM), powerful electron lenses are used to focus the beam from a very small bright source, formed by a field-emission gun, to form a small probe that is scanned across the specimen. Some selected part of the transmitted electron beam (part of the coherent convergent-

beam electron diffraction pattern produced) is detected to provide the image signal that is displayed or recorded in synchronism with the incident-beam scan. The principle of reciprocity suggests that, for equivalent lenses, apertures and column geometry, the resolution and contrast of STEM and TEM images will be identical (Cowley, 1969). Practical considerations of instrumental convenience distinguish particularly useful STEM modes.

Crewe & Wall (1970) showed that, if an annular detector is used to detect all electrons scattered outside the incident-beam cone, dark-field images could be obtained with high efficiency and with a resolution better than that of the bright-field mode by a factor of about 1.4. If the inner radius of the annular detector is made large (of the order of 10^{-1} rad for 100 kV electrons), the strong diffracted beams occurring for lower angles do not contribute to the resulting high-angle annular dark-field (HAADF) image (Howie, 1979), which is produced mainly by thermal diffuse scattering. The HAADF mode has important advantages for particular purposes because the contrast is strongly dependent on the atomic number, Z, of the atoms present but is not strongly affected by dynamical diffraction effects and so shows near-linear variation with Z and with the atom-number density in the sample. Applications have been made to the imaging of small high-Z particles in low-Z supports, such as in supported metal catalysts (Treacy & Rice, 1989) and to the high-resolution imaging of individual atomic rows in semiconductor crystals, showing the variations of composition across planar interfaces (Pennycook & Jesson, 1991).

The STEM imaging modes may be readily correlated with microchemical analysis of selected specimen areas having lateral dimensions in the nanometre range, by application of the techniques of electron energy-loss spectroscopy or X-ray energy-dispersive analysis (Williams & Carter, 1996; Section 4.3.4). Also, diffraction patterns (coherent convergent-beam electron diffraction patterns) may be obtained from any chosen region having dimensions equal to those of the incident-beam diameter and as small as about 0.2 nm (Cowley, 1992). The coherent interference between diffracted beams within such a pattern may provide information on the symmetries, and, ultimately, the atomic arrangement, within the illuminated area, which may be smaller than the projection of the crystal unit cell in the beam direction. This geometry has been used to extend resolution for crystalline samples beyond even the information resolution limit, d_i (Nellist, McCallum & Rodenburg, 1995), and is the basis for an exact, non-perturbative inversion scheme for dynamical electron diffraction (Spence, 1998).

The detection of secondary radiations (light, X-rays, low-energy 'secondary' electrons, *etc.*) in STEM or the detection of energy losses of the incident electrons, resulting from particular elementary excitations of the atoms in a crystal, in TEM or STEM, may be used to form images showing the distributions in a crystal structure of particular atomic species. In principle, this may be extended to the chemical identification of individual atom types in the projection of crystal structures, but only limited success has been achieved in this direction because of the relatively low level of the signals available. The formation of atomic resolution images using inner-shell excitations, for example, is complicated by the Bragg scattering of these inelastically scattered electrons (Endoh, Hashimoto & Makita, 1994; Spence & Lynch, 1982).

Reflection electron microscopy (REM) has been shown to be a powerful technique for the study of the structures and defects of crystal surfaces with moderately high spatial resolution (Larsen & Dobson, 1988), especially when performed in a specially built electron microscope having an ultra-high-vacuum specimen environment (Yagi, 1993). Images are formed by detecting strong diffracted beams in the RHEED patterns produced when kilovolt electron beams are incident on flat crystal surfaces at grazing incidence angles of a few degrees. The images suffer from severe foreshortening in the beam direction, but, in directions at right angles to the beam, resolutions approaching 0.3 nm have been achieved (Koike, Kobayashi, Ozawa & Yagi, 1989). Single-atom-high surface steps are imaged with high contrast, surface reconstructions involving only one or two monolayers are readily seen and phase transitions of surface superstructures may be followed.

The study of surface structure by use of high-resolution transmission electron microscopes has also been productive in particular cases. Images showing the structures of surface layers with near-atomic resolution have been obtained by the use of 'forbidden' or 'termination' reflections (Cherns, 1974; Takayanagi, 1984) and by phase-contrast imaging (Moodie & Warble, 1967; Iijima, 1977). The imaging of the profiles of the edges of thin or small crystals with clear resolution of the surface atomic layers has also been effective (Marks, 1986). The introduction of the scanning tunnelling microscope (Binnig, Rohrer, Gerber & Weibel, 1983) and other scanning probe microscopies has broadened the field of high-resolution surface structure imaging considerably.

4.3.8.8. *Combined use of HREM and electron diffraction*

For many materials of organic or biological origin, it is possible to obtain very thin crystals, only one or a few molecules thick, extending laterally over micrometre-size areas. These may give selected-area electron-diffraction patterns in electron microscopes with diffraction spots extending out to angles corresponding to d spacings as low as 0.1 nm. Because the materials are highly sensitive to electron irradiation, conventional bright-field images cannot be obtained with resolutions better than several nanometres. However, if images are obtained with very low electron doses and then a process of averaging over the content of a very large number of unit cells of the image is carried out, images showing detail down to the scale of 1 nm or less may be derived for the periodically repeated unit. From such images, it is possible to derive both the magnitudes and phases of the Fourier coefficients, the structure factors, out to some limit of d spacings, say d_m. From the diffraction patterns, the magnitudes of the structure factors may be deduced, with greater accuracy, out to a much smaller limit, d_d. By combination of the information from these two sources, it may be possible to obtain a greatly improved resolution for an enhanced image of the structure. This concept was first introduced by Unwin & Henderson (1975), who derived images of the purple membrane from *Halobacterium halobium*, with greatly improved resolution, revealing its essential molecular configuration.

Recently, several methods of phase extension have been developed whereby the knowledge of the relative phases may be extended from the region of the diffraction pattern covered by the electron-microscope image transform to the outer parts. These include methods based on the use of the tangent formula or Sayre's equation (Dorset, 1994; Dorset, McCourt, Fryer, Tivol & Turner, 1994) and on the use of maximum-entropy concepts (Fryer & Gilmore, 1992). Such methods have also been applied, with considerable success, to the case of some thin inorganic crystals (Fu *et al.*, 1994). In this case, the limitation on the resolution set by the electron-microscope images may be that due to the transfer function of the microscope, since radiation-damage effects are not so limiting. Then, the resolution achieved

by the combined application of the electron diffraction data may represent an advance beyond that of normal HREM imaging. Difficulties may well arise, however, because the theoretical basis for the phase-extension methods is currently limited to the WPO approximation. A summary of the present situation is given in the book by Dorset (1995).

references