Image resolution and image contrast in the electron microscope. I. Elastic scattering and coherent illumination

1973 J. Phys. A: Math. Nucl. Gen. 6 62

(http://iopscience.iop.org/0301-0015/6/1/007)

J. Phys. A: Math., Nucl. Gen., Vol. 6, January 1973. Printed in Great Britain.

Image resolution and image contrast in the electron microscope I. Elastic scattering and coherent illumination

D L Misell

Department of Physics, Queen Elizabeth College, Campden Hill Road, London W8 7AH, UK

MS received 24 July 1972

Abstract. The present work evaluates the effect of lens aberrations on image resolution and image contrast in the transmission electron microscope for the elastic component of the electron beam; coherent illumination of the specimen is assumed. The effects of lens aberrations on the image resolution are evaluated in terms of a resolution function in preference to the use of the transfer function. Although no precise figures can be given for image resolution, it is found that the condition for maximum contrast corresponds to the best resolution in the image. The fundamental limit in resolution, due to diffraction at the objective aperture, should be avoided by increasing the objective aperture size with a subsequent correction for the increased lens aberrations. For coherent illumination these corrections can, in general, only be made in bright field microscopy and not in dark field microscopy.

1. Introduction

The effect of the objective lens on the transmitted electron beam (wavefunction $\psi_0(\boldsymbol{r}_0)$) for elastic scattering is described by the convolution integral (eg Lenz 1965, 1971b, Hanszen 1971)

$$\psi_i(\boldsymbol{r}_i) = \int \psi_0(\boldsymbol{r}_0) G(\boldsymbol{r}_i - \boldsymbol{r}_0)\, \mathrm{d}\boldsymbol{r}_0 \tag{1}$$

for coherent illumination. $\psi_i(\boldsymbol{r}_i)$ is the electron wavefunction in the image plane, as affected by the lens aberration function $G(\boldsymbol{r})$ (in the isoplanatic approximation with magnification $M = 1$). In terms of the Fourier transforms of ψ_0 and G, equation (1) becomes

$$\psi_i(\boldsymbol{r}_i) = \int S_0(\mathbf{v}) T(\mathbf{v}) \exp(-2\pi i \mathbf{v} \cdot \boldsymbol{r}_i)\, \mathrm{d}\mathbf{v} \tag{2}$$

where $S_0(\mathbf{v}) = \delta(\mathbf{v}) + S(\mathbf{v})$ represents the scattered wave (+ unscattered $\delta(\mathbf{v})$) in the back focal plane of the objective lens. $T(\mathbf{v})$ is referred to as the transfer function for a spatial frequency $\mathbf{v}$; $\mathbf{v}$ is related to the angle of scattering $\boldsymbol{\theta}$ by $\lambda_0 \mathbf{v} = \boldsymbol{\theta}$ for electrons of wavelength λ_0. The phase shift introduced by $T(\mathbf{v})$ into $S_0(\mathbf{v})$ for a lens subject to spherical aberration (coefficient C_s) and defocusing (Δf) is given by (Hanszen 1971)

$$T(\mathbf{v}) = \tfrac{1}{2}(Q_1(\mathbf{v}) - iQ(\mathbf{v}))B(\mathbf{v}) = \{\cos(K_0 W(\mathbf{v})) - i \sin(K_0 W(\mathbf{v}))\} B(\mathbf{v}) \tag{3}$$

and

$$W(\mathbf{v}) = \tfrac{1}{4}C_s\lambda_0^4 v^4 + \tfrac{1}{2}\Delta f \lambda_0^2 v^2. \tag{4}$$

The factor of $\frac{1}{2}$ in equation (3) is introduced here only to simplify later analysis. $B(\mathbf{v})$ is the aperture function for the objective lens and $K_0 = 2\pi/\lambda_0$.

The use of $T(\mathbf{v})$ to describe the effect of lens aberrations on the image intensity $j_i(\boldsymbol{r}_i) = |\psi_i(\boldsymbol{r}_i)|^2$ has been in terms of the removal of certain spatial frequencies in $S_0(\mathbf{v})$, so eliminating certain structural information from the final image. The oscillatory behaviour of $T(\mathbf{v})$ in the integral (2) can be used to describe the contrast $C_i(\boldsymbol{r}_i)$ in the image plane; adjusting Δf to minimize the number of oscillations in $T(\mathbf{v})$ leads to a condition for maximum contrast (eg Zeitler and Thomson 1970). However, in order to discuss image resolution, the resolution function $G(\boldsymbol{r})$ is more informative;

$$G(\boldsymbol{r}) = \tfrac{1}{2}(q_1(\boldsymbol{r}) - \mathrm{i}q(\boldsymbol{r}))$$

where $q_1(\boldsymbol{r})$ and $q(\boldsymbol{r})$ are respectively the inverse Fourier transforms of $Q_1(\mathbf{v})$ and $Q(\mathbf{v})$. In particular, the present work examines the effect of lens aberrations on the image and the relationship between the object structure, defined by $\psi_0(\boldsymbol{r}_0)$, and $j_i(\boldsymbol{r}_i)$ for (i) a phase object and (ii) an amplitude object (Hanszen 1971) under conditions of spatially and chromatically coherent illumination; describing the phase shift by $\eta(\boldsymbol{r}_0)$ and the amplitude attenuation (absorption) by $\epsilon(\boldsymbol{r}_0)$ (Lenz 1971a):

$$\psi_0(\boldsymbol{r}_0) \simeq 1 + \mathrm{i}\eta(\boldsymbol{r}_0) - \epsilon(\boldsymbol{r}_0) \tag{5}$$

where $\eta(\boldsymbol{r}_0)$ is the projection of the potential distribution in the object onto the (x_0, y_0) plane, that is, (eg Grinton and Cowley 1971)

$$\eta(\boldsymbol{r}_0) = \frac{-2\pi me\lambda_0}{h^2}\int_0^t V(x_0, y_0, z)\,\mathrm{d}z \simeq \frac{-2\pi me\lambda_0}{h^2} t V_{\mathrm{mean}}(x_0, y_0) \tag{6}$$

for a specimen of thickness t; V_{mean} is a measure of the potential difference from the mean inner potential. For approximation (5) to be valid $\eta(\boldsymbol{r}_0) \ll \pi/2$, say $\eta \simeq 0{\cdot}1$, corresponding to $t_{\max} = 13{\cdot}0$–$1{\cdot}3$ nm ($E = 100$ keV) and $t = 5{\cdot}6$–$0{\cdot}6$ nm ($E_0 = 20$ keV) for $V_{\mathrm{mean}} \simeq -2$ V (for an amorphous specimen) to $V_{\mathrm{mean}} \simeq -20$ V (for a heavy metal staining material). The origin of the amplitude attenuation (absorption) term $\epsilon(\boldsymbol{r}_0)$ is not clear; ϵ is used to represent scattering outside the objective aperture (eg Grinton and Cowley 1971) or equivalently scattering contrast (Erickson and Klug 1971), the anomolous phase shift introduced by the failure of the first Born approximation (eg Frank 1972), inelastic scattering (eg Radi 1970). Empirically the parameter $\epsilon(\boldsymbol{r}_0)$ is required to interpret an electron micrograph (eg Erickson and Klug 1971) and the requirement that $\epsilon(\boldsymbol{r}_0) \ll 1$ leads to similar limits on $t_{\max}$ as given above for $\eta(\boldsymbol{r}_0)$. If inelastic scattering is omitted from the consideration of image formation, then $\epsilon(\boldsymbol{r}_0)$ is a necessary parameter, although $\epsilon(\boldsymbol{r}_0)$ takes account also of neglected second order terms, such as $\frac{1}{2}\eta^2(\boldsymbol{r}_0)$, omitted from equation (5).

In § 2 we give the basic equations for the image intensity and the model for the specimen structure. Initially we examine the effect of the diffraction limit, as described by $B(\mathbf{v})$ in equation (3), on the image resolution (§ 3). Section 4 examines the condition for maximum contrast at an origin corresponding to the centre of the specimen structure ($\boldsymbol{r}_0 = 0$). For the defocus values corresponding to maximum contrast, we examine the

effect of lens aberrations on the image intensity for electrons of incident energy $E_0 = 20$ keV and $E_0 = 100$ keV (§ 5).

2. The basic equations for the image intensity

In bright field microscopy the relationship between $j_i(\boldsymbol{r}_i)$ and $\psi_0(\boldsymbol{r}_0)$ is ($B(0) = 1$, eg Lenz 1971a)

$$j_i(\boldsymbol{r}_i) \simeq 1 + \int \eta(\boldsymbol{r}_0) q(\boldsymbol{r}_i - \boldsymbol{r}_0)\, d\boldsymbol{r}_0 - \int \epsilon(\boldsymbol{r}_0) q_1(\boldsymbol{r}_i - \boldsymbol{r}_0)\, d\boldsymbol{r}_0 \tag{7}$$

neglecting second order terms.

Evidently equation (7) may be solved for both η and ϵ from at least two micrographs (eg Hoppe *et al* 1970, Erickson and Klug 1971, Frank 1972, Hoenders 1972) but not from a single micrograph. In dark field microscopy ($B(0) = 0$)

$$j_i(\boldsymbol{r}_i) = \left| \int (i\eta(\boldsymbol{r}_0) - \epsilon(\boldsymbol{r}_0)) G(\boldsymbol{r}_i - \boldsymbol{r}_0)\, d\boldsymbol{r}_0 \right|^2 \tag{8}$$

which, in general, cannot be inverted to give $i\eta(\boldsymbol{r}_0) - \epsilon(\boldsymbol{r}_0)$, although both the real and imaginary parts of G are known. Equation (8) may be transformed to a convolution square integral (eg Hosemann and Bagchi 1962); writing $\phi(\boldsymbol{r}_0) = i\eta(\boldsymbol{r}_0) - \epsilon(\boldsymbol{r}_0)$:

$$j_i(\boldsymbol{r}_i) = \int\int \phi(\boldsymbol{r}_0) G(\boldsymbol{r}_i - \boldsymbol{r}_0) \phi^*(\boldsymbol{r}_0') G^*(\boldsymbol{r}_i - \boldsymbol{r}_0')\, d\boldsymbol{r}_0\, d\boldsymbol{r}_0'$$

giving for the Fourier transform of $j_i(\boldsymbol{r}_i)$, $J(\boldsymbol{\nu})$

$$J(\boldsymbol{\nu}) = \int K(\boldsymbol{\nu}) K^*(\boldsymbol{\nu} - \boldsymbol{\nu}')\, d\boldsymbol{\nu}' \tag{9}$$

with

$$K(\boldsymbol{\nu}) = F(\phi(\boldsymbol{r}))F(G(\boldsymbol{r})).$$

Only in the case of an object with a centre of symmetry may a formal solution be obtained for $K(\boldsymbol{\nu})$ (Hosemann and Bagchi 1962). Thus, in general, only bright field microscopy allows an inversion of the convolution integral to give information on the object structure.

We discuss now the model used for the object structure: both $\eta(\boldsymbol{r}_0)$ and $\epsilon(\boldsymbol{r}_0)$ are represented by a gaussian $A \exp(-br_0^2)$ of radial halfwidth 0·37 nm ($b = 20\,\text{nm}^{-2}$); the cylindrical symmetry of both η and ϵ, and the resolution function $G(\boldsymbol{r})$ simplifies the presentation of the results. The condition that $\eta \simeq 0{\cdot}1$ gives $A \simeq 0{\cdot}1$; estimates for A based on the total scattered intensity

$$\int |S(\boldsymbol{\nu})|^2\, d\boldsymbol{\nu}$$

give $A \simeq 0{\cdot}1$ for 10 % elastic scattering from the incident beam.

In the case of $\epsilon(\boldsymbol{r}_0)$ a realistic value for A was based on the value for the imaginary part of the complex potential $V^{(i)}$ (eg Radi 1970) also giving $A \simeq 0{\cdot}1$. The limits on the specimen thickness vary according to the constituent atoms and the incident electron energy: for $t/\Lambda_E = 0{\cdot}1$ (Λ_E is the mean free path for elastic scattering), corresponding to 10 % scattering, t_{max} varies from 13·0 nm for carbon to 0·6 nm for gold ($E_0 = 100$ keV)

in agreement with the values given in § 1. The stain distribution in a biological specimen is an important factor in estimating a t_{max} for a real specimen.

3. The effect of the diffraction limit

In this section we examine the effect of the objective aperture size on the image intensity $j_i(\boldsymbol{r}_i)$; this represents the best resolution that can be derived from an electron micrograph after correction for lens aberrations. If $C_s = 0$, $\Delta f = 0$, then the resolution function $G(\boldsymbol{r}) = \nu_{max} J_1(2\pi\nu_{max} r)/r$ for a circular aperture of semi-angle $\alpha = \lambda_0 \nu_{max}$. In terms of the transfer function $T(\boldsymbol{\nu})$, $T(\boldsymbol{\nu})$ is zero for $|\boldsymbol{\nu}| > \nu_{max}$ and all spatial frequencies greater than this are absent from the image.

If we require to resolve detailed molecular structure (detail about 0·15 nm) then the α values required are $\alpha_{20} > 0{\cdot}05$ rad and $\alpha_{100} > 0{\cdot}02$ rad; the correction of the lens aberrations that arise from such α values (with the normal $C_s = 2$ mm) would appear to be difficult, especially in the presence of noise. Thus in the present work, we limit the maximum value of α_{20} to 0·02 rad and α_{100} to 0·015 rad where the lens aberration effects, although severe (see § 5), may be corrected.

Figure 1 shows the behaviour of $G(\boldsymbol{r})$ for $\alpha = 0{\cdot}01$, 0·015 and 0·02 rad ($E_0 = 20$ keV, figure 1(*a*)) and for $\alpha = 0{\cdot}005$, 0·01, 0·015 rad ($E_0 = 100$ keV, figure 1(*b*)) and the corresponding image intensity in bright field microscopy ($E_0 = 20$ keV, figure 1(*c*); $E_0 = 100$ keV, figure 1(*d*)), where the zero value for $j_i(\boldsymbol{r}_i)$ corresponds to the background level of unity. Evidently if we wish to reconstruct an accurate representation, after correction for lens aberrations, of any structural feature of less than 0·4 nm the smaller objective aperture sizes (0·01, 0·015 rad for $E_0 = 20$ keV; 0·005 for $E_0 = 100$ keV) should not be used. For $\alpha = 0{\cdot}02$ ($E_0 = 20$ keV) and $\alpha = 0{\cdot}01, 0{\cdot}015$ ($E_0 = 100$ keV) the image intensity gives a good representation of the gaussian structure (dotted curves in figures 1(*c*) and 1(*d*)). If we assume that the diffraction effect is a fundamental limit on image information, then we must use the larger objective apertures and rely on a correction procedure for lens aberrations. The correction of a micrograph for lens aberrations with a smaller objective aperture, although a simpler problem, relies on the concept of analytic continuation for extending the structural information in $J(\boldsymbol{\nu})$ beyond $\nu_{max} = \alpha/\lambda_0$ (Peřina and Kvapil 1968, Peřina 1971).

In the present work we examine in more detail image contrast and image resolution for only these larger α values.

4. The conditions for maximum image contrast

If we consider a structural feature in the object at $\boldsymbol{r}_0 = 0$, then practically we defocus the objective lens to 'maximize' the image contrast at $\boldsymbol{r}_i = 0$, $C_i(0)$, or equivalently $j_i(0)$. Here we refer to the 'maximum' contrast arising from either a mathematical minimum or maximum in $j_i(0)$ as Δf adjusted or α is varied. Additional interest in the condition for 'maximum' contrast relates the achievement of the best relationship between the object structure and the image intensity. The dependence of $j_i(0)$ on α and Δf is complex and we consider below optimizing $j_i(0)$ with respect to α (§ 4.1) and Δf (§ 4.2) separately; in the former case we obtain analytic results but in the latter case, which is of more practical interest (since α is normally fixed by the lens geometry), only numerical results can be presented. Additionally we show that the condition for maximum contrast can be

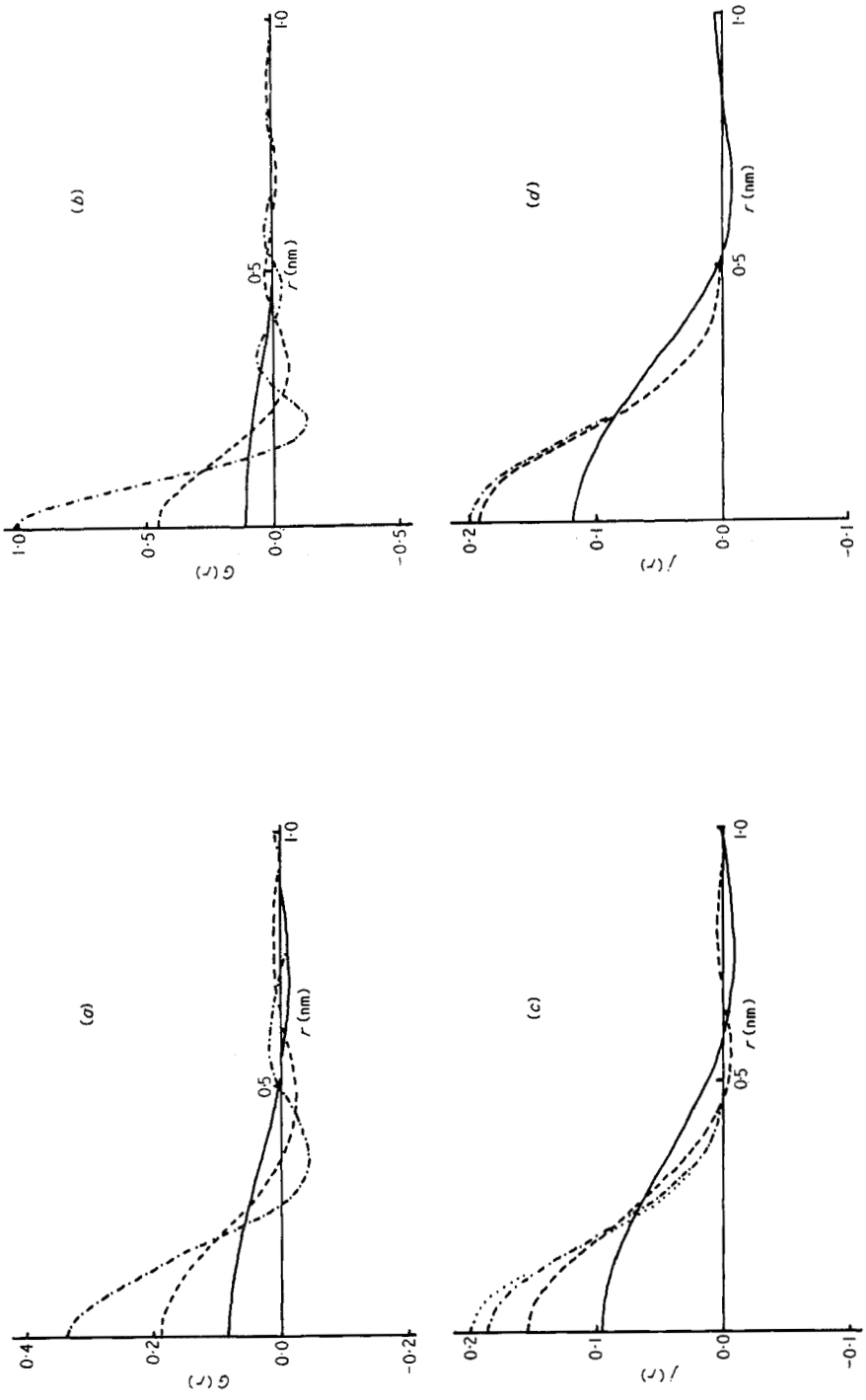


Figure 1. The effect of the diffraction limit on image resolution. (*a*) The resolution function $G(r)$ for $E_0 = 20$ keV: full curve, $\alpha = 0{\cdot}01$; broken curve, $\alpha = 0{\cdot}015$; chain curve, $\alpha = 0{\cdot}02$ rad. (*b*) $G(r)$ for $E_0 = 100$ keV: full curve, $\alpha = 0{\cdot}005$; broken curve, $\alpha = 0{\cdot}01$; chain curve, $\alpha = 0{\cdot}015$ rad. (*c*) and (*d*) are the corresponding image intensity distribution $j(r)$; the dotted curve is the gaussian structure ($\alpha = 0{\cdot}015$). $E_0 = 100$ keV curve coincides with the dotted curve.

related to the behaviour of the resolution functions $q(\boldsymbol{r})$ and $q_1(\boldsymbol{r})$ (§ 4.3). The spherical aberration constant $C_s = 2$ mm in all calculations.

4.1. Contrast variation with the objective aperture size

In the case of the weak phase-amplitude object (equation (7))

$$j_i(\boldsymbol{r}_i) = 1 + 2\int S(\mathbf{v})\sin(K_0 W(\mathbf{v}))B(\mathbf{v})\exp(-2\pi i\mathbf{v}\cdot\boldsymbol{r}_i)\,d\mathbf{v} - 2\int E(\mathbf{v})\cos(K_0 W(\mathbf{v}))B(\mathbf{v}) \times \exp(-2\pi i\mathbf{v}\cdot\boldsymbol{r}_i)\,d\mathbf{v} \tag{10}$$

where $S(\mathbf{v}) = F(\eta(\boldsymbol{r}_0))$ and $E(\mathbf{v}) = F(\epsilon(\boldsymbol{r}_0))$. Evidently it is not possible to 'maximize' $j_i(0)$ with respect to both $\eta(\boldsymbol{r}_0)$ and $\epsilon(\boldsymbol{r}_0)$ and we consider separately the criteria for maximum contrast for a phase object ($\epsilon(\boldsymbol{r}_0) = 0$) and an amplitude object ($\eta(\boldsymbol{r}_0) = 0$); we see that the two sets of conditions are, in general, incompatible.

4.1.1. The phase object. The terms of angular coordinates (θ, ϕ), equation (10) becomes with $\Psi_s(\theta, \phi) \equiv S(\mathbf{v})$:

$$j_i(\boldsymbol{r}_i) = 1 + \frac{K_0}{\pi}\int_0^{2\pi}\int_0^{\alpha}\Psi_s(\theta,\phi)\sin(K_0\chi(\theta))\exp\{-iK_0(x_i\theta\cos\phi + y_i\theta\sin\phi)\}\theta\,d\theta\,d\phi \tag{11}$$

for the normal circular aperture and $\chi(\theta) = \frac{1}{4}C_s\theta^4 + \frac{1}{2}\Delta f\,\theta^2$. For a local maximum or minimum at $j_i(0)$ with varying (Δf fixed) we require

$$\frac{\partial j_i(0)}{\partial\alpha} = \frac{K_0}{\pi}\int_0^{2\pi}\Psi_s(\alpha,\phi)\sin(K_0\chi(\alpha))\alpha\,d\phi = 0. \tag{12}$$

The general condition, satisfying equation (12) irrespective of the form of Ψ_s, is

$$K_0\chi(\alpha) = \pm n\pi \qquad \text{or} \qquad \Delta f = \frac{\pm 2n\pi}{\alpha^2 K_0} - \frac{C_s\alpha^2}{2} \tag{13}$$

and as α is varied we obtain local maxima and minima at Δf values given by equation (13), leading to a complete series of Δf for a given α. In order to determine whether the Δf given by equation (13) are maxima or minima we calculate $\partial^2 j_i(0)/\partial\alpha^2$:

$$\frac{\partial^2 j_i(0)}{\partial\alpha^2} = \frac{K_0}{\pi}\int_0^{2\pi}\Psi_s(\alpha,\phi)\{\sin(K_0\chi(\alpha)) + (C_s\alpha^4 + \Delta f\alpha^2)\cos(K_0\chi(\alpha))\}\,d\phi \tag{14}$$

ie, from the sign of $(C_s\alpha^2 + \Delta f)\cos(K_0\chi(\alpha))$. Since $\cos(K_0\chi(\alpha)) > 0$ for $K_0\chi(\alpha) = \pm 2n\pi$ and $\cos(K_0\chi(\alpha)) < 0$ for $\pm(2n+1)\pi$, we have a complex behaviour for the sign of $\partial^2 j_i(0)/\partial\alpha^2$. For $\eta(\boldsymbol{r}_0) = A\exp(-br_0^2)$, the value of n giving 'maximum' contrast is $n = 11$ for $\alpha = 0{\cdot}02$ rad, $E_0 = 20$ keV ($\Delta f = -163$ nm), $n = 0$ for $\alpha = 0{\cdot}01$ rad, $E_0 = 100$ keV ($\Delta f = -100$ nm) and $n = 8$ for $\alpha = 0{\cdot}015$ rad, $E_0 = 100$ keV ($\Delta f = -93$ nm). In order to determine whether these local minima are saddle points we also require $\partial j_i(0)/\partial(\Delta f) = 0$ (α fixed—see § 4.2). These local minima are independent of b for $b = 20$–2000 nm^{-2}, corresponding to structural features in the 0·4–0·04 nm range.

4.1.2. The amplitude object. A similar analysis to § 4.1.1 leads to the following condition for 'maximum' $j_i(0)$ with an amplitude object:

$$\cos(K_0\chi(\alpha)) = \tfrac{1}{2}(\pm 2n+1)\pi$$

or

$$\Delta f = (\pm 2n+1)\frac{\pi}{\alpha^2 K_0} - \frac{C_s\alpha^2}{2}$$

and the sign of $\partial^2 j_i(0)/\partial\alpha^2$ depends on $-(C_s\alpha^2+\Delta f)\sin(K_0\chi(\alpha))$. The maximum value of $|j_i(0)|$ occurs for: $n = 14$, $\alpha = 0{\cdot}02$ rad, $E_0 = 20$ keV ($\Delta f = -88$ nm); $n = 1$, $\alpha = 0{\cdot}01$ rad, $E_0 = 100$ keV ($\Delta f = -44$ nm); $n = 9$, $\alpha = 0{\cdot}015$ rad, $E_0 = 100$ keV ($\Delta f = -52$ nm). Evidently these conditions for 'maximum' contrast are imcompatible with the corresponding conditions obtained for a phase object.

4.2. Contrast variation with the objective lens defocus

In reality the value for α is fixed by the lens geometry at a series of discrete values and Δf is variable. The condition for 'maximum' $j_i(0)$ (in bright field microscopy) with α constant is thus of more practical importance; for the phase object the equation

$$\frac{\partial j_i(0)}{\partial(\Delta f)} = \frac{K_0}{\pi}\int_0^{2\pi}\int_0^{\alpha}\Psi_s(\theta,\phi)\,\Delta f\cos(K_0\chi(\theta))\theta^2\,\mathrm{d}\theta\,\mathrm{d}\phi = 0 \qquad (15)$$

cannot be solved analytically to give a relationship between Δf and α. Using the models for $\eta(r_0)$ and $\epsilon(r_0)$ given in § 2 (with $A = 0{\cdot}1$, $b = 20\ \mathrm{nm}^{-2}$), results for $C_i(0) = (j_i(0)-1)$ (representing the background subtraction) are presented in figure 2 for $E_0 = 20$ keV, $\alpha = 0{\cdot}02$ rad (figure 2(*a*)) and for $E_0 = 100$ keV, $\alpha = 0{\cdot}015$ rad (figure 2(*b*)); the full curve corresponds to $C_i(0)$ for a phase object and the broken curve $C_i(0)$ for an amplitude object. Examination of figure 2(*a*) shows that the minimum at $\Delta f = -150$ nm is the optimum defocus condition for a phase object. The other maxima and minima of $C_i(0)$ for $\Delta f < 0$ are clearly smaller and oscillate rapidly with Δf; the broad maximum at $\Delta f = 10$ nm should also be considered. If we set the precision of Δf to ± 50 nm, the only peaks that are sufficiently broad for this accuracy of defocus correspond to $\Delta f_{opt} = -150$ nm and $\Delta f_{opt} = 10$ nm. In figure 2(*b*) we estimate for the phase object $\Delta f_{opt} = -100$ nm and $\Delta f_{opt} = 40$ nm. There is only one Δf_{opt} for the amplitude object corresponding to $\Delta f_{opt} = -90$ nm for $\alpha = 0{\cdot}02$ rad (20 keV) and $\Delta f_{opt} = -50$ nm for $\alpha = 0{\cdot}015$ rad (100 keV). In § 5 we will examine in detail the image intensity distributions, $j_i(r_i)$, corresponding to $\Delta f_{opt} \pm 50$ nm; the condition for 'maximum' contrast at $r_i = 0$ corresponds to the best representation of the object structure.

We note that although the values for $C_i(0)$ vary with the gaussian structure parameters b, the Δf_{opt} given above vary only marginally (± 10 nm) with b in the range 20–2000 nm^{-2}. Thus the Δf_{opt} values are almost independent of the structural features in the object, particularly for the 0·4–0·04 nm range.

4.3. Contrast maximum based on the behaviour of the resolution functions

We can show that the condition for 'maximum' contrast in bright field microscopy can be determined directly from the behaviour of the resolution functions $q(r)$ and $q_1(r)$ in the particular case where the object structure corresponds to a set of 'point' atoms.

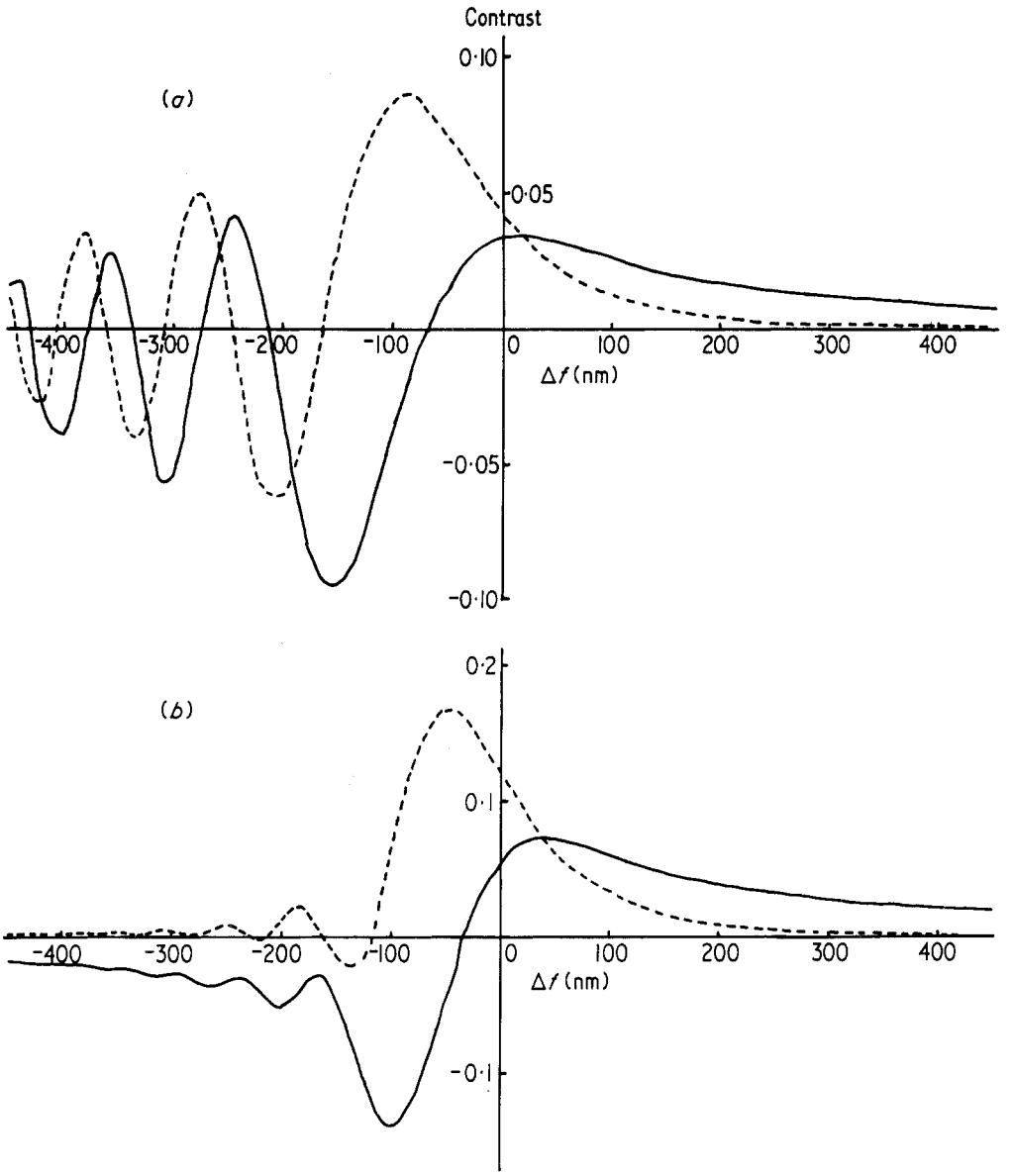


Figure 2. The image contrast $C_i(0)$ as a function of defocus, Δf nm, for an objective aperture: (*a*) $\alpha = 0{\cdot}02$ rad ($E_0 = 20$ keV) and (*b*) $\alpha = 0{\cdot}015$ rad ($E_0 = 100$ keV) for a phase object (full curve) and an amplitude object (broken curve). $C_s = 2$ mm.

Thus for a phase object with atom sites $\boldsymbol{R}_j$

$$\eta(\boldsymbol{r}_0) = \frac{-2\pi me\lambda_0}{h^2} \sum_{\text{all atoms}} V(\boldsymbol{r}_0 - \boldsymbol{R}_j) \tag{16}$$

where $V(\boldsymbol{r})$ describes the projected atomic potential distribution, and for 'point' atoms $V(\boldsymbol{r}) = \delta(\boldsymbol{r})$.

The image intensity is given by

$$j_i(\boldsymbol{r}_i) = 1 + \frac{4\pi m e \lambda_0}{h^2} \int \sum_{\text{all atoms}} \delta(\boldsymbol{r}_0 - \boldsymbol{R}_j) q(\boldsymbol{r}_i - \boldsymbol{r}_0)\, \mathrm{d}\boldsymbol{r}_0 \tag{17}$$

or on reversing the order of summation and integration

$$j_i(\boldsymbol{r}_i) = 1 + \frac{4\pi m e \lambda_0}{h^2} \sum_{\text{all atoms}} q(\boldsymbol{r}_i - \boldsymbol{R}_j) \tag{18}$$

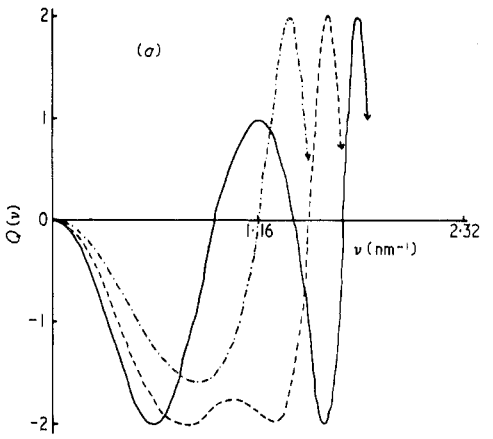
corresponding to the superposition of the resolution functions q each centred at atomic sites $\boldsymbol{R}_j$. For 'maximum' contrast at each $\boldsymbol{R}_j$ we require $q(0)$ = maximum or minimum. Thus we have a condition for 'maximum' contrast in terms of a lens dependent function q and no dependence on the particular structure giving $\eta(\boldsymbol{r}_0)$. We have verified that for several molecular structures (eg benzene and the phthalocyanines) the condition for $q(0)$ = maximum or minimum, leads to 'maximum' contrast and also the best conditions in the image for the resolution of molecular structure.

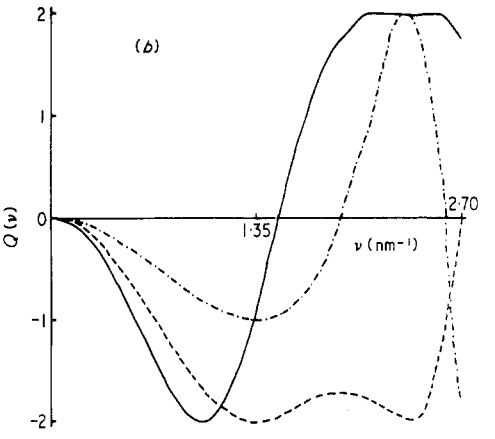
5. Image calculations

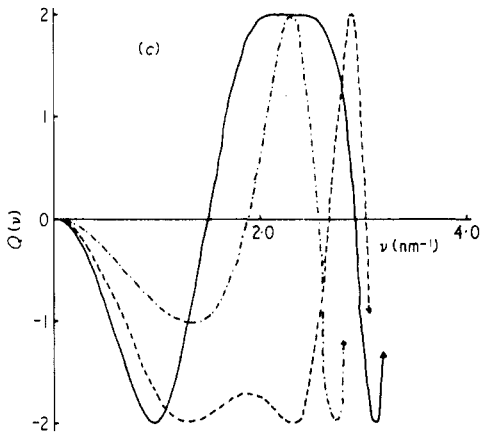
We examine in this section the results of image calculations based on the Δf_{opt} values derived in § 4. We are looking for the closest relationship between the object structure and the image intensity. This is a different aim from the analysis of Thon (1965, 1966, 1971) which described the conditions for resolving a particular structural feature, corresponding to a single spatial frequency, in the object. In figure 3 we show the results of image calculations $j_i(\boldsymbol{r}_i)$ from the convolution of $\eta(\boldsymbol{r}_0)$ with the resolution function $q(\boldsymbol{r})$, using the gaussian model for η. Figures 3(a, b, c) show the phase transfer function $Q(\boldsymbol{v}) = 2\sin(K_0 W(\boldsymbol{v}))B(\boldsymbol{v})$ for $\Delta f_{\text{opt}} \pm 50$ nm where in (a) $\alpha = 0{\cdot}02$ rad, $E_0 = 20$ keV ($\Delta f_{\text{opt}} = -150$ nm), (b) $\alpha = 0{\cdot}01$ rad, $E_0 = 100$ keV ($\Delta f_{\text{opt}} = -100$ nm), (c) $\alpha = 0{\cdot}015$ rad, $E_0 = 100$ keV ($\Delta f_{\text{opt}} = -100$ nm) corresponding to the Δf_{opt} which gave the largest minimum value for $j_i(0)$. The behaviour of the function $Q(\boldsymbol{v})$ is useful in determining which spatial frequencies $\boldsymbol{v}$ are absent from the image and Δf_{opt} corresponds to a $Q(\boldsymbol{v})$ which oscillates least rapidly. However, $Q(\boldsymbol{v})$ does not seem to be as informative as $q(\boldsymbol{r})$, the resolution function (figure 3(d, e, f)). It can be seen that $q(\boldsymbol{r})$ gives detailed information on image resolution in relation to the radial convergence and the radial halfwidth of $q(\boldsymbol{r})$. Clearly the $q(\boldsymbol{r})$ for Δf_{opt} (broken curve) correspond to the achievement of optimum resolution. The critical dependence of $q(\boldsymbol{r})$ on Δf shows that $q(\boldsymbol{r})$ is an ideal function to estimate the best conditions to record a micrograph for any given objective aperture value α radians. The image intensity distributions $j_i(\boldsymbol{r}_i)$ (minus the background value) shown in figures 3(g, h, i) verify that the behaviour of $q(\boldsymbol{r})$ is sufficient to determine the optimum image giving 'maximum' contrast and the closest relationship between the gaussian structure $\eta(\boldsymbol{r}_0)$ and $j_i(\boldsymbol{r}_i)$, although the image contrast $C_i(0)$ is negative. Having recorded the optimum image we can correct the image for lens aberrations, whereas images recorded some way from Δf_{opt} are more difficult to correct, particularly in the presence of noise, because of the slow convergence of $q(\boldsymbol{r})$ at large radial distances.

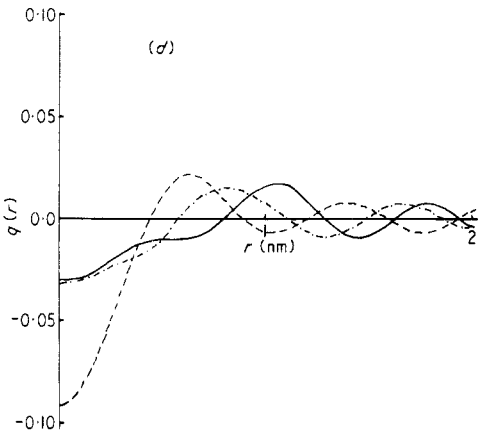
The corresponding results for an amplitude object, calculated for $\Delta f_{\text{opt}} \pm 50$ nm, are presented in figure 4. Figures 4(a, b, c) show the amplitude transfer function

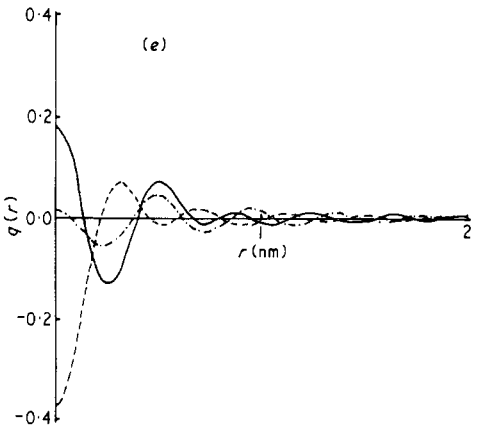
$$Q_1(\boldsymbol{v}) = 2\cos(K_0 W(\boldsymbol{v}))B(\boldsymbol{v})$$

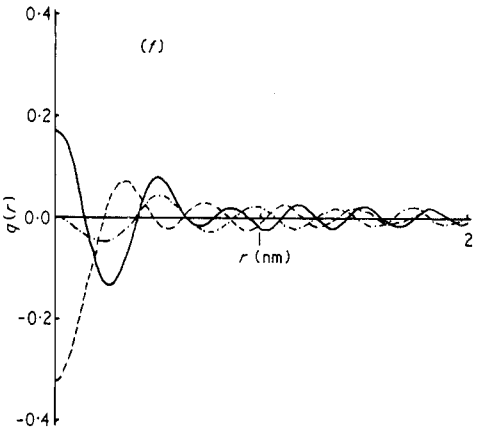

(a)
2
1
0
-1
-2
Q(ν)
1·16
2·32
ν (nm⁻¹)


(b)
2
1
0
-1
-2
Q(ν)
1·35
2·70
ν (nm⁻¹)


(c)
2
1
0
-1
-2
Q(ν)
2·0
4·0
ν (nm⁻¹)


(d)
0·10
0·05
0·0
-0·05
-0·10
q(r)
r (nm)
2


(e)
0·4
0·2
0·0
-0·2
-0·4
q(r)
r (nm)
2


(f)
0·4
0·2
0·0
-0·2
-0·4
q(r)
r (nm)
2

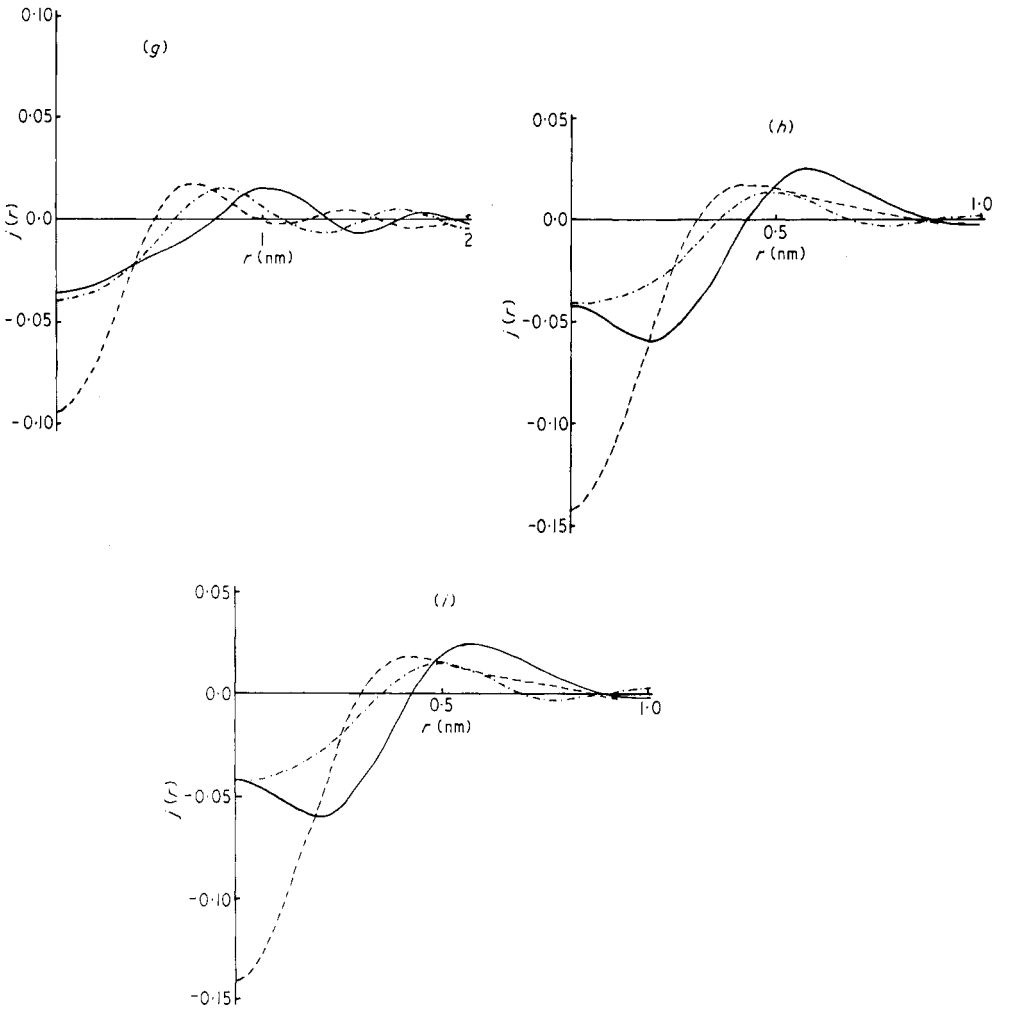
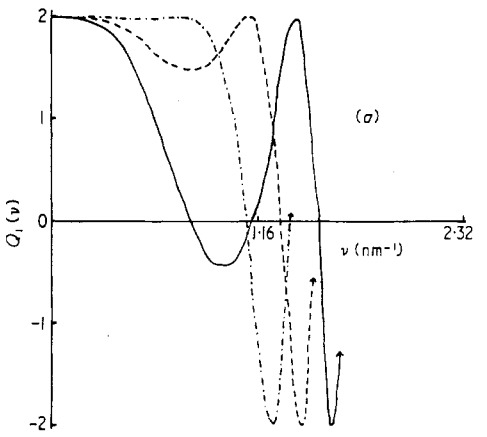
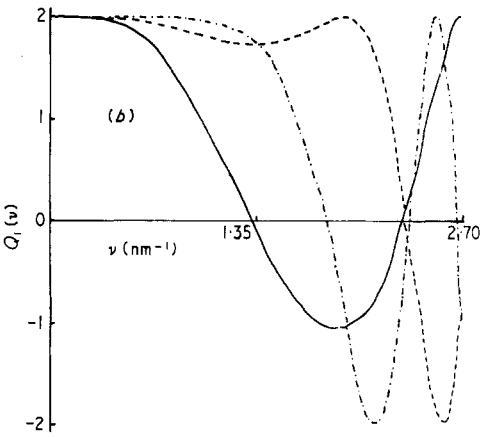


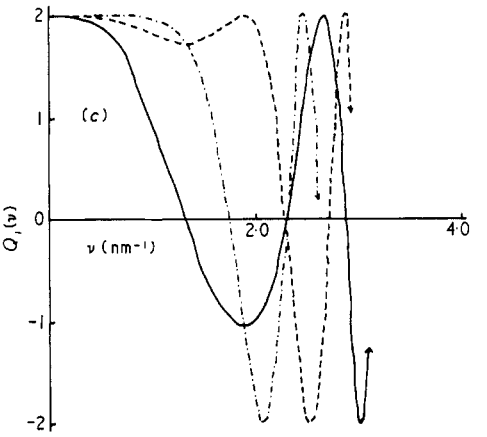
Figure 3. The image intensity distributions for a phase object. (*a*), (*b*), (*c*) show the phase contrast transfer function $Q(v)$ for : (*a*) $\alpha = 0{\cdot}02$ rad, $E_0 = 20$ keV ; (*b*) $\alpha = 0{\cdot}01$ rad, $E_0 = 100$ keV ; (*c*) $\alpha = 0{\cdot}015$ rad, $E_0 = 100$ keV for $(\Delta f_{opt} - 50)$ nm (full curves), Δf_{opt} (broken curves), $(\Delta f_{opt} + 50)$ nm (chain curves). (*d*), (*e*) and (*f*) show the corresponding resolution functions $q(r)$, and (*g*), (*h*) and (*i*) the image intensity distributions $j(r)$. In (*a*) $\Delta f_{opt} = -150$ nm, (*b*) $\Delta f_{opt} = -100$ nm, (*c*) $\Delta f_{opt} = -100$ nm. $C_s = 2$ mm.

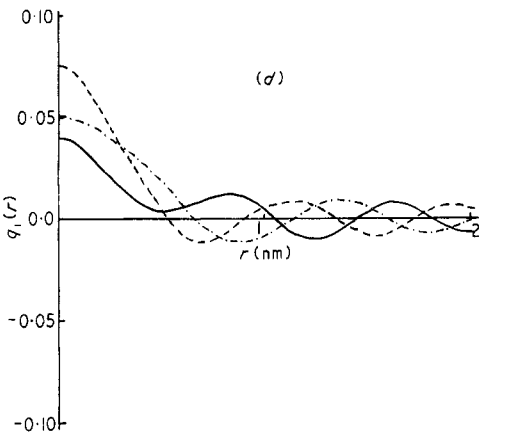
for the same α, E_0 values of figure 3 ; in (*a*) $\Delta f_{opt} = -90$ nm, (*b*) $\Delta f_{opt} = -50$ nm, (*c*) $\Delta f_{opt} = -50$ nm. Again for the defocus values of Δf_{opt} $q_1(0)$ attains its maximum value (figures 4(*d*, *e*, *f*)) and the image contrast is a maximum (figures 4(*g*, *h*, *i*)).

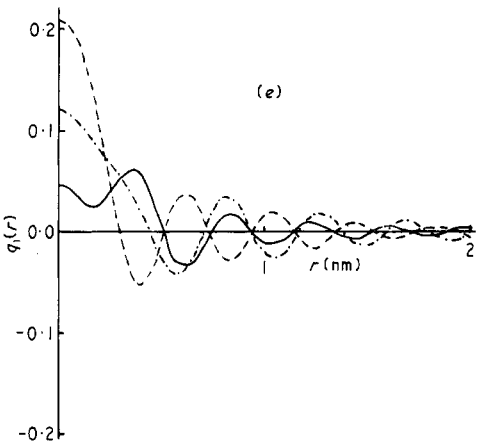
If we consider subsidiary maxima/minima of the $C_i(0)$ curve (see § 4), or equivalently the maxima/minima in $q(0)$ or $q_1(0)$, we do not obtain such good resolution in the image. This is clearly shown in figure 5 where the convergence of the resolution function $q(r)$ is extremely slow (figures 5(*a*, *b*, *c*) for (*a*) $\alpha = 0{\cdot}02$ rad, $E_0 = 20$ keV, $\Delta f_{opt} = 10$ nm, (*b*) $\alpha = 0{\cdot}01$ rad, $E_0 = 100$ keV, $\Delta f_{opt} = 40$ nm, (*c*) $\alpha = 0{\cdot}015$ rad, $E_0 = 100$ keV, $\Delta f_{opt} = 40$ nm) and the corresponding image resolution is noticeably inferior to the

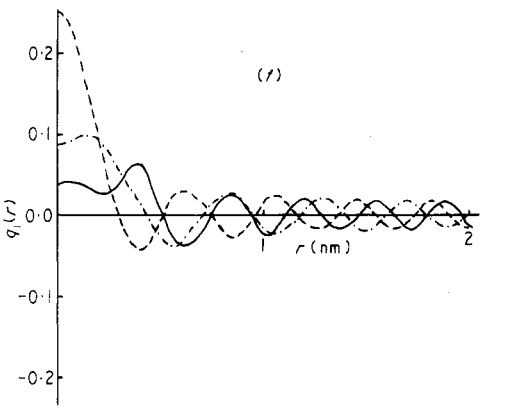

(a)
$Q_1(\nu)$
2
1
0
-1
-2
1·16
2·32
ν (nm^{-1})


(b)
$Q_1(\nu)$
2
1
0
-1
-2
1·35
2·70
ν (nm^{-1})


(c)
$Q_1(\nu)$
2
1
0
-1
-2
2·0
4·0
ν (nm^{-1})


(d)
$q_1(r)$
0·10
0·05
0·0
-0·05
-0·10
1
2
r (nm)


(e)
$q_1(r)$
0·2
0·1
0·0
-0·1
-0·2
1
2
r (nm)


(f)
$q_1(r)$
0·2
0·1
0·0
-0·1
-0·2
1
2
r (nm)

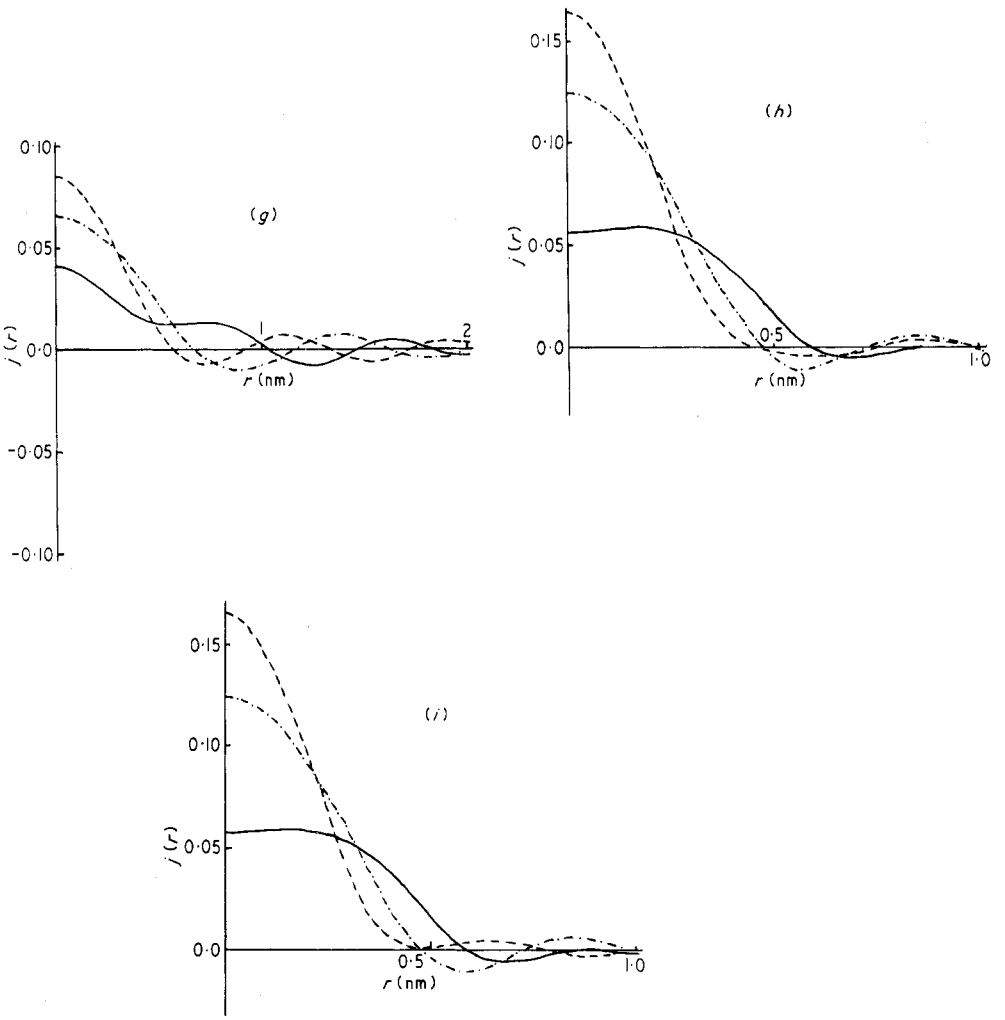


Figure 4. The image intensity distributions for an amplitude object. (*a*), (*b*), (*c*) show the amplitude contrast transfer function $Q_1(v)$ for: (*a*) $\alpha = 0{\cdot}02$ rad, $E_0 = 20$ keV; (*b*) $\alpha = 0{\cdot}01$ rad, $E_0 = 100$ keV; (*c*) $\alpha = 0{\cdot}015$ rad, $E_0 = 100$ keV for $(\Delta f_{opt} - 50)$ nm (full curves), Δf_{opt} (broken curves), $(\Delta f_{opt} + 50)$ nm (chain curves). (*d*), (*e*) and (*f*) show the corresponding resolution functions $q_1(r)$, and (*g*), (*h*) and (*i*) the image intensity distributions $j(r)$. In (*a*) $\Delta f_{opt} = -90$ nm, (*b*) $\Delta f_{opt} = -50$ nm, (*c*) $\Delta f_{opt} = -50$ nm. $C_s = 2$ mm.

results presented in figure 3. In this case with $\Delta f_{opt} > 0$ it is not surprising that the effect of spherical aberration is more serious than that for $\Delta f_{opt} < 0$, where the underfocus term $\frac{1}{2}\Delta f\theta^2$ partially cancels the phase shift term $\frac{1}{4}C_s\theta^4$ due to spherical aberration. However, a large underfocus value, corresponding to a subsidiary minima in the $C_1(0)$ curve for an amplitude object with $E_0 = 20$ keV (figure 2(*a*)), also leads to an inferior image resolution (figure 6). Although $q_1(r)$ (for figure 6(*a*), $\alpha = 0{\cdot}02$ rad, $E_0 = 20$ keV, $\Delta f_{opt} = -200$ nm) has quite a large value at $r = 0$, a large second oscillation in $q_1(r)$ at $r \simeq 0{\cdot}5$ nm is an undesirable feature for the optimum image resolution. Also it is noted that a ± 50 nm variation in Δf_{opt} is now more critical than for the less negative Δf_{opt} values given above.

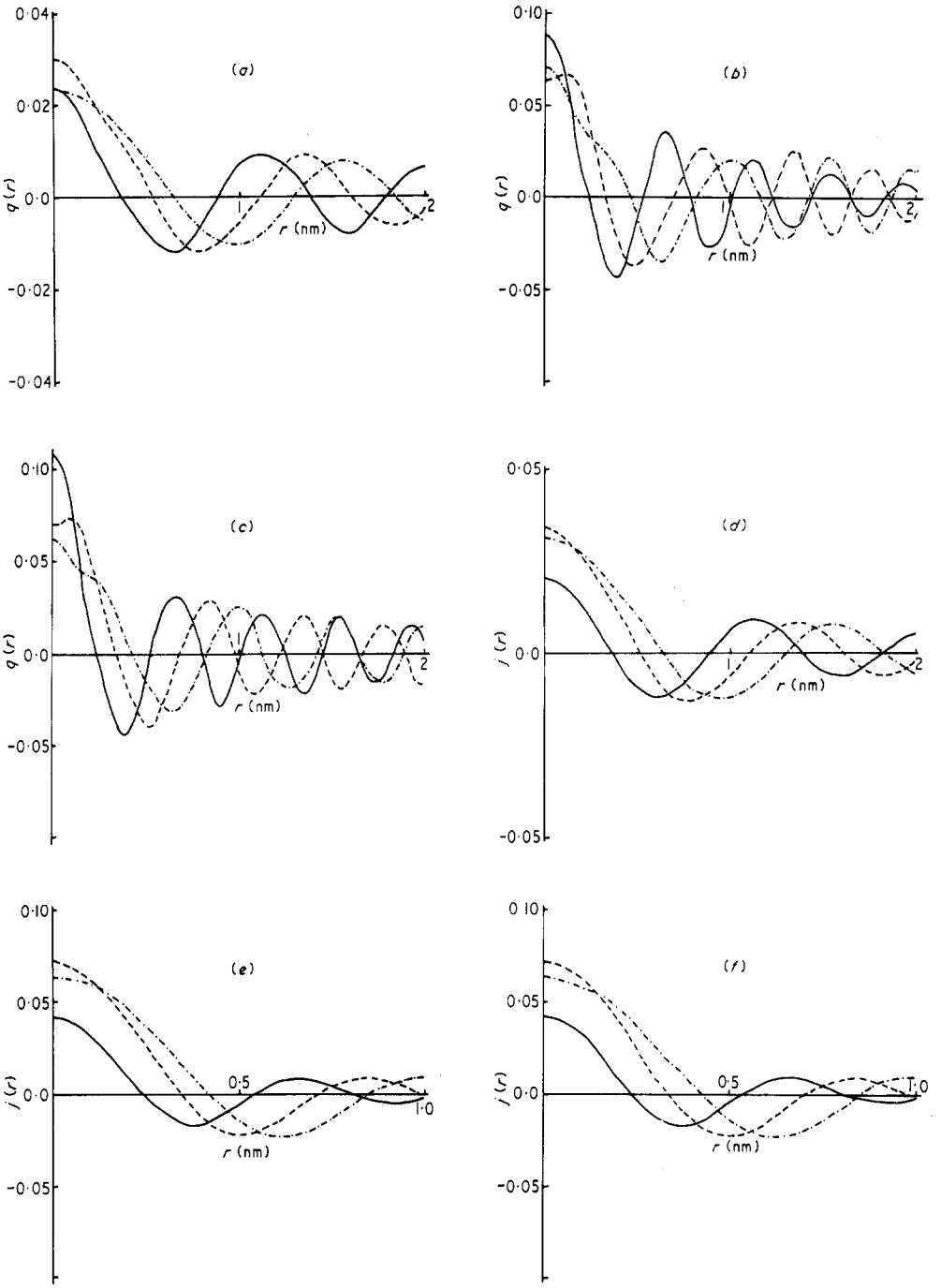


Figure 5. The image intensity distributions for a phase object. (*a*), (*b*), (*c*) show the resolution function $q(r)$ for: (*a*) $\alpha = 0{\cdot}02$ rad, $E_0 = 20$ keV; (*b*) $\alpha = 0{\cdot}01$ rad, $E_0 = 100$ keV; (*c*) $\alpha = 0{\cdot}015$ rad, $E_0 = 100$ keV for $(\Delta f_{\text{opt}} - 50)$ nm (full curves), Δf_{opt} (broken curves), $(\Delta f_{\text{opt}} + 50)$ nm (chain curves). (*d*), (*e*) and (*f*) show the corresponding image intensity distributions $j(r)$. In (*a*) $\Delta f_{\text{opt}} = 10$ nm, (*b*) $\Delta f_{\text{opt}} = 40$ nm, (*c*) $\Delta f_{\text{opt}} = 40$ nm. $C_s = 2$ mm.

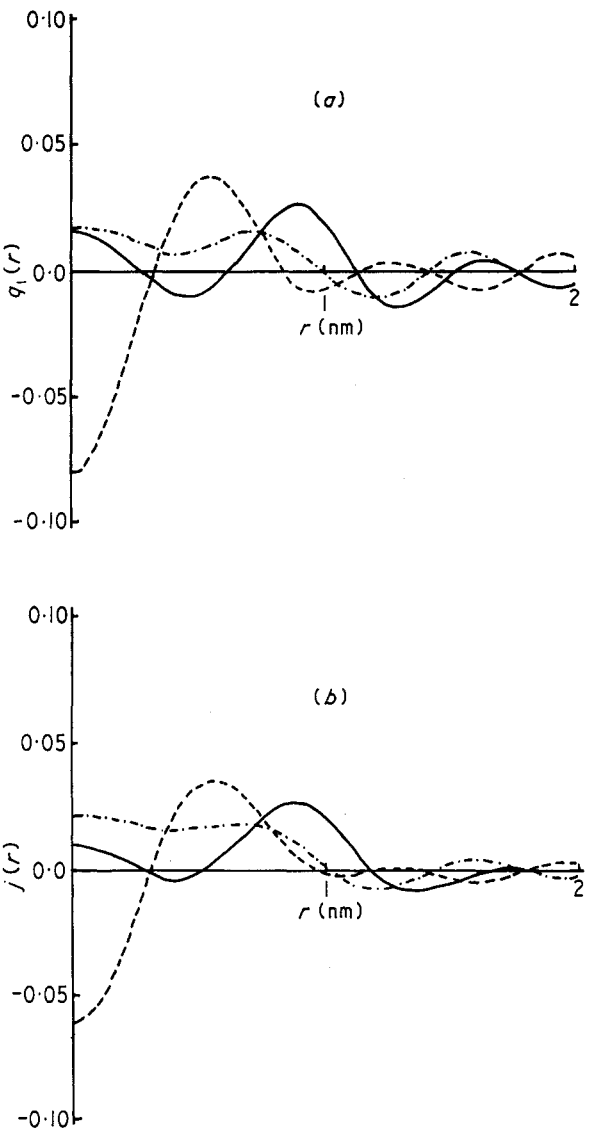


Figure 6. The image intensity distribution for an amplitude object. (*a*) shows the resolution function $q_1(r)$ for $\alpha = 0{\cdot}02$ rad, $E_0 = 20$ keV for $(\Delta f_{opt} - 50)$ nm (full curve), Δf_{opt} (broken curve), $(\Delta f_{opt} + 50)$ nm (chain curve). (*b*) shows the corresponding image intensity distributions $j(r)$. $\Delta f_{opt} = -200$ nm. $C_s = 2$ nm.

We conclude that the optimum defocus Δf_{opt} is, in general, significantly smaller than the value $\Delta f = -\frac{1}{2}C_s\alpha^2$ required to cancel the spherical aberration at α, and that Δf_{opt} for a phase object can be determined from the condition that $q(0)$ attains its largest value for a given α or more critically that $q(r)$ should converge rapidly to zero for large arguments. As the Δf_{opt} for a phase object and amplitude object are incompatible, the optimum defocus for an image will depend on the relative magnitude of η and ϵ, but this is then a structure dependent criterion.

6. Conclusion

The resolution functions for the phase object $q(r)$ and for the amplitude object $q_1(r)$ can be used to determine the optimum defocus value Δf_{opt} for 'maximum' contrast in the image and optimum image resolution. In table 1 we list Δf_{opt} for a series of objective

Table 1. The optimum defocus Δf_{opt} nm for 'maximum' contrast for a series of objective aperture sizes α rad, and for $E_0 = 20$ keV and $E_0 = 100$ keV. $C_s = 2$ mm.

		Δf_{opt} (nm)	
E_0 (keV)	α (rad)	phase object	amplitude object
20	0·01	−140	−70
20	0·015	−150	−70
20	0·02	−150	−90
100	0·005	−130	−20
100	0·01	−100	−60
100	0·015	−100	−60

aperture values α radians for the phase and amplitude objects, as determined from q and q_1 with the normal C_s value of 2 mm; these calculations may be reproduced for any given α and C_s values. The problem of determining $q(r)$ experimentally has been discussed by Thon and Siegel (1970), Thon (1971) and Frank (1972) using the optical diffraction pattern obtained from the micrograph of thin carbon film. Having calculated the optimum defocus Δf_{opt} we can evaluate, from the optical diffraction pattern of the substrate carbon film which micrograph corresponds nearest to the Δf_{opt}. Experimentally we know that a defocus value giving the 'maximum' contrast also corresponds to the best image resolution, and the deconvolution of the micrograph to give $\eta(r_0)$, or a diffraction limited $\eta(r_0)$, requires a determination of Δf from the corresponding optical diffraction pattern.

The calculations of $q(r)$ and $q_1(r)$ presented are relevant to both the conventional transmission and the scanning transmission electron microscopes, although in the latter case the Δf and C_s values correspond to the condenser lens that produces the scan spot (Crewe and Wall 1970, Zeitler and Thomson 1970).

Acknowledgments

The author is grateful for discussion with Professor R E Burge and the provision of excellent computing facilities by the University of London is acknowledged.

References

Crewe A V and Wall J 1970 *Optik* **30** 461–74
Erickson H P and Klug A 1971 *Phil. Trans. R. Soc.* B **261** 105–18
Frank J 1972 *Biophys. J.* **12** 484–511
Grinton G R and Cowley J M 1971 *Optik* **34** 221–33

Hanszen K-J 1971 *Advances in Optical and Electron Microscopy* vol 4, ed R Barer and V E Cosslett (New York: Academic Press) pp 1–84
Hoenders B J 1972 *Optik* **35** 116–33
Hoppe W, Langer R and Thon F 1970 *Optik* **30** 538–45
Hosemann R and Bagchi S N 1962 *Direct Analysis of Diffraction by Matter* (London: North-Holland) chap 4
Lenz F 1965 *Optik* **22** 270–88
—— 1971a *Proc. 25th Anniversary Meeting of EMAG* (London: The Institute of Physics) pp 224–9
—— 1971b *Electron Microscopy in Material Science* ed U Valdré (New York: Academic Press) pp 540–69
Peřina J 1971 *Czech. J. Phys.* B **21** 731–48
Peřina J and Kvalpil J 1968 *Optik* **28** 575–7
Radi G 1970 *Acta Crystallogr.* A **26** 41–56
Thon F 1965 *Z. Naturf.* **20a** 154–5
—— 1966 *Z. Naturf.* **21a** 476–8
—— 1971 *Electron Microscopy in Material Science* ed U Valdré (New York: Academic Press) pp 570–625
Thon F and Siegel B M 1970 *Ber. Bunsen-Ges. phys. Chem.* **74** 1116–20
Zeitler E and Thomson M G R 1970 *Optik* **31** 258–80