DISPERSIVE EQUATIONS FOR HIGH RESOLUTION IMAGING AND LATTICE FRINGE ARTIFACTS

L.D. MARKS *

Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK

The imaging equations for high resolution electron microscopy are analysed using an approach based upon expanding the electron wave as positionally modulated diffraction beams. A high-order semi-linear approximation is derived which includes the leading non-linear effects of a reciprocal space integration and can be conveniently expressed in real space. This allows a direct visualisation of imaging effects in real space using a conventional wave dispersion approach. Based on this, the primary sources of lattice fringe artifacts and errors in the $2\frac{1}{2}$D technique are briefly discussed, and an experimental technique for guarding against them employing dark field imaging is described.

1. Introduction

Perhaps one of the most confusing areas in high resolution electron microscopy is understanding how the various aberrations of the instrument affect the image. Theoretical understanding is well advanced based upon a reciprocal space integration [1–3], but the significance of this in the image plane is an awkward area; it is unreasonable to perform mental Fourier transforms. Contrast transfer theory and qualitative descriptions are possible in reciprocal space (e.g., ref. [4]), but the only existing theories for real space are the weak phase object and other linear approximations which neglect highly significant non-linear cross terms [5]. In this paper we deal specifically with an approach for imaging in real space that provides visual models for the process. By adopting a form for the electron wave after the specimen which includes crystalline diffraction, a highly accurate approximation is derived which has a simple interpretation in terms of wave dispersion. The approximation also includes the dominant non-linear effects and is consequently considerably more accurate than any of the linear models for almost all specimens of experimental interest. Employing this

model the basic source of lattice fringe artifacts and errors in the $2\frac{1}{2}$D technique (e.g., ref. [6]) are briefly discussed. It is pointed out that dark field images from the area of interest can be employed to assess the likelihood of artifacts in high resolution images.

2. Theory

We make our starting point the standard expression for the image intensity using a reciprocal space integration. If $\psi(r)$ is the electron wave leaving the specimen with a Fourier transform $\Psi(u)$, the image $I(r)$ can be written

$$I(r) = F^{-1} \int \Psi(u') \, \Psi^*(u - u') \exp(-i\chi(u'))$$
$$+ i\chi(u - u')) \, T(u', u - u') \, d^2u', \qquad (1)$$

where F^{-1} stands for an inverse Fourier transform, with a phase shift $\chi(u)$ due to lens defocus and spherical aberration (in reduced units)

$$\chi(u) = \pi u^2\left(\tfrac{1}{2}u^2 - D\right), \qquad (2)$$

and an envelope term $T(u', u-u')$

$$T(u', u - u') = \exp\left\{ - \frac{\pi^2 d_0}{2}\left[(u')^2 - (u - u)^2\right]^2 \right.$$
$$\left. - s_0^2\left[\nabla \chi(u') - \nabla \chi(u - u')\right]^2 \right\}, \qquad (3)$$

* Present address: Department of Physics, Arizona State University, Tempe, Arizona 85287, USA.

where s_0 and d_0 are the rms widths of the energy distribution and convergence respectively. (This equation neglects variations in the diffraction over the illumination aperture, which can be a serious error for thicker specimens. Employing the standard two-beam dynamical equations, we can estimate that eq. (1) will be valid for specimen thicknesses less than about the extinction distance, i.e. ~ 20 nm). Qualitative analyses of eq. (1) in the literature almost invariably concentrate on the integrand with $|\Psi(u)|$ a constant for all values of $u \neq 0$ (a non-linear contrast transfer function), or employ the weak phase object approximation to derive a point response function. However, these approaches can be misleading. It is only for a very small class of objects, primarily thin amorphous films, that $\Psi(u)$ extends over a large range of u values with roughly equal moduli; the vast majority of experimental specimens are crystalline, with the scattered intensity concentrated around diffraction spots. Hence the standard, qualitative analysis includes spatial frequencies which may not be present in the exit wave. To avoid this we choose a model for the exit wave which represents more closely the physical situation. We expand $\psi(r)$ as a Fourier series.

$$\psi(r) = \sum_g \phi_g(r) \exp(-2\pi i g \cdot r), \tag{4}$$

where the g vectors correspond to crystalline reciprocal lattice vectors. Provided that the coefficients $\phi_g(r)$ are not rapidly varying, they are the complex amplitudes calculated using, for instance, dynamical techniques with or without the column approximation (e.g., ref. [7]). Writing the Fourier transform of $\phi_g(r)$ as $\Phi_g(u)$, eq. (1) can be rewritten as

$$I(r) = F^{-1} \sum_{g,q} \delta(u - g + q) \otimes \int \Phi_g(u') \Phi_q^*(u - u')$$
$$\times \exp[-i\chi(g + u') + i\chi(q + u - u')]$$
$$\times T(g + u', q + u - u') \, d^2 u', \tag{5}$$

where $\otimes$ represents a convolution.

We now consider the magnitude of the terms in the integrand of eq. (5), in particular the magnitude of the scattering terms $\Phi_g(u)$ and $\Phi_q(u)$.

Provided that the real space terms $\phi_g(r)$ and $\phi_q(r)$ are not rapidly varying, they consist of large central peaks for $u = 0$ surrounded by weaker diffuse term for $u \neq 0$. The cross terms between the g and q beams can therefore be quantitatively ordered

$$|\Phi_g(0) \, \Phi_q^*(0)| > |\Phi_g(0) \, \Phi_q^*(u')|$$
$$= |\Phi_g(u') \, \Phi_q^*(0)| > |\Phi_g(u') \, \Phi_q^*(u - u')|, \tag{6}$$

with

$$u' \neq 0, \quad u \neq u'.$$

This suggests that an accurate approximation is one which takes into account the relative order of these terms. We therefore expand the non-linear envelope term in a quasi-linear fashion so that it is *exact* for the three largest terms in eq. (6), i.e.

$$T(g + u', q + u - u')$$
$$= T'(g + u', q) \, T'(q + u - u', g), \tag{7}$$

where

$$T'(g + u, q)$$
$$= \exp\left\{ -\frac{\pi^2 d_0^2}{2} (g + u)^2 \left[(g + u)^2 - q^2 \right] \right.$$
$$- \pi^2 s_0^2 \nabla \chi(g + u)$$
$$\left. \cdot \left[\nabla \chi(g + u) - \nabla \chi(q) \right] \right\}. \tag{8}$$

(This approach is philosophically similar to a weak object approximation, extended to include strong diffraction into certain beams.) The approximation can be refined one further stage by taking some account of the terms in the envelope function that have been neglected by including the first-order residual contributions in u' and $u - u'$. This leads to an effective envelope $T^{\text{eff}}(g + u, q)$ of form

$$T^{\text{eff}}(g + u, q)$$
$$= T'(g + u, q) \exp\left\{ -2\pi^2 d_0^2 g \cdot u q^2 \right.$$
$$\left. - \pi^3 s_0^2 \nabla \chi(q) \cdot \left[u g^2 + 2(g \cdot u) g \right] \right\}. \tag{9}$$

In the process of eqs. (6)–(9), we have separated the non-linear envelope term into effective envelope terms local to the g and q beams. The

method has been to take initially an analytical approximation justified by the magnitude of the terms in the integrand of eq. (5), and then add a first-order correction. Provided that either the envelope term is only slowly varying, or the diffuse–diffuse terms are small, the errors involved are small.

Defining now an amplitude transfer function $\tilde{A}_{g.q}(u)$ as

$$\tilde{A}_{g.q}(u) = T^{\text{eff}}(g + u, q) \exp[-i\chi(g + u)], \quad (10)$$

eq. (5) can be written in a semi-linear form

$$I(r) = F^{-1} \sum_{g.q} \delta(u - g + q) \otimes \left(\Phi_g(u) \, \tilde{A}_{g.q}(u) \right)$$

$$\otimes \left(\Phi_q(u) \, \tilde{A}_{q.g}(u) \right)^*, \quad (11)$$

which can also be written in real space as

$$I(r) = \sum_{g.q} \exp[-2\pi i(g - q) \cdot r] \left[\phi_g(r) \otimes A_{g.q}(r) \right]$$

$$\times \left[\left(\phi_q(r) \otimes A_{g.q}(r) \right) \right]^*, \quad (12)$$

where the inverse Fourier transform of $\tilde{A}_{g.q}(u)$, i.e. $A_{g.q}(r)$, will be referred to here as the amplitude response function. The final result in eqs. (11) and (12) includes the most significant non-linear imaging effects, is analytically accurate for the leading terms with a first-order correction for the higher terms, and has a very convenient and simple form in both real and reciprocal space.

The form of the semi-linear equation is analytically simple and can be tested numerically for any particular case without difficulty. Inaccuracies will arise if $\phi_g(r)$ is rapidly varying, but the approximation will always be more accurate than the standard linear approximations (e.g. weak object or weak phase object) because far fewer assumptions are made. (Linear approximations are inaccurate for all $\Phi_g(u')\Phi_g^*(u - u')$, u', $u \neq 0$, whilst the semi-linear equations are inaccurate only for the weakest diffuse–diffuse type terms.) As an illustration, a simple test object of the form

$$\psi(r) = \begin{cases} 1 + \exp(-2\pi i g \cdot r), & r \leqslant R \\ 0, & r > R \end{cases} \quad (13)$$

is imaged in fig. 1 where $R = 1.25$ nm, the lattice spacing is 0.235 nm (gold (111)), for Gaussian
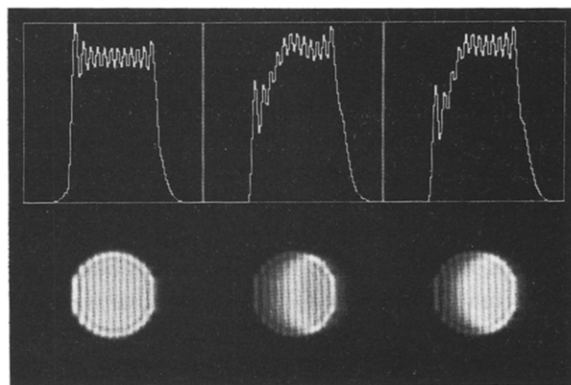

Fig. 1. A comparison of (from left to right) the weak object approximation, a reciprocal space integration and the semi-linear approximation. Below are the simulated images for the conditions described in the text, above which are line traces across the centre of the images. Only in the line traces can differences between the semi-linear approximation and the reciprocal space integration be detected, whilst the weak object approximation is inadequate, failing to reproduce the image asymmetry.

defocus and typical conditions of the Cambridge High Resolution High Voltage Electron Microscope, i.e. electrons of 500 kV, $C_s = 2.7$ mm, chromatic half-width of 16 nm and a convergence semi-angle of 0.5 mrad. The weak object approximation is inadequate, giving a much too favourable prediction, whilst the non-linear result is exceptionally close to that of the reciprocal space in integration.

3. Visualising the imaging process

With the simple semi-linear form in real space, it is possible to visualise imaging in relatively simple terms and not to rely upon a contrast transfer in reciprocal space whose effects in real space are not obvious. The main feature is that it is possible to analyse how the position of information is transferred; for a perfect image $A_{g.q}(r)$ would reduce to delta function, whereas in general the convolution will produce some form of weighted average in real space. It is important to point out that we are dealing with dispersive equations which are well known for wave systems. Considering the amplitude transfer function in equation (10), there is a non-linear phase shift

$\chi(g + u)$ and a damping term $T^{\text{eff}}(g + u)$. In real space this leads to damped, propagating waves with the image at any defocus reflecting a cross-section of the wave dispersion. A very simple, visual analogy for the variations with defocus is useful. Dropping a stone vertically into a pool of water produces surface waves which spread out from the point of impact, decaying in amplitude as they do so. This is the same qualitative behaviour as the central beam in $A_{0,0}(r)$ where time is equivalent to defocus. Alternatively, if the stone falls not vertically but at an inclined angle, this is roughly the form of $A_{g,q}(r)$.

Considering these dispersive equations in more detail, the modulation term $\phi_g(r)$ can be treated as a wave packet directed along the g diffraction vector. Carrying out the standard dispersive expansion for the phase term gives, according to eq. (2),

$$\chi(g + u) = \chi(g) + 2\pi g \cdot u D_g + \pi u^2 D_g + 2\pi(g \cdot u)^2$$
$$+ 2\pi(g \cdot u)u^2 + \tfrac{1}{2}\pi u^4, \tag{14}$$

where

$$D_g = D - g^2. \tag{15}$$

In addition to a phase shift $\chi(g)$ for the wave packet, there is a terminating series of higher-order dispersive terms. The leading of these is $2\pi g \cdot u D_g$, which is the "Group shift", i.e. the integrated value of the Group velocity. This produces in real space a nett shift of the information of $g D_g$, which is the well known Abbe or ray-diagram translation for a diffracted beam. Going to higher order there is a defocus term involving D_g rather than D, a third-order astigmatism of $2\pi(g \cdot u)^2$, a term resembling Siedel Coma and finally a spherical aberration. We note that both the translation and the defocus vanish for $D_g = 0$, which is the standard "overlap defocus" where convergence effects are minimised. As a general rule this defocus is stronger than simply what is good for minimising convergence effects – it minimises positional errors.

The general trend with defocus will therefore be a nett translation coupled with directional distortions of the wave packet $\phi_g(r)$. However, a little caution should be exercised here since the precise variations will depend upon the extent of $\Phi_g(u)$

(i.e. the rate of variation of $\phi_g(r)$). If $\Phi_g(u)$ is strongly localised around $u = 0$ (slow variations in the image) it is reasonable to neglect the higher-order terms in eq. (14), i.e. the comma-like and spherical aberration contributions. Here the image effects are a translation and an astigmatic defocus. If $\Phi_g(u)$ is not strongly localised, it is not justifiable to neglect the higher-order terms; eq. (14) is a terminating series and not a Taylor series with decaying terms. (In the limit as $\phi_g(r)$ becomes a delta function there is no nett translation.) Although the general behaviour is as described above, specific effects will depend upon the rate of variation of the wavepacket.

4. Discussion and lattice fringe artifacts

We have detailed here a form of the imaging equations in electron microscopy which allows direct access to effects in real space rather than a contrast transfer in reciprocal space whose consequences are not obvious. The main result is that in an electron microscope we lose an accurate picture of where in the object information is present, i.e. positional and not just contrast effects. It is useful to detail briefly here the relationship to artifacts in the spacing of lattice fringes.

We can write the complex modulations in an amplitude and phase form, i.e.

$$\phi_g(r) = R_g(r) \exp\left[-2\pi i \theta_g(r)\right]. \tag{16}$$

With a frequency modulation term $\theta_g(r)$, the local spacing (lattice parameter) is then

$$g + \nabla \theta_g(r). \tag{17}$$

If there are no phase effects associated with diffraction through the specimen, eq. (17) represents the true lattice spacing of the specimen, but this is unlikely. Inhomogeneous strains, thickness or mass variations couple to phase shifts by virtue of dynamical diffraction [7]. As a simple example, with a wedge thickness the beam split [8,9] and a shift in the fringe spacing occurs [10]. The phase shift in a two-beam approximation can be expressed as

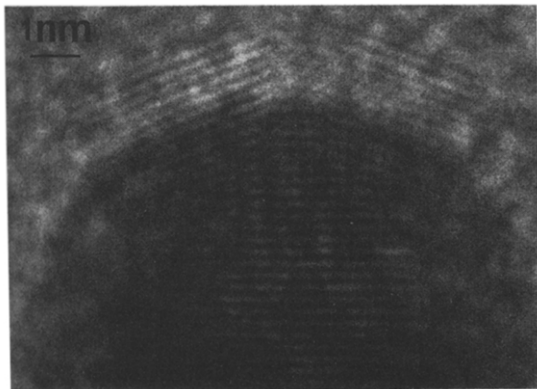$$\theta_q(r) = \alpha m \cdot r, \tag{18}$$

Fig. 2. High resolution image of a small silver particle at a large defocus showing lattice fringes well away from the particle.

where m is the thickness gradient and α is a parameter whose value depends upon the diffraction conditions [11,12]. The apparent lattice spacing is $g + \alpha m$, that is, wrong. We note in passing that the $2\frac{1}{2}$D technique (e.g., ref. [6]), which measures $\nabla \theta_g(r)$, will be susceptible to systematic errors from dynamical diffraction effects, and this has been recently confirmed experimentally [13].

Expanding the final image equation (12) in an amplitude and phase form, we may write

$$\left[\phi_g(r) \otimes A_{g,q}(r) \right] \left[\phi_q(r) \otimes A_{q,g}(r) \right]^*$$
$$= R_{g,q} \exp\left[-2\pi i \theta_{g,q}(r) \right]. \qquad (19)$$

It is apparent that artifacts arise from both the g and q beam modulations, so in standard high resolution bright field we must also worry about artifacts from the central beam, i.e. $q = 0$. One or two features of eq. (19) merit particular mention here, although as a general rule the effects are likely to be rather complicated. Firstly there is the simple group shift which leads to large positional errors as illustrated by the metal particle in fig. 2 where fringe structure is evident well away from the particle. Secondly, lattice variations can arise even if $\phi_g(r)$ and $\phi_q(r)$ are completely real, that is amplitude contrast but no real lattice changes or phase variations from dynamical diffraction. Finally, if the amplitude is constant (i.e. the modulus) but the phase of $\phi_g(r)$ is varying, imaging does not affect the linear lattice or phase variation

but does affect the curvature of the lattice (second derivative), excepting the positional shift. (The convolutions in eq. (12) can be considered as weighted averages.) We note that unless there is a fixed point of reference, for example a grain boundary, the positional error can be neglected so the linear effects are accurately imaged. However, there is no reason to assume *a priori* that these are not dynamical artifacts of the form described above.

Given that lattice fringe artifacts can so readily occur, it is important to point out that phase and amplitude variations in the scattering are implicitly coupled in dynamical diffraction, and that $A_{g,q}(r)$ is similar to $A_{g,0}(r)$. This suggests that dark field images can be used as a safety check; with a large objective aperture of diameter equal to the spot separation, the off-axis dark field image intensity is

$$I(r) = |\phi_g(r) \otimes A_{g,g}(r)|^2, \qquad (20)$$

whilst for tilted beam dark field with g down the optic axis

$$I(r) = |\phi_g(r) \otimes A_{0,0}(r)|^2 \approx |\phi_g(r)|^2. \qquad (21)$$

(We assume here that $\Phi_g(u)$ in reciprocal space decays sufficiently rapidly that the aperture cut-off can be neglected.) Hence the presence of artifacts which require detailed image simulations before they can be interpreted can be directly, experimentally checked by looking at the dark field images with tilted illumination to check on dynamical effects, and off-axis dark field for imaging effects. As a further refinement, it would be useful to record images at more than one specimen orientation, for example after a tilt of about 1°. Particularly for the $2\frac{1}{2}$D technique, this should guard against systematic errors by changing the diffraction conditions and consequently changing the magnitude of the shift in eqs. (17) and (18).

Acknowledgments

linear equations, and would also like to thank Dr. A. Howie for his valuable comments and the SERC, UK, for financial support.

References

[1] M.A. O'Keefe, in: Proc. 37th Annual EMSA Meeting, San Antonio, TX, 1979 (Claitor's, Baton Rouge, LA, 1979) p. 556.

[2] W.O. Saxton, J. Microsc. Spectrosc. Electron. 5 (1980) 55.

[3] K. Ishizuka, Ultramicroscopy 5 (1980) 55.

[4] J.C.H. Spence, Experimental High Resolution Electron Microscopy (Oxford University Press, 1980).

[5] M.A. O'Keefe and W.O. Saxton, in: Proc. 41st Annual EMSA Meeting, Phoenix, AZ, 1983 (Claitor's, Baton Rouge, LA, 1983) p. 288.

[6] E. Lindahl and R. Österlund, Ultramicroscopy 9 (1982) 355.

[7] P.B. Hirsch, A. Howie, R.B. Nicholson, D.W. Pashley and M.J. Whelan, Electron Microscopy of Thin Crystals (Butterworths, London, 1965).

[8] K. Molière and H. Wagenfeld, Z. Krist. 110 (1958) 3.

[9] G. Lehmpfuhl and A. Reissland, Z. Naturforsch. 23a (1968) 544.

[10] H. Hashimoto, M. Mannami and T. Naiki, Phil. Trans. Roy. Soc. London 253 (1961) 459.

[11] L.D. Marks, PhD Thesis (1980).

[12] A. Gomez, P. Schabes-Retchkiman, M. Jose-Yacaman and T. Ocana, Phil. Mag. 47 (1983) 169.

[13] A.J. Porter and R.A. Ricks, Electron Microscopy and Analysis 1983, in press.