

Computation in electron microscopy

Earl J. Kirkland*

School of Applied and Engineering Physics, Cornell University, Ithaca, NY 14853, USA. *Correspondence e-mail: ejk14@cornell.edu

Received 9 June 2015

Accepted 19 September 2015

Edited by J. Miao, University of California, Los Angeles, USA

Keywords: high-resolution transmission electron microscopy; HRTEM; multislice method; exit-wave reconstruction; deconvolution.

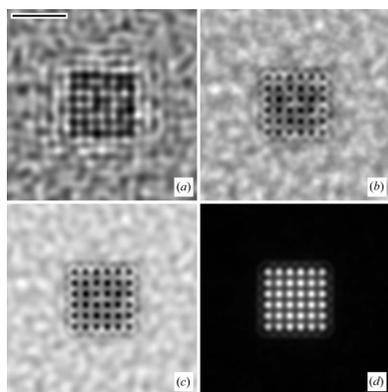
Some uses of the computer and computation in high-resolution transmission electron microscopy are reviewed. The theory of image calculation using Bloch wave and multislice methods with and without aberration correction is reviewed and some applications are discussed. The inverse problem of reconstructing the specimen structure from an experimentally measured electron microscope image is discussed. Some future directions of software development are given.

1. Introduction

Computers and computation have found many uses in high-resolution transmission electron microscopy (TEM; in this paper the abbreviation EM is used to mean both electron microscopy and electron microscope) over the last few decades (Hawkes, 1981). The theory of image formation at high resolution is sufficiently complex that it may require significant numerical calculation to implement. The electron is fundamentally quantum mechanical in nature. Manipulating the electron trajectories is adequately described by classical electric and magnetic fields and can be understood using wave or ray optics much like light optics in a conventional optical microscope. The interaction between the imaging electrons and the material in the specimen can be very quantum mechanical in nature. Simple analysis amenable to solution with pencil and paper provides some useful intuitive understanding but in many cases cannot go far enough for a detailed understanding of the images and other data recorded in TEM. Given an accurate understanding of how the TEM image is generated it is also possible to remove some of the artifacts introduced by the instrument and possibly improve the image to better understand the material being observed (the inverse problem). The goal of computation is to better understand the information produced in TEM, to better understand and control the materials being observed, and possibly improve the operation of the instrument. For a more complete, formal discussion including detailed theory and a longer list of references, see, for example, Kirkland (2010).

The operation of the microscope frequently utilizes computers for real-time control and data acquisition. Various modes of operating the microscope can generate vast amounts of data that require a large computer and data storage. Early user efforts at real-time computer control using Fortran, and later C/C++, are discussed and reviewed in Kirkland (1990). There have been recent efforts using open-source Python (Murfit *et al.*, 2013; Meyer *et al.*, 2014). Most of these uses are now more properly managed by the manufacturers of the instrument and will not be discussed in detail here.

The design and manufacture of modern instruments also require extensive computer-aided design for both the mechanical and electronic structure and the electron optical



design. The basic theory of a modern aberration corrector was understood half a century ago but was not practical until recently, largely due to improvements in computer technology. Multipole correctors are so complicated that by the time a human operator might get it sufficiently aligned to operate, other portions of the instrument have drifted out of alignment, making it unusable. A modestly fast computer (by today's standards) with sophisticated (computation) software can align the corrector fast enough for it to be usable, making the computer an indispensable component of current aberration-corrected TEM. The resolution and contrast in the image are fundamentally determined by the accuracy of tuning of the aberrations (Kirkland, 2011).

The fields of electron tomography (for example, Midgley & Weyland, 2003) and cryogenic electron microscopy of biological specimens (for example, Frank, 2006; Kourkoutis *et al.*, 2012) can also have a substantial computational component in averaging multiple low-dose images or three-dimensional reconstruction using tilt tomography or single-particle analysis. Off-axis electron holography (Lichte, 1986) may also involve a significant amount of computation, but these important topics will be left to an author more knowledgeable in these areas.

There are many general books on TEM (for example, Edington, 1976; Reimer, 1993; DeGraf, 2003; Williams & Carter, 2009; Erni, 2010; Peng *et al.*, 2011; Fultz & Howe, 2013; Spence, 2013), STEM (scanning transmission electron microscopy) (Keyse *et al.*, 1998; Pennycook & Nellist, 2011; Tanaka, 2015) and low-voltage electron microscopy (Bell & Erdman, 2013). The related scanning electron microscope (SEM) utilizing secondary electrons is equally worthwhile, but will not be discussed here; refer to Goldstein *et al.* (2003), Joy (1995) or Reimer (1998) for further discussion of the SEM.

2. Model of the instrument

There are two basic types of transmission electron microscope. The conventional fixed-beam microscope (CTEM) forms the whole image at one time and the scanning microscope (STEM) forms a small probe that is raster scanned across the specimen to form the image one element at a time. Some people name the CTEM an HRTEM (high-resolution transmission electron microscope) and the STEM an HAADF (high-angle annular dark field). However both HRTEM/CTEM and the HAADF/STEM are in fact high resolution (which makes HRTEM ambiguous) so the names CTEM and STEM will be used here.

The current generation of high-resolution TEMs have become very complex (and very expensive). It is helpful to have a simplified model of the instrument to think about the theory of image formation. Only the most important portions of the instrument will be included for simplicity.

A simple model of the CTEM is shown in Fig. 1. A uniform (or nearly uniform) beam of electrons is incident on the specimen, passes through the specimen and is imaged by the objective lens onto the detector plane. Each point on the specimen is imaged onto a different position on the detector

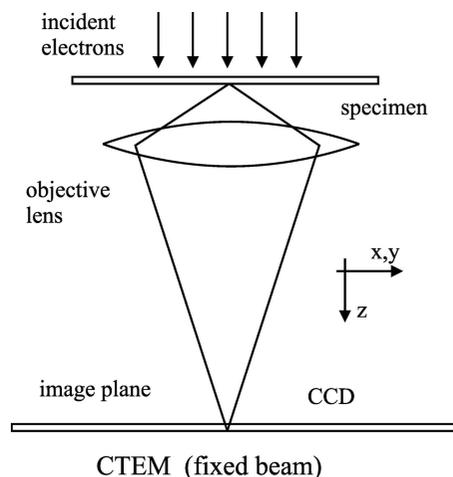


Figure 1
Simplified schematic of the conventional transmission electron microscope (CTEM). Incident electrons are traveling from top to bottom. (Drawing not to scale.)

(many points at the same time, one point shown). In practice, several condenser lenses form the incident parallel beam and several projector lenses further magnify the image produced by the objective. Because possible defects in the objective are magnified by the projector lenses the objective lens has the largest impact on the final image. The detector is typically a scintillator plus CCD with analog integration of the signal. Some new direct electron detectors are very sensitive and may be capable of electron counting of low-current signals in CTEM (McMullan *et al.*, 2009, 2014).

Similarly, a simplified model of the STEM is shown in Fig. 2. Here the incident electrons are focused into a small probe that is incident on the specimen. The electrons transmitted through the specimen form the image intensity at each position of the probe and the probe is scanned across the specimen. The electrons scattered at high angle are collected by the annular dark-field (ADF) detector and the unscattered electrons continue straight onto the bright-field (BF) detector. The BF

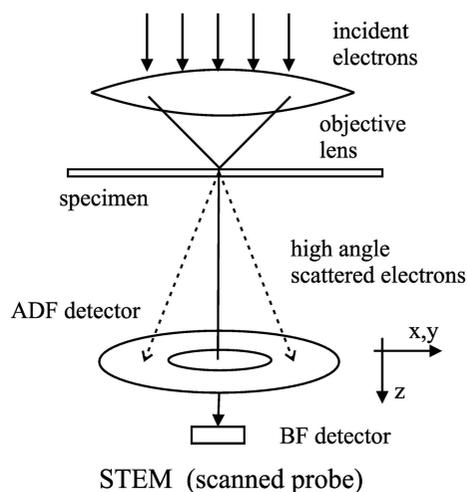


Figure 2
Simplified schematic of the scanning transmission electron microscope (STEM). Incident electrons are traveling from top to bottom. (Drawing not to scale.)

Table 1

Optimum values for aberration compensation when C_{S3} is fixed and is the main factor limiting resolution, for phase-contrast BF-CTEM (Scherzer, 1949) and ADF-STEM (Kirkland, 2010).

α_{\max} is the maximum objective aperture, Δf is defocus and d_{\min} is the approximate resolution.

Parameter	CTEM	STEM
Δf	$(1.5C_{S3}\lambda)^{1/2}$	$0.87(C_{S3}\lambda)^{1/2}$
α_{\max}	$(6\lambda/C_{S3})^{1/4}$	$1.34(\lambda/C_{S3})^{1/4}$
d_{\min}	$0.67(C_{S3}\lambda^3)^{1/4}$	$0.43(C_{S3}\lambda^3)^{1/4}$

signal in STEM is formally equivalent to BF in CTEM *via* the reciprocity theorem (Pogany & Turner, 1968; Cowley, 1969) which arises from symmetry arguments. If the BF detector is expanded to have the same outer angle as the objective aperture but with a hole in the middle (typically half the maximum angle) then this mode is referred to as annular bright field or ABF (Findlay *et al.*, 2009, 2010; Ishikawa *et al.*, 2011). The STEM detector is typically a scintillator plus photomultiplier tube (PMT) and can be integrated in analog for high signals and directly count electrons in low signal rates (Kirkland & Thomas, 1996; Findlay & LeBeau, 2013; Ishikawa *et al.*, 2014).

3. Aberrations and compensation

Electron lenses have much larger aberrations than their light optical counterparts, limiting them to small objective aperture angles. The resolution of a traditional electron lens is mainly limited by third-order spherical aberration C_{S3} . The phase error in the electron wavefunction (relative to a perfect spherical wave) or aberration function *versus* polar angle α in the objective aperture is

$$\chi(\alpha) = \frac{2\pi}{\lambda} \left(\frac{1}{4} C_{S3} \alpha^4 - \frac{1}{2} \Delta f \alpha^2 \right) \quad (1)$$

where λ is the electron wavelength and Δf is defocus (can be defined with either sign) determined by the strength of the objective lens. $\alpha = \lambda k$ where k is spatial frequency in the image. The aberration function can also be written as

$$\begin{aligned} \chi(k) &= \frac{2\pi}{\lambda} \left(\frac{1}{4} C_{S3} \lambda^4 k^4 - \frac{1}{2} \Delta f \lambda^2 k^2 \right) \\ &= \pi \lambda k^2 (0.5 C_{S3} \lambda^2 k^2 - \Delta f). \end{aligned} \quad (2)$$

For bright-field phase contrast this phase error should be $\pi/2$ for the scattered waves ($k > 0$) and zero for the unscattered wave ($k = 0$) to obtain phase contrast. For a small focused probe this phase error should be close to zero everywhere inside the objective aperture leading to two different focus criteria.

In a traditional objective lens, spherical aberration is fixed by the physical shape and geometry of the magnetic material of the lens. Scherzer (1949) realized that defocus (controlled by the current in the lens coils) can partially offset (or compensate) spherical aberration over a small range of angles

Table 2

Optimum values for aberration compensation when C_{S5} is fixed and C_{S3} is variable, for phase-contrast BF-CTEM (Scherzer, 1970) and ADF-STEM (Intaraprasongk *et al.*, 2008) (with modification).

α_{\max} is the maximum objective aperture, Δf is defocus and d_{\min} is the approximate resolution.

Parameter	CTEM	STEM
C_{S3}	$-3.2(\lambda C_{S5}^2)^{1/3}$	$-2.289(\lambda C_{S5}^2)^{1/3}$
Δf	$-2(\lambda^2 C_{S5})^{1/3}$	$-0.983(\lambda^2 C_{S5})^{1/3}$
α_{\max}	$\frac{7}{4}(\lambda/C_{S5})^{1/6}$	$1.513(\lambda/C_{S5})^{1/6}$
d_{\min}	$\frac{4}{7}(C_{S5}\lambda^5)^{1/6}$	$0.403(C_{S5}\lambda^5)^{1/6}$

for phase contrast to derive optimum values for defocus and the objective angle. Black & Linfoot (1957), Crewe & Salzman (1982) and Kirkland (2010) also found similar (but slightly different) requirements for a small focused probe in STEM. The results are summarized in Table 1. The definition of optimum may vary a little for different goals, so these results may also vary a little.

In an aberration-corrected instrument fifth-order spherical aberration $C_{S5} = C_{S0}$ becomes important because C_{S3} is controlled and can be reduced to near zero. The aberration function becomes

$$\chi = \frac{2\pi}{\lambda} \left(\frac{1}{2} C_1 \alpha^2 + \frac{1}{4} C_{S3} \alpha^4 + \frac{1}{6} C_{S5} \alpha^6 + \dots \right). \quad (3)$$

In the case where C_{S5} is fixed and defocus ($-\Delta f = C_1$) and third-order spherical aberration C_{S3} are controlled, Scherzer (1970) found the optimum values for phase contrast (see also Chang *et al.*, 2006; Lentzen, 2008; Erni, 2010). Intaraprasongk *et al.* (2008) derived the conditions for optimum compensation for a small probe in STEM; however the quoted results do not agree with numerical calculation. Intaraprasongk *et al.* (2008) assume a maximum phase error of $\lambda/4$ (quarter wavelength) instead of the traditional values of $\pi/4$ [$b = 4$ was used instead of $b = 8$ in Appendix A of Intaraprasongk *et al.* (2008)]. Changing this error yields the results in Table 2 which are more consistent with numerical calculation. However, small values of C_{S5} may predict rather large values for the objective angle which cannot be practically corrected (limit of accuracy of the corrector or chromatic aberration becomes dominant *etc.*) and require smaller objective angles. Changing the objective angle (α_{\max}) may also change the optimum values for Δf and C_{S3} . Lentzen (2008), Erni (2010) and Intaraprasongk *et al.* (2008) also consider the case where seventh-order spherical aberration C_{S7} becomes the limiting factor.

Aberration-corrected instruments have a very elaborate system of multipole lenses coupled to the objective lens to reduce the aberrations of the objective lenses. These devices can be very complicated and will not be discussed in detail here other than to summarize their influence on the final image (and are not shown in Figs. 1 or 2). Correctors using combinations of octupoles and quadrupoles may have on the order of 70 elements and those using hexapoles (Rose, 1981) can have about half as many elements but require more power (Müller *et al.*, 2006; Haider *et al.*, 2008, 2009; Krivanek *et al.*,

Table 3

Some axial aberration symbols through fifth order in two different systems of notation (first column: Krivanek *et al.*, 1999, 2008; second column: Haider *et al.*, 2000).

Aberrations with (a,b) have two components at different azimuthal angles or equivalently a single rotation angle and combined magnitude.

Symbol K	Symbol H	Description
C_{10}	C_1	Defocus
C_{12} (a,b)	A_1	Twofold astigmatism
C_{21} (a,b)	B_2	Axial coma
C_{23} (a,b)	A_2	Threefold astigmatism
$C_{30} = C_5$	C_3	Third-order spherical
C_{32} (a,b)	S_3	Axial star aberration
C_{34} (a,b)	A_3	Fourfold astigmatism
C_{41} (a,b)	B_4	Fourth-order axial coma
C_{43} (a,b)	D_4	Three-lobe aberration
C_{45} (a,b)	A_4	Fivefold astigmatism
$C_{50} = C_{55}$	C_5	Fifth-order spherical
C_{52} (a,b)	S_5	Fifth-order axial star
C_{54} (a,b)	R_5	Fifth-order rosette
C_{56} (a,b)	A_5	Sixfold astigmatism

2009). In the process of reducing spherical aberration of the objective a large collection of multipole aberrations are introduced and must also be corrected. Although this device is referred to as an aberration corrector (implying aberration identically equal to zero), in practice the aberration is merely reduced to small values and may still be present in the image. At high resolution only the axial aberrations near the axis need be considered (images not near the axis will have a much larger set of aberrations not listed here). There are two systems of aberration nomenclatures that are in common use illustrated by Krivanek *et al.* (1999, 2008) and Haider *et al.* (2000). Using α for the polar angle and φ for the azimuthal angle, the deviation of a spherical wave can be written as

$$\chi(\alpha, \varphi) = \frac{2\pi}{\lambda} \sum_{nm} \frac{\alpha^{n+1}}{n+1} [C_{nma} \cos(m\varphi) + C_{nmb} \sin(m\varphi)] \quad (4)$$

where n and m are positive integers and zero and λ is the wavelength of the electron. Some are listed in Table 3. Aberrations with $m \neq 0$ are sometimes referred to as parasitic because they are mostly introduced by the corrector itself.

4. Quick and simple image approximations

There is some value in simple methods that run quickly but may involve significant approximation (possibly suspect) either for rapid testing or to satiate the impatient. When first learning a new topic it may be helpful to be able to vary different parameters and get fast answers to develop an intuition of overall trends. With a general understanding it is then productive to run an accurate calculation that may take significant computing time. Also, testing the input specimen description quickly before a long calculation with a small mistake can be less frustrating. Two simple approximations, one for CTEM and one for STEM, are discussed next. Both of these can be computed fast enough to work interactively and

can even be implemented as a Java applet and run inside an internet browser.

4.1. Phase-grating approximation for CTEM

The main influence of the atoms in the specimen on the electron beam used to form an image arises from the interaction with the electrostatic potential of the atom nucleus screened by the outer electrons in the atom. The changes in the atomic electrons due to bonding with other atoms in the specimen are approximately ignored, making the total specimen potential just the sum over the potential of individual neutral atoms:

$$v_z(\mathbf{x}) = \sum_{j=1}^N v_{zj}(\mathbf{x} - \mathbf{x}_j) \quad (5)$$

where $\mathbf{x}_j = (x_j, y_j)$ is the position of atom j and $v_{zj}(\mathbf{x})$ is its projected atomic potential,

$$v_{zj}(x, y) = \int_{-\infty}^{+\infty} V_{aj}(x, y, z) dz \quad (6)$$

where V_{aj} is the atomic potential of the atom j . Both the atomic potential and its projection are tabulated in Kirkland (2010) using a Dirac–Fock calculation for the whole periodic chart. In practice, the potentials are saved in a look-up table for computational efficiency. There have been many tabulations of atomic potentials which are reviewed in more detail in Kirkland (2010). Lobato & Van Dyck (2014, 2015a) have given a new set of parameters fit to the data in Kirkland (1998).

The phase-grating approximation models the specimen as a very thin object whose main effect is a small phase shift of the incident electron wave (wide parallel beam) ignoring inelastic interactions. The incident electrons are usually a much higher energy than the electrons in the specimen so pass through the specimen with only a small change in phase. The imaging electrons mainly interact with the specimen *via* the electrostatic potential of the atoms in the specimen and are modulated by the specimen transmission function:

$$t(\mathbf{x}) = \exp[i\sigma v_z(\mathbf{x})] \quad (7)$$

where $\sigma = 2\pi m e \lambda / h^2$ is the interaction parameter, $m = \gamma m_0$ is the relativistic mass of the electron (γ being the relativistic correction factor), e is the magnitude of the charge on the electron, λ is the wavelength of the electron and h is Planck’s constant. \mathbf{x} is position in a plane perpendicular to the optical axis of the microscope. The transmission function is in some ways a phase grating giving name to this approach. If $\psi_0 \sim 1$ is the electron wavefunction incident on the specimen then the wave transmitted through the specimen is approximately

$$\psi_t(\mathbf{x}) = t(\mathbf{x})\psi_0 = t(\mathbf{x}). \quad (8)$$

The objective lens forms an image of the wave exiting the specimen into the image detector plane,

$$g(\mathbf{x}) = |t(\mathbf{x}) \otimes h_0(\mathbf{x})|^2, \quad (9)$$

where \otimes represents convolution, $h_0(\mathbf{x})$ is the complex point-spread function of the objective lens which is easier to state in Fourier space:

$$H_0(\mathbf{k}) = \text{FT}[h_0(\mathbf{x})] = \exp[-i\chi(\mathbf{k})]A(\mathbf{k}) \quad (10)$$

where $A(\mathbf{k})$, the objective aperture function, is 1 inside the aperture and 0 elsewhere. FT[] signifies a Fourier transform. The Fourier transform in two dimensions (continuous or discretely sampled) is central to understanding the electron microscope image (as is also true for optical images). There are several possible placements of minus signs and constants that can be used to define the Fourier transform (every author seems to define these a little differently). The definitions as given in Kirkland (2010) are used here. The expression above is a perfectly coherent image which is not actually true. In reality, the image is at best partially coherent. In practice, this expression [equation (9)] must also be integrated over a small range of defocus values to account for small instabilities in the high voltage and lens current supplies and a small range of illumination angles (size of the condenser aperture), either numerically or analytically using the transmission cross coefficient (more below) (Kirkland, 2010).

Silicon has a low atomic number (small phase shift) so thin specimens of silicon are reasonable candidates for a phase-grating calculation. Low atomic number specimens are usually more affected by radiation damage so a low beam energy should be used. Fig. 3 shows calculated images of a few atomic layers of silicon in the 110 projection for a non-corrected instrument and an aberration-corrected instrument. White represents a larger positive value (larger electron intensity). Atoms should appear dark in (a) and bright in (b) (more discussion below). In this projection there are pairs of atoms 1.4 Å apart (the so-called dumbbells), which are not resolved in (a) but are resolved in (b). This calculation actually used the transmission cross coefficient (described below) for the partial coherence, for ease of use with existing programs.

One step further in this approximation is the weak phase object approximation where the atomic potential is assumed to be very small. Expanding the transmission function and

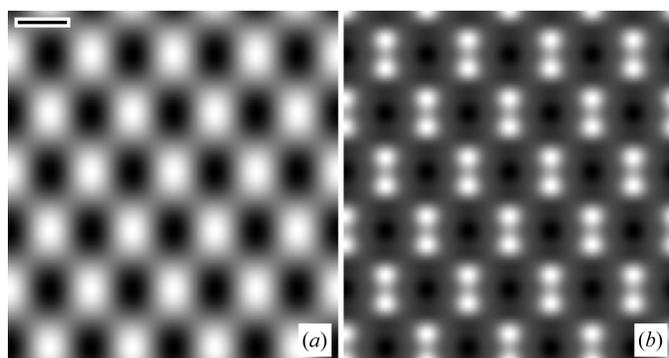


Figure 3
Calculated phase-grating images of 110 silicon at 100 kV and condenser angle of 0.1 mrad. (a) $C_{S3} = 0.7$ mm, $\Delta f = 623$ Å, objective aperture of 14 mrad and defocus spread of 50 Å. (b) With aberration correction, $C_{S5} = 50$ mm, $C_{S3} = -0.06718$ mm, $\Delta f = -178$ Å, objective aperture of 40 mrad and defocus spread of 20 Å. Scale bar is 2 Å.

keeping only the lowest term the transmission function and the recorded image become

$$\begin{aligned} t(\mathbf{x}) &= \exp[i\sigma v_z(\mathbf{x})] \sim 1 + i\sigma v_z(\mathbf{x}) \\ g(\mathbf{x}) &= 1 + 2\sigma v_z(\mathbf{x}) \otimes h_{\text{WP}}(\mathbf{x}) \end{aligned} \quad (11)$$

and the transfer function becomes

$$H_{\text{WP}}(\mathbf{k}) = \text{FT}[h_{\text{WP}}(\mathbf{x})] = \sin[\chi(\mathbf{k})]. \quad (12)$$

To include partial coherence this expression should be integrated over relevant small instabilities in the instrument. Defocus spread and illumination (condenser) angle are the most commonly included terms (neglecting other instabilities) (Frank, 1973; Fejes, 1977; Wade & Frank, 1977). The transfer function should then be modified as

$$H_{\text{WP}}(\mathbf{k}) = \int \sin[\chi(\mathbf{k} + \delta\mathbf{k}, \Delta f + \delta f)] p(\delta\mathbf{k}) p(\delta f) d^2\delta\mathbf{k} d\delta f \quad (13)$$

where $p(\delta\mathbf{k})$ and $p(\delta f)$ are the probability distributions of illumination angle and defocus spread, respectively. A straightforward but tedious calculation (Kirkland, 2010) through C_{S5} produces

$$\begin{aligned} H_{\text{WP}}(k) &= \frac{1}{(1 + \varepsilon k^2)^{1/2}} \times \sin\left\{ \frac{\pi\lambda k^2}{1 + \varepsilon k^2} \right. \\ &\quad \times \left. \left[\frac{1}{3} C_{S5} (1 - 2\varepsilon k^2) \lambda^4 k^4 + 0.5 C_{S3} (1 - \varepsilon k^2) \lambda^2 k^2 - \Delta f \right] \right\} \\ &\quad \times \exp\left(- \left\{ [\pi\lambda k_s k (C_{S5} \lambda^4 k^4 + C_{S3} \lambda^2 k^2 - \Delta f)]^2 \right. \right. \\ &\quad \left. \left. + 0.25(\pi\lambda \Delta_0 k^2)^2 \right\} / (1 + \varepsilon k^2) \right) \end{aligned} \quad (14)$$

where Δ_0 is the spread in defocus values, λk_s is approximately the condenser illumination angle and $\varepsilon = (\pi\lambda k_s \Delta_0)^2$.

A graph of the transfer function in the weak phase object approximation for the image conditions in Fig. 3 is shown in Fig. 4. The traditional (low-resolution) image produces a negative contrast (positive C_{S3} , top graph) and the aberration-

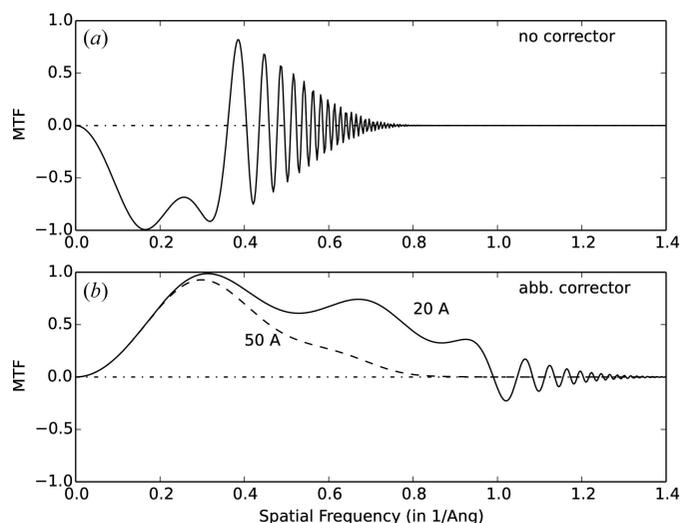


Figure 4
Transfer function for the weak phase object approximation image of Fig. 3 at 100 kV. Two values of defocus spread (20 Å and 50 Å) are shown for the aberration-corrected curve.

corrected image produces a positive contrast (negative C_{S3} , bottom graph). The transfer function for the aberration-corrected image goes to much higher spatial frequency (scattering angle) but is also very sensitive to defocus spread and would probably require a chromatic aberration corrector to achieve this resolution (with defocus spread of less than 20 Å needed). This suggests that an aberration-corrected CTEM may be very sensitive to instabilities, consistent with Schramm *et al.* (2012) and Barthel & Thust (2013).

4.2. Incoherent STEM approximation

Many authors have considered imaging approximations for thin specimens in ADF-STEM (for example, Misell *et al.*, 1974; Cowley, 1976; Spence & Cowley, 1978; Jesson & Pennycook, 1993; Treacy & Gibson, 1993; Loane *et al.*, 1992; Hawkes & Kasper, 1994; Bosch & Lazić, 2015). If the STEM probe is small compared to the spacing between the atoms in the specimen and the specimen is thin enough so that the incident electrons do not scatter more than once while going through the specimen, and only electrons scattered to angles larger than the objective aperture are collected (*i.e.* no interference between scattered and unscattered electrons), then the image is approximately incoherent and approximately the same as a linear incoherent optical image as described by Black & Linfoot (1957). The recorded image intensity $g(\mathbf{x})$ is

$$g(\mathbf{x}) = f(\mathbf{x}) \otimes h(\mathbf{x}) = \int f(\mathbf{x}')h(\mathbf{x} - \mathbf{x}') d\mathbf{x}' = \text{FT}^{-1}[F(\mathbf{k})H(\mathbf{k})] \quad (15)$$

where $\mathbf{x} = (x, y)$ is a position vector in the image plane, \otimes represents convolution, $h(\mathbf{x})$ is the point-spread function or PSF and $f(\mathbf{x})$ is the ideal image of the specimen. The convolution is best performed using an FFT (fast Fourier transform) and FT represents a Fourier transform. The PSF is approximately the probe intensity and the ideal image is the scattering strength or partial cross section for scattering onto the ADF detector angles from each atom. The probe wavefunction when deflected to position \mathbf{x}_p is calculated from the phase error across the objective aperture as

$$\psi_p(\mathbf{x}, \mathbf{x}_p) = A_p \int_0^{k_{\max}} \exp[-i\chi(\mathbf{k}) - 2\pi i\mathbf{k} \cdot (\mathbf{x} - \mathbf{x}_p)] d^2\mathbf{k} \quad (16)$$

where A_p is a normalization constant chosen to yield

$$\int |\psi_p(\mathbf{x}, \mathbf{x}_p)|^2 d^2\mathbf{x} = 1. \quad (17)$$

The point-spread function in this incoherent ADF-STEM approximation is approximately the intensity in the probe:

$$h_{\text{ADF}}(\mathbf{x}) = |\psi_p(\mathbf{x})|^2 = A_p^2 \left| \int_0^{k_{\max}} \exp[-i\chi(\mathbf{k}) - 2\pi i\mathbf{k} \cdot \mathbf{x}] d^2\mathbf{k} \right|^2. \quad (18)$$

In practice equation (15) should also be convolved with the source size demagnified to the specimen plane because the probe is really just an image of the electron source.

$$h_{\text{ADF}}(\mathbf{x}) = |\psi_p(r)|^2 \otimes p_s(\mathbf{x}) \quad (19)$$

where $p_s(\mathbf{x})$ is the source shape function (usually Gaussian) in the image plane. This equation assumes that each point on the source produces a ray that is incoherent with all other points on the source. It is not clear whether a cold field emitter should be summed coherently or incoherently. An incoherent Gaussian is assumed here.

$$p_s(\mathbf{x}) = \exp\left(-\frac{x^2}{d_s^2} \ln 2\right) \quad (20)$$

$$P_s(\mathbf{k}) = \exp[-\pi^2 d_s^2 k^2 / (\ln 2)] \quad (21)$$

where $2d_s$ is the full width at half-maximum. Measurements of the source distribution (Dwyer *et al.*, 2010) seem to indicate a Lorentzian-shaped source size (long probe tails). However there are many different ways to produce a Lorentzian-like probe. For example, the accumulation of many small tuning errors leads to a Lorentzian-like probe (Kirkland, 2011). Before this can be identified as a source effect it has to be shown to vary with source magnification (condenser strength) which has not yet been done systematically. More work is needed here.

Also in this approximation the specimen is viewed as a collection of isolated atoms, each with a given scattering strength $S_j(\mathbf{x}_j)$ that is the partial cross section $\partial\sigma/\partial\theta$ for scattering onto the ADF detector angles:

$$f(\mathbf{x}) \sim \sum_j S_j(\mathbf{x}_j) \quad (22)$$

$$S(\mathbf{x}) \sim \int_{\theta_{\min}}^{\theta_{\max}} \frac{\partial\sigma}{\partial\theta} d\theta \quad (23)$$

where subscript j denotes different atoms and the ADF detector covers polar angles θ_{\min} to θ_{\max} . The partial cross sections are calculated from the atomic potentials in the Moliere or Eikonal approximation (equation 5.18 of Kirkland, 2010).

Although the incoherent ADF-STEM approximation may not be quantitatively accurate for most specimens, it is fairly accurate for the new class of materials that are only one or two atoms thick such as graphene. An image of graphene calculated in this approximation is shown in Fig. 5. A specimen this thin composed of low atomic number atoms such as carbon (low scattering power) produces a very small signal on the detector requiring very sensitive detectors. The image shown in Fig. 5(b) includes random Poisson distributed noise for the probe current and dwell time shown. This is fundamentally the best that can be achieved. An actual experimental image will usually have more noise unless it uses single electron counting, which is possible in STEM (Kirkland & Thomas, 1996).

An aberration corrector really only reduces the aberrations and does not correct them to zero. The accuracy of the corrector in some ways determines the final resolution. A multipole corrector introduces a large number of new parasitic aberrations. The accumulation of many small errors in these aberrations can produce a probe (in STEM) with large Lorentzian-like tails (Kirkland, 2011), which will be hard to distinguish from a Lorentzian-like source distribution (Dwyer

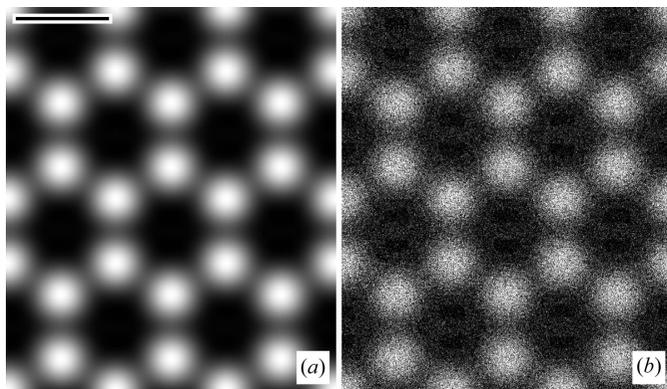


Figure 5
Calculated incoherent ADF-STEM image for a single layer of graphene at 60 kV with aberration correction ($C_{S3} = \Delta f = 0$, $C_{S5} = 10$ mm, objective aperture of 30 mrad, defocus spread 80 Å, detector angles 60 to 200 mrad, source size 0.5 Å). (a) Ideal, no noise and (b) Poisson electron counting noise (60 pA for 30 μ s, yielding a maximum of about 10 to 20 electrons in a pixel). Scale bar is 2 Å.

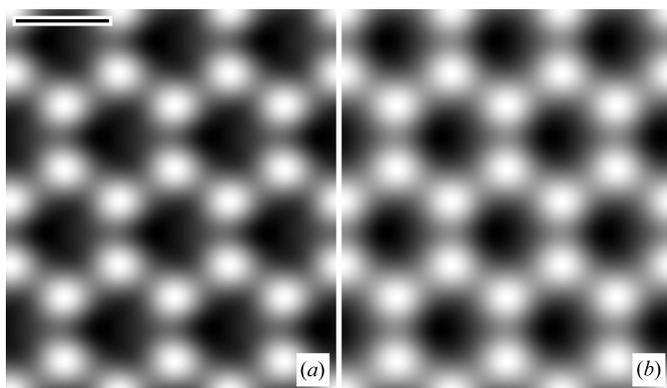


Figure 6
Calculated incoherent ADF-STEM image for a single layer of graphene at 60 kV with aberration correction ($C_{S3} = \Delta f = 0$, objective aperture of 30 mrad, defocus spread 80 Å, detector angles 60 to 200 mrad, source size 0.5 Å). (a) $C_{45a} = 0.4$ mm. (b) $C_{45a} = 0.4$ mm and $C_{56a} = 20$ mm. Scale bar is 2 Å.

et al., 2010). A third-order corrector will be able to compensate many of the fourth- and fifth-order aberrations if they are small except for C_{45} and C_{56} (no low-order aberrations with similar azimuthal symmetry). These aberrations and the small errors in other aberrations may be a final limiting factor in such a corrector. Fig. 6 shows graphene images with medium amounts of these two aberrations, which start to appear similar to the observed graphene image (Huang *et al.*, 2011; Krivanek *et al.*, 2010; Ercius *et al.*, 2012; Lupini & Pennycook, 2012) (there are several ways to produce similar image artifacts).

5. Propagation through thick specimens

As the electron passes through the specimen it strongly interacts with the specimen and most likely scatters more than once, which complicates the calculation of the resulting electron wavefunction exiting the specimen (the exit wave) and

interpretation of the micrograph. There are two popular methods of calculation, the Bloch wave and multislice methods. The real-space method (Van Dyck & Coene, 1984; Coene & Van Dyck, 1984*a,b*) has some interesting features but usually takes more computer time, is less commonly used and not discussed in detail here. The Bloch wave or eigenvalue method (Bethe, 1928) (and, for example, Humphreys, 1979; Spence, 2013; Spence & Zuo, 1992; DeGraf, 2003) solves for the eigenvectors and values inside the specimen. This method was originally used to solve for electron diffraction in crystals (electron microscopes had not yet been invented) but is equally valid for electron-microscope images. Matching the wavefunction at the entrance surface of the specimen to a sum of eigenvectors gives the wavefunction at the exit surface. The second method was originally based on a physical optics perspective and is called the multislice method (Cowley & Moodie, 1957; Goodman & Moodie, 1974), later extended to STEM by Kirkland *et al.* (1987). Reviews of multislice have been given by Van Dyck (1985), Ishizuka (2004) and Kirkland (2010). Multislice divides the specimen into a sequence of small slices. Each slice is thin enough to produce only a small phase shift of the high-energy electron wavefunction. The electron wave alternately passes through a thin slice and propagates to the next slice. Both methods ignore any back-scattered electrons consistent with a high-energy electron approximation in which the electron is only slightly perturbed by the specimen in addition to a variety of similar small approximations. Both end up with a similar level of accuracy. However there is a large difference in the computational efficiency or required computer time. The multislice method can conveniently be written in terms of successive Fourier transforms (Ishizuka & Uyeda, 1977; Bursill & Wilson, 1977). The FFT (Cooley & Tukey, 1965; Brigham, 1988; Walker, 1996) is one of the most efficient numerical algorithms available with many well developed subroutine libraries. Both methods need about the same number of Fourier terms or beams to achieve the same level of accuracy so the multislice method has a dramatic advantage in terms of computer time (more below).

In the special case of propagation in a simple crystal along a high-symmetry zone axis the electrons can be thought of as channeling along columns of atoms. The electrons seem to be bound in a two-dimensional plane in the specimen in atomic like states centered on each atomic column (Kambe *et al.*, 1974; Buxton *et al.*, 1978). Hovden *et al.* (2012) considered the case in which adjacent columns can constructively and destructively interfere to produce oscillation between columns *versus* depth in the specimen. This approach can provide some useful intuition but will not be considered further here.

Even at a kinetic energy of 100 keV the electron is traveling at about half of the speed of light so there are significant relativistic effects. A full relativistic treatment of electron optics (Fujiwara, 1961) is not an easy task but has been found to approximately agree with a simple non-relativistic Schrödinger equation using the relativistic electron mass and wavelength [see equations 2.2 and 2.5 of Kirkland (2010)]. This approximation will be used here as well (electron spin is

also usually ignored). The full wavefunction $\psi(x, y, z)$ as a function of three spatial coordinates (x, y, z) in an electrostatic potential $V(x, y, z)$ of the specimen satisfies the Schrödinger equation:

$$\left[-\frac{\hbar^2}{2m} \nabla^2 - eV(x, y, z) \right] \psi(x, y, z) = E\psi(x, y, z) \quad (24)$$

where $\hbar = h/2\pi$ is Planck's constant divided by 2π , E is the kinetic energy of the electron and $-eV$ is the potential energy of the electron.

5.1. Bloch waves: the eigenvalue approach

In the Bloch wave approach the electrostatic potential inside the specimen (usually a crystal) and the electron wavefunction are expanded in a three-dimensional Fourier series with the same periodicity as the crystal specimen. The components of the resulting wavefunction are referred to as Bloch waves (named after a similar construction in solid-state physics). This form of the potential and wavefunction are inserted into the Schrödinger equation for the region inside the specimen.

The resulting Schrödinger equation requires a variety of approximations [see Kirkland (2010) for more details] to get an equation that can be solved (very similar to the approximations to be used in the multislice method in the next section). In particular, half of the solutions will be lost which correspond to the back-scattered wave (DeGraf, 2003). The Bloch waves must be eigenvectors of the resulting (large) matrix equation.

Solving this eigenvalue equation for the eigenvectors and eigenvalues yields the mode of propagation inside the specimen. Then matching the wavefunction at the entrance surface of the specimen (plane wave for CTEM and probe wavefunction for STEM) yields the electron wavefunction at the exit surface of the specimen. If the crystal (specimen) is centrosymmetric then the matrix is real and symmetric, otherwise it is Hermitian. The math has been written out in detail by Humphreys (1979), Spence (2013), Spence & Zuo (1992), DeGraf (2003) and Kirkland (2010).

5.2. Multislice: the FFT approach

In principle, a finite difference solution should start with the time-dependent Schrödinger equation and a sampling size that is a small fraction of the electron wavelength (for example $\lambda/10$). However, the electron wavelength is on the order of 10^{-12} m which would require a prodigious amount of memory for a specimen size of several 100's of Å's in all three dimensions and many extremely small time steps for electrons that are moving near the speed of light. Some clever approximations are needed to make this approach practical.

The multislice method originated from concepts in physical optics (Cowley & Moodie, 1957; Goodman & Moodie, 1974), but can be derived more rigorously from the Schrödinger equation and is similar to the (independently developed) split-step method used to calculate the propagation of light in fibers (for example, Feit & Fleck, 1978). The imaging electrons have

a very small wavelength and the atoms in the specimen have only a small effect in most cases so factor the electron wavelength into a slowly varying component $\psi_S(x, y, z)$ and a fast varying component (varies as the vacuum wavelength). The full electron wavefunction is

$$\psi(x, y, z) = \psi_S(x, y, z) \exp(2\pi iz/\lambda). \quad (25)$$

Substitute this expression into the Schrödinger equation. Dropping a small second-derivative term ($\lambda|\partial^2\psi_S/\partial z^2|$) the expression for the slowly varying portion of the wavefunction becomes

$$\left[\nabla_{xy}^2 + \frac{4\pi i}{\lambda} \frac{\partial}{\partial z} + \frac{2meV(x, y, z)}{\hbar^2} \right] \psi(x, y, z) = 0 \quad (26)$$

where the subscript S has been dropped for simplicity, and ∇_{xy}^2 is the sum of second derivatives with respect to x and y . This approximation is roughly equivalent to neglecting back-scattered electrons and limiting the results to the paraxial or small-angle approximation (similar to the approximation used in the Bloch wave eigenvalue approach in the previous section). A formal operator solution is

$$\psi(x, y, z + \Delta z) = \exp \left\{ \int_z^{z+\Delta z} \left[\frac{i\lambda}{4\pi} \nabla_{xy}^2 + i\sigma V(x, y, z') \right] dz' \right\} \times \psi(x, y, z). \quad (27)$$

The wavefunction propagates for a slice thickness of Δz . With some tedious math this complicated operator can be approximately split into two factors (Kirkland, 2010) to accuracy $\mathcal{O}(\Delta z^2)$ leaving

$$\begin{aligned} \psi(x, y, z + \Delta z) &= \exp \left(\frac{i\lambda\Delta z}{4\pi} \nabla_{xy}^2 \right) \exp [i\sigma v_{\Delta z}(x, y, z)] \psi(x, y, z) \\ &\quad + \mathcal{O}(\Delta z^2) \\ &= \exp \left(\frac{i\lambda\Delta z}{4\pi} \nabla_{xy}^2 \right) t(x, y, z) \psi(x, y, z) + \mathcal{O}(\Delta z^2) \end{aligned} \quad (28)$$

where $t(x, y, z)$ is the transmission function of the specimen between z and $z + \Delta z$:

$$t(x, y, z) = \exp \left[i\sigma \int_z^{z+\Delta z} V(x, y, z') dz' \right]. \quad (29)$$

The slice thickness Δz is usually much larger than the size of the atomic potential so the range of the integral is usually changed to be from $-\infty$ to $+\infty$ for each atom in the slice (with summation over atoms in the slice) and this transmission function is the same as that in the phase-grating approximation as in equations (6) and (7). The remaining operator factor can be identified as a convolution with the propagation function (Fresnel diffraction), which is easier to state in Fourier space as

$$P(k, \Delta z) = \exp(-i\pi\lambda k^2 \Delta z). \quad (30)$$

The multislice equation can be written in compact form from layer n to layer $n + 1$ (from z to $z + \Delta z$) of the slowly varying portion of the electron wavefunction as

$$\psi_{n+1}(x, y) = p_n(x, y, \Delta z_n) \otimes [t_n(x, y)\psi_n(x, y)] + \mathcal{O}(\Delta z^2). \quad (31)$$

Convolutions can conveniently be performed using Fourier transforms as

$$\psi_{n+1}(x, y) = \text{FT}^{-1}\{P_n(k_x, k_y, \Delta z_n)\text{FT}[t_n(x, y)\psi_n(x, y)]\} + \mathcal{O}(\Delta z^2). \quad (32)$$

This expression should be bandwidth limited (typically 2/3 of maximum bandwidth) (Kirkland, 2010) to avoid aliasing. Generally speaking, if calculating a crystal specimen there should be an integer number of slices per unit cell to avoid sampling problems (otherwise a false first-order Laue zone or FOLZ ring is generated corresponding to the slice thickness). The FFT is one of the most efficient algorithms available which makes this form of multislice computationally very fast and likely accounts in part for the popularity of the multislice method.

Each step in multislice is unitary which makes it numerically stable (stability and accuracy are not the same thing) and the total integrated intensity of the wavefunction should remain constant. Monitoring the total integrated intensity $\int |\psi|^2 dx dy$ is a simple test for convergence. If started normalized to unity, then the value of this integral should stay at least above 0.9. Values > 0.95 are good and values < 0.90 are usually bad (may be qualitatively correct but not quantitatively correct). Although simple to implement this test is not very rigorous and is only a first step [see Kirkland (2010) for a more detailed discussion of convergence tests].

The frozen phonon method (Loane *et al.*, 1991; Hillyard & Silcox, 1995; Kirkland, 2010) uses multislice to calculate the effect of thermal vibrations of the atoms in the specimen. The atom positions are randomly displaced consistent with the known thermal vibration amplitudes (usually from the Debye-Waller factor) and a multislice calculation is performed to produce an image or diffraction pattern. Then another set of atom positions with different random displacements is produced and another multislice calculation performed. The intensity (not amplitude) of the image or diffraction pattern from each set of random atomic displacements is then averaged from many sets of atomic coordinates. This procedure has been shown to accurately reproduce thermal diffuse scattering (Muller *et al.*, 2001; LeBeau *et al.*, 2008). There is another method in condensed-matter theoretical calculations also called the frozen phonon method (Martin, 2004) used to calculate phonon dynamics. Although there is a similarity in the underlying physics these two methods of the same name are different and appear to have been independently invented at about the same time.

5.3. Relative performance

It is more than just vain competition to ask which method is faster. Some significant calculations may take days or weeks to

perform. A faster method may make some calculations practical instead of impractical. The computer time for a Bloch wave eigenvalue calculation scales as N^3 where N is the number of beams or Fourier coefficients used and an FFT multislice calculation scales as $N \log N$. The memory storage requirements scale as N^2 for a Bloch wave calculation and N for a multislice calculation. The counting of beams (or Fourier coefficients) is different in these two methods so a precise comparison is difficult, but there is enough similarity for an approximate comparison. This comparison has been given previously in Fig. 6.2 of Kirkland (2010) with older versions of eigenvalue and FFT subroutines. It is perhaps worth repeating this comparison here with newer improved software that has become available. The computer time to compute a Bloch wave solution and an FFT multislice solution *versus* the number of 'beams' is shown in Fig. 7, using the *Eigen* C++ linear algebra library (<http://eigen.tuxfamily.org>) and the *FFTW* (Frigo & Johnson, 2005) software package. Only the non-aliased beams were counted in the multislice calculation. Both software packages are highly developed and at least representative of the best that can be currently achieved although there may be some variation with computer hardware and software.

Both programs were run on the same small computer and were compiled with the same compiler, so the relative performance is a reasonable representation of the relative performance of these two approaches. One specimen is centrosymmetric (aluminium) so the eigenvalue matrix is real and symmetrical which improves the speed and the other is not symmetric (silicon) so the eigenvalue matrix is complex and Hermitian. Two different thicknesses are shown for the FFT multislice program which unlike the Bloch wave method increases in time with the thickness. The eigenvalue method is clearly much slower and does not scale well as the number of

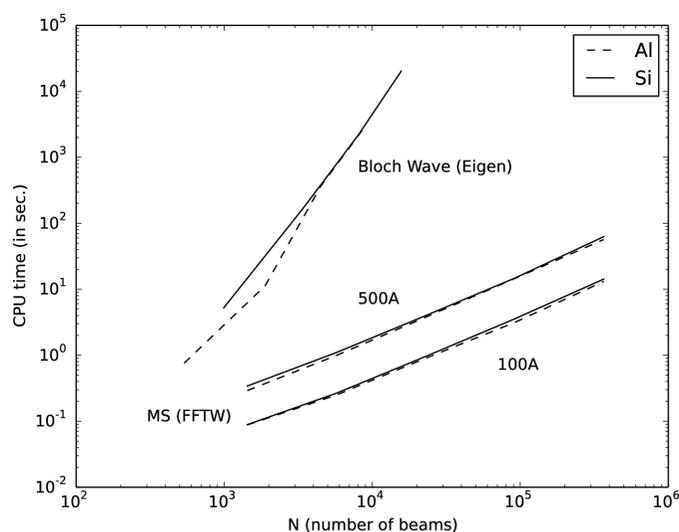


Figure 7

Computer speed comparison between the Bloch wave eigenvalue approach and the FFT multislice approach. Both programs were compiled and run on the same small computer in single-thread mode. Two specimens were tested, silicon (non-symmetric) and aluminium (symmetric) and two specimen thicknesses are shown for the FFT multislice program.

beams increases. Generally speaking, both methods require a similar number of beams for a similar accuracy although the beam counting is a little different in each method. The Bloch wave (eigenvalue) method may have some uses in theoretical investigation of small unit-cell specimens but is otherwise not competitive and will not be discussed further here.

6. CTEM, phase contrast and partial coherence

In the CTEM the incident electron wavefunction is unity (a plane wave), which is then propagated through the specimen using either Bloch waves or multislice (used here). The electron wave transmitted through the specimen, $\psi_t(\mathbf{x})$, is imaged by the objective lens and its intensity is recorded (typically on a CCD). A perfectly coherent image would be

$$g(\mathbf{x}) = |\psi_t(\mathbf{x}) \otimes h_0(\mathbf{x})|^2. \quad (33)$$

In practice, there is a small range of incident angles from the condenser angles (β) and a small spread in defocus values from small instabilities in the high voltage and lens current instabilities. The recorded image should be integrated over these small instabilities (assumed to be on a time scale much shorter than the image recording time):

$$g(\mathbf{x}) = \int |\psi_t(\mathbf{x}) \otimes h_0(\mathbf{x}, \Delta f + \delta_f, \mathbf{k}_\beta)|^2 \times p(\mathbf{k}_\beta) p(\delta_f) d\delta_f d^2\mathbf{k}_\beta \quad (34)$$

where δ_f is defocus fluctuation and \mathbf{k}_β is an illumination angle. $p(\delta_f)$ and $p(\mathbf{k}_\beta)$ are the probability distributions of these fluctuations. This expression can be evaluated by direct numerical integration. Some further analytical simplification is also possible as a Fourier transform. The Fourier transform of the coherent image is

$$G(\mathbf{k}) = [\Psi_t(\mathbf{k})H_0(\mathbf{k})] \otimes [\Psi_t^*(-\mathbf{k})H_0^*(-\mathbf{k})] = \int T_{cc}(\mathbf{k}', \mathbf{k}' + \mathbf{k}) \Psi_t^*(\mathbf{k}') \Psi_t(\mathbf{k}' + \mathbf{k}) d^2\mathbf{k}' \quad (35)$$

where $T_{cc}(\mathbf{k}', \mathbf{k}' + \mathbf{k})$ is the transmission cross coefficient that is similar to the function of the same name in light optics [for example, section 10.5.3 of Born & Wolf (1980)]. In the perfectly coherent case

$$T_{cc}^{\text{coh}}(\mathbf{k}', \mathbf{k}' + \mathbf{k}) = \exp[i\chi(\mathbf{k}') - i\chi(\mathbf{k}' + \mathbf{k})] A(\mathbf{k}') A(\mathbf{k}' + \mathbf{k}). \quad (36)$$

Now integrate over instabilities:

$$T_{cc}(\mathbf{k}', \mathbf{k}' + \mathbf{k}) = \int \exp[i\chi(\mathbf{k}' + \mathbf{k}_\beta, \Delta f + \delta_f) - i\chi(\mathbf{k}' + \mathbf{k} + \mathbf{k}_\beta, \Delta f + \delta_f)] \times A(\mathbf{k}' + \mathbf{k}_\beta) A(\mathbf{k}' + \mathbf{k} + \mathbf{k}_\beta) p(\delta_f) p(\mathbf{k}_\beta) d\delta_f d^2\mathbf{k}_\beta. \quad (37)$$

Expanding small terms to lowest order O'Keefe (1979), Ishizuka (1980), Pulvermacher (1981) and Kirkland (2010) obtained a result through C_{S3} . Adding terms through C_{S5} and keeping only symmetrical aberration ($m = 0$):

$$T_{cc}(\mathbf{k}', \mathbf{k}' + \mathbf{k}) = T_{cc}^{\text{coh}} \frac{1}{(1 + \pi^2 \beta^2 \Delta_0^2 k^2)^{1/2}} \times \exp \left[-\frac{\beta^2}{4\lambda^2} W_{C1}^2 + \frac{\Delta_0^2 (\pi \beta^2 \mathbf{k} \cdot \mathbf{W}_{C1} / \lambda - i W_{C2})^2}{4(1 + \pi^2 \beta^2 \Delta_0^2 k^2)} \right] \quad (38)$$

$$= T_{cc}^{\text{coh}} T_{cc}^{\text{PC}} \quad (39)$$

where

$$\begin{aligned} \mathbf{W}_{C1} &= 2\pi\lambda^5 C_{S3} [|\mathbf{k}'|^4 \mathbf{k}' - |\mathbf{k}' + \mathbf{k}|^4 (\mathbf{k}' + \mathbf{k})] \\ &\quad + 2\pi\lambda^3 C_{S3} [|\mathbf{k}'|^2 \mathbf{k}' - |\mathbf{k}' + \mathbf{k}|^2 (\mathbf{k}' + \mathbf{k})] \\ &\quad + 2\pi\lambda \Delta f \mathbf{k} \\ W_{C2} &= -\pi\lambda (|\mathbf{k}'|^2 + |\mathbf{k}' + \mathbf{k}|^2). \end{aligned} \quad (40)$$

Equation (35) with equation (40) is sadly not separable so cannot be done with FFTs but must be calculated as a direct weighted convolution in two dimensions which may require significant computer time.

7. STEM

The STEM forms a focused probe on the entrance of the specimen [equation (16)] and then it travels through the specimen. If the specimen is thicker than a single atom, a multislice or Bloch wave calculation must be performed at each position of the probe.

After passing through the specimen the electron wavefunction $\psi_t(\mathbf{x}, \mathbf{x}_p)$ hits the detector and the integrated intensity forms the image signal at each position \mathbf{x}_p of the probe,

$$g(\mathbf{x}_p) = \int |\Psi_t(\mathbf{k}, \mathbf{x}_p)|^2 D(\mathbf{k}) d^2\mathbf{k} \quad (41)$$

$$\Psi_t(\mathbf{k}, \mathbf{x}_p) = \text{FT}[\psi_t(\mathbf{x}, \mathbf{x}_p)] \quad (42)$$

where $D(\mathbf{k})$ is the detector function:

$$D(\mathbf{k}) = 1 \text{ for } k_{D\text{min}} \leq |\mathbf{k}| \leq k_{D\text{max}} \quad (43)$$

$$= 0 \text{ otherwise} \quad (44)$$

where $\alpha_{D\text{min}} = \lambda k_{D\text{min}}$ and $\alpha_{D\text{max}} = \lambda k_{D\text{max}}$ are the minimum and maximum angles of the detector. In ADF these are large angles and in BF or ABF these are small angles. The function $|\Psi_t(\mathbf{k}, \mathbf{x}_p)|^2$ versus \mathbf{k} is also called the CBED (convergent-beam electron diffraction) pattern and can also be recorded separately on most instruments.

In the Bloch wave approach all forms of propagation through the specimen are found at once (with one set of eigenvectors). All that is necessary is to match each probe position wavefunction to a set of Bloch waves, then the exit wavefunction is known. This is an advantage of some kind; however in the multislice approach each probe may be propagated independently so is fairly easy to parallelize and run on multiple CPUs (central processing units) at the same time, which tends to give multislice an advantage in overall speed.

8. Examples

A simple example of image simulation is shown in Fig. 8. This is a small crystal of gold [face-centered cubic (f.c.c.) lattice size 4.08 Å] on the center of a thin carbon support. The atomic coordinates for the amorphous carbon support were generated by selecting uniformly distributed random values within the corresponding rectangular volume until the density of carbon was reached with the constraint that atoms cannot be closer than 1 Å. The transmission function, probe wavefunction and STEM images were sampled with 512 by 512 pixels. Fig. 8(*a*) is a BF-CTEM phase-contrast image (defocus spread of 50 Å and illumination angle 0.1 mrad) and (*d*) is an ADF-STEM image (detector 30 to 200 mrad). Figs. 8(*b*) and 8(*c*) are ABF images with the BF detector covering the outer half of the objective aperture angle. White is a larger positive value (electron intensity).

The BF-CTEM image in Fig. 8(*a*) has only a weak dependence on atomic number so shows both the carbon support and the heavy gold atoms (dark spots) in the middle. The ADF-STEM image in Fig. 8(*d*) has a strong dependence on atomic number (*Z* contrast) so the gold atoms in the center (white spots) stand out sharply and the carbon support has disappeared. ADF-STEM can be used to image single heavy atoms on a carbon (Isaacson *et al.*, 1976) or silicon support (Loane *et al.*, 1988) or inside a silicon crystal (Voyles *et al.*, 2003, 2004). The ABF images in Figs. 8(*b*), 8(*c*) are something in between. When focused as BF (*b*) there is a little of the

carbon support and sharp gold atoms. When focused as ADF-STEM (*c*) the carbon support is weaker and the gold atoms are sharp.

Multislice is very adept at dealing with unusual specimen structure (non-crystalline). All that is needed is a collection of three-dimensional coordinates and atomic numbers. As an example a multislice calculation of a protein structure of immunoglobulin from the PDB (PDB code 1igt, Harris *et al.*, 1997) on a thin carbon support similar to that in Fig. 8 is shown in Fig. 9. In practice, this specimen would be quickly damaged in the beam so is unlikely to form an image. This example is what might happen in the absence of radiation damage. This image was calculated with a transmission function of 2048 by 2048 pixels and a probe wavefunction of 512 by 512 pixels and a slice thickness of 1.5 Å. STEM images were calculated for an image size of 512 by 512 pixels. There is almost nothing visible in a traditional BF-CTEM image (needs to be stained to be visible); however the ABF (*b*) and ADF-STEM (*d*) show some possible structure.

There is a large collection of multislice and Bloch wave results in the literature. Usually a large array of images of a crystalline specimen with small increments in defocus and thickness are given, which will not be repeated here. The effects of specimen tilt (Yu *et al.*, 2008) and strain (Yu *et al.*, 2004) at an interface between amorphous and crystalline silicon and tilt in a bulk crystal (Maccagnano-Zacher *et al.*, 2008) and amorphous layers on crystalline specimens

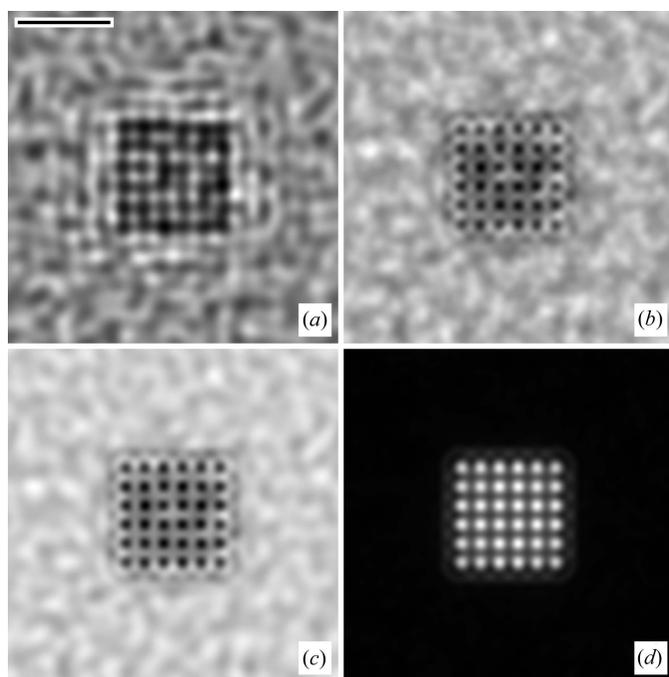


Figure 8

Calculated images of a small gold crystal (3 by 3 by 3 unit cells near the center) on a thin amorphous carbon support (20 Å thick) at 200 kV and $C_{S3} = 0.7$ mm. (*a*) BF-CTEM, $\Delta f = 520$ Å, objective aperture 12 mrad, (*b*) ABF-STEM, $\Delta f = 520$ Å, objective aperture 12 mrad, (*c*) ABF-STEM, $\Delta f = 365$ Å, objective aperture 10.5 mrad, (*d*) ADF-STEM, $\Delta f = 365$ Å, objective aperture 10.5 mrad. Scale bar is 10 Å.

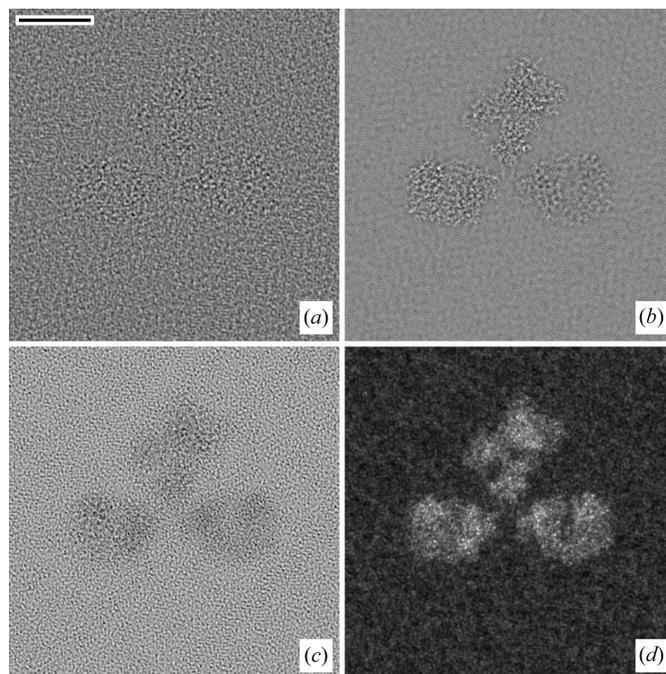


Figure 9

Calculated images of immunoglobulin (PDB code 1igt) on a thin amorphous carbon support (30 Å thick) at 200 kV and $C_{S3} = 0.7$ mm. Radiation damage has not been included but in practice would likely prevent this image from being recorded. (*a*) BF-CTEM, $\Delta f = 520$ Å, objective aperture 12 mrad, (*b*) ABF-STEM, $\Delta f = 520$ Å, objective aperture 12 mrad, (*c*) ABF-STEM, $\Delta f = 365$ Å, objective aperture 10.5 mrad, (*d*) ADF-STEM, $\Delta f = 365$ Å, objective aperture 10.5 mrad. Scale bar is 50 Å.

(Mkhoyan *et al.*, 2008) have been investigated for ADF STEM using multislice.

9. Accuracy

It is important to quantitatively compare the results of theoretical calculations to real experimental measurements occasionally to verify that the calculation is actually correct. Early comparisons between theory and experiment of a BF-CTEM image were found to differ by about a factor of two or three which has become known as the ‘Stobbs factor’ (Hýtch & Stobbs, 1994; Boothroyd, 1998). There may be several explanations, including the transfer function (MTF) of the image recording device (film, CCD *etc.*) (Thust, 2009), amorphous contamination layers *etc.*, and there is still some controversy. Meyer *et al.* (2011) obtained good agreement with aberration-corrected BF-CTEM imaging of graphene when the detector MTF and charge redistribution around defects were included. Krause *et al.* (2013) obtained remarkably good agreement between aberration-corrected CTEM images and multislice calculations for small objective apertures and careful measurements including the detector MTF, and small deviations ($1.2\times$) for larger apertures (most likely due to the difficulty of measuring the aberrations accurately enough to match the rapidly oscillating portion of the transfer function). Imaging in the ADF-STEM has been quantitatively tested by LeBeau *et al.* (2008, 2009), LeBeau & Stemmer (2008) and Findlay & LeBeau (2013) with very good results, and Muller *et al.* (2001) have obtained good quantitative agreement of theory and measured values in CBED including thermal diffuse scattering. Kourkoutis *et al.* (2011) have shown good agreement between calculated and measured STEM signal for various channeling conditions in silicon. Koch & Zuo (2000) have compared the results of various multislice and Bloch wave programs and found good agreement. Multislice theory seems to be substantially correct when experiment and calculations are carefully performed.

Aberration correctors create a big problem for quantitative comparisons. A large set of new aberrations is generated with a multipole corrector. Spherical aberration is greatly reduced but not exactly zero. Instead of a few large aberrations that can be accurately measured there are a large number of aberrations with small random measurement errors which are unknown by definition. It is quite easy to lose a factor of two in contrast from small measurement errors (Kirkland, 2011). The only hope of being quantitative at high precision in image intensity with an aberration corrector is if a stochastic average of small errors in many aberrations tends to produce the same result (more work is needed here) or aberration measurement improves significantly.

The multislice method is only accurate to second order (locally) or first order (globally) which is not that great. It can be shown that the standard multislice can be factored slightly differently and interpreted as a globally second-order solution that is offset by one half of one slice (Van Dyck, 1985). In some ways it is more accurate than it seems at first glance. Both multislice and Bloch wave solutions ignore the back-

scattered electrons and have other similar approximations. Watanabe *et al.* (1988), Chen *et al.* (1997), Chen & Van Dyck (1997), Kirkland (1998, 2010), Cai & Chen (2012) and Ming & Chen (2013) have proposed more accurate numerical multislice formulations, some of which may include back-scattered electrons in some way. Dulong *et al.* (2008) have done a similar study for Bloch wave calculations. However, most advanced methods are complicated and require large amounts of computation, making them difficult to use for routine calculations.

One straightforward but perhaps inelegant approach to test the accuracy of multislice is to reduce the slice thickness to smaller and smaller values. However, the transmission function, equation (29), is usually calculated by integrating the projected atomic potential from $-\infty$ to $+\infty$ because there is a convenient analytical result which is easily tabulated. It is a programming convenience to use atomic potentials that have been analytically integrated through the atom, so the minimum slice thickness is approximately the size of the atomic potential, which is about 1 Å. The atomic potentials tabulated in Kirkland (2010) also include a full three-dimensional potential permitting a direct numerical integration of the atomic potential in the transmission function [equation (29)] from z_n to $z_n + \Delta z$, for slice n . Each slice may be small enough to have more than one slice in a single atom although the slice must be large compared to the size of the nucleus because the atomic potentials usually do not properly include a non-zero nuclear size. This approach also has numerical problems. The atomic potential for single atoms is very narrow and sharp. The integrand must be sampled on a very fine scale to avoid missing a whole atom altogether, effectively setting a small maximum slice thickness. A small slice thickness with many samples of the potential in between is required. Needless to say, this procedure is very computer intensive, and not competitive for everyday use, but is an interesting test of accuracy. Using an analytically integrated potential as in the traditional multislice is a big advantage (in programming and computation time).

A single gold atom has a phase shift approaching unity near its center at 100 kV so might be suspect even at 100 kV. In general, the heavier the atoms in the specimen the larger the phase shift and less accurate the multislice might be. Fig. 10 compares a standard traditional 100 kV multislice calculation (dots) of an incident uniform plane wave propagating in a bulk gold crystal along the z or (001) direction. Gold is an f.c.c. structure with a unit-cell size of $c = 4.078$ Å. It is best to match the unit-cell size to be an integer number of slices to avoid producing artifacts so the slice thickness is set to $c/4 = 1.0195$ Å, which is about as small as a standard multislice calculation can go. A special version of multislice has been implemented to directly integrate the three-dimensional potential [equation C.19 of Kirkland (2010)] between z_n and $z_n + \Delta z$ using ten-point Gauss quadrature, permitting a much smaller slice thickness and greater accuracy. The results for a slice thickness of $c/40$ and $c/80$ are shown in Fig. 10 as a dashed and solid line, respectively. The results for $c/40$ and $c/80$ agree well so there is not likely to be any further

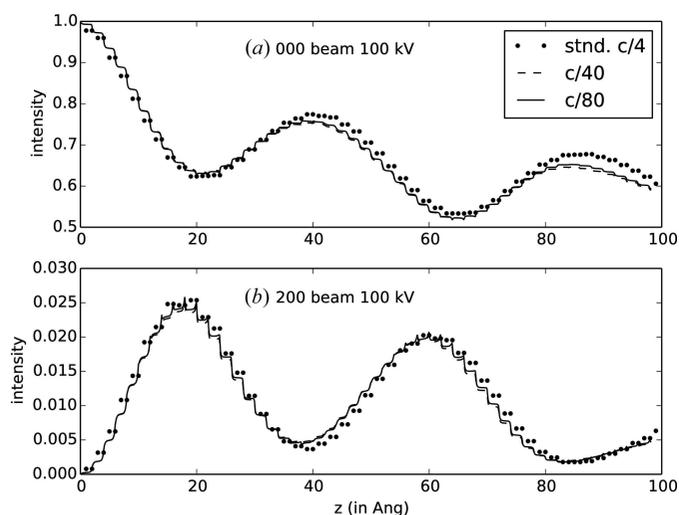


Figure 10
Calculated intensity in two directions in gold at 100 kV using a standard multislice calculation (stdn) and a special multislice calculation permitting arbitrarily small slices with increased accuracy to test accuracy.

improvement in accuracy by using a smaller slice thickness. All three calculations used 1024 by 1024 pixels and a supercell size of 7 by 7 by 25 unit cells of gold. Fig. 10(a) shows the intensity transmitted in the unscattered (000) direction and Fig. 10(b) shows the intensity scattered in the (200) direction. There is no significant difference between these three calculations, implying that this 100 kV multislice calculation is accurate except for ignoring back-scattered electrons (same approximation in all three calculations). Multislice calculations seem to be more accurate than might initially be expected which is good.

A similar calculation at 60 kV (not shown) also produced a small enough difference to be negligible. Fig. 11 shows a similar calculation (with 2048 by 2048 pixels) at a lower beam energy where multislice should become less accurate for heavy

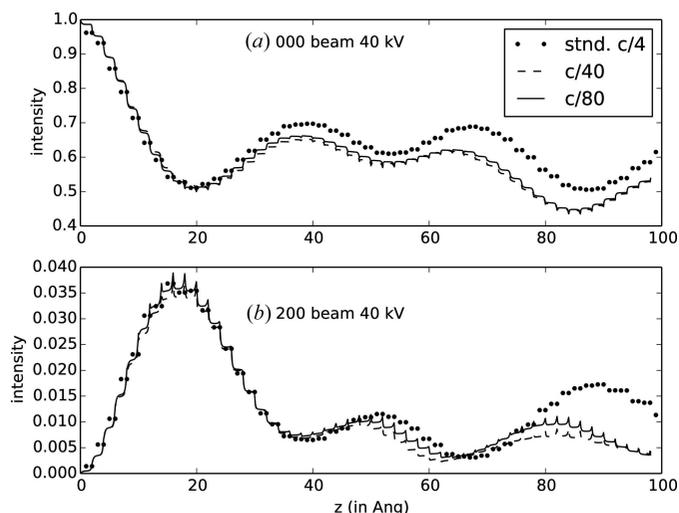


Figure 11
Calculated intensity in two directions in gold at 40 kV using a standard multislice calculation (stdn) and a special multislice calculation permitting arbitrarily small slices with increased accuracy to test accuracy.

atoms such as gold. There is a significant deviation as thickness increases. In general, multislice should get less accurate as the beam energy decreases (or the atomic number of the atoms in the specimen increases) as is evident here. Beam energies below about 50 to 60 kV for heavy atoms are probably not accurate using standard multislice, consistent with the results of Ming & Chen (2013). New instruments going to 20 kV (Kaiser *et al.*, 2011) or 15 kV (Sasaki *et al.*, 2014) will need significant improvements in the theory for theoretical calculations.

10. The inverse problem

It is useful to understand how a high-resolution TEM image (CTEM or STEM) is formed and to separate what information is related to the specimen and what artifacts may be produced from practical limitations of the instrument. Calculating an expected result from first principles, as discussed so far, can be labeled a forward problem. It might be even more useful to start from a real recorded image and work backwards to extract more detailed information about the specimen, which can be referred to as the inverse problem (or image restoration or deconvolution). This concept of the *forward* and *inverse* problem applies to many types of measured experimental data but will be limited to TEM images here. The general topic of image restoration has been reviewed by Andrews & Hunt (1977), Jansson (1997), Bertero & Boccacci (1998), Gonzalez & Woods (2008) and Gunturk & Li (2013). Other simple methods of image enhancement other than deconvolution are also discussed in Gonzalez & Woods (2008) and Jain (1989) for example.

There has been a large amount of computational effort to invert an X-ray or electron diffraction pattern to a specimen structure, for example Fienup (1982, 1987), Millane (1990), Miao *et al.* (1998, 2004, 2015), Elser (2003a,b), Hau-Riege *et al.* (2004), Wu *et al.* (2004), Marchesini (2007), Rodriguez *et al.* (2013). The Gerchberg–Saxton algorithm (Gerchberg & Saxton, 1971, 1973) reconstructs an image from an image and diffraction pattern, although it is experimentally difficult to record an electron micrograph and diffraction pattern from exactly the same area of the specimen. Ptychography (Rodenburg & Bates, 1992; Nellist & Rodenburg, 1998; Plamann & Rodenburg, 1998; D’Alfonso *et al.*, 2014) records the whole diffraction pattern at each position of the STEM probe with the goal of numerically reconstructing a super-resolution image of the specimen. This may require a rather large amount of data storage. These important topics are however outside the scope of this article.

The ambitious project of inverting the whole three-dimensional scattering (an inverse multislice or Bloch wave calculation) is somewhat difficult and has been considered by Gribelyuk (1991), Beeching & Spargo (1993, 1998), Spargo *et al.* (1994), Allen *et al.* (1998, 1999, 2000), Spence (1998), Spence *et al.* (1999), O’Leary & Allen (2005) and likely others. Although important, more work is needed on this topic and will not be considered in detail here. Instead the simpler problem of reconstructing only the exit wave (after passing

through the specimen) prior to the effects of the objective lens will be discussed.

Noise removal (denoising) is a related but different subject. Simple Fourier filtering may help if the noise and signal occupy different spatial frequencies. A low-pass filter may remove high-frequency noise and leave the low-frequency signal but also smoothes out edges which are frequently of interest. The median filter (Frieden, 1976) and the adaptive median filter (Hwang & Haddad, 1995; Chan *et al.*, 2005; Gonzalez & Woods, 2008) can remove so-called ‘salt-and-pepper’ or impulsive noise while leaving edges unchanged. Buades *et al.* (2005), Chatterjee & Milanfar (2010) and Gunturk & Li (2013) have recently reviewed various methods of noise removal. Buades *et al.* (2005) have introduced a non-local means (NLM) algorithm, averaging similar but non-adjacent patches in the image leading to the BM3D (block matching three-dimensional) method of Dabov *et al.* (2007). Wei & Yin (2010) have applied noise cleaning to low-dose cryo-EM images and Mevenkamp *et al.* (2015) have applied a periodic BM3D method to STEM images.

Usually the inverse problem is much harder than the forward problem. The inverse problem is said to be ill conditioned, frequently requiring some mathematical incongruity such as solving a singular matrix or dividing by zero. Complicating the issue, the starting data are almost always corrupted by some amount of noise. The signal-to-noise ratio, SNR, largely determines how well the solution to the inverse problem works. A high SNR of 1000 or more may permit significant improvement in image qualities such as resolution. A low SNR of ten or less usually only permits some small noise cleaning in most cases. Intermediate SNR may do something in between, which is more typical for TEM images. van Kempen *et al.* (1997) have given a quantitative comparison of several image deconvolution methods *versus* SNR for confocal microscopy.

10.1. CTEM exit-wave reconstruction

Many parameters change the recorded image in some way that may produce a different subset of information about the specimen. For example there is usually an optimal defocus that produces an image with the highest resolution that is easily interpreted. However, different defocus values may change the transfer function so that the image is not easily interpreted, but contains some small amount of information not in the best focus image. If the imaging parameters such as defocus, magnification, aberrations *etc.* are well characterized quantitatively then the information in several images can be combined into one image with higher resolution or other information, that is easily interpreted. Together the images in a defocus series may contain more information than a single image in the series. Reconstructing a single image from a defocus series may produce an improved image of the specimen. Exit-wave reconstruction has been reviewed by Kirkland & Meyer (2004) and Ophus & Ewalds (2012).

Defocus is the easiest parameter to vary but parameters other than defocus might be varied to achieve similar results.

Improved images have also been reconstructed from images recorded with different tilts (Kirkland *et al.*, 1995; Meyer *et al.*, 2002, 2004).

Schiske (1968) [and later translated to English (Schiske, 2002)] was the first to propose reconstructing an improved image of the specimen from a series of images recorded at different defocus (a defocus series). Frank (1972) extensively studied Schiske’s method. Each BF-CTEM image in the image is modeled as a linearized multicomponent image similar to the weak phase object approximation. The real and imaginary parts of the exit wave passing through the specimen are found free of aberrations with a least-squares fit to each Fourier component (pixel in the Fourier transform) *versus* defocus. The real portion of the exit wave (or imaginary portion of the image in the notation of these papers) results from the so-called anomalous scattering and should have a strong dependence on atomic number ($\propto Z^{1.3}$ instead of $\propto Z^{0.7}$ as in a typical BF-CTEM image) with the goal of discriminating heavy atoms (Kirkland & Siegel, 1981). In BF-CTEM the transfer function can have many zeros in the range of interest (oscillating as $\sin \chi$) and will always go to zero at high angle (spatial frequency) resulting in division by zero (singularity), making this approach impractical in most cases. Dividing noise by a small value near zero amplifies the noise, rendering the image unusable. This singularity problem was solved by adapting a multicomponent form of the Wiener filter or minimum mean-square error approach (Kirkland *et al.*, 1980). This linearized approach is still limited because it does not include the nonlinear properties of the recorded image [equation (9)]. The nonlinear nature of the image has been approximately solved using an adaptation of Newton’s method for finding square roots (Kirkland, 1982; Kirkland *et al.*, 1982) with a Wiener filter to deconvolve the remaining aberration function (called NLIP below). Although the nonlinear treatment is improved the partial coherence of the image is not accurate. Allen *et al.* (2004a) used a similar approach with approximate partial coherence and proposed a justification for this approximation (Allen *et al.*, 2004b). An improved method for nonlinear images using accurate partial coherence, called MIMAP, was developed by Kirkland (1984, 1988a) and Kirkland *et al.* (1985). Linear and nonlinear methods have been compared (Chang & Kirkland, 2006). Erni *et al.* (2010) have considered the effect of residual aberrations on exit-wave reconstruction using an aberration-corrected instrument.

The maximum *a posteriori* or MAP approach (Hunt, 1977; Trussell & Hunt, 1979; Trussell, 1980) for deconvolving (or restoring) a single image includes a nonlinear sensor or detector function which might be adapted to the square-law image detection process [intensity from wavefunction, equation (9)] of the CTEM. Recorded electron micrographs may have significant amounts of noise (like most images), which suggests treating the image signal using a statistical or probabilistic interpretation. If $f(\mathbf{x})$ is the ideal image and $g(\mathbf{x})$ is the actual recorded image that has been degraded by the instrument (CTEM) in some way, MAP maximizes the *a posteriori* conditional probability

of the original image given a measurement of the degraded image:

$$p(f|g) = \max. \quad (45)$$

This expression can be generalized to include a defocus series of m micrographs as (Kirkland, 1984; Kirkland *et al.*, 1985)

$$p(f|g_1)p(f|g_2) \cdots p(f|g_m) = \prod_{\mu=1}^m p(f|g_\mu) = \max \quad (46)$$

where $g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_m(\mathbf{x})$ represent the recorded image intensity distributions in the series. This approach has been labeled MIMAP for multiple input MAP (Kirkland, 1984).

If the electron image was collected by counting electrons in each pixel then the noise in the image might be Poisson distributed. However, if recorded with a CCD (or film in the past) the noise comes from many sources and it is more likely Gaussian in nature (from the central limit theorem) which is easier to manipulate mathematically. Assuming the noise in the image is Gaussian, the probability of the ideal image (the exit wave) is

$$p(f) = \frac{1}{\sigma_f(2\pi)^{1/2}} \exp \left\{ -\frac{1}{2A\sigma_f^2} \int_A [f(\mathbf{x}) - f_M(\mathbf{x})]^2 d^2\mathbf{x} \right\} \quad (47)$$

where A is the area of the image, σ_f is its standard deviation and $f_M(\mathbf{x})$ is its *a priori* mean. In general this *a priori* mean is unknown as well, but the results below will turn out not to be very sensitive to this exact value. The probability of the recorded image is

$$p(g) = \text{independent of } f; \quad (48)$$

however the conditional probability of the recorded image given the ideal image is

$$p(g|f) = \frac{1}{\sigma_n(2\pi)^{1/2}} \exp \left\{ -\frac{1}{2A\sigma_n^2} \int_A [g(\mathbf{x}) - g_{\text{theo}}(\mathbf{x})]^2 d^2\mathbf{x} \right\} \quad (49)$$

where σ_n is the standard deviation of the noise ($\sigma_f^2/\sigma_n^2 \sim S/N =$ signal-to-noise ratio) and $g_{\text{theo}}(\mathbf{x})$ is the theoretical model of the image calculated from $f(\mathbf{x})$ and the sensor or detector function:

$$g_{\text{theo}}(\mathbf{x}) = s[f(\mathbf{x}) \otimes h(\mathbf{x})] \quad (50)$$

$$= \text{FT}^{-1} \int T_{\text{cc}}(\mathbf{k}', \mathbf{k}' + \mathbf{k}) F^*(\mathbf{k}') F(\mathbf{k}' + \mathbf{k}) d^2\mathbf{k}' + c_0 \quad (51)$$

$$f(\mathbf{x}) = \text{FT}^{-1}[F(\mathbf{k})] = \psi_1(\mathbf{x}). \quad (52)$$

c_0 is a possible background constant or offset. Probabilities may be transformed with Bayes theorem as

$$p(f|g) = \frac{p(g|f)p(f)}{p(g)}. \quad (53)$$

Taking the logarithm of equation (46) and applying Bayes theorem produces

$$\ln \prod_{\mu=1}^m p(f|g_\mu) = \sum_{\mu} \ln p(f|g_\mu) = \max \quad (54)$$

$$\sum_{\mu} \ln p(g_\mu|f) + m \ln p(f) - \sum_{\mu} p(g_\mu) = \max. \quad (55)$$

With a Gaussian distribution:

$$\begin{aligned} & -\frac{1}{2A\sigma_n^2} \sum_{\mu=1}^m \int_A [g_\mu(\mathbf{x}) - g_{\mu, \text{theo}}(\mathbf{x})]^2 d^2\mathbf{x} \\ & -\frac{m}{2A\sigma_f^2} \int_A [f(\mathbf{x}) - f_M(\mathbf{x})]^2 d^2\mathbf{x} + [\text{independent of } f(\mathbf{x})] \\ & = \max. \end{aligned} \quad (56)$$

Simplifying

$$\begin{aligned} & \int_A \sum_{\mu=1}^m [g_\mu(\mathbf{x}) - g_{\mu, \text{theo}}(\mathbf{x})]^2 d^2\mathbf{x} \\ & + \eta m \int_A [f(\mathbf{x}) - f_M(\mathbf{x})]^2 d^2\mathbf{x} = \min \end{aligned} \quad (57)$$

where $\eta = \sigma_n^2/\sigma_f^2$ is the approximate noise-to-signal power ratio. Parseval's theorem allows this equation to be stated in Fourier space as

$$\begin{aligned} & \int_A \sum_{\mu=1}^m [G_\mu(\mathbf{k}) - G_{\mu, \text{theo}}(\mathbf{k})]^2 d^2\mathbf{k} \\ & + \eta m \int_A [F(\mathbf{k}) - F_M(\mathbf{k})]^2 d^2\mathbf{k} = \min. \end{aligned} \quad (58)$$

In practice, the images are digitized as discrete pixels. Restating this result for discrete pixel images:

$$J[\mathbf{F}] = \sum_{\mathbf{k}} \left(\sum_{\mu} |\xi_{\mu\mathbf{k}}|^2 + \eta m |F_{\mathbf{k}} - F_{M\mathbf{k}}|^2 \right) = \min \quad (59)$$

where

$$\xi_{\mu\mathbf{k}} = G_{\mu\mathbf{k}} - \sum_{\mathbf{k}'} T_{\mu}(\mathbf{k}', \mathbf{k}' + \mathbf{k}) F_{\mathbf{k}'}^* F_{\mathbf{k}'+\mathbf{k}} - N_x N_y c_0 \delta_{\mathbf{k}} \quad (60)$$

and $G_{\mu\mathbf{k}}$ is the value at one point in the Fourier transform of one image in a defocus series. N_x and N_y are the number of pixels in the image. The first term in equation (59) is just a familiar least-squares fitting function figure of merit. The second term containing η is a protection term to keep the ill-conditioned nature of this problem from causing the solution to diverge. In practice the *a priori* mean image is not known but can be approximated from the best focus image (as well as the starting image), and η can be used as a control factor. η is approximately the inverse of the SNR, and can be used as a free parameter and set as low a value as produces a result that is not excessively noisy or obviously far from the expected result. A variety of other stabilizing or regularization functions may be used (for example, Karayiannis & Venetsanopoulos, 1990; Reeves & Mersereau, 1990) except the popular maximum entropy form because $f(x)$ may be negative and positive.

To minimize $J[\mathbf{F}]$ first form its derivative:

$$\begin{aligned} \frac{\partial J}{\partial F_1^*} &= \sum_{\mu} \sum_{\mathbf{k}} \left(\xi_{\mu\mathbf{k}} \frac{\partial \xi_{\mu\mathbf{k}}^*}{\partial F_1^*} + \xi_{\mu\mathbf{k}}^* \frac{\partial \xi_{\mu\mathbf{k}}}{\partial F_1^*} \right) + \eta m(F_1 - F_M) \\ &= \sum_{\mu} \sum_{\mathbf{k}} \left[-T_{\mu}^*(\mathbf{l} - \mathbf{k}, \mathbf{l}) F_{1-\mathbf{k}} \xi_{\mu\mathbf{k}} \right. \\ &\quad \left. - T_{\mu}(\mathbf{l}, \mathbf{l} + \mathbf{k}) F_{1+\mathbf{k}} \xi_{\mu\mathbf{k}}^* \right] + \eta m(F_1 - F_M) \end{aligned} \quad (61)$$

using $\mathbf{l} = \mathbf{k}' + \mathbf{k}$ or $\mathbf{k}' = \mathbf{l} - \mathbf{k}$ in the first term and $\mathbf{l} = \mathbf{k}'$ in the second term. Using symmetry

$$T_{\mu}(\mathbf{l}, \mathbf{l} - \mathbf{k}) = T_{\mu}^*(\mathbf{l} - \mathbf{k}, \mathbf{l}) \quad (62)$$

and because ξ is the Fourier transform of a real function

$$\xi_{\mu, \mathbf{k}} = \xi_{\mu, -\mathbf{k}}^* \quad (63)$$

it can be shown that

$$\begin{aligned} \frac{\partial J}{\partial F_1^*} &= - \sum_{\mu} \sum_{\mathbf{k}} [2T_{\mu}(\mathbf{l}, \mathbf{l} - \mathbf{k}) F_{1-\mathbf{k}} \xi_{\mu\mathbf{k}}] \\ &\quad + \eta m(F_1 - F_M). \end{aligned} \quad (64)$$

To find the optimum ideal image $f(\mathbf{x})$ or equivalently $F(\mathbf{k})$ this function $J[\mathbf{F}]$ must be minimized. Kirkland (1984) and Kirkland *et al.* (1985) used the obvious approach of successive one-dimensional numerical searches in the direction opposite to the gradient [equation (64)], which is called the method of steepest descent. Steepest descent initially produces large improvements but then convergence slows or is nonexistent. The conjugate gradient method (Scales, 1985; Fletcher, 1987; Press *et al.*, 2007) searches along successive directions that are mutually conjugate which can dramatically improve convergence. In summary, given a search direction \mathbf{S} find the scalar parameter γ to minimize the figure of merit by performing a numerical line search:

$$J[\mathbf{F} + \gamma\mathbf{S}] = \min. \quad (65)$$

The initial search direction is the negative of the gradient. The conjugate gradient method uses the results of previous searches and a new gradient to calculate a new conjugate search direction without requiring excessive CPU time or memory to calculate the underlying Hessian matrix and is one of the few (or only) optimization methods appropriate for optimizing large systems. Although there may be thousands of variables (pixels in the image) convergence may be achieved in a few dozen iterations. There may also be some advantages to periodically restarting the search direction to be the negative gradient direction (particularly because the numerical searches may not be exact).

Early applications of MIMAP took a large amount of computer time on the small computers available at that time. However MIMAP is easily multithreaded on modern multiple-core processors and runs in reasonable amounts of computer time on current computers.

Coene *et al.* (1992) adapted MIMAP to use the special case of $\eta = 0$, approximations to partial coherence that may be achieved with a high-coherence FEG (field emission gun) source and a minimum search method similar to steepest descent. They have named it the MAL (maximum likelihood)

method although it is essentially the same as MIMAP. With $\eta = 0$ the estimator is actually termed a maximum likelihood estimator and should be called the MIML method. MAL has been tested successfully by Coene *et al.* (1996) and Thust *et al.* (1996). Although MAL is only appropriate for very small illumination angles it has been (incorrectly) applied to large illumination angles by other authors.

A real experimental defocus series will have measurement errors in defocus and alignment (δx_{μ} , δy_{μ}) between images in the series. The series should be recorded sequentially close in time on the same detector so should be close to aligned but there still may be a small drift during recording. The accuracy of the image parameters is important for an accurate reconstruction (Kirkland & Siegel, 1979). Denoting the current error in the parameter with a subscript e and the current value with a subscript j , the actual values are

$$\begin{aligned} \Delta f_{\mu} &= \Delta f_{\mu j} + \Delta f_{\mu j e} \\ \delta x_{\mu} &= \delta x_{\mu j} + \delta x_{\mu j e} \\ \delta y_{\mu} &= \delta y_{\mu j} + \delta y_{\mu j e}. \end{aligned} \quad (66)$$

The transmission cross coefficient [equation (39)] needs to be modified to include alignment:

$$T_{\mu} = T_{\mu}^{\text{coh}} T_{\mu}^{\text{PC}} T_{\mu}^{\text{A}} \quad (67)$$

where

$$T_{\mu}^{\text{A}} = \exp(2\pi i \mathbf{k} \cdot \delta \mathbf{x}_{\mu}). \quad (68)$$

The current estimate of the recorded image including small errors in the image parameters (in $T_{\mu j e}$) at iteration j becomes

$$G_{\mu\mathbf{k}} = \sum_{\mathbf{k}'} (T_{\mu j} + T_{\mu j e}) F_{\mathbf{k}'}^* F_{\mathbf{k}'+\mathbf{k}} + N_x N_y c_0 \delta_{\mathbf{k}}; \quad (69)$$

rearranging terms and Taylor expanding the error terms to lowest order in the errors gives

$$\begin{aligned} G_{\mu\mathbf{k}} &- \sum_{\mathbf{k}'} T_{\mu j} F_{\mathbf{k}'}^* F_{\mathbf{k}'+\mathbf{k}} - N_x N_y c_0 \delta_{\mathbf{k}} \\ &= \sum_{\mathbf{k}'} T_{\mu j e} F_{\mathbf{k}'}^* F_{\mathbf{k}'+\mathbf{k}} \\ &= \Delta f_{\mu j e} R_{\mu j \mathbf{k}} + \delta x_{\mu j e} Q_{\mu j \mathbf{k}} + \delta y_{\mu j e} P_{\mu j \mathbf{k}} \end{aligned} \quad (70)$$

where

$$\begin{aligned} R_{\mu j \mathbf{k}} &= \sum_{\mathbf{k}'} \frac{\partial T_{\mu j}}{\partial \Delta f_{\mu}} F_{\mathbf{k}'}^* F_{\mathbf{k}'+\mathbf{k}} \\ P_{\mu j \mathbf{k}} &= \sum_{\mathbf{k}'} \frac{\partial T_{\mu j}}{\partial \delta x_{\mu}} F_{\mathbf{k}'}^* F_{\mathbf{k}'+\mathbf{k}} \\ Q_{\mu j \mathbf{k}} &= \sum_{\mathbf{k}'} \frac{\partial T_{\mu j}}{\partial \delta y_{\mu}} F_{\mathbf{k}'}^* F_{\mathbf{k}'+\mathbf{k}}. \end{aligned} \quad (71)$$

If the difference between the recorded image and the current estimate is large compared to the noise ($\int |\xi_{\mu\mathbf{k}j}|^2 d^2k > \eta \int |G_{\mu\mathbf{k}}|^2 d^2k$) then the current errors in defocus and alignment parameters may be estimated by performing a least-squares fit over all relevant \mathbf{k} points to equation (70) (Kirk-

land, 1984, 1988a). Then update these parameters every few iterations as

$$\Delta f_{\mu}^{(j+1)} = 0.5[\Delta f_{\mu}^{(j)} + \Delta f_{\mu}^{(j-1)}] \quad (72)$$

$$\delta x_{\mu}^{(j+1)} = 0.5[\delta x_{\mu}^{(j)} + \delta x_{\mu}^{(j-1)}] \quad (73)$$

$$\delta y_{\mu}^{(j+1)} = 0.5[\delta y_{\mu}^{(j)} + \delta y_{\mu}^{(j-1)}]. \quad (74)$$

If the error is small then no further improvement is necessary. If there are N images in the defocus series then only $N - 1$ alignments and defocus values may be refined in this manner. The absolute defocus cannot be determined self-consistently. The remaining error in the optical parameters (hopefully small) remains with the final reconstructed exit wave.

The background constant c_0 can also be refined as

$$c_0^{(j+1)} = 0.5c_0^{(j-1)} + \frac{1}{2mN_xN_y} \times \sum_{\mu} \text{Real} \left(G_{\mu 0} - \sum_{k'} T_{\mu} F_{k'} F_{k'}^* \right). \quad (75)$$

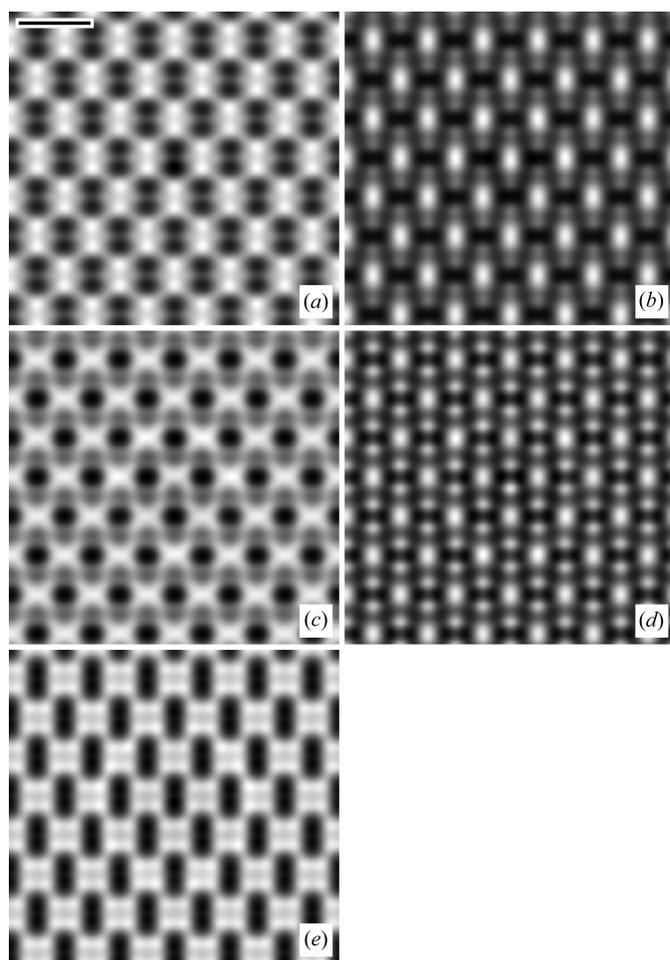


Figure 12 Calculated defocus series of a single Sb atom ($Z = 51$) on the exit surface of 110 silicon 50 Å thick (256 by 256 pixels). Beam energy 200 kV, $C_{S3} = 0.5$ mm, objective angle 20 mrad, condenser angle 0.1 mrad, defocus spread 50 Å, and defocus of (a) 434 Å, (b) 662 Å, (c) 830 Å, (d) 970 Å, (e) 1091 Å. Scale bar 5 Å.

See Kirkland (1984) for a complete derivation.

The MIMAP method has been successfully demonstrated on an experimental defocus series taken on the Kyoto 500 kV microscope (Kirkland, 1984, 1988a; Kirkland *et al.*, 1985). An example application to a calculated defocus series will be shown next.

Fig. 12 shows a calculated defocus series of thin silicon with a single heavy atom on the exit surface. The defocus values are chosen as $\Delta f = [(2n - 0.5)C_{S3}\lambda]^{1/2}$ ($n = 1, 2, \dots$) to produce wide flat bands at progressively higher angles inspired by early work on zone plates (Eisenhandler & Siegel, 1966a,b; Kirkland, 2010). The wide bands are also conveniently not very sensitive to instabilities. Typically four or five images with appropriate defocus values in a series are sufficient to reconstruct an exit wave (in principle two is sufficient). With more images the specimen may change or become damaged during recording.

Fig. 13 shows the linear approximation of the transfer function for the first and last image in the series. Higher defocus values produce information at higher resolution but are mixed up in some way. A careful reconstruction with accurate values for the image parameters including the aberrations can put all of this information back together in an interpretable form. Portions with a value in the transfer function greater than the noise ($\sim 1/\text{SNR}$) can be recovered and portions much less than the noise will be rejected. The NLIP method has several significant approximations but the math is simple enough to identify an effective transfer function of the reconstructed exit wave as shown in part (c). The best focus image would give a resolution of about 0.5 \AA^{-1} but the reconstructed image may have information to about 0.8 \AA^{-1} giving a modest improvement in resolution and, more importantly in some ways, separates the real and imaginary parts of the exit wave. An accurate nonlinear reconstruction may also recover the very low frequencies.

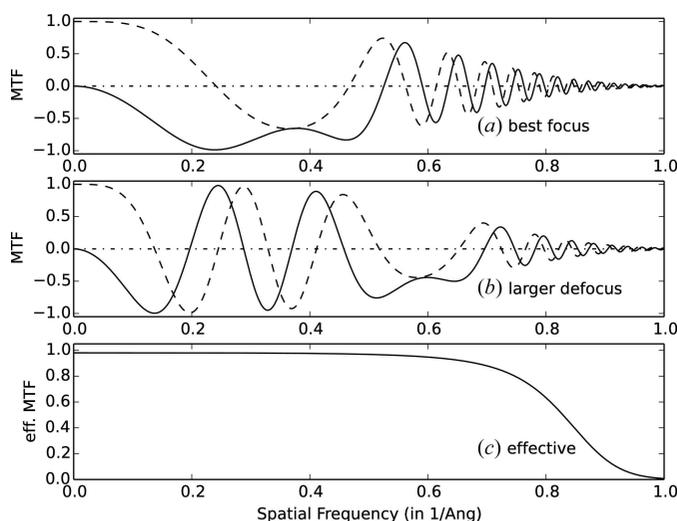


Figure 13 Transfer function for a beam energy of 200 kV, $C_{S3} = 0.5$ mm, condenser angle 0.1 mrad, defocus spread 50 Å, and defocus of (a) 434 Å, (b) 1091 Å. The solid line is $\sin(\chi)$ and the dashed line is $\cos(\chi)$. (c) The effective transfer function for the exit wave reconstructed from the defocus series of five images in Fig. 12 using the NLIP approximation.

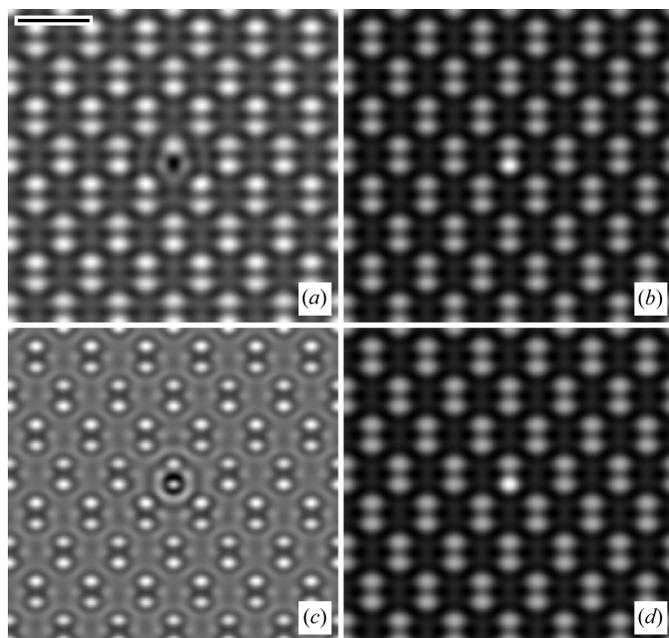


Figure 14
MIMAP exit wave reconstructed from defocus series in Fig. 12. (a), (b) Real and imaginary parts, and (c), (d) amplitude and phase. NLIP yields a visually indistinguishable result. Scale bar 5 Å.

Fig. 14 shows the exit wave reconstructed using the MIMAP method. It is shown as both a real and imaginary part and an amplitude and phase. There is a small improvement in resolution and the single heavy atom in the middle has been revealed.

For comparison the NLIP exit wave reconstructed with the nonlinear method in Kirkland (1982) and Kirkland *et al.* (1982) produced a result that is visually indistinguishable from Fig. 14, and is not shown to save space. This method uses Newton's method for finding square roots and a Wiener filter to deconvolve the aberration function (see Appendix A). Although this method treats the partial coherence in an approximate manner it gets nearly the same result, but has the advantage of being much faster (about a factor of 100× in this example corresponding to a fraction of a second on an inexpensive laptop computer) because it can take advantage of the FFT. An appealing approach is to use NLIP first for a few iterations to get close to the correct answer and then switch to MIMAP to refine the last few details.

10.2. ADF-STEM

The simple incoherent ADF-STEM model in §4.2 is only an approximation for simple thin specimens and is not very accurate for thick specimens. However, it is essentially the same linear image model as most incoherent light optical images. There is a large amount of work on deconvolving images described by this image model which can be immediately exploited for ADF-STEM in this approximation (Kirkland, 1988*b*; Nellist & Pennycook, 1998).

A simple ADF-STEM image model starts with equation (15) and adds experimental, random noise $n(\mathbf{x})$. Assuming

simple additive noise (other models possible) the recorded image is

$$g(\mathbf{x}) = f(\mathbf{x}) \otimes h_{\text{ADF}}(\mathbf{x}) + n(\mathbf{x}) \quad (76)$$

where $f(\mathbf{x})$ is the ideal image and $h_{\text{ADF}}(\mathbf{x})$ is the point-spread function calculated from the probe wavefunction [equations (18) and (19)]. The Fourier transform of this expression (lower-case variable names denote real-space and upper-case their Fourier transforms) is

$$G(\mathbf{k}) = H_{\text{ADF}}(\mathbf{k})F(\mathbf{k}) + N(\mathbf{k}) \quad (77)$$

where $H_{\text{ADF}}(\mathbf{k}) = \text{FT}[h_{\text{ADF}}(\mathbf{x})]$ is the transfer function of the image. It is tempting to simply divide the transform of the recorded image $G(\mathbf{k})$ by the calculated transfer function $H_{\text{ADF}}(\mathbf{k})$ (using measured aberrations *etc.*) to obtain the unaberrated ideal image $F(\mathbf{k})$. However, $H_{\text{ADF}}(\mathbf{k})$ inevitably becomes small or zero in regions where the noise $N(\mathbf{k})$ is still significant, resulting in a large increase in noise in the final image due to the ill-conditioned nature of this problem.

In the presence of random noise the best that can be done is to seek an optimal estimate of the ideal image $\hat{f}(\mathbf{x})$ in a statistical sense. The well known Wiener filter (for example, Gonzalez & Woods, 2008) tries to find an estimate of the ideal image to minimize:

$$\varepsilon = \left\langle \int |\hat{f}(\mathbf{x}) - f(\mathbf{x})|^2 d^2x \right\rangle = \min \quad (78)$$

where $\langle \dots \rangle$ represents a statistical average over an ensemble of possible measurements of the image. Parseval's theorem allows this integral to be transformed as

$$\varepsilon = \left\langle \int |\hat{F}(\mathbf{k}) - F(\mathbf{k})|^2 d^2k \right\rangle = \min. \quad (79)$$

The Wiener filter criterion is sometimes referred to as a minimum mean-squared error criterion. The optimal solution can be found as

$$\hat{F}(\mathbf{k}) = \frac{H_{\text{ADF}}^*(\mathbf{k})}{\eta + |H_{\text{ADF}}(\mathbf{k})|^2} G(\mathbf{k}) \quad (80)$$

where $\eta = \langle |N|^2 \rangle / \langle |F|^2 \rangle$ is the noise-to-signal power ratio (inverse of SNR). When $|H_{\text{ADF}}|^2 \gg \eta$ the result is approximately G/H_{ADF} which is the desired result. When $|H_{\text{ADF}}|^2 \ll \eta$ the result is approximately zero, because there is no real signal left. The Wiener filter tries to join these two regions in an optimal way. η is a small positive value that in practice can be adjusted by the user to control how much noise is tolerable in the resulting image. With a small SNR the Wiener filter is mostly a noise cleaning operation and with a high SNR it can deconvolve the point-spread function in some small amount.

The Wiener filter is a linear method. Most linear methods have a tendency to produce a negative undershoot (or ringing) at edges (or points) which are usually not physically correct. There are a variety of nonlinear methods that include some form of *a priori* constraint or information. One popular constraint is positivity which forces the resulting image to remain positive, appropriate for many types of images including ADF-STEM images formed by electron counting. It

is thought that positivity improves resolution and some authors report significant increases in resolution.

The Richardson–Lucy (RL) method (Richardson, 1972; Lucy, 1974) is a popular nonlinear restoration method including a positivity constraint, that is easy to implement. It was independently discovered by Shepp & Vardi (1982) in the context of tomography. RL has been successfully applied to images ranging from astronomy (Hanisch *et al.*, 1997) to confocal optical microscopy (van Kempen *et al.*, 1997). RL is formed from the maximum-likelihood estimator for a Poisson distributed signal (Hanisch *et al.*, 1997) resulting in a fixed-point iteration solution for a general point-spread function $h(\mathbf{x}) = h_{\text{ADF}}(\mathbf{x})$:

$$\hat{f}_{n+1}(\mathbf{x}) = \hat{f}_n(\mathbf{x}) \left\{ h(-\mathbf{x}) \otimes \left[\frac{g(\mathbf{x})}{\hat{g}_n(\mathbf{x})} \right] \right\} \quad (81)$$

$$\hat{g}_n(\mathbf{x}) = h(\mathbf{x}) \otimes \hat{f}_n(\mathbf{x}) = \text{FT}^{-1}[H\hat{F}]. \quad (82)$$

Typically the point-spread function is symmetrical $h(-\mathbf{x}) = h(\mathbf{x})$ although this may not be true with some aberrations such as coma, and there is an assumption that $\int h(\mathbf{x}) d^2x = 1$. Convolving the ratio with $h(-\mathbf{x})$ also serves to control the noise. If the original estimate of $f(\mathbf{x})$ and the point-spread function $h(\mathbf{x})$ are positive then the result is also positive. However, for a significant improvement over the Wiener filter the real zero level of the image (background level) must be known and the original image must have a fairly high SNR.

RL is typically controlled by limiting the number of iterations. If allowed to go too far it has a tendency to amplify the noise. The noise can be further controlled with a prefilter (van Kempen *et al.*, 1997) or with the damped-RL method (Hanisch *et al.*, 1997).

The maximum-entropy method (MEM) is another popular method that constrains the image to be positive, and has been considered by many authors [for example, Frieden (1972), Andrews & Hunt (1977), Wernecke & D’Addario (1977), Trussell (1980), Burch *et al.* (1983)]. MEM seeks to maximize an entropy-like function summed over values f_i in each pixel of the image,

$$E = - \sum_i f_i \ln f_i \quad (83)$$

subject to the constraint on the residual R :

$$R = \sum_i |g_i - (h \otimes f)_i|^2. \quad (84)$$

The total image intensity $F = \sum_i f_i$ is held constant. MEM is equivalent to minimizing the figure of merit:

$$W(f) = -E + \lambda R \quad (85)$$

$$= \sum_i f_i \ln f_i + \lambda \sum_i |g_i - (h \otimes f)_i|^2. \quad (86)$$

The entropy term E makes the image smooth (and forces positivity) while the residual term R proportional to λ forces it to agree with the original data (λ balances between the two). The entropy term can be recognized as one of many different forms of regularization that might be used. This expression can be solved using a direct conjugate gradient minimization

method, for example Wernecke & D’Addario (1977), Fletcher (1987) or Press *et al.* (2007). High values of λ produce higher resolution and probably more noise. In practice different values of the parameter λ should be tested to get a reasonable compromise between noise and resolution. Meinel (1988) has also given a recursive solution. MEM is harder to program and typically does not yield better results than RL.

An aberration corrector works by subtracting a large negative aberration from a large positive aberration (principally C_{53}) in hopes of eliminating this aberration. In practice many new multipole aberrations are added that must be corrected. This process is not perfect and there will be small residual aberrations remaining. One large aberration (C_{53}) has been traded for a large number of small aberrations with unknown values, which can be an improvement but is not perfect. It is misleading to say this is an aberration corrector; it is more accurate to describe it as an aberration reducer. There are 22 aberrations (Table 3) through fifth-order that need to be accurately measured and adjusted (not including focus and astigmatism which might be adjusted by the user). In practice all of these aberrations will have some small random measurement error that limits the probe size (Kirkland, 2011). In the microscope used here, the aberrations must be measured every hour or so in a process called ‘tuning’ the instrument. The fourth- and fifth-order aberrations are adjusted manually and do not change much over time. Lower-order aberrations are adjusted frequently using automated computer software. To deconvolve an ADF-STEM image, in principle, requires a value for all of the remaining small aberrations which are not known exactly by definition due to this random measurement error. There are a large number of small random aberrations which might tend to stochastically average to a reproducible probe shape in some approximation, which is the only hope of accurately including this effect until aberration correctors become exact (unlikely).

The traditional tuning criterion is to reduce each aberration until the maximum phase error (at the maximum objective aperture angle) is $\pi/4$ rad. A statistical probe shape may be estimated in a Monte Carlo style calculation by adding random tuning errors about this criterion. Treating just the nine second- and third-order aberrations with random values at 0.9 times the maximum allowed value and a standard deviation of 0.35 of this maximum value yields a mean probe shape and standard deviation shown in Fig. 15. The solid line is the mean and the dashed lines are the range due to the standard deviation. This calculation used an ensemble of 50 probes of 1024 by 1024 pixels. All curves are normalized to be unity at the origin. The main effect is a slightly larger probe with long tails consistent with observed images (mostly from the errors in the second- and third-order errors).

This probe calculation may be estimated in a simple manner with a more practical calculation as follows. The NION UltraSTEM 100 used here (Krivanek *et al.*, 1999, 2008, 2009), operated at a beam energy of 100 kV, has been observed to have a statistical measurement error of approximately 2 or 3 μm for the third-order aberrations and about 1 to 5 mm for the fifth-order aberration. The actual systematic errors are not

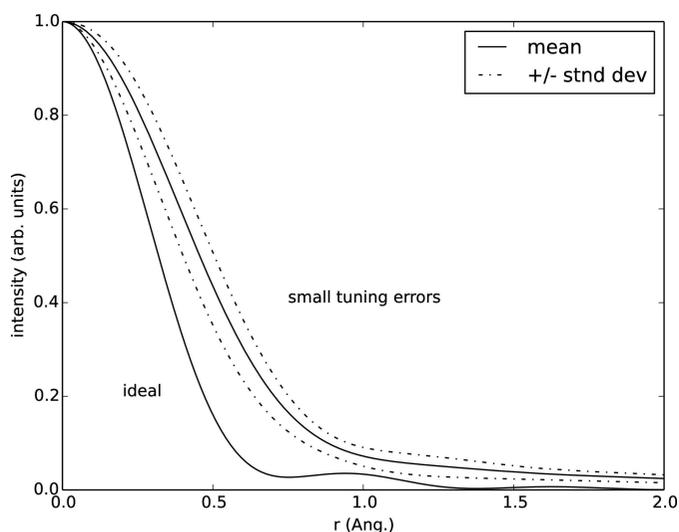


Figure 15

Calculated STEM probe, 100 kV, 30 mrad objective aperture and 50 Å defocus spread (from chromatic aberration), Gaussian source size 0.5 Å, $C_{45} = 0.2$ mm, $C_{56} = 10$ mm and with small random residual measurement errors (see text). The solid line is the mean shape and the dashed line is the range from its standard deviation (azimuthally averaged). Also shown is a probe with an ideal aberration corrector (all geometric aberrations zero).

known so will not be included. Without knowing the azimuthal orientation of all of the residual aberrations it is probably safer to include only symmetrical ($m = 0$) aberrations. Therefore the C_{2m} and C_{4m} aberrations will be ignored and only C_{S3} and C_{S5} will be included. All of the residual aberrations will be approximated as being part of C_{S3} and C_{S5} for simplicity. There are five C_{3m} terms which will be added in quadrature and approximated as $C_{30} = C_{S3} = 5^{1/2}(3 \mu\text{m}) = 7 \mu\text{m}$ and seven C_{5m} terms which will similarly be approximated as $C_{50} = C_{S5} = 7^{1/2}(5 \text{ mm}) = 13 \text{ mm}$. (Remember that these are place holders for a larger collection of small aberrations and not real aberrations that could be corrected.) There are many effects being ignored for simplicity, some positive and some negative which may tend to cancel out. (For example, C_{10} and $C_{12a,b}$ might compensate C_{30} and $C_{32a,b}$ but all of C_{2m} and C_{4m} are being ignored.) The objective lens starts out with C_{S3} of about 1 mm which is reduced to about 7 μm or about a 0.7% error. The corrector is really doing a good job, it's just not perfect. The resulting probe is very close to the mean probe in Fig. 15.

Fig. 16(a) shows an experimentally recorded ADF-STEM image of a gold particle on an amorphous carbon support using a NION UltraSTEM 100 at 100 kV with 512 by 512 pixels, a dwell time of 16 μs per pixel and objective aperture of 30 mrad. The results of deconvolution using a defocus spread of 50 Å, a source size of 0.5 Å and an approximation to the remaining small measurement errors in the aberration as discussed above are shown in (b), (c). The Wiener filter (b) used $\text{SNR} = 20$ and RL (c) had 40 iterations. RL also has a small low-pass prefilter limiting it to about twice the objective aperture with a Gaussian filter (included in the transfer function). The Wiener filter produces some small improvement in contrast and RL produces a little more contrast.

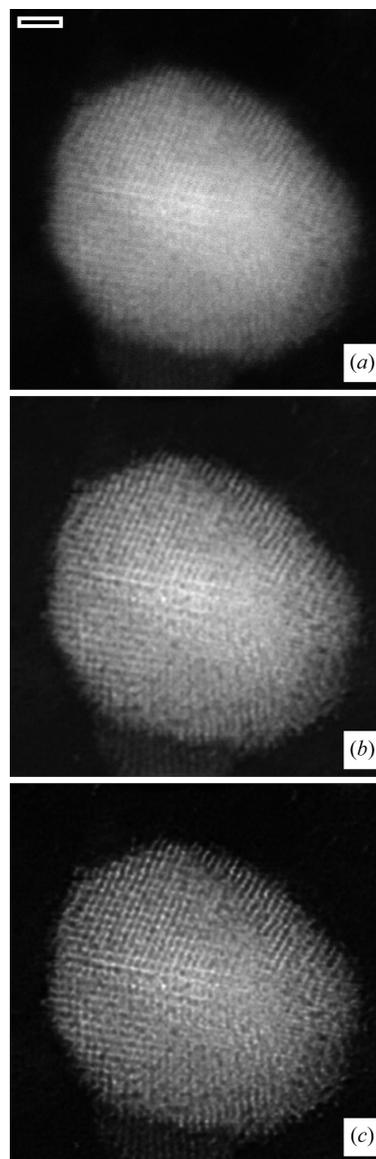


Figure 16

Experimental ADF-STEM image of a gold particle on an amorphous carbon support, recorded on an aberration-corrected NION UltraSTEM100 at 100 kV. Atoms should be white. (a) Original, (b) Wiener filter, (c) Richardson-Lucy (scale bar is 10 Å).

Deconvolution offers some small improvement if carefully applied. This approach is somewhat less expensive than buying a new microscope so might have some small value in practice.

11. Software and implementations

There have been many versions of software implementing various methods of electron microscope image simulations and analysis. Some of these are listed in Table 4. Smith & Carragher (2008) have given a recent review of software developments for analysis of biological electron microscope images. An online listing is maintained at Wikipedia (see https://en.wikibooks.org/wiki/Software_Tools_For_Molecular_Microscopy).

Table 4

Some available image simulation and analysis software packages for electron microscopy.

Method M is multislice, B is Bloch wave, IP is image processing, BIP is biological image processing, DA is diffraction analysis. Some may be commercial and other private programs likely exist. Some older programs may no longer be available. Apologies to the authors of many other programs that may have been left out.

Program	Author(s)	Method
<i>SHRLI</i>	O'Keefe & Buseck (1979)	M
<i>TEMPAS</i>	Kilaas (1987)	M
<i>EMS</i>	Stadelmann (1987)	B
<i>NCEMSS</i>	O'Keefe & Kilaas (1988)	M
<i>MacTEMPAS</i>	Kilaas (2006)	M
<i>MULTIS</i>	Spence & Zuo (1992)	M
<i>TEMSIM</i>	Kirkland (1998)	M
?	Ishizuka (2006)	B, M
?	DeGraf (2003)	B
<i>JEMS</i>	Stadelmann (2004)	B, M
<i>WebEMAPS</i>	Spence & Zuo (1992), Zuo (2009)	B
<i>EDM</i>	Marks & Kilass (2006)	B, M
<i>SimulaTEM</i>	Gómez-Rodríguez <i>et al.</i> (2010)	M
<i>TEM Simulator</i>	Rullgård <i>et al.</i> (2011)	M
<i>MEGACELL</i>	Stroppa <i>et al.</i> (2011)	
<i>STEM CELL</i>	Grillo & Rotunno (2013)	M
	Grillo & Rossi (2013)	
<i>STEMSIM</i>	Krause <i>et al.</i> (2013)	M
<i>computem</i>	Kirkland (2013)	M
<i>QSTEM</i>	Koch (2015)	M
<i>MULTEM</i>	Lobato & Van Dyck (2015 <i>b</i>)	M
μ <i>STEM</i>	Allen <i>et al.</i> (2015)	B, M
<i>FDES</i>	van den Broek <i>et al.</i> (2015)	M
<i>IMPROC</i>	Saxton (1978)	IP
<i>SEMPER</i>	Saxton <i>et al.</i> (1979)	IP
	Saxton (1996)	
<i>SPIDER</i>	Frank <i>et al.</i> (1981)	BIP
<i>SPIDER/WEB</i>	Frank <i>et al.</i> (1996)	BIP
<i>EM</i>	Hegerl & Altbauer (1982)	IP
	Hegerl (1996)	
<i>CRISP</i>	Hovmöller (1992)	DA
<i>IMAGIC</i>	van Heel & Keegstra (1981)	BIP
	van Heel <i>et al.</i> (1996)	IP
<i>Xmipp</i>	Marabini <i>et al.</i> (1996)	BIP
<i>EMAN</i>	Ludtke <i>et al.</i> (1999)	BIP
<i>EDM</i>	Kilaas <i>et al.</i> (2005)	DA
<i>EMAN2</i>	Tang <i>et al.</i> (2007)	BIP
<i>SPIRE</i>	Baxter <i>et al.</i> (2007)	BIP
<i>imageJ</i>	Schneider <i>et al.</i> (2012)	IP

11.1. Fast Fourier transforms

At the heart of the multislice algorithm is the fast Fourier transform or FFT. There are a great many well developed FFT subroutines available. Each version may have some advantage for a different size FFT or for a specific type of computer hardware. In the past there were many arguments about whose FFT was faster, with conflicting results because each party was running on a different computer. Current computers are more memory bound than compute bound for large multidimensional FFTs. The memory system may be slower than the floating-point calculation itself because memory access may have to go to the outside bus. In a multi-dimensional FFT the order in which the data are stored in memory may be as important as the number of actual floating-point operations. The current leader in speed is *FFTW* (the fastest Fourier transform in the west) distributed by Frigo &

Johnson (2005). Internally it contains a large number of different styles of FFTs. At run time it tests some subset of these on the specific type and size of FFT being run and selects the fastest one for later use. There is a significant CPU time required for this test so it does not help if the FFT is run only a few times, but if the successive FFT is run a great many times then there is a net benefit. This is a good fit for multislice as the same size FFT is run a great many times and results in an overall improvement of about a factor of $2\times$ or $3\times$ in speed (including the measure overhead). The code used here uses *FFTW*.

11.2. Parallel computation

In the beginning of the integrated circuit (IC) era Moore observed that the number of transistors on an integrated circuit was doubling about every two years [this article has been lost and reprinted (Moore, 1998)]. This rate of growth has persisted till this day and has become known as Moore's law. Interestingly, the year 2015 will be the 50th anniversary of Moore's law. For many years the number of transistors on an IC was increased primarily by making them smaller. However the size of today's transistors is approaching the spacing between atoms and requires TEM to diagnose transistor manufacturing processes. Transistor speed and hence computer speed approximately scale with size, so the speed of individual CPUs was steadily increasing for many years and now is only slowly increasing and may be leveling off in some respects. Moore's law has taken the form of an increasingly larger IC with increasingly more transistors and CPUs on a single IC. Computation is in the middle of a transition to massively parallel processing from the historic calculation in a (von Neumann) single thread of instructions on a single processor to calculating many things at the same time on multiple CPUs (multithreading), which may require significant rethinking and reorganization of many cherished algorithms. A skilled programmer must also partition the program into multiple independent tasks that can be performed independently at the same time and then synchronize these tasks with a single result at the end.

In recent years new software and hardware tools for parallel computing have become standard across multiple computing platforms and are easier to use than previous proprietary methods, for example Quinn (2004). Most computers today have two or more CPUs in a shared memory processor (SMP) architecture, in which different CPUs communicate through shared memory on the same circuit board (motherboard). Communication between CPUs is relatively fast but this architecture is limited to a small number of CPUs (about four to eight). The *openMP* software model is intended for use with SMP and the software used here uses *openMP*. Alternatively, a distributed memory or cluster architecture is made by connecting a large number of independent computers through a fast network. In principle there is no limit to the maximum number of CPUs and memory that may be connected together; however communication between CPUs is usually much slower than in an SMP architecture. A cluster requires more

expertise to set up and use and is less common, but may have more capability. The *openMPI* (message passing interface) software standard is intended for a cluster. Robertson *et al.* (2006), Carlino *et al.* (2008), Grillo & Rotunno (2013) and Grillo & Rossi (2013) have described using a cluster in electron-microscope image simulations. In both styles of multithreading it is challenging to produce code that scales with the number of CPUs (*i.e.* N CPUs run N times faster). The overall code only runs as fast as the slowest thread and there can be a significant CPU overhead to manage each thread. The best results usually come from a coarse-grain program in which each thread (CPU) does a large amount of computation before being reunited with the other threads. An easy example is to perform a STEM calculation in which each CPU or thread calculates the signal from each different position of the scanned probe at the same time (used in current software). Each thread is independent and does a lot of computation by itself.

Computer video gaming is currently popular and supports a large industry to supply video graphics display cards (circuit boards). Generating complex two-dimensional images with millions of pixels representing realistic three-dimensional scenes in real time requires significant numerical processing. Different parts of the image can be calculated independently using separate computing units in parallel. Video cards have evolved into graphical display processors (GPUs) with a thousand or more simple numeric processors (usually slower than the CPU in the host). Some manufacturers have recognized the advantages of these GPUs for scientific calculations and provide software tools to program the GPU directly as well as numerical subroutine libraries (including multidimensional FFTs). GPUs can be difficult to program but the prospect of thousands of processors at very low cost is definitely intriguing, and have been adapted to electron microscopy image simulation by Dwyer (2010), Lobato & Van Dyck (2015*b*), diffraction simulation (Eggeman *et al.*, 2013) and Bloch wave calculations (Pennington *et al.*, 2014). The GPU program of van den Broek *et al.* (2015) also proposes an improved method of calculating the atomic potentials labeled the hybrid method that may be faster as compared to the 'real-space' potential calculation (as used here). However, their implementation of the real-space approach does not use a look-up table (slower) so it is not clear how much improvement in speed there really is. Generally speaking, a calculation on the graphics card is relatively fast (a factor of 20 or so faster than the host computer) but moving the data between the card and the host computer is rela-

tively slow. Simply moving the data to the card to perform an FFT and back to the host only gives a modest improvement in speed (about a factor of 4 or so). (Execution speeds may vary dramatically with different hardware; times quoted are for simplicity of discussion only.) To get the full benefit of a GPU card requires keeping the data on the graphics card and doing most of the calculations there. Many cards only have a small amount of memory so rewriting existing programs to use a GPU effectively may not be easy. Programming a GPU is a lot like programming an array processor of several decades ago (Rez, 1985) for those of us who can remember such things. Loane *et al.* (1988) also used a large array processor although only the name of the computer facility was quoted and the type of hardware was not described. Array processors have been almost completely forgotten now. Hopefully GPUs will last longer.

11.3. The user interface

A simple menu-driven graphical user interface (GUI) can make the software easier to use for both a computer novice and a computer expert. It is already difficult enough to program the numerical portion of any significant numerical calculation but there is about the same amount of effort to add a simple GUI to make the program accessible to more people. Each platform (operating system) has a different application programming interface (API) so in principle would require a huge programming effort to port the numerical code plus GUI

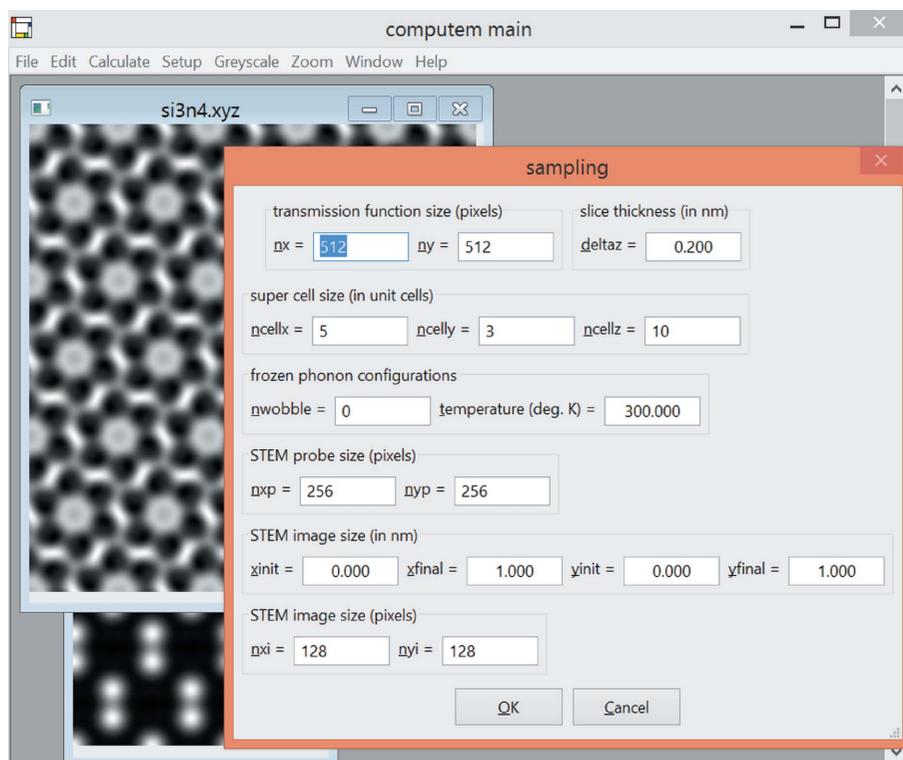


Figure 17

Screen shot of the multislice simulation program *computem* using the cross-platform GUI software package *wxWidgets*.

to several different platforms in common use. There are several dozen well developed cross-platform GUI packages that allow the programmer to code the GUI only once and simply recompile it on several different computer platforms (or operating system). *wxWidgets* (Smart *et al.*, 2006) and *QT* (Blanchette & Summerfield, 2008) are two popular examples. The *FLTK* (Fast and Light Tool Kit) package has much fewer features but is notable for being easier to use.

An example of a portion of the user interface is shown in Fig. 17 using the *wxWidgets* package. This example is on the Microsoft Windows platform. It can be recompiled on several other different platforms. The look and feel are similar but vary a little between platforms to conform to normal usage on each different platform.

12. Future directions

Computer simulation continues to be useful in understanding image artifacts at the limit of resolution and to investigate new instrument configurations before investing large amounts of effort and money to build them.

Computer technology is vigorously moving in the direction of parallel processing on many CPUs at the same time (multithreading). Many new algorithms and methods have been developed to utilize multithreading hardware and more improved methods of simulation in electron microscopy need to be developed further to keep up with the available computers. It might also be possible with faster computers to move simulation from an offline activity separated from the microscope to a more interactive calculation while the specimen is being observed in the microscope.

Aberration correctors have been developed to correct both the geometric aberrations (spherical aberration) and chromatic aberrations (defocus spread *etc.*). Corrector technology works well enough to form high-resolution images at very low beam voltages (20 kV in some cases). Both the multislice and Bloch wave approach have significant approximations suitable for high beam voltage but neither is very accurate at the low voltages now possible. Theory needs to be improved in some way to accurately calculate these low-voltage images.

Most image simulations are based on a simple superposition of neutral atoms. However, many solids are ionic in nature. Electron scattering from ions contains some divergent results and might be expected to significantly change the image results. However, simply superposing isolated ions does not agree with experiment or a more rigorous calculation from the results of a sophisticated electronic structure calculation (Gemming *et al.*, 1998) which conveniently is very similar to a superposition of neutral atoms perhaps due in part to cooperative screening effects. Some work on integrating pseudo-potentials with multislice has also been done (Kambe & Stampfl, 1994). This is a somewhat harder calculation but more work might be helpful in integrating current electronic structure modeling in condensed-matter physics with TEM image simulations (Meyer *et al.*, 2011). Simply superposing neutral atoms seems to work better than it should at first glance.

Electron energy-loss spectroscopy (EELS) and other forms of analytical microscopy are very useful for analyzing the chemical structure of the specimen. There is a large amount of computation involved in calculating and analyzing the resulting spectra that has not been discussed here. More work is needed in combining electronic structure calculations and image simulation. Dwyer (2005) and Kirkland (2005) have discussed combining multislice and EELS.

APPENDIX A The NLIP method

The NLIP (nonlinear image processing) method (Kirkland, 1982; Kirkland *et al.*, 1982) uses an approximate form of partial coherence that may not be very accurate (or equivalently should only be used for nearly coherent images), but yields simpler math that is easier to work with. It uses an adaptation of Newton's iterative method for finding square roots to solve for the complex exit wave in BF-CTEM. Representing a defocus series of m images similar to the phase-grating approximation as

$$g_{\mu}(\mathbf{x}) = |f(\mathbf{x}) \otimes h_{\mu}(\mathbf{x})|^2 + n_{\mu}(\mathbf{x}) \quad (87)$$

where $\mu = 1, 2, \dots, m$, the transfer function is taken from the weak phase object partial coherence model (not justified in all cases):

$$\begin{aligned} H_{\mu}(k) = & \frac{1}{(1 + \varepsilon k^2)^{1/2}} \\ & \times \exp \left\{ \frac{-i\pi\lambda k^2}{1 + \varepsilon k^2} \left[\frac{1}{3} C_{SS}(1 - 2\varepsilon k^2)\lambda^4 k^4 \right. \right. \\ & \left. \left. + 0.5 C_{S3}(1 - \varepsilon k^2)\lambda^2 k^2 - \Delta f_{\mu} \right] \right\} \\ & \times \exp \left(- \left\{ [\pi\lambda k_s k (C_{SS}\lambda^4 k^4 + C_{S3}\lambda^2 k^2 - \Delta f_{\mu})]^2 \right. \right. \\ & \left. \left. + 0.25(\pi\lambda\Delta_0 k^2)^2 \right\} / (1 + \varepsilon k^2) \right) \quad (88) \end{aligned}$$

where Δ_0 is the spread in defocus values, λk_s is approximately the condenser illumination angle and $\varepsilon = (\pi\lambda k_s \Delta_0)^2$.

The initial starting image is estimated from the best focus image $g_1(\mathbf{x})$ for iteration $j = 0$ as

$$f_0(\mathbf{x}) = [g_1(\mathbf{x})]^{1/2} + i\{[g_1(\mathbf{x})]^{1/2} - (\max g_1)^{1/2}\}. \quad (89)$$

Next estimate the linear portion of the defocus series from this estimate:

$$a_{\mu j}(\mathbf{x}) = f_j(\mathbf{x}) \otimes h_{\mu}(\mathbf{x}) = \text{FT}^{-1}[F_j(\mathbf{k})H_{\mu}(\mathbf{k})] \quad (90)$$

for all images in the series ($\mu = 1, 2, \dots, m$). Now form an improved estimate of the linear defocus series from the original images using Newton's method of finding square roots with a small protection term (inspired by the minimum mean-square error of the Wiener filter approach) to avoid dividing by zero:

$$a_{\mu_{j+1/2}}(\mathbf{x}) = \frac{1}{2} \left[a_{\mu_j}(\mathbf{x}) + \frac{a_{\mu_j}(\mathbf{x})g_{\mu}(\mathbf{x})}{\eta_0 + |a_{\mu_j}(\mathbf{x})|^2} \right] \quad (91)$$

where η_0 is a small positive number approximately the inverse of SNR. In the limit $\eta_0 \rightarrow 0$ this becomes Newton's method. This step assumes that any extra background constant has been removed (a value of zero really means zero electron intensity). Finally form a revised estimate of the ideal image using a multicomponent Wiener filter (Kirkland & Siegel, 1979; Schiske, 1973):

$$F_{j+1} = \frac{\sum_{\mu} H_{\mu}^*(\mathbf{k})A_{\mu_{j+1/2}}(\mathbf{k})}{\eta_1 + \sum_{\mu} |H_{\mu}(\mathbf{k})|^2} \quad (92)$$

where η_1 is another small positive number similar to η_0 (may be the same in some cases). Repeat the above steps until the final image changes by less than some small specified amount consistent with the observed noise in the image.

Convergence may be judged using a figure of merit:

$$\varepsilon_{\text{rms}}^2 = \frac{\sum_{\mu=1}^m \int | |f_j(\mathbf{x}) \otimes h_{\mu}(\mathbf{x})|^2 - g_{\mu}(\mathbf{x})|^2 d^2\mathbf{x}}{\int |g_{\mu}(\mathbf{x})|^2 d^2\mathbf{x}}. \quad (93)$$

The effective transfer function, as estimated from equation (92), is approximately

$$H_{\text{eff}} \sim \frac{\sum_{\mu} |H_{\mu}(\mathbf{k})|^2}{\eta_1 + \sum_{\mu} |H_{\mu}(\mathbf{k})|^2}. \quad (94)$$

If the initial estimate is reasonably close to the correct result, then Newton's method converges very fast. Most of the calculation uses FFTs which also are a fast calculation, resulting in a very fast method with an approximate treatment of partial coherence.

Acknowledgements

A small portion (the NION UltraSTEM 100 images, Fig. 16) of this work made use of the Cornell Center for Materials Research Facilities supported by the National Science Foundation under award No. DMR-1120296.

References

- Allen, L. J., D'Alfonso, A. J. & Findlay, S. D. (2015). *Ultramicroscopy*, **151**, 11–22.
- Allen, L. J., Faulkner, H. M. L. & Leeb, H. (2000). *Acta Cryst.* **A56**, 119–126.
- Allen, L. J., Josefsson, T. W. & Leeb, H. (1998). *Acta Cryst.* **A54**, 388–398.
- Allen, L. J., Leeb, H. & Spargo, A. E. C. (1999). *Acta Cryst.* **A55**, 105–111.
- Allen, L. J., McBride, W., O'Leary, N. L. & Oxley, M. P. (2004a). *Ultramicroscopy*, **100**, 91–104.
- Allen, L. J., McBride, W., O'Leary, N. L. & Oxley, M. P. (2004b). *J. Microsc.* **216**, 70–75.
- Andrews, H. C. & Hunt, B. R. (1977). *Digital Image Restoration*. New Jersey: Prentice Hall.
- Barthel, J. & Thust, A. (2013). *Ultramicroscopy*, **134**, 6–17.
- Baxter, W. T., Leith, A. & Frank, J. (2007). *J. Struct. Biol.* **157**, 56–63.
- Beeching, M. J. & Spargo, A. E. C. (1993). *Ultramicroscopy*, **52**, 243–247.
- Beeching, M. J. & Spargo, A. E. C. (1998). *J. Microsc.* **190**, 262–266.
- Bell, D. C. & Erdman, N. (2013). Editors. *Low Voltage Electron Microscopy, Principles and Applications*. Chichester: Wiley.
- Bertero, M. & Boccacci, P. (1998). *Introduction to Inverse Problems in Imaging*. London: Institute of Physics Publishing.
- Bethe, H. (1928). *Ann. Phys.* **87**, 55–129.
- Black, G. & Linfoot, E. H. (1957). *Proc. R. Soc. London Ser. A*, **239**, 522–540.
- Blanchette, J. & Summerfield, M. (2008). *C++ GUI Programming with Qt 4*, 2nd ed. Westford: Prentice Hall. <http://www.qt.io>.
- Boothroyd, C. B. (1998). *J. Microsc.* **190**, 99–108.
- Born, M. & Wolf, E. (1980). *Principles of Optics*, 6th ed. Oxford: Pergamon Press.
- Bosch, E. G. T. & Lazić, I. (2015). *Ultramicroscopy*, **156**, 59–72.
- Brigham, E. O. (1988). *The Fast Fourier Transform and its Applications*. Upper Saddle River, New Jersey: Prentice-Hall.
- Broek, W. van den, Jiang, X. & Koch, C. T. (2015). *Ultramicroscopy*, **158**, 89–97.
- Buades, A., Coll, B. & Morel, J. M. (2005). *Multiscale Model. Simul.* **4**, 490–530.
- Burch, S. F., Gull, S. F. & Skilling, J. (1983). *Computer Vision Graphics and Image Processing*, **23**, 113–128.
- Bursill, L. A. & Wilson, A. R. (1977). *Acta Cryst.* **A33**, 672–676.
- Buxton, B. F., Loveluck, J. E. & Steeds, J. W. (1978). *Philos. Mag. A*, **38**, 259–278.
- Cai, C. & Chen, J. (2012). *Micron*, **43**, 374–379.
- Carlino, E., Grillo, V. & Palazzari, P. (2008). *Microscopy of Semiconducting Materials 2007*, Vol. 120, *Springer Proceedings in Physics*, edited by A. G. Cullis & P. A. Midgley, pp. 177–180. Dordrecht: Springer.
- Chan, R. H., Ho, C.-W. & Nikolova, M. (2005). *IEEE Trans. Image Process.* **14**, 1479–1485.
- Chang, L. Y. & Kirkland, A. I. (2006). *Microsc. Microanal.* **12**, 469–475.
- Chang, L. Y., Kirkland, A. I. & Titchmarsh, J. M. (2006). *Ultramicroscopy*, **106**, 301–306.
- Chatterjee, P. & Milanfar, P. (2010). *IEEE Trans. Image Process.* **19**, 895–911.
- Chen, J. H. & Van Dyck, D. (1997). *Ultramicroscopy*, **70**, 29–44.
- Chen, J. H., Van Dyck, D., Op de Beck, M. & van Landuyt, J. (1997). *Ultramicroscopy*, **69**, 219–240.
- Coene, W. & Van Dyck, D. (1984a). *Ultramicroscopy*, **15**, 41–50.
- Coene, W. & Van Dyck, D. (1984b). *Ultramicroscopy*, **15**, 287–300.
- Coene, W., Janssen, G., Op de Beck, M. & Van Dyck, D. (1992). *Phys. Rev. Lett.* **69**, 3743–3746.
- Coene, W. M. J., Thust, A., Op de Beck, M. & Van Dyck, D. (1996). *Ultramicroscopy*, **64**, 109–135.
- Cooley, J. W. & Tukey, J. W. (1965). *Math. Comput.* **19**, 297–301.
- Cowley, J. M. (1969). *Appl. Phys. Lett.* **15**, 58–59.
- Cowley, J. M. (1976). *Ultramicroscopy*, **2**, 3–16.
- Cowley, J. M. & Moodie, A. F. (1957). *Acta Cryst.* **10**, 609–619.
- Crewe, A. V. & Salzman, D. B. (1982). *Ultramicroscopy*, **9**, 373–378.
- Dabov, K., Foi, A., Katkovnik, V. & Egiazarian, K. (2007). *IEEE Trans. Image Process.* **16**, 2080–2095.
- D'Alfonso, A. J., Morgan, A. J., Yan, A. W. C., Wang, P., Sawada, H., Kirkland, A. I. & Allen, L. J. (2014). *Phys. Rev. B*, **89**, 064101.
- DeGraf, M. (2003). *Introduction to Conventional Transmission Electron Microscopy*. Cambridge University Press.
- Dulong, B. J., Haynes, R. D. & Robertson, M. D. (2008). *Ultramicroscopy*, **108**, 415–425.
- Dwyer, C. (2005). *Ultramicroscopy*, **104**, 141–151.
- Dwyer, C. (2010). *Ultramicroscopy*, **110**, 195–198.
- Dwyer, C., Erni, R. & Etheridge, J. (2010). *Ultramicroscopy*, **110**, 952–957.
- Edington, J. W. (1976). *Practical Electron Microscopy in Materials Science*. New York: Van Nostrand Reinhold.
- Eggeman, A. S., London, A. & Midgley, P. A. (2013). *Ultramicroscopy*, **134**, 44–47.

- Eisenhandler, C. B. & Siegel, B. M. (1966a). *J. Appl. Phys.* **37**, 1613–1620.
- Eisenhandler, C. B. & Siegel, B. M. (1966b). *Appl. Phys. Lett.* **8**, 258–260.
- Elser, V. (2003a). *Acta Cryst.* **A59**, 201–209.
- Elser, V. (2003b). *J. Opt. Soc. Am. A*, **20**, 40–55.
- Ercius, P., Boese, M., Duden, T. & Dahmen, U. (2012). *Microsc. Microanal.* **18**, 676–683.
- Erni, R. (2010). *Aberration-corrected Imaging in Transmission Electron Microscopy*. London: Imperial College Press.
- Erni, R., Rossel, M. D. & Nakashima, P. N. H. (2010). *Ultramicroscopy*, **110**, 151–161.
- Feit, M. D. & Fleck, J. A. (1978). *Appl. Opt.* **17**, 3990–3998.
- Fejes, P. L. (1977). *Acta Cryst.* **A33**, 109–113.
- Fienup, J. R. (1982). *Appl. Opt.* **21**, 2758–2769.
- Fienup, J. R. (1987). *J. Opt. Soc. Am. A*, **4**, 118–123.
- Findlay, S. D. & LeBeau, J. M. (2013). *Ultramicroscopy*, **124**, 52–60.
- Findlay, S. D., Shibata, N., Sawada, H., Okunishi, E., Kondo, Y. & Ikuhara, Y. (2010). *Ultramicroscopy*, **110**, 903–923.
- Findlay, S. D., Shibata, N., Sawada, H., Okunishi, E., Kondo, Y., Yamamoto, T. & Ikuhara, Y. (2009). *Appl. Phys. Lett.* **95**, 191913.
- Fletcher, R. (1987). *Practical Methods of Optimization*, ch. 4, 2nd ed. New York: Wiley.
- Frank, J. (1972). *Biophys. J.* **12**, 484–511.
- Frank, J. (1973). *Optik*, **38**, 519–536.
- Frank, J. (2006). *Three Dimensional Electron Microscopy of Macromolecular Assemblies*. New York: Oxford University Press.
- Frank, J., Radermacher, M., Penczek, P. & Zhu, J. (1996). *J. Struct. Biol.* **116**, 190–199.
- Frank, J., Shimkin, B. & Dowse, H. (1981). *Ultramicroscopy*, **6**, 343–358.
- Frieden, B. R. (1972). *J. Opt. Soc. Am.* **62**, 511–518.
- Frieden, B. R. (1976). *J. Opt. Soc. Am.* **66**, 280–283.
- Frigo, M. & Johnson, S. G. (2005). *Proc. IEEE*, **93**, 216–231. <http://www.fft.org>.
- Fujiwara, K. (1961). *J. Phys. Soc. Jpn*, **16**, 2226–2238.
- Fultz, B. & Howe, J. (2013). *Transmission Electron Microscopy and Diffractometry of Materials*, 4th ed. Berlin: Springer-Verlag.
- Gemming, T., Möbus, G., Exner, M., Ernst, F. & Rühle, M. (1998). *J. Microsc.* **190**, 89–98.
- Gerchberg, R. W. & Saxton, W. O. (1971). *Optik*, **34**, 275–284.
- Gerchberg, R. W. & Saxton, W. O. (1973). *Image Processing and Computer-aided Design in Electron Optics*, edited by P. W. Hawkes, pp. 66–81. London and New York: Academic Press.
- Goldstein, J. I., Newbury, D. E., Echlin, P., Joy, D. C., Lyman, C. E., Lifshin, E., Sawyer, L. & Michael, J. R. (2003). *Scanning Electron Microscopy and X-Ray Microanalysis*, 3rd ed. New York: Springer.
- Gómez-Rodríguez, A., Beltrán-del-Río, L. M. & Herrera-Becerra, R. (2010). *Ultramicroscopy*, **110**, 95–104.
- Gonzalez, R. C. & Woods, R. E. (2008). *Digital Image Processing*, 3rd ed. Upper Saddle River: Pearson/Prentice Hall.
- Goodman, P. & Moodie, A. F. (1974). *Acta Cryst.* **A30**, 280–290.
- Gribelyuk, M. A. (1991). *Acta Cryst.* **A47**, 715–723.
- Grillo, V. & Rossi, F. (2013). *Ultramicroscopy*, **125**, 112–129.
- Grillo, V. & Rotunno, E. (2013). *Ultramicroscopy*, **125**, 97–111.
- Gunturk, B. K. & Li, X. (2013). Editors. *Image Restoration, Fundamentals and Advances*. New York: CRC Press.
- Haider, M., Hartel, P., Müller, H., Uhlemann, S. & Zach, J. (2009). *Philos. Trans. R. Soc. A*, **367**, 3665–3682.
- Haider, M., Müller, H. & Uhlemann, S. (2008). *Aberration-corrected Electron Microscopy*, edited by P. W. Hawkes, Vol. 153 of *Advances in Imaging and Electron Physics*, pp. 43–119. Amsterdam: Academic Press.
- Haider, M. S., Uhlemann, S. & Zach, J. (2000). *Ultramicroscopy*, **81**, 163–175.
- Hanisch, R. J., White, R. L. & Gilliland, R. L. (1997). *Deconvolution of Images and Spectra*, 2nd ed., edited by P. A. Jansson, pp. 310–360. San Diego: Academic Press.
- Harris, L. J., Larson, S. B., Hasel, K. W. & McPherson, A. (1997). *Biochemistry*, **36**, 1581–1597.
- Hau-Riege, S. P., Szoke, H., Chapman, H. N., Szoke, A., Marchesini, S., Noy, A., He, H., Howells, M., Weierstall, U. & Spence, J. C. H. (2004). *Acta Cryst.* **A60**, 294–305.
- Hawkes, P. W. (1981). *J. Phys. E*, **14**, 1353–1367.
- Hawkes, P. W. & Kasper, E. (1994). *Principles of Electron Optics*, Vol. 3, *Wave Optics*. San Diego: Academic Press.
- Heel, M. van, Harauz, G. & Orlova, E. V. (1996). *J. Struct. Biol.* **116**, 17–24.
- Heel, M. van & Keegstra, W. (1981). *Ultramicroscopy*, **7**, 113–130.
- Hegerl, R. (1996). *J. Struct. Biol.* **116**, 30–34.
- Hegerl, R. & Altbauer, A. (1982). *Ultramicroscopy*, **9**, 109–116.
- Hillyard, S. & Silcox, J. (1995). *Ultramicroscopy*, **58**, 6–17.
- Hovden, R., Xin, H. L. & Muller, D. A. (2012). *Phys. Rev. B*, **86**, 195415.
- Hovmöller, S. (1992). *Ultramicroscopy*, **41**, 121–135. <http://www.calidris-em.com>.
- Huang, P. Y., Ruiz-Vargas, C. S., van der Zande, A. M., Whitney, W. S., Levendorf, M. P., Kevek, J. W., Garg, S., Alden, J. S., Hustedt, C. J., Zhu, Y., Park, J., McEuen, P. L. & Muller, D. A. (2011). *Nature (London)*, **469**, 389–392.
- Humphreys, C. J. (1979). *Rep. Prog. Phys.* **42**, 1825–1887.
- Hunt, B. R. (1977). *IEEE Trans. Comput.* **C-26**, 219–229.
- Hwang, H. & Haddad, R. A. (1995). *IEEE Trans. Image Process.* **4**, 499–502.
- Hýtch, M. J. & Stobbs, W. M. (1994). *Ultramicroscopy*, **53**, 191–203.
- Intaraprasong, V., Xin, H. L. & Muller, D. A. (2008). *Ultramicroscopy*, **108**, 1454–1466.
- Isaacson, M. S., Langmore, J., Parker, N. W., Kopf, D. & Utlaut, M. (1976). *Ultramicroscopy*, **1**, 359–376.
- Ishikawa, R., Lupini, A. R., Findlay, S. D. & Pennycook, S. J. (2014). *Microsc. Microanal.* **20**, 99–110.
- Ishikawa, R., Okunishi, E., Sawada, H., Kondo, Y., Hosokawa, F. & Abe, E. (2011). *Nat. Mater.* **10**, 278–281.
- Ishizuka, K. (1980). *Ultramicroscopy*, **5**, 55–65.
- Ishizuka, K. (2004). *Microsc. Microanal.* **10**, 34–40.
- Ishizuka, K. (2006). <http://www.hremresearch.com>.
- Ishizuka, K. & Uyeda, N. (1977). *Acta Cryst.* **A33**, 740–749.
- Jain, A. K. (1989). *Fundamentals of Digital Image Processing*. Englewood Cliffs: Prentice Hall.
- Jansson, P. A. (1997). Editor. *Deconvolution of Images and Spectra*, 2nd ed. San Diego: Academic Press.
- Jesson, D. E. & Pennycook, S. J. (1993). *Proc. R. Soc. London Ser. A*, **441**, 261–281.
- Joy, D. C. (1995). *Monte Carlo Modeling for Electron Microscopy and Analysis*. New York: Oxford University Press.
- Kaiser, U., Biskupek, J., Meyer, J. C., Leschner, J., Lechner, L., Rose, H., Stöger-Pollach, M., Khlobystov, A. N., Hartel, P., Müller, H., Haider, M., Eyhusen, S. & Benner, G. (2011). *Ultramicroscopy*, **111**, 1239–1246.
- Kambe, K., Lehmpfuhl, G. & Fujimoto, F. (1974). *Z. Naturforsch. Teil A*, **29**, 1034–1044.
- Kambe, K. & Stampfl, C. (1994). *Ultramicroscopy*, **55**, 221–227.
- Karayannis, N. B. & Venetsanopoulos, A. N. (1990). *IEEE Trans. Acoust. Speech Signal Process.* **38**, 1155–1179.
- Kempen, G. M. van, van Vliet, L. J., Verveer, P. J. & van der Voort, H. T. M. (1997). *J. Microsc.* **185**, 354–365.
- Keyse, R. J., Garratt-Reed, A. J., Goodhew, P. J. & Lorimer, G. W. (1998). *Introduction to Scanning Transmission Electron Microscopy*. New York: Springer-Verlag.
- Kilaas, R. (1987). *Proceedings of the 45th Annual Meeting of the Microscopy Society of America*, edited by G. W. Bailey, pp. 66–69. San Francisco Press.
- Kilaas, R. (2006). <http://www.totalresolution.com/index.html>.
- Kilaas, R., Marks, L. D. & Own, C. S. (2005). *Ultramicroscopy*, **102**, 233–237.

- Kirkland, A. I. & Meyer, R. R. (2004). *Microsc. Microanal.* **10**, 401–413.
- Kirkland, A. I., Saxton, W. O., Chau, K.-L., Tsuno, K. & Kawasaki, M. (1995). *Ultramicroscopy*, **57**, 355–374.
- Kirkland, E. J. (1982). *Ultramicroscopy*, **9**, 45–64.
- Kirkland, E. J. (1984). *Ultramicroscopy*, **15**, 151–172.
- Kirkland, E. J. (1988a). *Image and Signal Processing in Electron Microscopy, Scanning Microscopy*, Supplement 2, edited by P. W. Hawkes, F. P. Ottensmeyer, W. O. Saxton & A. Rosenfeld, pp. 139–147. Chicago: Scanning Microscopy International.
- Kirkland, E. J. (1988b). *Proceedings of the 46th Annual Meeting of the Electron Microscopy Society of America*, edited by G. W. Bailey, pp. 832–833. San Francisco Press.
- Kirkland, E. J. (1990). *Ultramicroscopy*, **32**, 349–364.
- Kirkland, E. J. (1998). *Advanced Computing in Electron Microscopy*. New York: Plenum.
- Kirkland, E. J. (2005). *Ultramicroscopy*, **102**, 199–207.
- Kirkland, E. J. (2010). *Advanced Computing in Electron Microscopy*, 2nd ed. New York: Springer.
- Kirkland, E. J. (2011). *Ultramicroscopy*, **111**, 1523–1530.
- Kirkland, E. J. (2013). <http://sourceforge.net/projects/computem>.
- Kirkland, E. J., Loane, R. F. & Silcox, J. (1987). *Ultramicroscopy*, **23**, 77–96.
- Kirkland, E. J. & Siegel, B. M. (1979). *Optik*, **53**, 181–196.
- Kirkland, E. J. & Siegel, B. M. (1981). *Ultramicroscopy*, **6**, 169–180.
- Kirkland, E. J., Siegel, B. M., Uyeda, N. & Fujiyoshi, Y. (1980). *Ultramicroscopy*, **5**, 479–503.
- Kirkland, E. J., Siegel, B. M., Uyeda, N. & Fujiyoshi, Y. (1982). *Ultramicroscopy*, **9**, 65–74.
- Kirkland, E. J., Siegel, B. M., Uyeda, N. & Fujiyoshi, Y. (1985). *Ultramicroscopy*, **17**, 87–104.
- Kirkland, E. J. & Thomas, M. G. (1996). *Ultramicroscopy*, **62**, 79–88.
- Koch, C. (2015). <http://qstem.org>.
- Koch, C. & Zuo, J. M. (2000). *Microsc. Microanal.* **6** (Suppl. 2), 126–127.
- Kourkoutis, L. F., Parker, M. K., Vaithyanathan, V., Schlom, D. G. & Muller, D. A. (2011). *Phys. Rev. B*, **84**, 075485.
- Kourkoutis, L. F., Plitzko, J. M. & Baumeister, W. (2012). *Annu. Rev. Mater. Res.* **42**, 33–58.
- Krause, F. F., Müller, K., Zillmann, D., Jansen, J. & Schowalter, M. (2013). *Ultramicroscopy*, **134**, 94–101.
- Krivanek, O. L., Corbin, G. J., Dellby, N., Elson, B. F., Keyse, R. J., Murfitt, M. F., Own, C. S., Szilagi, Z. S. & Woodruff, J. W. (2008). *Ultramicroscopy*, **108**, 179–195.
- Krivanek, O. L., Dellby, N. & Lupini, A. R. (1999). *Ultramicroscopy*, **78**, 1–11.
- Krivanek, O. L., Dellby, N. & Murfitt, M. F. (2009). *Handbook of Charged Particle Optics*, 2nd ed., edited by J. Orloff, ch. 12, pp. 601–640. Boca Raton: CRC Press/Taylor and Francis.
- Krivanek, O. L., Dellby, N., Murfitt, M. F., Chisholm, M. F., Pennycook, T. J., Suenga, K. & Nicolosi, V. (2010). *Ultramicroscopy*, **110**, 935–945.
- LeBeau, J. M., D'Alfonso, A. J., Findlay, S. D., Stemmer, S. & Allen, L. J. (2009). *Phys. Rev. B*, **80**, 174106.
- LeBeau, J. M., Findlay, S. D., Allen, L. J. & Stemmer, S. (2008). *Phys. Rev. Lett.* **100**, 206101.
- LeBeau, J. M. & Stemmer, S. (2008). *Ultramicroscopy*, **108**, 1653–1658.
- Lentzen, M. (2008). *Microsc. Microanal.* **14**, 16–26.
- Lichte, H. (1986). *Ultramicroscopy*, **20**, 293–304.
- Loane, R. F., Kirkland, E. J. & Silcox, J. (1988). *Acta Cryst.* **A44**, 912–927.
- Loane, R. F., Xu, P. & Silcox, J. (1991). *Acta Cryst.* **A47**, 267–278.
- Loane, R. F., Xu, P. & Silcox, J. (1992). *Ultramicroscopy*, **40**, 121–138.
- Lobato, I. & Van Dyck, D. (2014). *Acta Cryst.* **A70**, 636–649.
- Lobato, I. & Van Dyck, D. (2015a). *Ultramicroscopy*, **155**, 11–19.
- Lobato, I. & Van Dyck, D. (2015b). *Ultramicroscopy*, **156**, 9–17.
- Lucy, L. B. (1974). *Astron. J.* **79**, 745–754.
- Ludtke, S. J., Baldwin, P. R. & Chiu, W. (1999). *J. Struct. Biol.* **128**, 82–97.
- Lupini, A. R. & Pennycook, S. J. (2012). *Microsc. Microanal.* **18**, 699–704.
- Maccagnano-Zacher, S. E., Mkhoyen, K. A., Kirkland, E. J. & Silcox, J. (2008). *Ultramicroscopy*, **108**, 718–726.
- Marabini, R., Masegosa, I. M., San Martín, M. C., Marco, S., Fernández, J. J., de la Fraga, L. G., Vaquerizo, C. & Carazo, J. M. (1996). *J. Struct. Biol.* **116**, 237–240.
- Marchesini, S. (2007). *Rev. Sci. Instrum.* **78**, 011301.
- Marks, L. & Kilass, R. (2006). <http://www.numis.northwestern.edu/edm/documentation/edm.htm>.
- Martin, R. M. (2004). *Electronic Structure, Basic Theory and Practice*. New York: Cambridge University Press.
- McMullan, G., Clark, A., Turchetta, R. & Faruqi, A. R. (2009). *Ultramicroscopy*, **109**, 1411–1416.
- McMullan, G., Faruqi, A. R., Clare, D. & Henderson, R. (2014). *Ultramicroscopy*, **147**, 156–163.
- Meinel, E. S. (1988). *J. Opt. Soc. Am. A*, **5**, 25–29.
- Mevenkamp, N., Binev, P., Dahman, W., Voyles, P. M., Yankovich, A. B. & Berkels, B. (2015). *Adv. Struct. Chem. Imaging*, **1**, 3.
- Meyer, C. E., Dellby, N., Dellby, Z., Lovejoy, T. C., Sarahan, M. C., Skone, G. S. & Krivanek, O. L. (2014). *Microsc. Microanal.* **20** (Suppl. 3), 1108–1109.
- Meyer, J. C., Kurasch, S., Park, H. J., Skakalova, V., Künzel, D., Gross, A., Chuvilin, A., Algara-Siller, G., Roth, S., Iwasaki, T., Starke, U., Smet, J. H. & Kaiser, U. (2011). *Nat. Mater.* **10**, 209–215.
- Meyer, R. R., Kirkland, A. I. & Saxton, W. O. (2002). *Ultramicroscopy*, **92**, 89–109.
- Meyer, R. R., Kirkland, A. I. & Saxton, W. O. (2004). *Ultramicroscopy*, **99**, 115–123.
- Miao, J., Chapman, H. N., Kirz, J., Sayre, D. & Hodgson, K. O. (2004). *Annu. Rev. Biophys. Biomol. Struct.* **33**, 157–176.
- Miao, J., Ishikawa, T., Robinson, I. K. & Murnane, M. M. (2015). *Science*, **348**, 530–535.
- Miao, J., Sayre, D. & Chapman, H. N. (1998). *J. Opt. Soc. Am. A*, **15**, 1662–1669.
- Midgley, P. A. & Weyland, M. (2003). *Ultramicroscopy*, **96**, 413–431.
- Millane, R. P. (1990). *J. Opt. Soc. Am. A*, **7**, 394–411.
- Ming, W. Q. & Chen, J. H. (2013). *Ultramicroscopy*, **134**, 135–143.
- Misell, D. L., Stroke, G. W. & Halioua, M. (1974). *J. Phys. D Appl. Phys.* **7**, L113–L117.
- Mkhoyan, K. A., Maccagnano-Zacher, S. E., Kirkland, E. J. & Silcox, J. (2008). *Ultramicroscopy*, **108**, 791–803.
- Moore, G. E. (1998). *Proc. IEEE*, **86**, 82–85. Reprinted from *Electronics*, April 19, 1965, pp. 114–117.
- Muller, D. A., Edwards, B., Kirkland, E. J. & Silcox, J. (2001). *Ultramicroscopy*, **86**, 371–380.
- Müller, H., Uhleman, S., Hartel, P. & Haider, M. (2006). *Microsc. Microanal.* **12**, 442–455.
- Murfitt, M. F., Meyer, C. E., Skone, G. S., Dellby, N. & Krivanek, O. L. (2013). *Microsc. Microanal.* **19** (Suppl. 2), 782–783.
- Nellist, P. D. & Pennycook, S. J. (1998). *J. Microsc.* **190**, 159–170.
- Nellist, P. D. & Rodenburg, J. M. (1998). *Acta Cryst.* **A54**, 49–60.
- O'Keefe, M. A. (1979). *Proceedings of the 37th Annual Meeting of the Electron Microscopy Society of America*, edited by G. W. Bailey, pp. 556–557. Cincinnati: San Francisco Press.
- O'Keefe, M. A. & Buseck, P. R. (1979). *Trans. Am. Crystallogr. Assoc.* **15**, 27–46.
- O'Keefe, M. A. & Kilaas, R. (1988). *Image and Signal Processing in Electron Microscopy, Scanning Microscopy*, Supplement 2, edited by P. W. Hawkes, F. P. Ottensmeyer, W. O. Saxton & A. Rosenfeld, pp. 225–244. Chicago: Scanning Microscopy International.
- O'Leary, N. L. & Allen, L. J. (2005). *Acta Cryst.* **A61**, 252–259.
- Ophus, C. & Ewalds, T. (2012). *Ultramicroscopy*, **113**, 88–95.
- Peng, L.-M., Dudarev, S. L. & Whelan, M. J. (2011). *High-Energy Electron Diffraction and Microscopy*. Oxford University Press.

- Pennington, R. S., Wang, F. & Koch, C. T. (2014). *Ultramicroscopy*, **141**, 32–37.
- Pennycook, S. J. & Nellist, P. D. (2011). Editors. *Scanning Transmission Electron Microscopy, Imaging and Analysis*. New York: Springer.
- Plamann, T. & Rodenburg, J. M. (1998). *Acta Cryst.* **A54**, 61–73.
- Pogany, A. P. & Turner, P. S. (1968). *Acta Cryst.* **A24**, 103–109.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (2007). *Numerical Recipes*, 3rd ed. Cambridge University Press.
- Pulvermacher, H. (1981). *Optik*, **60**, 45–60.
- Quinn, M. J. (2004). *Parallel Programming in C with MP and open MP*. New York: McGraw Hill.
- Reeves, S. J. & Mersereau, R. M. (1990). *Opt. Eng.* **29**, 446–454.
- Reimer, L. (1993). *Transmission Electron Microscopy. Springer Series in Optical Sciences*, Vol. 36, 3rd ed. New York: Springer-Verlag.
- Reimer, L. (1998). *Scanning Electron Microscopy. Springer Series in Optical Sciences*, Vol. 45, 2nd ed. New York: Springer-Verlag.
- Rez, P. (1985). *Ultramicroscopy*, **16**, 255–260.
- Richardson, W. H. (1972). *J. Opt. Soc. Am.* **62**, 55–59.
- Robertson, M. D., Bennett, J. C., Burns, M. M. J. & Currie, D. (2006). *Microscopy and Microanalysis 2006*, Vol. 12 (Suppl. 2), edited by P. Kotula, M. Marko, J.-H. Scott, R. Gauvin, D. Beniac, G. Lucas, S. McKernan & J. Shields, pp. 714–715. Cambridge University Press.
- Rodenburg, J. M. & Bates, R. H. T. (1992). *Philos. Trans. R. Soc. London Ser. A*, **339**, 521–553.
- Rodriguez, J. A., Xu, R., Chen, C.-C., Zou, Y. & Miao, J. (2013). *J. Appl. Cryst.* **46**, 312–318.
- Rose, H. (1981). *Nucl. Instrum. Methods Phys. Res.* **187**, 187–199.
- Rullgård, H., Öfverstedt, L.-G., Masich, S., Daneholt, B. & Öktem, O. (2011). *J. Microsc.* **243**, 234–256.
- Sasaki, T., Sawada, H., Hosokawa, F., Sato, Y. & Suenaga, K. (2014). *Ultramicroscopy*, **145**, 50–55.
- Saxton, W. O. (1978). *Computer Techniques for Image Processing in Electron Microscopy*. In *Advances in Electronics and Electron Physics*, Supplement 10. New York: Academic Press.
- Saxton, W. O. (1996). *J. Struct. Biol.* **116**, 230–236.
- Saxton, W. O., Pitt, T. J. & Horner, M. (1979). *Ultramicroscopy*, **4**, 343–354.
- Scales, L. E. (1985). *Introduction to Non-Linear Optimization*. New York: Springer-Verlag.
- Scherzer, O. (1949). *J. Appl. Phys.* **20**, 20–29.
- Scherzer, O. (1970). *Ber. Bunsenges. Phys. Chem.* **74**, 1154–1167.
- Schiske, P. (1968). *Proceedings of the Fourth Regional Congress on Electron Microscopy*, Vol. I, edited by D. S. Bocciarelli, pp. 145–146.
- Schiske, P. (1973). *Image Processing and Computer-aided Design in Electron Optics*, edited by P. W. Hawkes, pp. 82–90. London, New York: Academic Press.
- Schiske, P. (2002). *J. Microsc.* **207**, 154.
- Schneider, C. A., Rasband, W. S. & Eliceiri, W. K. (2012). *Nat. Methods*, **9**, 671–75. <http://imagej.nih.gov/ij/>.
- Schramm, S. M., van der Molen, S. J. & Tromp, R. M. (2012). *Phys. Rev. Lett.* **109**, 163901.
- Shepp, L. A. & Vardi, Y. (1982). *IEEE Trans. Med. Imag.* **1**, 113–122.
- Smart, J., Hock, K. & Csomor, S. (2006). *Cross Platform GUI Programming with wxWidgets*. Upper Saddle River: Prentice Hall. <http://wxwidgets.org/>.
- Smith, R. & Carragher, B. (2008). *J. Struct. Biol.* **163**, 224–228.
- Spargo, A. E. C., Beeching, M. J. & Allen, L. J. (1994). *Ultramicroscopy*, **55**, 329–333.
- Spence, J. C. H. (1998). *Acta Cryst.* **A54**, 7–18.
- Spence, J. C. H. (2013). *High-Resolution Electron Microscopy*, 4th ed. New York: Oxford University Press.
- Spence, J. C. H., Calef, B. & Zuo, J. M. (1999). *Acta Cryst.* **A55**, 112–118.
- Spence, J. C. H. & Cowley, J. M. (1978). *Optik*, **50**, 129–142.
- Spence, J. C. H. & Zuo, J. M. (1992). *Electron Microdiffraction*. New York: Plenum Press.
- Stadelmann, P. A. (1987). *Ultramicroscopy*, **21**, 131–146.
- Stadelmann, P. A. (2004). *JEMS – EMS Java version*, <http://cimewww.epfl.ch/people/stadelmann/jemsWebSite/jems.html>.
- Stroppa, D. G., Righetto, R. D., Montoro, L. A. & Ramirez, A. J. (2011). *Ultramicroscopy*, **111**, 1077–1082.
- Tanaka, N. (2015). Editor. *Scanning Transmission Electron Microscopy of Nanomaterials*. London: Imperial College Press.
- Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I. & Ludtke, S. J. (2007). *J. Struct. Biol.* **157**, 38–46.
- Thust, A. (2009). *Phys. Rev. Lett.* **102**, 220801.
- Thust, A., Coene, W. M. J., Op de Beck, M. & Van Dyck, D. (1996). *Ultramicroscopy*, **64**, 211–230.
- Treacy, M. M. J. & Gibson, J. M. (1993). *Ultramicroscopy*, **52**, 31–53.
- Trussell, H. J. (1980). *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-28**, 114–117.
- Trussell, H. J. & Hunt, B. R. (1979). *IEEE Trans. Comput.* **C-27**, 7–62.
- Van Dyck, D. (1985). *Advances in Electronics and Electron Physics*, Vol. 65, edited by P. W. Hawkes, pp. 295–355. Orlando: Academic Press.
- Van Dyck, D. & Coene, W. (1984). *Ultramicroscopy*, **15**, 29–40.
- Voyles, P. M., Grazul, J. L. & Muller, D. A. (2003). *Ultramicroscopy*, **96**, 251–273.
- Voyles, P. M., Muller, D. A. & Kirkland, E. J. (2004). *Microsc. Microanal.* **10**, 291–300.
- Wade, R. H. & Frank, J. (1977). *Optik*, **49**, 81–92.
- Walker, J. S. (1996). *Fast Fourier Transforms*, 2nd ed. Boca Raton: CRC Press.
- Watanabe, K., Kikuchi, Y., Hiratsuka, K. & Yamaguchi, H. (1988). *Phys. Status Solidi A*, **109**, 119–126.
- Wei, D.-Y. & Yin, C.-C. (2010). *J. Struct. Biol.* **172**, 211–218.
- Wernecke, S. J. & D’Addario, L. R. (1977). *IEEE Trans. Comput.* **C-26**, 351–364.
- Williams, D. B. & Carter, C. B. (2009). *Transmission Electron Microscopy, a Textbook for Materials Science*, 2nd ed. New York: Springer.
- Wu, J. S., Weierstall, U., Spence, J. C. H. & Koch, C. T. (2004). *Opt. Lett.* **29**, 2737–2739.
- Yu, Z., Muller, D. A. & Silcox, J. (2004). *J. Appl. Phys.* **95**, 3362–3371.
- Yu, Z., Muller, D. A. & Silcox, J. (2008). *Ultramicroscopy*, **108**, 494–501.
- Zuo, J. M. (2009). *Web Electron Microscopy Applications Software (WebEMAPS)*, <http://emaps.mrl.uiuc.edu/>.